



*DIFU*100ci@

REVISTA DE DIFUSIÓN CIENTÍFICA
INGENIERÍA Y TECNOLOGÍAS

ISSN 2007-3585

UAZ Universidad
Autónoma de
Zacatecas

"Francisco García Salinas"



Vol. 7, No.2, septiembre-diciembre 2013

Número especial dedicado al

Ci²Comp

Congreso Internacional en
Ciencias Computacionales

www.cicomp.org

ISSN 2007-3585



PRODUCCIÓN
Universidad Autónoma de Zacatecas

PRODUCCIÓN Y DISEÑO
Jorge Flores Troncoso

D.R. de la Presente Edición
Jorge Flores Troncoso
Universidad Autónoma de Zacatecas
López Velarde 801, Centro
98010 Zacatecas, Zac. México

ISSN 2007-3585

DIFU100ci@ (léase difucencia) Vol. 7, No.2, septiembre-diciembre 2013, es una publicación cuatrimestral editada por la Universidad Autónoma de Zacatecas, "Francisco García Salinas", Jardín Juárez 147, Col. Centro Zacatecas, Zac. C.P. 98000. <http://www2.uaz.edu.mx/web/www/publicaciones>. email: jflorest@uaz.edu.mx. Reservas de Derechos al Uso Exclusivo del Título expedido por el INDAUTOR, Reserva: 04-2010-110314331900-102. Responsable de la última actualización Jorge Flores Troncoso, López Velarde 801, Zona Centro, Zacatecas, Zac. C.P. 98000. Fecha de última modificación 31 de diciembre de 2013.

Hecho en México
Made in Mexico

Directorio

I. Q. Armando Silva Cháirez Rector
Lic. Cuauhtémoc Rodríguez Aguirre Secretario General
Dr. Miguel Rodríguez Jáquez Secretario Académico
M. en A Emilio Morales Vera Secretario Administrativo
Dr. Marco Antonio Salas Luévano Coord. Investigación y Posgrado
Dra. Georgia Aralu González Pérez Coord. Depto. Editorial

Grupo Editor Ejecutivo

Jorge Flores Troncoso, Editor en Jefe, UAZ México
Manuel Reta Hernández, UAZ México
Hamurabi Gamboa Rosales, UAZ México
Claudia Sifuentes Gallardo, UAZ México

Grupo Revisor

Oliver Joekisch, University of Applied Sciences, Germany
Ding Hongwei, Tongji University, China
Robert Weissbach, Penn State Erie, The Behrend College
Naim Logic, Arizona Public Service
Leonardo Acho Zuppa, U Politècnica de Catalunya, España
Miguel Andrés, U. de Valencia, España
Manuel Hernández Calviño, Universidad de La Habana, Cuba
Luis C. García Santander, Universidad de Concepción, Chile
Ramón Parra Michel, CINVESTAV, México
Leonel Soriano Equigua, FIME-UCol, México
Geminiano D. Martínez Ponce, CIO, México
Carmen Maya Sánchez, CICESE México
Jaime Sánchez García, CICESE México
David H. Covarrubias Rosales, CICESE, México
Enrique Pacheco Cabrera, AEM, México
Remberto Sandoval Aréchiga, UAZ México
Salvador Ibarra Delgado, UAZ México
Rodrigo Soulé de Castro, UAZ México
Ernesto García Domínguez, UAZ México
Claudia Reyes Rivas, UAZ México
Rafael Villela Varela, UAZ México
Jorge de la Torre y Ramos, UAZ México
Guillermo Romo Guzmán, UAZ México
Rodrigo Castañeda Miranda, UAZ México
Luis Octavio Sánchez Solís, UAZ México

Contenido

Editorial	p. 1
Collaborative Educational Experience with Smartphone Support: An Exploratory Analysis	pp. 2-7
Optimización en la Aproximación de Curvas NURBS con Algoritmos Genéticos Paralelos en un GPU	pp. 8-16
Estimación de la Función de Iluminación No Uniforme para la Restauración de Imágenes en un Procesador Gráfico	pp. 17-24
Procesando el Reconocimiento de Múltiples Objetos en Tiempo Real Hacia la Implementación para el Reconocimiento de Actividades Humanas Concurrentes Utilizando GPU	pp. 25-32
Solución Paralela en un GPU para la Ecuación de Advección-Difusión-Reacción mediante Diferencias Finitas Explícitas e Implícitas	pp. 33-40
Modelo Conceptual Inicial para Determinar la Calidad en Objetos de Aprendizaje Integrando un Enfoque a Servicios	pp. 41-45
Implementation of an Intelligent Algorithm in a PLC: A Practical Approach	pp. 46-53
Análisis de Factores que Afectan el QoS que Ofrecen las WSN Aplicado a Entornos de Salud	pp. 54-62
Arquitectura de Agentes ARSEC-AMS para Diseño de Sistemas de Seguridad	pp. 63-70
Reconocimiento Biométrico de la Palma de la Mano Basado en el Promediado de Filtros de Correlación Simples	pp. 71-76
Medidor Modular Inalámbrico de la Tensión Hídrica del Suelo	pp. 77-83
Uso de Matlab para la Síntesis en Lenguaje VHDL de Decodificadores Viterbi	pp. 84-87
Sistema Inmerso de Bajo Costo para la Localización y Segmentación de Uvas	pp. 88-92

Determinación de Heurísticas de diseño en Objetos de Aprendizaje como elemento en la definición de un Modelo de Preferencias de Usuario	pp. 93–97
Control de Asistencia y Tiempo de Permanencia de Personas en un Área Laboral Mediante la Detección de Smartphones	pp. 98–103
Aplicación Móvil como Asistente en la Comunicación de Personas Discapacitadas y con Limitaciones del Habla	pp. 104–109
Análisis Térmico para un Circuito SoC de Disparo de un Rectificador Controlado de Alta Eficiencia	pp. 110–115
Implementación de una Comunidad Virtual de Aprendizaje en los Subsistemas de los Tecnológicos Descentralizados del Estado de Tabasco para el Fortalecimiento Académico	pp. 116–121
Selección de Características para Análisis de Sentimientos Basado en Computación Evolutiva: Resultados Preliminares	pp. 122–128
Sistema Multi-agente para la Detección de Fraudes en el Correo Electrónico	pp. 129–134

Editorial

La revista *DIFU*₁₀₀*ci*@ es una revista cuatrimestral que comenzó su publicación oficial en 2005. En mayo del 2012, la revista *DIFU*₁₀₀*ci*@ adquirió el ISSN.

Desde entonces, se pretende contribuir a la difusión del conocimiento de la comunidad académica tanto nacional como internacional mediante la difusión de resultados de investigación de alta calidad. La Revista se centra en obras originales, que incluyen principalmente los estudios experimentales, análisis numéricos, estudios de casos y revisiones bibliográficas que proporcionan una significativa contribución a las áreas de ingeniería y tecnología en todas las disciplinas (Electrónica, Eléctrica, Ciencias de la Computación, Mecatrónica, Robótica, Telecomunicaciones, Procesamiento de señales, Ingeniería Industrial, Ingeniería de Control, y Bioingeniería).

Desde el comienzo, la revista ha buscado la mejora de los artículos aceptados para su publicación por un proceso de evaluación por pares o árbitro de los manuscritos recibidos. Estas evaluaciones son llevadas a cabo por expertos de reconocido prestigio por sus conocimientos y logros académicos, con el objetivo de asegurar que las publicaciones seleccionadas están contribuyendo al estado del arte en diferentes áreas de interés. Además, desde su inicio, la revista se ha abierto a los estudiantes y académicos a través del Sistema Open Journal, facilitando todo el proceso de presentación y publicación.

En este número, se publican trabajos seleccionados por el Comité Organizador de la **sexta edición del Congreso Internacional en Ciencias Computacionales CiComp 2013** (www.cicomp.org). La organización del CiComp 2013 esta a cargo de la Licenciatura en Ciencias Computacionales de la Facultad de Ciencias de la Universidad Autónoma de Baja California (UABC), México. Agradecemos al Comité por la confianza para la publicación de sus trabajos. La selección y revisión de los manuscritos aquí presentados quedan bajo responsabilidad del Comité Revisor del CiComp 2013.

Por último, quiero dar las gracias al Dr. Gerardo Miramontes, fundador de la revista, por permitirme continuar con la noble causa de difundir la ciencia y la tecnología; al equipo editorial, a los autores y revisores, que se esfuerzan para mejorar la calidad de los manuscritos. Además, me gustaría aprovechar la oportunidad para animar a todos los investigadores, académicos y estudiantes en las áreas de ingeniería y tecnología para que continúen sometiendo sus artículos en nuestra revista.

Jorge Flores Troncoso
Editor en Jefe, Revista *DIFU*₁₀₀*ci*@
Universidad Autónoma de Zacatecas

Collaborative Educational Experience with Smartphone Support: An Exploratory Analysis

Javier Organista Sandoval^a, Arturo Serrano Santoyo^b

^aIIDE, Universidad Autónoma de Baja California.
Carr. Tijuana-Ensenada, Km. 103, Ensenada, B.C., México, 22830.
javor@uabc.edu.mx

^bCentro de Investigación Científica y Educación Superior de Ensenada
Carr. Ensenada-Tijuana 3918, Zona Playitas, Ensenada, B.C., México, 22860.
serrano@cicese.edu.mx

2013 Published by *DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Abstract

The educational context has become fertile ground for exploring the potential application of mobile handheld devices to support learning. The aim of this paper is to present the results of the implementation of a collaborative educational activity (m.Activity) using smartphones. This paper identifies the key elements that support ubiquitous communication and interaction of groups of university students in a context that requires an intense level of interactivity and communication for the delivery of short-term assignments.

Keywords: Educational technology, Human-computer interaction, M-learning

1. Introduction

The convergence of global wireless communications and mobile devices with Internet access capabilities [1] and the technological development of wireless networks and their explosive penetration in practically all regions of the world have created opportunities, not previously available, for exploring applications of mobile devices in social, cultural, economic and educational contexts. The deployment of 3G and 4G (third and fourth generation technologies) is advancing and fueled by converging applications that include multi-functional features which enable the user to take advan-

tage of services such as MMS (Multimedia Messaging), mobile video and more importantly, mobile Internet, that allow access to social networking applications and other related services. With the emergence of 3G and its evolution to 4G, the mobile device is seen more as a platform rather than just a telephone, creating challenges for its usability and adoption by ordinary users. Nevertheless, service providers and device manufacturers are seeking strategies to facilitate the adoption of mobile telephones as universal terminals for communication and access to the Internet.

Interaction and collaboration, crucial elements in educational environments, can be supported with powerful

handheld devices such as smartphones. The purpose of our paper is to identify key factors in the implementation of a collaborative educational activity using smartphones from an exploratory perspective. The advanced technological capabilities embedded in smartphones and the development of content for educational emphasis have become key elements in the discipline known as m-learning, a discipline that has recently caught the attention of scholars and educators to look on the one hand, strategies for increasing the coverage, quality and efficiency of educational services and on the other alternative learning opportunities outside the classroom [2]-[3]. In view of the authors, m-learning is considered as the set of processes and technologies to support learning through mobile devices. Not only is the concept of exploring ways to support learning in a ubiquitous or nomadic fashion. Aspects of interaction, interactivity, immediacy, intention of use and other ergonomic, usability and adoption factors are part of such an approach, which aims to comprehensively address the deployment and implementation of smartphones in the educational context [4].

Our approach in this paper focuses on describing the experience of conducting a group activity designed for educational purposes (*m.Activity*) with the support of a Mobile Collaboration Space (MCS) hosted on a web server, where each of the participants in the research experience could be accessed from a handheld device (see Figure 1). The paper is organized as follows: In section 2 we describe the development of the MCS, the functional characteristics of smartphones and the description and procedure to perform the *m.Activity*. In section 3 we describe the *m.Activity* and in sections 4, 5 and 6 we present respectively the results of the interactive experience, a final discussion highlighting the most significant findings of our contribution and the conclusions. The participants in this experience were graduate students with computer science, engineering and psychology backgrounds.

2. Development of the MCS around the *m.Activity*: Purpose and structure of the MCS

The MCS is a collaboration tool hosted on a webserver (<http://m.educa.ens.uabc.mx/>). For the MCS design, we consider sociocultural elements involved in the application of mobile devices in educational contexts around the framework of convergence of mobile and ubiquitous computing. We started from the premise that college students show a natural tendency and willingness to use technology in their learning process, ie digital natives, from the perspective of Prensky in [5].



Figure 1. Components of the collaborative experience with smartphone support.

Thus the design of our MCS focuses on harnessing the familiarity and inclination of students to the use of social networks in their social interaction and extrapolates those experiences to an educational context. The MCS objective is to support university student groups to perform tasks that require interaction and communication among all the group members. The design provides a virtual space for collaboration and communication for 3G/4G access through mobile devices (smartphones).

We designed the MCS with the objective of hosting the collaborative activities of various student groups simultaneously. The *m.Activity* was limited to a maximum of six participants per group, this to facilitate the flow and display of visual information with smartphones. Groups are created in classroom sessions for the initial organization of the task goals and scope. The course facilitator defines the core issues and delivery date of the assignment. In turn, each group appoints its respective project leader, whose job is to coordinate and monitor the progress of the group. At the start of activities, all group members access the MCS to register their profiles, usernames and passwords. Based on this record, the dynamics of mobile and ubiquitous interaction of the participants is enabled. The development of collaborative activities to fulfill the goals is highly dependent on sociocultural, technological and usability factors regarding the use of mobile devices, as well as on the participants' interrelations and quality of interactions [6].

The MCS comprises seven modules (see Figure 2). The first, called *Progress*, provides information about the task or project to be developed. The next module, *Notices*, shows information regarding the development of the activity to be undertaken. These notices must be short messages for adequate display on small screen



Figure 2. MCS main menu

devices such as smartphones. The third module called *Repository* contains supporting information (text, images, multimedia) to develop the *m.Activity*. Each group sets its repository independently and for their exclusive use. The next module, called *Chat*, allows real-time communication. The fifth module is a *Wiki* to facilitate document edition. This encourages the collaborative construction of documents and agreements between the participants. The next module, called *Contact*, contains profiles and key information necessary to locate the members of each group. The last module, *Help*, contains the MCS user manual and some answers to frequent questions.

To structure the MCS, three key aspects were considered. The first was to allow interaction and communication of students through a smartphone device regardless of type or model, operating system and browser. The second aspect was to use the Internet connectivity offered by cellular technology. The third key aspect of the MCS is the webserver architecture that integrates services for content storage and their associated databases to support learning and to accomplish the tasks. This architecture enables operation and access to MCS without having to develop specific applications for each type and model of smartphone.

To achieve a good user experience in the collaborative activity, the capabilities and limitations of smartphones must be considered. It is essential to develop applications that consider the limitation of screen size, storage capacity and processing of mobile devices. The design of the screens and the definition of content are crucial for successful collaborative interaction. Another element to consider is to contract the lowest cost rate plans and payment schemes of cellular phone providers in order to create a cost-effective learning experience in a real environment.

3. Description of the *m.Activity*

As a starting point, we identified the main features, aspects of connectivity, applications and functionality of the six smartphones used. From the selected dimensions and variables, we developed a questionnaire for assessing the functionality of such smartphones. To do this, we considered 24 items with five response options:

not applicable, bad, fair, good and excellent (Consolidation 0-4). Regarding the operating system of the six smartphones, two of them use the operating system Windows Mobile 6.0 Professional, two operate under Symbian S60 and the other two work with I-phone OS.

The objective of the *m.Activity* was to develop an essay in digital format that included the definition of smartphone devices, their characteristics and pedagogical potential. The development of this essay required the collaborative participation of the team to elaborate on the educational potential of mobile technology. As originally planned, we configured the MCS to enabling smartphone access from each team member.

Undoubtedly, the educational potential of mobile technologies is enormous. To mention some examples, students can locate information in various formats, to communicate with experts or with peers, exchange information, among many other activities. However, they also are exposed to distractions such as games, movies and social networks, among others. We observed that in some instances students were unaware of special features, capabilities and potential educational applications of smartphones. Under these circumstances, it was considered pertinent for the *m.Activity* to develop an essay in digital and multimedia format in order to realize the advantages and limitations of smartphones. This essay could in turn serve as a white paper to introduce the student community to the m-learning discipline. The collaborative group was formed with six members (two researchers and four graduate students). The idea was to integrate a group whose participants had equal roles to perform the task (*m.Activity*). By agreement of the group, one student was designated as project leader.

As part of our experience, regarding the *m.activity*, a focus group was implemented to gather information about: i) technical aspects of the server that housed the mobile collaboration space, ii) development of the MCS, iii) feedback on the smartphone as an access point to the server, iv) opinion about the *m.Activity* and finally, and v) the advantages and disadvantages of mobile technology. The focus group participants were the six members of the collaborative research group. For the quantitative statistical analysis we used SPSS Ver. 17.0[®].

We kept track of the development and progress of the *m.Activity* as well as the quality of the working groups contributions in each of the collaboration spaces of the MCS. As it was mentioned before, at the start of the *m.Activity*, a kick-off meeting is organized to gather information about the activities and arrangements for the preparation of the essay and to coordinate the group members' interactions. A period of ten business days was set for delivery and publication of the essay.

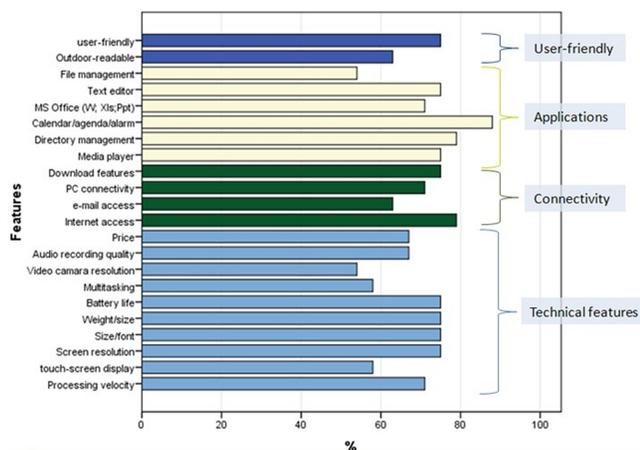


Figure 3. Features of Smartphones used in the research.

4. Results

The questionnaire considered an equivalent percentage of the scale used (0-4). Items were grouped into four categories: characteristics of the device, connectivity, applications and functionality (see Figure 3). For the characteristics of the device, associated items achieved moderate rates (~ 75%) for attributes such as size and weight, battery life and screen resolution. The attributes most penalized (~ 55%) were the ability to process multiple tasks, the quality of video camera and touch screen handling. Regarding connectivity, the best evaluated attribute (79%) was the ability to access the Internet, while the worst score is related to email management (63%). With regard to the handling of applications, smartphones had a moderate value (~ 80%). Of note is the attribute that has to do with the organization of files by their low value (54%). Ease of use was the best aspect evaluated for the functionality of the device (75%), while the ability to read text in outdoor spaces was poorly evaluated (63%).

Upon entering the *MCS PROGRESS* option, a description, main features and specific questions about the task, as well as the deadline of the assignment are displayed (see Figure 4).

For ten days, the research team collaboratively developed a paper outlining the advantages and limitations of smartphones as a teaching tool. As a result, an essay of 11 pages in Word format was produced and incorporated into the MCS repository.

5. Discussion

In this section we present our findings and a brief analysis about the smartphones used as well as the imple-

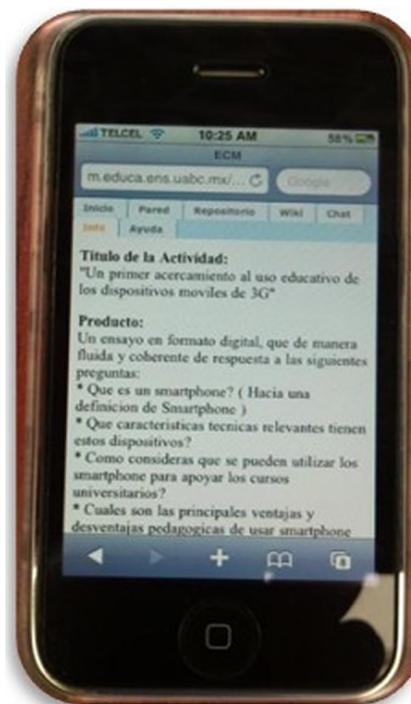


Figure 4. MCS: main features of the task

mentation of the *m.Activity*. Finally, some reflections on the pedagogical potential of portable devices are raised.

With regard to the benefits of using smartphones for the development of the *m.Activity*, portability, connectivity, personalization and user friendly features were identified. Internet connectivity was available to all members of the group at all times. The group members noted also some specific problems, among which, the reduced visibility of small screen displays in outdoors scenarios and the limited amount of text that can be displayed on the screen at a given time. An important aspect to consider is the limitation for inputting information in the smartphone and the tradeoffs of using haptic or qwerty keyboards. Additionally, programs or applications for smartphones available on the Internet may show unpredictable performance due to compatibility issues arising from the different operating systems available for such devices. The applications more frequently used during the *m.Activity* were electronic agendas, alarm and calendar, contact management and communication.

Our Mobile Collaboration Space was installed in a specially configured webserver for handling and identification of the smartphones employed. For the development of the software applications built on the MCS, we used public domain modules in addition to our own coding. The MCS was accessed in an efficient and stable fashion. The choice of bulletin boards for the NOTICES option was particularly useful to publish short messages

and keep a historical record of messages sent. Thus, any member of the group readily knew the status of the activity and the agreements reached. For its part, the *CHAT* option did not function as expected, mainly for the difficulty of reaching agreements when six participants interacted simultaneously. Added to this, it was difficult to manage historical records of text on the small screen. It is recommended to form smaller groups, perhaps with a maximum of three or four participants and consider some mechanism to moderate the communication and interactions among all participants.

Regarding the *REPOSITORY* option, we highlight the importance of this shared storage space. This option allowed the management and storage of important and useful files for the group. One of the main benefits and design features of the MCS, noted by the participants, was the ability to record the evolution of collaborative activity. We were able to keep records of progress and to implement eventually remedial actions. Nevertheless, the participants felt the need to improve the design of the MCS to optimize its performance and to include applications and features of new generation smartphones and tablets.

In relation to the collaborative activity undertaken with support of the MCS and smartphone Internet connectivity, we highlight the following. The collaborative work required a minimum of knowledge and skills on the part of the participants. Seeing the *m.Activity* as a process, some key steps such as: motivation, socialization, agreements, information flows and synthesis were identified. The participants mentioned the difficulty of constructing the document of the essay using the smartphone, for this reason the members of the working groups used their desktop computer or laptop to accomplish this task. The project leader noted that his work required a great effort to ensure the participation of all working group members in a timely manner. The coordination requires collaborative skills not usually available or taught, so it is suggested to train participants about collaborative work and their roles before or at the inception of the mobile collaborative projects.

6. Conclusions

In closing, we stress the enormous potential of smartphones for educational activities. Undoubtedly this requires the knowledge of its core capabilities and functions, as well as a clear pedagogical strategy. Several studies [7] have pointed out that the frequent contact of students with mobile technology encourages the development of skills to operate these devices, specially regarding the search and query of information from large

data banks and digital libraries on the web. According to [8], handheld devices are conducive to students' integration in social networking. The possibility of instant communication gives the owner of such devices, the perception of belonging to a community. Generally, social networks establish a virtual meeting point for students with common academic interests and responsibilities. By the same token, social networks support the functions of information sharing, collaboration and communication among network members. These functions are key to educational applications. One of the biggest challenges we have in our university community is the very conception and realization on the part of our students that smartphones are potential teaching and learning tools. Fortunately, a growing number of educational applications are currently deployed or under development. Though many of such applications are deliberately intended to arouse the recreational interests of students, the communication and entertainment applications are still more pervasive.

The MCS design was conceived to stimulate the collaboration of small groups of students in a ubiquitous fashion. According to our findings and interviews, the students responded expeditiously and enthusiastically to the assignment. Our aim was to design a collaboration space with a simple and user-friendly smartphone interface with the basic components to accomplish successful interaction. The MCS contributed to accomplish the objectives of our educational research. It must be acknowledged that new models of smartphones and new versions of mobile operational systems might have a significant impact on the results shown in section 4, particularly on the quality of the information displayed as well as in the perceptions of the MCS users. This is part of our future research in which we will incorporate some elements of cognitive nature and a more detailed analysis of the quality of the interactions among the participants.

References

- [1] Portio Research Ltd, "Mobile Factbook," (Online, February, 2013), <http://www.portioresearch.com/media/3986/Portio%20Research%20Mobile%20Factbook%202013.pdf>.
- [2] S. Berger, R. Mohr, H. Nösekabel, and K. J. Schäfer, "Mobile Collaboration Tool for University Education", *Proceedings of the Twelfth IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'03)*, Paris, France, pp. 77-78, June 2003.
- [3] J. Brown, "Exploring Mobile Learning: Part One of the mLearning Series." Paper on mobile learning, April, 2009.
- [4] A. Serrano Santoyo and J. Organista-Sandoval, "Implications of 4G Connectivity Related to M-Learning Contexts," *Journal of*

- the Research Center for Educational Technology*, Vol. 6, No. 1, pp. 129-135, 2010.
- [5] M. Prensky, "Digital natives, Digital immigrants. On the Horizon," *On the Horizon*, MCB University Press, Vol. 9, no. 5, pp. 1-6, October 2001.
- [6] D. Álvarez and M. E. Edwards-Schachter, "El teléfono móvil: una herramienta eficaz para el aprendizaje activo," *Current Developments in Technology-Assisted Education*, pp. 23+, 2006.
- [7] L. Naismith, P. Lonsdale, G. Vavoula, and M. Sharples, "Literature Review in Mobile Technologies and Learning, FutureLab Series 2009," University of Birmingham, Report 11, 2009.
- [8] M. Sharples, J. Taylor, and G. Vavoula, "A Theory of Learning for the Mobile Age," *Medienbildung in neuen Kulturräumen*, Bachmair, Ben., VS Verlag für Sozialwissenschaften, pp. 87-99, 2010.

Optimización en la Aproximación de Curvas NURBS con Algoritmos Genéticos Paralelos en un GPU

Leopoldo Noel Gaxiola Sánchez^a, Juan José Tapia Armenta^a, Julio César Rolón Garrido^a

^aCentro de Investigación y Desarrollo de Tecnología Digital del Instituto Politécnico Nacional
Av. del Parque 1310, Mesa de Otay, Tijuana, B.C., CP 22510

<http://fiad.ens.uabc.mx/>

^bCICESE, Laboratorio de Control, Departamento de Electrónica y Telecomunicaciones.
Carr. Ensenada-Tijuana 3918, Zona Playitas, Ensenada, B.C., México, 22860.

lgaxiolas1100@alumno.ipn.mx, jtapiaa@ipn.mx, jcrolon@ipn.mx

2013 Published by *DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>

Selection and peer-review under responsibility of the Organizing Committee of the CCOMP-2013, www.cicomp.org

Resumen

En este trabajo se desarrollan dos versiones de algoritmos genéticos paralelos, un algoritmo genético paralelo simple y un algoritmo genético con modelo de islas; los dos algoritmos se implementaron en un GPU. El algoritmo genético paralelo basado en el modelo de islas que se propone en este trabajo tiene la característica principal de que no se hace migración entre las islas, en su lugar se crea una isla élite con los mejores individuos de cada una de las islas para compartir los mejores individuos. El algoritmo se aplica a un problema de aproximación de curvas NURBS a un conjunto de puntos de una imagen médica, con este algoritmo, además de reducir el tiempo de ejecución se obtiene soluciones más cercanas al óptimo que con los resultados obtenidos con el algoritmo genético paralelo simple.

Palabras clave: Algoritmos genéticos paralelos, Curvas NURBS, GPU.

1. Introducción

Los algoritmos genéticos (AGs) son técnicas de búsqueda y optimización inspirados en el proceso de la evolución natural de las especies [1]. En este caso, un individuo es una solución potencial del problema, y su representación se define de acuerdo al problema a resolver. Al igual que ocurre en la naturaleza, los AGs se basan en la supervivencia de los individuos más aptos de una población. El AG inicia con

un conjunto de individuos que constituye una población que evoluciona de una generación a otra, a través de la creación de nuevos individuos con mejor función de aptitud y la eliminación en la población de individuos con baja función de aptitud.

En los AGs las poblaciones de individuos evolucionan mediante la aplicación de operadores genéticos, tales como selección, cruce y mutación, cuya funcionalidad e implementación depende del problema a resolver. Una de las cualidades principales de los algoritmos gené-

ticos es su facilidad de paralelización, ya que están basados en poblaciones con individuos independientes, de esta manera, el cálculo de la función de aptitud y de los operadores genéticos en un individuo no depende del cálculo en los demás individuos [2].

Típicamente, la población inicial se genera de manera aleatoria con distribución de probabilidad uniforme. El tamaño de la población inicial es importante porque influye en si el AG puede encontrar buenas soluciones y en el tiempo que se tarda en llegar a ellas. Si la población es demasiado pequeña, puede que no haya una muestra adecuada de individuos, y será difícil identificar buenas soluciones [2]. Si la población es demasiado grande, es necesario usar una gran cantidad de recursos de cómputo y el tiempo de procesamiento puede ser muy grande. En cada iteración del AG se crea una nueva generación de individuos en función de sus antecesores, teniendo más probabilidad de reproducirse los que tengan una mejor función aptitud.

Los AGs generalmente son capaces de aproximar la solución a un problema de optimización combinatoria en un tiempo razonable. Sin embargo, cuando se está cerca de la solución y se desea reducir el error de aproximación el tiempo de ejecución tiene un incremento muy grande. Como consecuencia, se han hecho múltiples esfuerzos para implementar AGs más rápidos, y una de las opciones más prometedoras es el uso de implementaciones paralelas, con lo que se puede obtener una reducción substancial del tiempo de procesamiento. En años recientes el desarrollo de procesadores gráficos más potentes ha tenido un fuerte impulso, como consecuencia es posible contar con plataformas de cómputo de alto rendimiento de bajo costo.

En este trabajo se resuelve un problema de optimización combinatoria por medio de algoritmos genéticos paralelos (AGPs). El problema consiste en encontrar la combinación de posiciones y coeficientes de los pesos de los puntos de control de una curva NURBS (del inglés, *Non-Uniform Rational B-spline*) que generarán la mejor aproximación posible a un conjunto de puntos que describen un contorno en imágenes médicas de resonancia magnética (puntos dato). La complejidad computacional de este problema es NP-Completo [3]. El espacio solución de este problema consiste en una gran cantidad de combinaciones de las posiciones y los coeficientes de los pesos de los puntos de control de la curva paramétrica, por lo tanto no es apropiado usar un algoritmo de búsqueda exhaustiva. Para este fin se desarrollaron dos versiones de algoritmos genéticos paralelos implementados en un GPU, un algoritmo genético paralelo simple y un algoritmo genético paralelo basado en el modelo de islas. Se hace una comparación

entre ellos y se muestran los resultados obtenidos.

2. Curva B-spline racional no uniforme (NURBS)

Una curva NURBS está completamente determinada por sus puntos de control, lo cual implica que la curva cambia de manera predecible conforme lo hacen sus puntos de control [4]. A esto se le conoce como la propiedad de soporte local y permite que el movimiento de un punto de control sólo afecte a la curva localmente. Una curva NURBS está definida como [5]:

$$C(t) = \frac{\sum_{i=1}^{n+1} w_i P_i N_{i,k}(t)}{\sum_{i=1}^{n+1} w_i N_{i,k}(t)} \quad a \leq t \leq b \quad (1)$$

donde:

w_i son factores de ponderación o pesos,
 P_i son los puntos de control,
 $N_{i,k}$ son las funciones base de orden k (grado $k - 1$).

Las funciones base son determinadas por la recurrencia:

$$N_{i,1}(t) = \begin{cases} 1 & \text{si } x_i \leq t < x_{i+1} \\ 0 & \text{en otro caso} \end{cases} \quad (2)$$

$$N_{i,k}(t) = \frac{t - x_i}{x_{i+k-1} - x_i} N_{i,k-1}(t) + \frac{x_{i+k} - t}{x_{i+k} - x_{i+1}} N_{i+1,k-1}(t) \quad (3)$$

donde el vector de nodos está determinado por $X = (x_0, \dots, x_{n+k+1})$ y satisface la relación $x_i \leq x_{i+1}$.

La parametrización tiene el objetivo de asociar las coordenadas (x, y) de los puntos dato a un parámetro de t en el dominio de la curva, es decir es una función de $\mathfrak{R}^2 \rightarrow \mathfrak{R}$. Si los puntos a ser aproximados son Q_0, \dots, Q_n , entonces se deben encontrar los parámetros τ_0, \dots, τ_n , uno para cada punto dato. En este trabajo se utiliza el método de parametrización de longitud de cuerda (del inglés, *chord length*) que se define por medio de la siguiente recurrencia [6]:

$$\tau_0 = 0 \quad (4)$$

$$\tau_i = \tau_{i-1} + \frac{\|Q_i - Q_{i-1}\|}{\sum_{i=1}^{n+1} \|Q_i - Q_{i-1}\|} \quad (5)$$

donde τ_{i+1} representa la suma de las distancias de los puntos dato anteriores y Q_i los puntos dato, por lo que cada valor τ_i representa un punto dato en el dominio de

la curva.

El vector de nodos se construye en base al método de parametrización de longitud de cuerda de la siguiente manera [4]:

$$x_{i+k-1} = \frac{i(\tau_{i+1} - \tau_i)}{n - k + 1} + \tau_i \quad 0 \leq i \leq k - 1 \quad (6)$$

3. Algoritmos genéticos paralelos

Los algoritmos genéticos paralelos surgen ante la necesidad de cómputo requerida por problemas de extrema complejidad cuyo tiempo de ejecución al utilizar los algoritmos genéticos secuenciales es una limitante [7]. Es por eso que se busca la manera de poder adaptar este tipo de heurísticas a distintas configuraciones de cómputo paralelo. La aplicación de los algoritmos genéticos paralelos tiene como finalidad dividir un problema en varios sub-problemas y resolverlos de manera simultánea en varios procesadores, lo cual permite mejorar el desempeño y la calidad de búsqueda de los algoritmos genéticos.

En general cuando se paraleliza un algoritmo su comportamiento es el mismo que el del algoritmo secuencial. Sin embargo este no es necesariamente el caso en la paralelización de algoritmos genéticos. En la estructura de un algoritmos genético paralelo una tarea puede dividirse en sub-tareas de forma que exista una distribución equilibrada de actividades; los miembros de las poblaciones de un algoritmo genético puede ser divididos en sub-poblaciones que son distribuidos en diferentes procesadores mediante mecanismos de comunicación y control que ayudarán a generar una estructura sólida a nivel de los algoritmos genéticos en un ambiente paralelo.

Existen varias formas de paralelizar un algoritmo genético. La primera y más intuitiva es la global, que consiste básicamente en paralelizar la evaluación de la función de aptitud de los individuos manteniendo una sola población. Una mejor opción de paralelización de los algoritmos genéticos es dividir la población en subpoblaciones que evolucionan por separado e intercambian individuos cada cierto número de generaciones [7]. A continuación se describen cada una de las etapas de un algoritmo genético.

3.1. Representación de los individuos

Una consideración importante que se debe tener al diseñar un AG es definir una representación para los individuos de la población que modelan la solución del

problema [1]. Los individuos son representados por su cromosoma, que contiene una cadena de genes que simbolizan los datos que queremos optimizar. Algunas formas de representar el cromosoma de cada individuo son, por medio de una cadena binaria, una lista de números. En este trabajo cada gen del cromosoma de los individuos contiene tres valores, dados por la posición (x, y) y el peso de cada punto de control de la curva NURBS.

3.2. Función de aptitud

La función de aptitud es la que determina el potencial de solución que tienen los individuos dentro de la población [1] En este trabajo la función de aptitud está determinada en base a la sumatoria de las distancias de cada punto dato y su correspondiente punto de la curva, entre menor sean estas distancias mejor será la solución que ofrece el individuo y está dada por:

$$\lambda = \sum_{i=1}^n \sqrt{(x_i + \hat{x}_i)^2 + (y_i + \hat{y}_i)^2} \quad (7)$$

donde (x_i, y_i) son los puntos de los datos y (\hat{x}_i, \hat{y}_i) son los puntos de la curva.

El problema de definir la función de aptitud debe considerarse cuidadosamente para que se pueda alcanzar una buena aproximación al problema planteado. Si se elige mal una función aptitud o se define de manera inexacta, puede que el AG sea incapaz de encontrar una solución al problema, o puede acabar resolviendo un problema equivocado.

3.3. Operador de selección

Por medio del operador de selección se escogen los individuos que aplicarán el operador de cruce. Los individuos seleccionados heredarán sus características a la siguiente población de individuos de las posibles soluciones [1]. El operador de selección no produce puntos nuevos en el espacio de búsqueda, sino que determina qué individuo dejará descendencia y en qué cantidad lo hará para la próxima generación. Existen varios métodos de selección que se utilizan en los algoritmos genéticos, como por ejemplo selección por ruleta, selección por torneo, selección proporcional.

En esta investigación se implementa selección por torneo en el algoritmo genético paralelo simple, en la cual se elige aleatoriamente una pequeña muestra de la población y de ella se selecciona el individuo con mejor valor en su función aptitud (todos los números aleatorios

que se utilizan en este trabajo se generan con una distribución uniforme). Para el algoritmo genético paralelo con modelo de islas se selecciona un individuo de manera aleatoria de la misma isla o de la isla élite, ya que el tener una isla élite se tiene una mayor probabilidad de seleccionar un individuo de mayor función aptitud, este tipo de cruce se propone en este trabajo debido a la ventaja que le da el tener la isla élite al algoritmo. En ambos caso solamente se selecciona un individuo para ser cruza ya que el otro individuo está definido por el hilo ya que en ambos algoritmos cada individuo esta mapeado a un hilo en el GPU.

3.4. Operador de cruce

Para mejorar los individuos de la población se utiliza el operador genético de cruce. Para llevar a cabo la operación de cruce, se seleccionan dos individuos para ser combinados y el individuo resultante tiene la posibilidad de reemplazar a uno de los padres o al individuo con menor aptitud en la población [1]. El nuevo individuo va a sustituir a uno de los padres siempre y cuando tenga mejor función aptitud. En caso de que el nuevo individuo no sea mejor que los padres se puede intentar sustituirlo por el peor individuo de la población. El proceso está orientado a las sub-regiones del espacio de búsqueda, donde se supone que existe una solución óptima. Existen varias formas de aplicar el operador de cruce tales como cruce por un único punto, cruce por puntos múltiples, cruce uniforme, entre otras.

En este trabajo cada individuo de la población hace cruce con otro individuo escogido por el operador de selección en cada iteración y se utiliza el cruce por dos puntos, en donde dos posiciones en la cadena de genes de los cromosomas de ambos padres se seleccionan aleatoriamente. Uno de los padres aporta la información que queda fuera del rango entre las dos posiciones y el otro proporciona la información que queda dentro de las dos posiciones.

3.5. Operador de mutación

El operador de mutación crea un individuo realizando algún tipo de alteración, usualmente pequeña, en un individuo de la población escogido de forma aleatoria con una probabilidad pequeña [1]. La mutación tiene el objetivo de dispersar la búsqueda del algoritmo ya que da origen a individuos con material genético nuevo. En este trabajo, en el proceso de mutación algunos genes del cromosoma del individuo se seleccionan aleatoriamente para ser reemplazados por otros valores, que son seleccionados de manera aleatoria dentro del rango de

las soluciones posibles.

4. Unidades de procesamiento gráfico

La Unidad de Procesamiento Gráfico (GPU, del inglés *Graphics Processing Unit*) es un dispositivo de cómputo capaz de ejecutar una gran cantidad de procesos en paralelo, a través del uso de hilos (*threads*, en inglés). La GPU funciona como un coprocesador en una computadora de propósito general cuyo procesador central es un CPU (del inglés *Central Processing Unit*) [8].

El modelo de programación utilizado en los procesadores gráficos según la taxonomía de Flynn es el modelo SIMD (del inglés, *Single Instruction Multiple Data*), en este modelo todas las unidades de procesamiento ejecutan la misma instrucción y operan sobre diferentes datos. Esto reduce la complejidad en el diseño de la unidad de control y permite que se dedique un mayor espacio en el procesador gráfico para el procesamiento de datos.

La ventaja de utilizar un procesador gráfico en lugar de un procesador de propósito general consiste en que, en aquellos casos en los que es posible paralelizar las operaciones que se efectúan sobre un conjunto de datos, la GPU supera ampliamente a la CPU [9].

La función destinada para ejecutarse en la GPU se llama *kernel*. Cuando una función *kernel* se invoca desde el CPU, se indica el número de bloques e hilos que van a ejecutarla [8]. Cada bloque de hilos se ejecuta por los núcleos de un multiprocesador. Dependiendo de la cantidad de multiprocesadores de la GPU, el *kernel* puede ser completado en una, o varias re-distribuciones de bloques de hilos a los multiprocesadores.

5. Implementación de algoritmos genéticos paralelos en una GPU

En esta sección se explica cómo se implementa un algoritmo genético paralelo en una GPU. Por lo que se describe a detalle las dos versiones de algoritmos genéticos paralelos que se implementaron. Es importante resaltar que la cantidad de hilos es igual a la cantidad de individuos, ya que cada individuo de la población se mapea en un hilo del GPU.

5.1. Algoritmo genético paralelo simple

En el algoritmo genético paralelo simple (AGPS), sólo hay una población, en la que la evaluación de los individuos y los operadores genéticos se paralelizan de forma explícita [2]. Puesto que sólo hay una población,

Tabla 1. **Algoritmo 1.** Seudocódigo para el AGPS

```

main( ) {
1. asignación de memoria en el CPU y en el GPU
2. calcular parametrización de los puntos dato
3. copia de memoria del CPU a GPU
4. hilos = NUM_IND
5. bloques = NUM_ISLAS
6. GenPoblIni <<<bloques,hilos>>>(parametros)
7. while condición_terminar == FALSO do
8.   AGPS<<<bloques,hilos>>>(parametros)
9. end while
10. copia de memoria de población del GPU a CPU
11. obtener mejor solución
12. liberar memoria del GPU y CPU
}
__global__ GenPoblIni( ){
13. GenPob(P)
14. Evaluar(P)
}
__global__ AGPS( ){
15. seleccionar al mejor individuo
16. SelecPadres(P)
17. P' ← CrucePadres(P)
18. if prob_mut <5% then
19.   P' ← Mutacion(P)
20. end if
21. Evaluar(P')
22. if Aptitud(P') < Aptitud(P) then
23.   P = P'
24. end if
}

```

la selección considera a todos los individuos y cada individuo tiene oportunidad de aparearse con cualquier otro.

El AGPS es un método relativamente fácil de implementar y puede obtenerse una mejora significativa con respecto a la versión secuencial, si los costos de comunicación no dominan a los costos de procesamiento. Al algoritmo genético paralelo simple también se le conoce como algoritmo genético paralelo panmítico, pues se cuenta con un solo depósito de material genético (*gene pool*), es decir, con una sola población.

La implementación del AGPS en la GPU se describe en el pseudocódigo del Algoritmo 1. El primer paso es asignar memoria en el CPU y en la GPU, luego se calcula en el CPU la parametrización de los puntos de los datos para llevarlos al dominio de la curva, después se copia de la memoria del CPU a la memoria del GPU los valores de la parametrización y los demás datos que se necesitan.

Enseguida se llama al *kernel GenPoblIni* donde cada hilo genera el individuo que va a procesar, esto es, para cada punto dato se generan las coordenadas de la posición y los pesos de su correspondiente punto de control. Las coordenadas de la posición se obtienen de manera aleatoria dentro de un círculo con centro en el

punto dato y un radio de 10 unidades de distancia. Se propone el valor del peso como un número aleatorio con distribución uniforme en un intervalo de (0,1) con incremento de 0.1 que se asigna a cada punto de control. Al final en este *kernel* se evalúa la función aptitud para cada individuo, la cual es la sumatoria de la distancia que existe entre los puntos de la curva y los puntos dato. Para evaluar la función aptitud es necesario generar la curva NURBS que se define por los puntos de control y los pesos de cada individuo.

Los individuos generados en el *kernel GenPoblIni* son almacenados en una variable llamada población vieja P , donde cada hilo tiene espacio para guardar las posiciones (x, y) y los pesos w de los puntos de control. Después entra en un ciclo que procesa el *kernel AGPS*, en general el criterio de paro está determinado por el número de iteraciones o alcanzar un cierto valor de la función aptitud, en este trabajo se eligió que termine cuando se alcanza un cierto número de iteraciones.

El *kernel AGPS* lo primero que hace es seleccionar al mejor individuo de toda la población con el fin de mantenerlo y no sustituirlo si no es por otro mejor, para ello se aplica un algoritmo de reducción [8], luego se selecciona un individuo con el que se hará cruce por medio del operador de selección de torneo, el cual compara dos individuos y selecciona el que tenga mayor función aptitud. Como se puede apreciar, todos los hilos entran a hacer el cruce en cada iteración. Luego se aplica el operador de cruce entre los dos padres. Posteriormente se aplica el operador de mutación, cada individuo tiene una probabilidad de 5% de ser mutado. En la mutación cada uno de los genes del individuo que se alterarán es cambiado por un punto aleatorio dentro de un círculo de radio 5 y para el peso se sigue con las mismas consideraciones que cuando se genera la población inicial.

El individuo generado por cruce se guarda en la variable P' que representa la población nueva, donde cada hilo tiene espacio para guardar las posiciones (x, y) y los pesos w de los puntos de control del individuo generado. En caso de que el hilo entre también a hacer mutación, el individuo generado entra directamente en la población vieja sin reemplazar al mejor individuo.

Al final, en el *kernel AGPS*, a los individuos generados se les evalúa su función aptitud y en caso de que el individuo de la población nueva tenga una mejor aptitud que el individuo de la población vieja, se sustituye. Cuando las iteraciones terminan se copia la población vieja al CPU. Luego se obtiene la mejor solución generada por el AG y finalmente se liberan las memorias del CPU y el GPU.

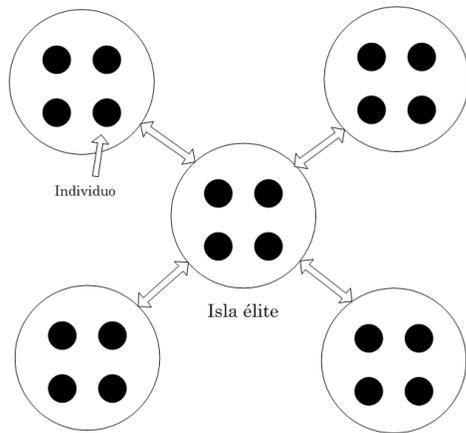


Figura 1. Organización de los individuos de la población del AGPI.

5.2. Algoritmo genético paralelo con modelo de islas

En el algoritmo genético paralelo con modelo de islas (AGPI) la población se divide en subconjuntos, también conocidos como poblaciones locales (en inglés se les conoce como *demes*), que evolucionan de manera aislada, aunque intercambian individuos con la frecuencia que se indica en el algoritmo [2]. A este intercambio de individuos se le llama migración, y se considera como un nuevo operador genético. El AGPI que se propone en este trabajo tiene una modificación, en lugar de migrar los individuos entre las islas, se construye una isla con los mejores individuos de cada una de las subpoblaciones, denominada isla élite. Esta isla de mejores individuos es compartida con todas las demás islas. En la Fig. 1 se puede observar la organización de los individuos de la población del AGPI.

La implementación del AGPI en la GPU es muy similar a la manera que se implementa el AGPS. La diferencia en este caso es que como se manejan subpoblaciones aisladas es necesario mapear las islas en los bloques de hilos, por lo que cada bloque representa una isla. En el Algoritmo 2 se presenta el pseudocódigo para el AGPI, el proceso es el mismo que en el algoritmo del AGPS hasta la ejecución del *kernel GenPoblNi*.

En el *kernel GenPoblNi* los individuos se crean de la misma manera como se crearon en el AGPS, con la diferencia de que después de la evaluación de la función aptitud se selecciona el mejor individuo de cada isla y se migra a la isla élite, para esto fue necesario asignar dos variables en memoria compartida un arreglo para guardar todas las aptitudes de la isla para utilizarlo en el algoritmo de reducción y una variable para guardar el identificador del hilo del mejor individuo, con el fin de que solo los individuos de una isla puedan acceder a

Tabla 2. Algoritmo 2. Seudocódigo para el AGPI

```

main( ){
1. asignación de memoria en el CPU y en el GPU
2. calcular parametrización de los puntos dato
3. copia de memoria del CPU a GPU
4. hilos = NUM_IND
5. bloques = NUM_ISLAS
6. GenPoblNi<<<bloques,hilos>>>(parametros)
7. while condición_terminar == FALSO do
8.  AGPI<<<bloques,hilos>>>(parametros)
9. end while
10. copia de memoria de isla élite del GPU a CPU
11. obtener mejor solución
12. liberar memoria del GPU y CPU
}
__global__ GenPoblNi( ){
13. GenPob(I)
14. Evaluar(I)
15. seleccionar al mejor individuo de cada isla
16. MigrarIndElite(E)
}
__global__ AGPI( ){
17. SelecPadres(I,E)
18. I'← CrucePadres(I,E)
19. Evaluar(I')
20. if Aptitud(I')<Aptitud(I) then
21.  I = I'
22. end if
23. if prob_mut < 5% then
24.  I'← Mutacion(I)
25. end if
26. seleccionar al mejor individuo de cada isla
27. MigrarIndElite(E)
}
    
```

estas variables y mantener las islas aisladas.

Los individuos generados en el *kernel GenPoblNi* son almacenados en la variable I que representa la población vieja. Además hay una variable E llamada élite, en donde hay un espacio asignado para guardar el mejor individuo de cada isla. Después se entra en un ciclo que procesa al *kernel AGPI*, que termina cuando se alcanza un cierto número de iteraciones. Al momento de hacer el cruce solo se selecciona aleatoriamente un individuo de la misma isla I o de la isla élite E . Después, a los individuos generados se les evalúa con la función de aptitud. En caso de que el individuo de la población nueva I' tenga mejor aptitud que el individuo de la población vieja I , se sustituye.

Luego se aplica la mutación de manera muy similar a la manera que se aplicó en AGPS, con la diferencia de que ahora los individuos generados tienen la excepción de reemplazar al mejor individuo de la isla. Al final del *kernel AGPI* se selecciona el mejor individuo de cada isla y se migra cada uno a la isla élite. Cuando las iteraciones terminan sólo se copia la isla élite al CPU. Luego se obtiene la mejor solución generada por el AG y finalmente se liberan las memorias del CPU y el GPU.

6. Resultados

La implementación del algoritmo de aproximación se realiza en el sistema operativo GNU/Linux y lenguaje CUDA C en una computadora con un CPU de intel i7 de 2.8 GHz. El algoritmo se procesa en un GPU GeForce GTX 670.

Se presentan algunos experimentos para optimizar la aproximación de una curva NURBS a un conjunto de puntos. En los experimentos se utilizaron imágenes de resonancia magnética de diferentes cortes, tomadas de [10]. Las imágenes tienen un tamaño de 217×181 píxeles con 8 bits de profundidad. Para estos experimentos se utilizó el algoritmo de aproximación con optimización de parámetros por el algoritmo genético paralelo con modelos de islas. La entrada y la salida de las imágenes se manejan por medio de la biblioteca de funciones para visión por computadora OpenCV, acoplada con CUDA.

Los puntos que se obtienen de la imagen son generados por el algoritmo de segmentación que se desarrolló en el trabajo de tesis [11], el cual emplea el método de modelos deformables.

En todos los experimentos se usaron 512 hilos por bloque, un porcentaje de mutación del 5%, a cada individuo seleccionado para la mutación se le muta el 30% de los genes y las curvas NURBS usadas son de orden 3 (grado 2). El tiempo (σ) es el tiempo que tarda en ejecutarse el programa completo y el error de

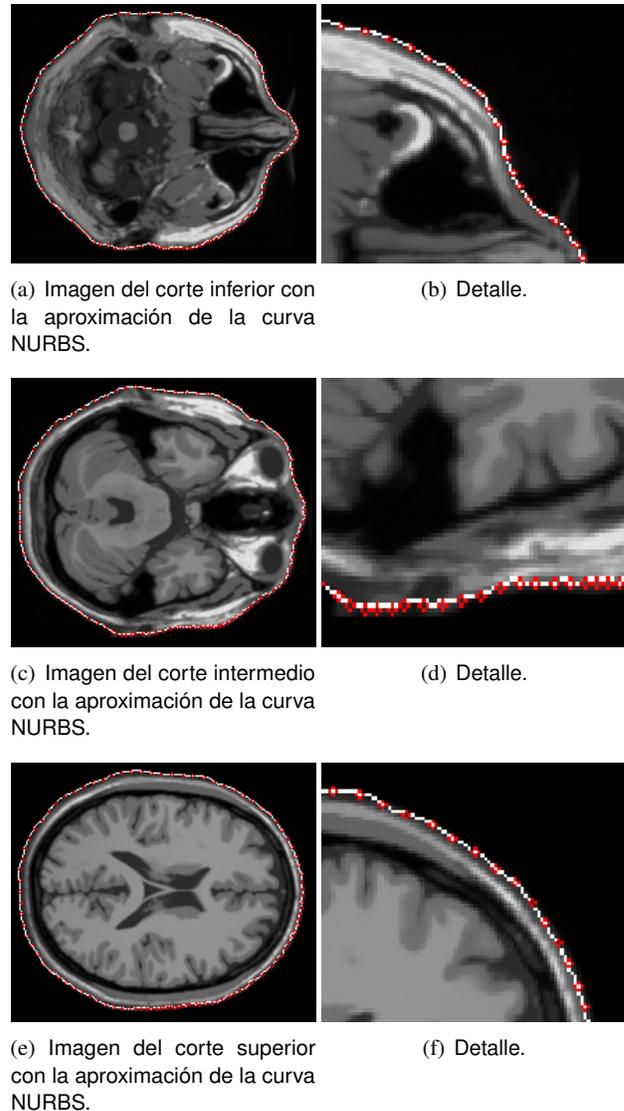


Figura 2. Resultado de la aproximación.

la aproximación (ϵ) es el valor de la función aptitud del mejor individuo que proporciona el algoritmo genético (el mejor individuo es la solución que presenta la distancia mínima entre los puntos a aproximar y la curva NURBS). Ver Ec. (7).

Para los experimentos de la Fig. 2, se utilizaron imágenes de resonancia magnética. En estos tres experimentos los parámetros que se utilizaron fueron 1200 iteraciones, el número de puntos que describen el contorno exterior de la imagen médica es de 100 y el número de bloques de hilos igual a 80. La imagen de la Fig. 2(a) es el corte inferior, la imagen de la Fig. 2(c) es el corte intermedio y la imagen de la Fig. 2(e) es el corte superior. En las Figs. 2(a), 2(c) y 2(e) se observa la aproximación final que se obtuvo y en las Figs. 2(b), 2(d) y 2(f) se observa una ampliación de una parte de la

imagen para mostrar la forma en que las curvas NURBS se ajustan a los puntos dato.

En el experimento de la Fig. 2(a) los resultados de la aproximación que se obtuvieron fueron $\epsilon = 10.89$ y $\sigma = 1959.30 s$. En la Fig. 2(b) se observa que la curva pasa exactamente por los puntos de dato de esa sección, ya que describen una curvatura suave que puede ser descrita por el grado de ese tramo de la curva.

Los resultados de la aproximación obtenidos para la Fig. 2(c) fueron $\epsilon = 10.76$ y $\sigma = 1960.65 s$. En la Fig. 2(d) se observa que la curva no pasa exactamente por los puntos dato esto se debe a que el contorno es muy irregular en esa área por lo que el grado de la curva que se utiliza no es suficiente para interpolarlos, esta es una limitación del algoritmo de aproximación que se desarrolló.

En el experimento de la Fig. 2(e) los resultados de la aproximación que se obtuvieron fueron $\epsilon = 12.90$ y $\sigma = 1961.22 s$. En la Fig. 2(f) se puede ver que la curva pasa exactamente por los puntos de datos, ya que en esa sección el contorno es suave y puede ser descrito por ese tramo de la curva.

6.1. Comparación de algoritmos genéticos paralelos

Se hace una comparación entre las implementaciones del algoritmo genético paralelo simple (AGPS) y el algoritmo genético paralelo con modelo de islas (AGPI). Se utilizaron los puntos de la imagen de resonancia magnética del corte inferior Fig. 2(a). Los parámetros definidos que se utilizaron para ambos algoritmos son porcentaje de mutación = 5 %, hilos por bloque = 512, cantidad de puntos a mutar de cada individuo = 30 %, número de puntos de la imagen médica = 100 y orden de la curva = 3, es necesario aclarar que en el caso del AGPS se procesa una sola población es decir 512 individuos equivalentes a la población de una isla.

Los resultados de los experimentos del AGPI se pueden ver en la Tabla 3 y los resultados del AGPS se pueden ver en la Tabla 4. Para los experimentos del AGPI los parámetros que se varían el número de iteraciones y el número de isla, mientras que para los experimentos del AGPS sólo se varía el número de iteraciones ya que la población es constante para este caso. Cada uno de los resultados mostrados en la tabla es el resultado de la media de 10 pruebas realizadas. Se observa que el AGPI obtiene mejores soluciones de aproximación que el AGPS esto se debe a la ventaja que ofrece el tener una isla élite para compartir los mejores individuos entre las islas, además de que se tienen varias islas que

Tabla 3. Algoritmo genético paralelo con modelo de islas (AGPI) en el GPU.

Iteraciones	Número de islas			
		40	60	80
300	ϵ	25.25	25.23	24.81
	$\sigma (s)$	248.12	384.89	490.23
600	ϵ	18.64	18.05	16.27
	$\sigma (s)$	490.86	768.83	977.27
900	ϵ	15.25	15.1	13.93
	$\sigma (s)$	739.54	1151.47	1464.34
1200	ϵ	13.74	12.86	11.56
	$\sigma (s)$	986.87	1534.31	1954.4
1500	ϵ	13.37	12.16	11.45
	$\sigma (s)$	1237.53	1919.61	2442.79

Tabla 4. Algoritmo genético paralelo simple (AGPS) en el GPU.

	Número de iteraciones			
	10000	20000	40000	60000
ϵ	25.15	16.44	14.07	12.44
$\sigma (s)$	832.72	1645.55	3593.67	4940.27

evolucionan al mismo tiempo por lo que tiene un campo de exploración mucho mayor y se necesita un tiempo menor para encontrar una solución mejor. Al tener una sola población se hace necesario que se incremente el número de iteraciones para obtener el mejor resultado provocando que el tiempo necesario para encontrar una solución sea muy grande y el tener una sola población produce que el campo de exploración no sea tan grande como el caso del AGPI.

7. Conclusiones

En este trabajo se implementaron algoritmos genéticos paralelos en un GPU para la optimización de la aproximación de una curva NURBS a un conjunto de puntos de una imagen médica. Se desarrollaron dos versiones de algoritmos genéticos paralelos, un algoritmo genético paralelo simple y un algoritmo genético paralelo con modelo de islas.

El algoritmo genético paralelo con modelo de islas da mejores resultados que el algoritmo paralelo simple, debido a que da una mejor aproximación y lo hace en la mitad del tiempo del que le toma al algoritmo genético paralelo simple. Esta ventaja se la proporciona el tener la isla élite que ayuda al programa a dar mejores resultados ya que se comparten los individuos más aptos entre todas las islas, y además ayuda a hacer más rápida la migración ya que de otra manera se tendría que hacer migración entre las islas. También, al evolucionar simultáneamente varias islas de individuos contribuyen a que el campo de exploración de soluciones sea mayor.

Al implementar el algoritmo genético paralelo con modelo de islas en el GPU se aprovecha su arquitectura al mapear las islas y los individuos en los bloques de hilos y los hilos, respectivamente. También la manera en la que se lleva a cabo la selección para hacer el cruzamiento es distinta a como se hace comúnmente en los algoritmos genéticos secuenciales, ya que al mapear un individuo a cada uno de los hilos se tiene la seguridad de que ese individuo entra al operador de cruce y el otro individuo se escoge de manera aleatoria, esta forma de seleccionar al otro individuo se debe a la característica esencial de este algoritmo que es la isla élite que aporta una mayor cantidad de mejores individuos dentro de la población. Además, en este algoritmo se asegura que en cada iteración se crea una generación completa, porque todos los individuos hacen cruzamiento en cada iteración.

Agradecimientos

Se agradece el apoyo recibido por parte del proyecto SIP-20130489.

Referencias

- [1] S. N. Sivanandam and S. N. Deepa, *Introduction to Genetic Algorithms*, Springer, 2008.
- [2] G. Luque and E. Alba, *Parallel Genetic Algorithm: Theory and Real World Applications*. Springer, 2011.
- [3] T. H. Cormen *et al*, *Introduction to Algorithms*, Third edition, MIT Press, 2009.
- [4] D. F. Rogers, *An Introduction to NURBS*, Morgan Kaufmann Publishers, 2001.
- [5] L. Piegl and W. Tiller, *The NURBS Book*, Springer, 1997.
- [6] J. J. Fang and C. L. Hung, "An improved parametrization method for B-spline curve and surface interpolation," *Comput.-Aided Des.*, Vol. 45, No. 6, pp. 1005-1028, 2013.
- [7] Erick Cantú-Paz. *Efficient and Accurate Parallel Genetic Algorithm*, Kluwer Academic Publishers, 2000.
- [8] J. Sanders and E. Kandrot, *CUDA by Example: An Introduction to General-Purpose GPU Programming*, Addison-Wesley Professional, 2010.
- [9] D. B. Kirk and W. W. Hwu, *Programming Massively Parallel Processors: A Hands-on Approach*, Morgan Kaufmann, 2010.
- [10] C. Cocosco *et al.*, "BrainWeb: Online Interface to a 3D MRI Simulated Brain Database," *Third Intl. Conf. on Functional Mapping of the Human Brain*, Vol. 5, No. 4, Copenhagen, Denmark, May 1997, p. S425.
- [11] R. Alvarado, "Procesamiento de imágenes aplicando modelos deformables con computo de alto rendimiento," Tesis de maestría, Centro de Investigación y Desarrollo de Tecnología Digital del IPN, 2012.

Estimación de la Función de Iluminación No Uniforme para la Restauración de Imágenes en un Procesador Gráfico

Kenia Picos Espinoza^a, Víctor Hugo Díaz Ramírez^a, Juan José Tapia Armenta^a

^aInstituto Politécnico Nacional - CITEDI
Ave. del Parque 1310, Mesa de Otay, Tijuana, B.C., 22510, México
kpicos@citedi.mx vdiazr@ipn.mx, jtapiaa@ipn.mx

2013 Published by *DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CCOMP-2013, www.cicomp.org

Resumen

Un sistema en tiempo real es propuesto para restaurar imágenes degradadas por iluminación y ruido mediante los parámetros estadísticos de la escena de entrada. Se diseña un sistema de restauración de imágenes mediante el diseño de filtros, implementado en una unidad de procesamiento gráfico. El sistema es capaz de restaurar la escena de entrada que consiste en una imagen afectada por degradaciones en iluminación no uniforme y ruido aditivo con distribución normal. El filtro diseñado es capaz de adaptarse a las condiciones de la escena de entrada para estimar los parámetros estadísticos de la función de la fuente de iluminación. Los resultados obtenidos en simulación en CPU/GPU con el sistema propuesto se presentan en términos de desempeño de restauración de imágenes, complejidad computacional y tolerancia a iluminación de la escena de entrada y ruido aditivo.

Palabras clave: Cómputo paralelo, Iluminación no uniforme, Procesador gráfico, Procesamiento de imágenes.

1. Introducción

En años recientes, el reconocimiento de objetos ha recibido mayor atención debido a la necesidad de desempeñar sistemas de procesamiento de imágenes que mejoren las actividades de vigilancia, navegación en vehículos, y seguimiento de objetos, entre otros. El reconocimiento de objetos consiste básicamente de dos características: (a) identificación del objeto dentro de una escena observada, y (b) estimación de

las coordenadas exactas del objeto. El reconocimiento de objetos basado en filtros por correlación ha probado ser una técnica efectiva en la detección y localización de objetos [1]. En este enfoque, la detección se lleva a cabo buscando los mejores niveles de la correlación en el sistema de salida, brindando una localización medianamente la estimación de la posición de un objeto dentro de una escena de entrada. Sin embargo, estos sistemas son muy sensibles a distorsiones y ruido [1]. La respuesta de los sensores (por ejemplo, cámaras o el ojo

humano) es proporcional a la radiación incidente sobre ellos, donde la constante de proporcionalidad depende de la geometría del sensor [2]. Los algoritmos de iluminación global calculan la cantidad de radiación estimada que incide sobre un objeto. La elección de un algoritmo adecuado es importante debido a que este afectará la exactitud de la restauración resultante de la imagen. Dichos algoritmos pueden presentar distintos enfoques tales como: (a) orientados por pixel u orientados por escena, (b) superficie difusa o superficie especular, (c) integración determinística o integración por Monte Carlo, entre otras [2]. Las distorsiones a la respuesta de los sensores pueden afectar la escena de entrada por degradaciones debido a iluminación no uniforme y ruido aditivo. Dado a que se asume que estas degradaciones son procesos no estacionarios, es necesario mejorar el proceso de filtrado con el fin de adaptar sus parámetros para estimar ruido aditivo y condiciones de iluminación mediante los parámetros estadísticos de la escena de entrada.

En el área de procesamiento de imágenes, la necesidad de aplicar los algoritmos en un sistema de tiempo real se muestra cada vez más evidente. Un sistema de tiempo real, requiere precisión en sus cálculos y que se produzcan dentro de un periodo de tiempo especificado.

Se ha vuelto indispensable implementar algoritmos de procesamiento de imágenes con un enfoque en paralelo, debido a la creciente popularidad del diseño de los procesadores modernos con las estructuras jerárquicas de memoria. Se ha demostrado que esta técnica de procesamiento de imágenes puede ser ejecutada para plataformas con estructuras de memoria compartida, produciendo ventajas de rendimiento significativas [3]. El trabajo actual se centra en mejorar el rendimiento de la estimación de ruido e iluminación, particularmente la porción no estacionaria de la aplicación GPU.

En este proyecto, los resultados son obtenidos en simulación por computadora empleando un procesador gráfico (GPU) para el sistema propuesto. La aplicación de un sistema CPU/GPU es utilizado para obtener un desempeño alto y eficiente en la implementación, obteniendo un cumplimiento de las restricciones temporales para el tiempo real. Estos resultados se presentan y discuten en términos de desempeño de restauración de imágenes complejidad computacional, tolerancia al ruido aditivo y tolerancia a iluminación no uniforme.

El trabajo se presenta con la siguiente organización. En la sección 2 se presenta el estudio del modelo de iluminación no uniforme. En la sección 3 se presenta el análisis propuesto para la estimación de los coeficientes de normalización de la función de iluminación. En la sección 4 se presenta la metodología realizada para su

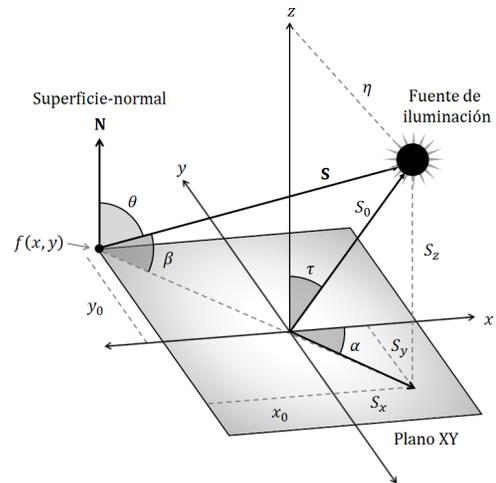


Figura 1. Geometría del modelo de iluminación en un plano $x - y$

implementación. Por último, en las secciones 5 y 6, se presentan los resultados obtenidos, y las conclusiones, respectivamente.

2. Modelo de iluminación no uniforme

Los modelos de reflectancia describen la relación que existe entre la dirección de iluminación y la forma de la superficie. Las superficies pueden ser Lambertiana, especular o híbrida [1].

La Fig. 1 muestra un modelo geométrico de una superficie que está iluminada por una fuente de luz, en el cual, el eje z es el punto de observación.

En el modelo Lambertiano, la superficie refleja la luz en todas las direcciones con amplitudes iguales a $d_L = \cos(\theta)$, donde θ es el ángulo incidente, por ejemplo, el ángulo entre el vector \mathbf{N} normal a la superficie y el vector de la dirección de iluminación \mathbf{S} .

Las superficies del tipo especular incluyen las reflexiones de espejo, refracciones, y cáustica [4]. Esta superficie puede describirse como $d_S = \delta(\theta_S - \theta)$, donde $\delta(x)$ es la función delta de Dirac, $\theta_S = \theta_N - \theta_v$, θ_N es el ángulo entre \mathbf{N} y el eje z , y θ_v es el ángulo entre \mathbf{N} y la dirección del observador.

Las superficies híbridas son combinaciones de componentes de superficies de tipo Lambertianas y especulares expresadas como $d_H = k_1 d_L + k_2 d_S$, donde $k_1 + k_2 = 1$.

En la Fig. 1 se observa que las direcciones de la iluminación se determinan por los parámetros τ , α y ρ , los cuales indican, el ángulo de rotación, el ángulo de inclinación y la magnitud del vector s_0 , respectivamente.

Estos parámetros definen la posición de la fuente de iluminación con respecto a su superficie de origen [1].

Para una superficie Lambertiana, la luz reflejada desde un punto $\mathbf{x} = (x, y)$ está dada por:

$$d(x, y) = \cos(\theta) = \cos\left(\frac{\pi}{2} - \beta\right) \quad (1)$$

donde β [1] es:

$$\beta = \arctan\left[\frac{s_z}{\left[(s_x - x)^2 + (s_y - y)^2\right]^{1/2}}\right] \quad (2)$$

Sustituyendo la Ec. (2) en la Ec. (1), la luz reflejada está dada por:

$$d(x, y) = \cos\left(\frac{\pi}{2} - \arctan\left[\frac{s_z}{\left[(s_x - x)^2 + (s_y - y)^2\right]^{1/2}}\right]\right) \quad (3)$$

donde los términos s_x , s_y y s_z se definen por:

$$s_x = \rho \tan \tau \cos \alpha \quad (4)$$

$$s_y = \rho \tan \tau \sin \alpha \quad (5)$$

$$s_z = \frac{\rho}{\cos \tau} \quad (6)$$

Notemos que $d(x, y)$ depende de los parámetros de la dirección de la iluminación τ , α y ρ , indicando el ángulo de inclinación, el ángulo de rotación, y la magnitud del vector s_0 , respectivamente. Estos parámetros pueden conocerse mediante un proceso de estimación para aplicaciones con escenas reales.

La función de iluminación de la Ec. (3) puede degradar la escena de entrada mediante el siguiente modelo de señal [5]:

$$f(\mathbf{x}) = s(\mathbf{x})d(\mathbf{x}) + n\mathbf{x} \quad (7)$$

donde $d(\mathbf{x})$ es el modelo de es el modelo de iluminación construido en el dominio espacial \mathbf{x} que representa dos dimensiones.

3. Estimación de coeficientes de la función de iluminación no uniforme

De manera general, analizar imágenes capturada por un sensor optodigital, presenta sensibilidad en la respuesta cuando las escenas de entrada contienen iluminación no uniforme. Para ello, se implementa un análisis de aproximación de la función de iluminación que degrada la escena.

Consideremos que la región de soporte $w_t(x)$ es suficientemente pequeña, y que en ella se encuentra una

porción de la señal de entrada tal que se puede considerar uniformemente iluminada. Ya que la iluminación de la escena de entrada es desconocida, consideremos el siguiente modelo de señal [6]:

$$\hat{f}(\mathbf{x}) = a_1(\mathbf{x})f(\mathbf{x}) + a_0\mathbf{x} \quad (8)$$

donde a_0 y a_1 son coeficientes de normalización, los cuales son considerados para aproximar la función de iluminación desconocida de la escena de entrada dentro de la región de soporte de la imagen del objeto w_t alrededor del k -ésimo pixel. Estos coeficientes se pueden obtener resolviendo un sistema de dos ecuaciones y dos incógnitas.

Minimizando el error cuadrático promedio (MSE) se obtiene:

$$MSE(k) = \sum_{\mathbf{x} \in w_t} [a_1(k) s(k + \mathbf{x}) + a_0(k) - t(\mathbf{x})]^2 \quad (9)$$

Entonces, resolviendo para $\frac{\partial}{\partial a_1} MSE(k) = 0$, la expresión se obtiene como:

$$a_1(k) = \frac{\sum_{\mathbf{x} \in w_t} t(\mathbf{x}) s(k + \mathbf{x}) - a_0(k)}{\sum_{\mathbf{x} \in w_t} s^2(k + \mathbf{x})} \quad (10)$$

De igual manera, resolviendo la expresión mediante $\frac{\partial}{\partial a_0} MSE(k) = 0$, obtenemos:

$$a_0(k) = \frac{1}{|w_t|} \sum_{\mathbf{x} \in w_t} t(\mathbf{x}) - a_1(k) \frac{1}{|w_t|} \sum_{\mathbf{x} \in w_t} s(k + \mathbf{x}) \quad (11)$$

donde $|w_t|$ es la suma de los elementos de $s(\mathbf{x})$.

Sustituyendo la Ec. (11) en (10), el coeficiente $a_1(k)$ se expresa como:

$$a_1(k) = \frac{\sum_{\mathbf{x} \in w_t} t(\mathbf{x}) s(k + \mathbf{x}) - \bar{t} \sum_{\mathbf{x} \in w_t} s(k + \mathbf{x})}{\sum_{\mathbf{x} \in w_t} s^2(k + \mathbf{x}) - \bar{s}(k) \sum_{\mathbf{x} \in w_t} t(\mathbf{x}) s(k + \mathbf{x})} \quad (12)$$

Simplificando las Ec. (11) y (12), los coeficientes se pueden estimar explícitamente mediante [6]:

$$a_1(k) = \frac{\sum_{\mathbf{x} \in w_t} t(\mathbf{x}) s(k + \mathbf{x}) - |w_t| \bar{t} \bar{s}(k)}{\sum_{\mathbf{x} \in w_t} [s(k + \mathbf{x})]^2 - |w_t| [\bar{s}(k)]^2} \quad (13)$$

$$a_0(k) = \bar{t} - a_1(k) \bar{s}(k) \quad (14)$$

donde \bar{t} y $\bar{s}(k)$ son los promedios del objeto y la ventana local sobre w_t en la k -ésima posición de la ventana, respectivamente.

4. Implementación

El procesamiento de imágenes en tiempo real y el cálculo de gráficos de alta definición requieren una gran cantidad de recursos de cómputo. Para satisfacer esta demanda de cómputo se usa una unidad de procesamiento gráfico (GPU) ya que presenta ventajas para la implementación por la capacidad de ejecución multiproceso altamente en paralelo, con gran poder de cómputo y ancho de banda de memoria muy alta [7].

Con la creciente capacidad de programación de los GPU, los cuales son capaces de realizar cálculos mediante operaciones en paralelo para grandes cantidades de datos, específicamente utilizados para gráficos, para los cuales han sido diseñados. En la actualidad hay procesadores potentes, que alcanzan una alta velocidad que pueden procesar distintas aplicaciones que se pueden expresar como en el cálculo de datos paralelos [8].

La arquitectura CUDA, por sus siglas en inglés, *Compute Unified Device Architecture*, es aplicada en hardware y software para la emisión y administración de operaciones en GPU como un dispositivo de computación de datos en paralelo sin la necesidad de asignar a una API de gráficos. La API de CUDA comprende una extensión al lenguaje de programación C, por lo tanto, con la ayuda de CUDA, programar un GPU es mucho más accesible [8]-[9].

Se implementa un algoritmo para la reconstrucción de una escena degradada por iluminación. La implementación se ha desarrollado en una computadora CPU con sistema operativo Linux (Debian 6.0.5). Este cuenta con una unidad de procesamiento gráfico (GPU) GeForce GTX 580 de NVIDIA, el cual se basa en lenguaje C con CUDA 5.0.

Se utilizan bibliotecas de funciones (APIs) de CUDA, tales como CURAND y CUDART. El manejo de imágenes se ha utilizado con OpenCV y C++.

Los pasos de operación se describen en la Tabla 1 (Algoritmo 1).

- 1. Inicialización.** Se captura una escena de entrada de tamaño 800×608 pixeles mostrada en la Fig. 2(a). Se utiliza OpenCV para cargar las imágenes y dimensionarlas al tamaño original [10]-[11].
- 2. Definición de parámetros de iluminación** La escena de entrada se degrada con una función de iluminación no homogénea, la cual contiene ruido aditivo, tal como se muestra en la Fig. 2(b) y 2(c), respectivamente. La función de iluminación se genera mediante un *kernel* (función en paralelo), en donde cada hebra procesa un pixel de la imagen.

Tabla 1. **Algoritmo 1.** Procedimiento para restauración de imágenes

Entrada:

$s(\mathbf{x})$ (Escena de entrada)
 var (Varianza del ruido con distribución normal)
 mt (Valor esperado del ruido con distribución normal)
 τ (Ángulo de inclinación de la fuente de iluminación)
 α (Ángulo de rotación de la fuente de iluminación)
 ρ (Magnitud del vector de la fuente de iluminación)

Salida:

$f(\mathbf{x})$ (Escena restaurada)

Pasos del Algoritmo:

Paso 1. Inicialización. Capturar escena de entrada $s(\mathbf{x})$.
Paso 2. Definición de parámetros de iluminación. Establecer valores constantes de los parámetros que definen la función de iluminación τ , α y ρ .
Paso 3. Definición de parámetros de ruido. Establecer valores constantes de los parámetros de varianza y valor esperado, var y mt , respectivamente, de la función del ruido aditivo con distribución normal.
Paso 4. Degradación de la señal. Calcular el modelo de señal degradada por la función de iluminación no uniforme y ruido aditivo $f(\mathbf{x}) = s(\mathbf{x})d(\mathbf{x}) + n(\mathbf{x})$.
Paso 5. Normalización. Calcular los coeficientes de normalización $a_0(\mathbf{x})$ y $a_1(\mathbf{x})$.
Paso 6. Restauración. Estimar la restauración de la imagen mediante los coeficientes de normalización y el modelo de señal $f(\mathbf{x}) = a_1(\mathbf{x})f(\mathbf{x}) + a_0(\mathbf{x})$.

3. Definición de parámetros de ruido. La función del ruido aditivo se implementa con la biblioteca CURAND, en la que se utiliza un generador de números aleatorios con distribución normal [12], tal como se muestra en el Algoritmo 2, asumiendo que el ruido aditivo contiene una respuesta Gaussiana. La implementación de CURAND se realiza en dos etapas. La primera etapa consiste en generar números aleatorios en el *device* (GPU), llamándolos con la librería `/include/curand.h` que ocurre dentro del *host* (CPU). La segunda etapa consiste en definir las funciones dentro del *device* para generar secuencias de números aleatorios, mediante la librería `/include/curand_kernel.h` esto permite que los números generados se puedan utilizar directamente en las funciones *kernel*, sin requerir la escritura y luego la lectura desde la memoria global [12].

4. Degradación de la señal. La degradación de la escena de entrada se obtiene utilizando el modelo multiplicativo y aditivo de la Ec. (7), tomando en cuenta las funciones resultantes de la iluminación y ruido, obtenidas en el Paso 2 y Paso 3.

5. Normalización. Se asume que la función de iluminación es desconocida, por lo cual este sistema

Tabla 2. **Algoritmo 2.** Generación de la función de ruido con distribución normal.

Paso 1. Inicialización.
Inicializar la variable del generador.

Paso 2. Crear el generador aleatorio.
Especificar tipo de generación de valores aleatorios.

Paso 3. Crear semilla.
Determinar la semilla del generador que sea dependiente del reloj.

Paso 4. Reservar memoria.
Utilizar variable de device del tamaño de la señal de entrada.

Paso 5. Generación ruido Gaussiano.
Utilizar la función de generación de números aleatorios con distribución normal, utilizando los parámetros de valor esperado y varianza var y mt , respectivamente.

restaura la imagen de entrada que ha sido degradada, mediante el cálculo de coeficientes a_0 y a_1 , representados en las Ecs. (13) y (14). Para la obtención de los coeficientes se realiza el cálculo de $\bar{s}(x)$, que es la imagen promedio de la escena de entrada en la k -ésima posición dentro de la región de soporte w_i . Este cálculo incluye un algoritmo de reducción [13] en una función de memoria global [9], tal como se muestra en la Fig. 3. La implementación del algoritmo de reducción se detalla en el Algoritmo 3.

6. Restauración. La restauración de la imagen se obtiene mediante el ajuste de los coeficientes de normalización, dada por la expresión de la Ec. (8). La calidad de restauración de la imagen se obtiene mediante el error cuadrático promedio (MSE) y el pico máximo de la relación señal a ruido (PSNR), obtenidas a continuación:

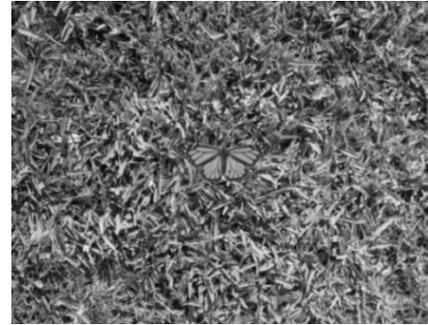
$$MSE = \sum_{x,y} (s(x,y) - \hat{f}(x,y))^2 \quad (15)$$

$$PSNR_{dB} = 10 \log_{10} \left(\frac{\max(s(x,y)^2)}{MSE} \right) \quad (16)$$

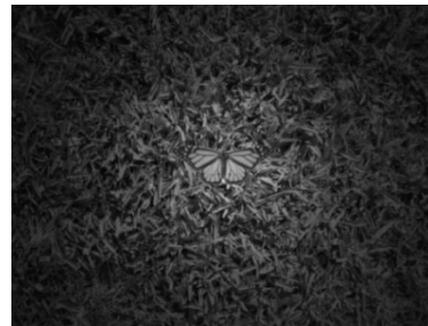
donde $s(x,y)$ es la escena sin degradación, y $\hat{f}(x,y)$ es la imagen estimada de la restauración.

5. Resultados

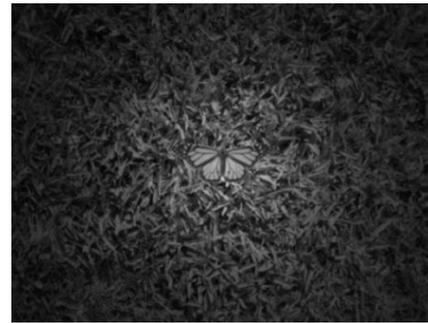
En esta sección se presentan los resultados en cuestiones de calidad de restauración y tiempos de ejecución. Se ha implementado el algoritmo de restauración de imágenes en paralelo, en donde las tareas de cada hilo se distribuyen dependiendo del número de elementos y el número de hebras. En esta técnica de procesamiento de imágenes, la operación ejecutada en cada pixel es



(a) Escena de entrada.



(b) Degradación de la imagen de entrada mediante una función de iluminación.



(c) Imagen iluminada que contiene ruido aditivo.

Figura 2. Degradación de imagen

Tabla 3. **Algoritmo 3.** Sistema de reducción para el cálculo de imagen promedio.

Paso 1. Inicialización de kernel.
Determinar índices para hebras en dos dimensiones.

Paso 2. Inicializar bloque.
Especificar que el valor del bloque k inicialmente valdrá la mitad de las dimensiones de la imagen de entrada.

Paso 3. Realizar reducción. 3.1 Mientras el bloque sea divisible entre dos, sumar los valores de cada bloque en orden con sus índices respectivos.
3.2 Sincronización de las hebras.
3.3 Reducir el bloque a la mitad.

Paso 4. Suma resultante.
Si el índice del bloque es el primero, la operación resultante se obtiene sumando los primeros índices de cada bloque.

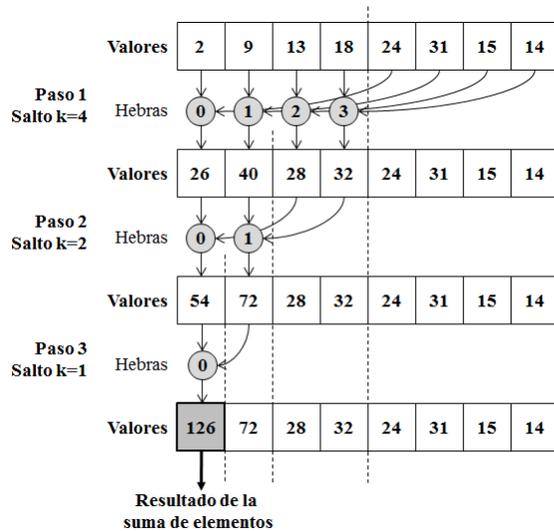


Figura 3. Desarrollo del algoritmo de reducción para la suma de elementos de un vector.

la misma y es independiente de la operación en otros pixeles, además su resultado no depende del resultado de la operación de otro pixel, por lo tanto se puede realizar en toda la imagen de manera simultánea [14].

Por ejemplo, para un CPU de 4 núcleos el número máximo de operaciones que pueden ocurrir en paralelo es 4. El uso del GPU está orientado para este tipo de aplicaciones de paralelización masiva.

Por ejemplo, el GPU utilizado en este trabajo NVIDIA GTX 580 tiene 512 multiprocesadores (16 por 32 núcleos), cada uno de los cuales soportan 1536 hilos de ejecución en paralelo. Combinados, estos multiprocesadores permiten que el GPU pueda operar más de 700.000 ejecuciones simultáneas de instrucciones individuales. Así, para un algoritmo que tiene un alto grado de paralelismo, una aplicación GPU es mucho más rápida que una implementación en CPU.

En la Tabla 4 se muestran los tiempos de ejecución resultantes con respecto a la utilización del algoritmo con distintos tipos de memoria. Se puede observar que la memoria compartida presenta ventajas ante la memoria global en la facilidad del acceso y procesamiento de la información. Se realizan comparaciones con respecto a distintas dimensiones de la imagen de entrada, las cuales presentan las mismas condiciones de degradación.

En la Fig. 4 se muestra la curva resultante del speedup para distintas dimensiones de la imagen de entrada, tomando en cuenta que es una imagen cuadrada, donde N representa el valor de los renglones o columnas.

En la Tabla 5 se muestra la evaluación de la calidad

Tabla 4. Tiempos de ejecución resultante con respecto al uso de memoria para distintas dimensiones de la imagen de entrada.

Tamaño	Tiempo de ejecución de GPU (ms)		Speedup
	Memoria compartida	Memoria Global	
32	0.190176	0.196992	1.04
64	0.20272	0.22	1.09
128	0.490624	0.622624	1.27
256	0.433856	1.208224	2.78
512	0.915136	4.4688	4.88
1024	1.876384	17.944576	9.56

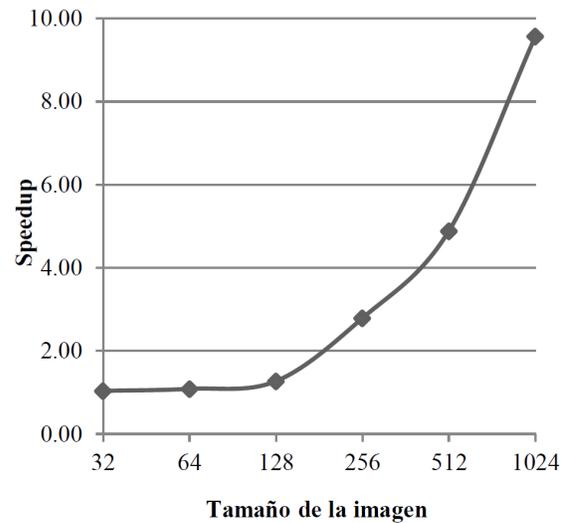


Figura 4. Speedup resultante con respecto a la dimensión de la imagen de entrada.

de restauración de la imagen de entrada mediante las métricas de MSE y PSNR de las Ecs. (16) y (16). La imagen utilizada es de dimensiones 800×608 pixeles, contiene un objeto dentro de un fondo, y es degradada utilizando distintos valores de los parámetros de iluminación, con valor de 100 dB de SNR de ruido aditivo.

En la Tabla 6 se muestran los resultados de las métricas de calidad con respecto a distintos niveles de SNR de ruido aditivo. Se utiliza una imagen que contiene un objeto sin fondo. Esta imagen es cuadrada que contiene dimensiones de 512×512 pixeles para todas las muestras, y con parámetros de iluminación de $\rho = 150$, $\alpha = 45$ y $\tau = 0$.

En las Figs. 5(a) y 5(b) se muestran las imágenes resultantes de la iluminación y restauración de la escena original sin degradación y libre de ruido aditivo presentada en la Fig. 2(a).

Tabla 5. Resultados de la calidad de restauración de imagen que contiene un objeto dentro de un fondo para diferentes parámetros de iluminación con dimensiones 800×608 .

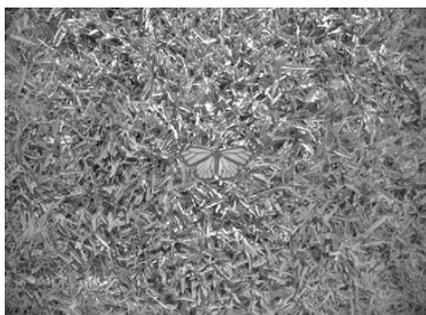
Parámetros de iluminación	MSE (%)	PSNR (dB)
$\rho = 40, \alpha = 45, \tau = 0$	15.1358	34.2206
$\rho = 80, \alpha = 45, \tau = 0$	14.1379	35.9241
$\rho = 160, \alpha = 45, \tau = 0$	14.1237	36.2839

Tabla 6. Evaluación de restauración para distintos niveles de ruido aditivo para una imagen que contiene un objeto sin fondo de dimensiones 512×512 .

Nivel de SNR de ruido aditivo (dB)	MSE (%)	PSNR (dB)
100	15.5137	8.09183
50	15.5176	8.09176
20	15.5722	8.07651
10	16.9016	7.72071
0	27.7875	5.56151



(a) Escena de entrada. Imagen resultante de la degradación por iluminación con niveles $\rho = 80, \alpha = 45, \tau = 0$ y 40 dB de SNR de ruido aditivo.



(b) Imagen resultante de la restauración.

Figura 5. Restauración de imagen.

6. Conclusiones

Mediante el desarrollo de este trabajo, se puede demostrar que el desempeño de restauración de imágenes puede incrementarse significativamente en condiciones cambiantes de iluminación y ruido, al utilizar algoritmos implementados en cómputo paralelo. Estos contribuyen a estimar los parámetros de la función de degradación en tiempo real.

Se encuentra un área de oportunidad en trabajos futuros, con el fin de obtener mayores valores de speedup al usar algoritmos de reducción en funciones atómicas, además de aprovechar de manera más eficiente los recursos de memoria compartida y de textura en la generación de la función de iluminación. El uso de bibliotecas de funciones APIs de CUDA podrá facilitar la programación de funciones, e incrementar la eficiencia en la ejecución.

En este trabajo se han presentado las mejoras de procesamiento de imagen aplicadas en procesador gráfico. Al realizar el procesamiento de imágenes en un GPU, se obtienen mejoras significativas con respecto al tiempo de ejecución. Además, proporciona una base de para poner en práctica trabajos futuros en algoritmos de procesamiento con mayor complejidad computacional, tales como el filtrado por correlación en el dominio de la frecuencia, que implica el uso de diferentes funciones para obtener resultados óptimos con respecto a la calidad de restauración de la imagen y respuesta en tiempo real. Los trabajos futuros realizados con este enfoque serán usados en migrar la aplicación de reconocimiento de patrones en paralelo utilizando de manera óptima los recursos computacionales de la tarjeta GPU.

Agradecimientos

Se agradece el apoyo recibido de los proyectos SIP-20130489 y SIP-20131209.

Referencias

- [1] V. H. Diaz-Ramirez, y V. Kober, "Target recognition under non-uniform illumination conditions," *Appl. Opt.*, Vol. 48, pp. 1408-1418, 2009.
- [2] P. Dutré, K. Bala, y P. Bekaert, *Advanced global illumination*, 2a. edición, Wellesley: A K Peters, Ltd., 2006.
- [3] M. Sherry, y A. Shearer, "IMPAIR: massively parallel deconvolution on the GPU," *Proc. SPIE, Image Processing: Algorithms and Systems XI*, Vol. 8655, Febrero, 2013.
- [4] L. Szirmay-Kalos, L. Szécsi, y M. Sbert. *GPU-based techniques for global illumination effects*, Hungary: Morgan & Claypool, 2008.
- [5] V. H. Diaz-Ramirez, y V. Kober, "Illumination invariant adaptive joint transform correlator," *Proc. SPIE, In Optics and Photonics*

- for *Information Processing*, Vol. 6695, pp. 66951B1-66951B7, 2007.
- [6] S. Martinez-Diaz, y V. Kober. "Nonlinear synthetic discriminant function filters for illumination-invariant pattern recognition," *Optical Engineering*, Vol. 47(6), pp. 067201-067201-9, 2008.
- [7] NVIDIA Corporation, "NVIDIA CUDA C Programming Guide," PG-02829-001 v5.5, Julio, 2013.
- [8] C. Jiang, P. Li, y Q. Luo, "High speed parallel processing of biomedical optics data with PC graphic hardware," *Proc. SPIE, Optical Sensors and Biophotonics*, Vol. 7634, Nov., 2009.
- [9] NVIDIA Corporation, "CUDA C best practices guide, Design Guide," DG-05603-001 v5.5, Julio, 2013.
- [10] G. Bradski, y A. Kaebler. *Learning OpenCV, Computer Vision with the OpenCV Library*, Sebastopol: O'Reilly, 2008.
- [11] R. Laganière, *OpenCV 2 Computer Vision Application Programming Cookbook*, Birmingham: Packt Publishing, 2011.
- [12] NVIDIA Corporation, "NVIDIA CUDA Toolkit V5.5 RN-06722-001 V.5.5," Mayo, 2013.
- [13] J. Sanders, y E. Kandrot. *CUDA by example, An Introduction to General-Purpose GPU Programming*, Addison-Wesley, 2011.
- [14] S. N. Swetadri Vasan *et al*, "Graphics processing unit (GPU) implementation of image processing algorithms to improve system performance of the control acquisition, processing, and image display system of the micro-angiographic fluoroscope," *Proc. SPIE 8313, Medical Imaging 2012: Physics of Medical Imaging*, Vol. 8313, Febrero, 2012.

Procesando el Reconocimiento de Múltiples Objetos en Tiempo Real Hacia la Implementación para el Reconocimiento de Actividades Humanas Concurrentes Utilizando GPU

Octavio I. Rentería-Vidales^a, Francisco E. Martínez Pérez^a, Juan C. Cuevas Tello^a, Omar Rodríguez González^a, Sandra E. Nava Muñoz^a

^aFacultad de Ingeniería, Universidad Autónoma de San Luis Potosí
Av. Dr. Manuel Nava No. 8, Zona Universitaria, San Luis Potosí, S.L.P., México, 78290
octavio.renteria@alumnos.uaslp.edu.mx, {[eduardo.perez](mailto:eduardo.perez@uaslp.edu.mx), [cuevas](mailto:cuevas@uaslp.edu.mx), [omarg](mailto:omarg@uaslp.edu.mx), [senavam](mailto:senavam@uaslp.edu.mx)}@uaslp.mx 2013 Published by

*DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Resumen

El presente artículo muestra los resultados preliminares de la implementación de una técnica de visión utilizando la unidad de procesamiento gráfica (GPU), aplicada al reconocimiento de actividades humanas entrelazadas y concurrentes. Se presenta la metodología de la implementación en pruebas realizadas en la unidad central de procesamiento (CPU) y utilizando GPU. Se muestra un ejemplo con el uso de cuatro filtros de correlación con el propósito de reconocer objetos en paralelo. Los resultados de los filtros de correlación son utilizados por el concepto Roaming Beat para dar un significado al comportamiento de los objetos dentro del desarrollo de actividades humanas. Se concluye que el uso de GPU es una opción factible en este tipo de sistemas.

Palabras clave: Procesamiento en paralelo, Reconocimiento de actividades, Reconocimiento de objetos.

1. Introducción

El reconocimiento de actividades humanas ha surgido como un problema de investigación en las últimas décadas. En el desarrollo de ambientes inteligentes, existen áreas tales como el cómputo ubico,

el cómputo móvil, el cómputo consciente de contexto, entre otras, que requieren del reconocimiento de actividades y su entorno para responder a las necesidades de los usuarios [1]-[3].

En la actualidad una técnica para el reconocimiento de actividades es aquella basada en la utilización de

sensores vestibles tales como acelerómetros o identificadores por radiofrecuencias, entre otros [4]. Este tipo de tecnología es utilizada para el reconocimiento de actividades físicas tales como caminar, correr, sentarse, pararse, o bien actividades que tienen patrones de movimiento.

Otra de las técnicas que es utilizada en el reconocimiento de actividades es aquella basada en visión. Dentro de esta técnica se destaca la vigilancia visual inteligente que se refiere al proceso de monitorización visual automatizado, el cual involucra el análisis e interpretación del comportamiento de los objetos. Esto es realizado mediante el reconocimiento y seguimiento de objetos con el propósito de entender los eventos de la escena de una forma visual [5].

Ambas técnicas para el reconocimiento de actividades proporcionan patrones característicos de las señales reales, cuando una actividad se realiza o cuando se identifica una situación específica de un escenario. Esto ocurre para el reconocimiento de actividades físicas como para aquellas que se basan en las interacciones del uso de objetos [6]; por ejemplo, el reconocimiento de las actividades tales como el cepillarse los dientes, el prepararse un café en el hogar o el tomar la presión sanguínea en una residencia de cuidados.

Cualquiera que sea la actividad a reconocer, es necesario tomar en cuenta que las actividades no se realizan de forma aislada sino que se entrelazan o concurren con otras actividades. Este hecho genera una dificultad en el reconocimiento de las actividades debido a las diversas variantes en las que se pueden desarrollar. Es decir, las actividades concurrentes comienzan en un mismo instante por una o más personas y con uno o más objetos para interactuar, lo que lleva a realizar el reconocimiento de varios objetos así como el reconocimiento de su comportamiento en la interacción que se realiza. Lo mismo sucede con las actividades entrelazadas, en las cuales una actividad comienza a ejecutarse en un determinado tiempo, y antes de que termine esta actividad otra actividad comienza a ejecutarse. Este hecho lleva a realizar el reconocimiento de varios objetos y su comportamiento en un lapso de tiempo entrelazado.

El propósito de este artículo es mostrar los procedimientos y resultados preliminares en el reconocimiento de objetos en tiempo real, paralelizando los procesos, con el objetivo de obtener la inferencia de actividades entrelazadas y concurrentes. El proceso se muestra mediante una técnica de visión, que hace uso de filtros de correlación compuestos, y se presenta la descripción de la implementación de forma secuencial en una computadora y la implementación en paralelo de los filtros de correlación utilizando una Unidad Gráfica de

Procesamiento (GPU por sus siglas en inglés *Graphics Processing Unit*).

El resto de este artículo está estructurado de la siguiente manera: en la sección 2 se muestra el trabajo relacionado, en la sección 3 se presenta la metodología de este artículo. En la sección 4 se muestran nuestros resultados preliminares y un ejemplo del desarrollo de actividades concurrentes utilizando nuestra implementación; por último se presentan nuestras conclusiones y las referencias.

2. Trabajo relacionado

El surgimiento de nuevos sistemas en ambientes reales ha sido posible gracias al avance de las nuevas tecnologías. Tal es el caso del uso de las unidades gráficas de procesamiento (GPU) que han propiciado la implementación del paralelismo como una alternativa para el futuro de la computación [7]. Algunos autores presentan un comparativo de la velocidad del CPU, OpenGL y CUDA utilizando una tarjeta de vídeo específica [7]. El propósito de nuestro trabajo no es el comenzar una discusión con respecto a las velocidades sino mostrar una implementación práctica y hacer uso del GPU como una alternativa en el reconocimiento de actividades entrelazadas y concurrentes.

Otros autores muestran una implementación en la cual hacen uso de filtrado de partículas, cuya característica es el costo computacional que se tiene [8]. Sin embargo, mencionan haber conseguido el desempeño en tiempo real necesario para su implementación. Otro ejemplo de implementación de la GPU es el desarrollo de una estrategia para el uso del GPU para la detección de objetos en movimiento [9].

Por otro lado, poco trabajo se ha reportado para el reconocimiento de actividades utilizando filtros de correlación y en menor proporción en el reconocimiento de actividades utilizando GPU. En este tipo de reconocimientos de actividades existen trabajos enfocados en el procesamiento de actividades entrelazadas o concurrentes entre los que destacan [10]-[12]. Éstos resaltan una técnica en particular denominada Patrones Emergentes (*Emerging Patterns - EP*). Esta técnica hace una interpretación de los datos de los sensores haciendo uso de una ventana deslizante en forma secuencial. Por lo que la captura y procesamiento se realiza de la misma manera.

Específicamente en sistemas para el reconocimiento de actividades se sugiere que los sistemas estén conformados por tres componentes tales como: i) un módulo de adquisición de información de bajo nivel; ii) un procesamiento y selección de características; y iii)

un módulo de clasificación que se encargue de utilizar las características para inferir las actividades [2]. Con este propósito, nuestro trabajo se basa en un concepto denominado Roaming Beat (RB) [13]. Este concepto hace uso de una metodología con la cual es posible obtener los tres componentes antes mencionados, y posteriormente llegar a reconocer las actividades entrelazadas y concurrentes, lo anterior será descrito en la siguiente sección.

3. Métodos

En esta sección se describen los pasos que fueron realizados en la implementación de los filtros de correlación, con el propósito de mostrar una alternativa para el reconocimiento de actividades entrelazadas y concurrentes.

3.1. Información Técnica

Los métodos implementados en este trabajo se realizaron en una computadora con procesador AMD FX-6100 con una frecuencia de reloj de 3.3/3.9 GHz; 6 MB en cache L2 y 8 MB en L3. El equipo contiene una tarjeta gráfica NVIDIA GeForce GTX 680, con una frecuencia 1006 /1058 MHz, 1536 núcleos CUDA; 2048 MB de memoria; una interfaz de memoria de 256-bit GDDR5 y un ancho de banda de 6 Gbps. La cámara utilizada fue una webcam Creative modelo VF0400 con una resolución de 640 × 480 píxeles.

Los métodos fueron realizados en C++ bajo la plataforma de Linux (Kernel 3.8.0) y con la ayuda de OpenCV, Armadillo y OpenCL¹ (para la implementación paralela).

La programación de GPUs se puede realizar utilizando CUDA u OpenCL. En este artículo se utiliza OpenCL porque permite su ejecución en cualquier tarjeta aceleradora gráfica como ATI Radeon (AMD), GeForce (NVIDIA) entre otras. Por otro lado, el código CUDA sólo se ejecuta en tarjetas de NVIDIA.

3.2. Implementación

El desarrollo de la implementación se realizó en dos etapas 1) Secuencial y 2) Paralela. Para ambas etapas se necesita el desarrollo de los filtros de correlación para el reconocimiento de objetos como se describe en [14]. En esta referencia se plantea el proceso de entrenamiento y el de las imágenes empleadas para considerar la traslación y rotación.

Además las dos etapas requirieron de la captura de la imagen que se obtiene del flujo de vídeo de la cámara web. Posteriormente se realiza un preprocesamiento a la imagen con el propósito de transformarla a escala de grises; recortarlas y adaptarlas al tamaño de los filtros (512 × 512) y aplicar un filtro pasa bajas mediante un blur Gaussiano con una matriz de 3 × 3. Este filtro fue utilizado con el propósito de suavizar la imagen de entrada, por lo que otro tipo de procesamiento tales como la detección de bordes queda fuera del alcance de este artículo, por el momento.

Un término que es utilizado en la programación de la GPU es *kernel* [15]. Esta palabra se refiere a un identificador interno de la GPU que ejecuta un fragmento de código en paralelo mediante el uso de hilos (*Threads* o *work items*). Por lo que este término será utilizado en este trabajo para hacer referencia a la implementación en paralelo.

3.2.1. Implementación secuencial

La implementación realizada en esta etapa fue en forma secuencial. Es decir, se realizaron cuatro funciones que contenían lo siguiente:

1. Un filtro por objeto a reconocer realizado con base en el algoritmo presentado en [14].
2. Tomar una imagen de entrada de la cámara de vídeo.
3. Aplicar la Transformada de Fourier (TF) a la imagen de entrada.
4. Aplicar la Ley K a la imagen de entrada. que es un filtro compuesto SDF no lineal [17]-[18] que tiene tolerancia a la deformación (rotación y traslación) de algunos objetos y un buen desempeño en la presencia de diferentes tipos de ruido, todo esto en el reconocimiento de objetos.
5. Realizar la correlación del filtro con la imagen de entrada.
6. Aplicar al resultado de la correlación la TF inversa

Dentro de un ciclo se ejecutaban las cuatro funciones. En el CPU cada píxel es procesado secuencialmente hasta completar la imagen.

3.2.2. Implementación en paralelo

Para la paralelización de procesos se empleó OpenCL para la programación de la GPU. No todos los pasos fueron paralelizados, sólo aquellos que eran críticos y consumían mucho tiempo como los antes mencionados en la implementación secuencial.

¹<http://www.khronos.org/opencl/>

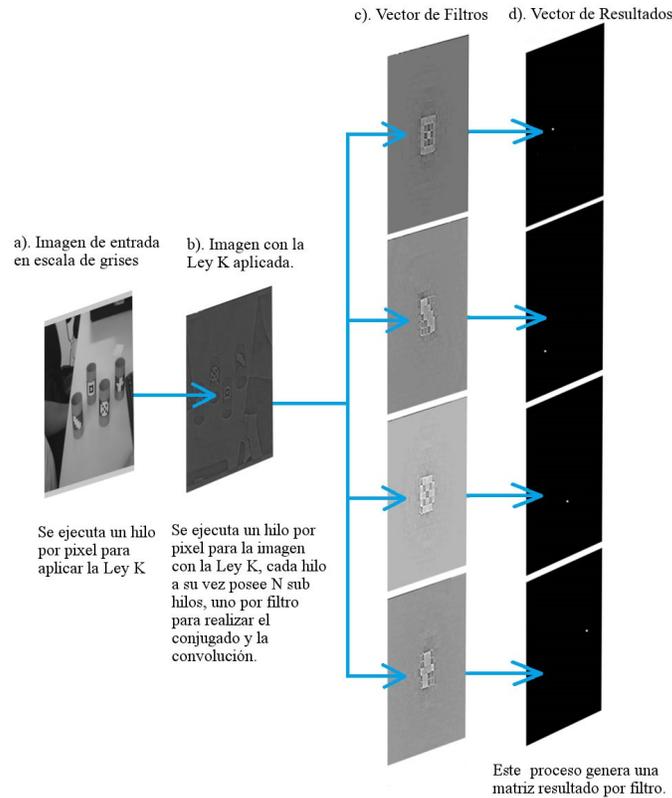


Figura 1. Representación de cuatro procesos paralelos implementados en GPU

El desarrollo consistió en implementar las siguientes funciones:

1. Se implementó un kernel para ejecutar la TF. En este módulo, fue necesario realizar las ecuaciones de la TF para implementarlas en OpenCL. Este procesamiento permitió realizar el código portable; es decir, poder implementar el mismo código en las tarjetas de vídeo AMD® o INTEL®. Esta implementación se basó en el algoritmo de la TF *Cooley-Tukey* [16]. Es importante mencionar que existen algoritmos de la TF. Sin embargo, hasta lo que se ha investigado, estos algoritmos están diseñados para ser implementados en CPU y no en GPU para el procesamiento en paralelo.
2. Un kernel para la función discriminante sintética llamada Ley K;
3. Un kernel para realizar la correlación.
4. Un kernel para implementar la transformada inversa de Fourier

Estos métodos son muy críticos por el procesamiento computacional que se requiere. Las imágenes fueron vectorizadas para aplicar la TF.

Una vez desarrollados los kernels se realizaron los siguientes pasos: Primeramente se copia la imagen

adquirida de la cámara en la memoria de vídeo de la GPU, esto con el propósito de evitar costosos tiempos de transferencia de datos entre el CPU y la GPU. Una vez en ella se ejecuta el kernel de cada función comenzando con aplicar la TF a la imagen de entrada. Posteriormente se transfiere el resultado al siguiente kernel que es la Ley K, en la cual se le aplica a la misma imagen (Figura 1 b)). Después, se ejecuta la correlación del filtro del objeto a reconocer con los resultados del kernel anterior. En este proceso, aunque se plantea de forma secuencial, en realidad se ejecuta la correlación de cada filtro con la imagen en forma paralela. En las Figuras 1 b) y 1 c) se muestra el proceso, en donde la imagen con la Ley K aplicada, se correlaciona con cada uno de los filtros en forma paralela.

Una vez realizada la correlación, se obtiene un vector por cada filtro con los resultados como lo muestra la Figura 1 d). A estos resultados se les aplica la transformada inversa de Fourier en forma paralela. La implementación del procesamiento en paralelo es realizado internamente por la tarjeta de vídeo.

Finalmente, el vector resultante de cada filtro es convertido a una matriz cuadrada, posteriormente se realiza un intercambio de los cuadrantes de la imagen (ifftshift) para poderla visualizarla.



Figura 2. Resultados de cuatro filtros de correlación implementados en la GPU

La Figura 2 muestra los resultados de correlación realizados con la GPU en tiempo real. En esta figura, se pueden observar cuatro objetos con diferentes etiquetas. Cada una de las ventanas ilustra el reconocimiento de su objeto, mostrando un pico muy alto, el cual indica la presencia del objeto en la escena. El valor alto representa las coordenadas (x, y) , en donde se localiza el objeto en la imagen entrante. Filtro A con un valor de reconocimiento de 1.1; filtro B con un valor de reconocimiento de 1.1; filtro C con un valor de reconocimiento de 1.2 y filtro D con un valor de reconocimiento de 0.68. Se puede observar que cada filtro reconoce a su objeto en específico y que la presencia de otras etiquetas no produce un valor significativo que pueda causar confusión. Esto se debe al umbral que se considera, con el cual se clasifica como objeto encontrado si excede el umbral, de lo contrario se descarta.

En la GPU, al implementar el kernel, todos los píxeles adquieren un hilo independiente, por lo que se ejecutan simultáneamente dependiendo de la implementación y de las capacidades internas de la tarjeta utilizada. La forma en que lo realiza la tarjeta produce una disminución del tiempo de procesamiento. Por ejemplo, considere un vector de 16 elementos, para el cual se tiene que definir un número de hilos a ejecutar. Cada hilo resolverá 4 elementos del vector, por lo cual es necesario dividir el vector entre 4. En este caso se utilizarán 4 hilos. Por lo que cada uno de los hilos se encarga de ejecutar un proceso a 4 elementos de forma independiente. De esta

forma se optimiza internamente el procesamiento. Para mayor detalle de la programación de OpenCL para GPU consultar [19].

3.3. Procesamiento de la información

Una vez realizada la correlación de las imágenes en la secuencia de vídeo, el siguiente paso es realizar un procesamiento y la selección de características. Para la interpretación de la selección de las características se utilizó el concepto Roaming Beat el cual se define como:

“La habilidad de un artefacto para dar una marca de tiempo (hora y fecha) para cambiar de un estado sin movimiento a uno con movimiento, de una localización base a otra posición errante [20]”

Este concepto hace uso de una metodología con la cual es posible inferir actividades entrelazadas y concurrentes. El concepto se basa en la obtención de un comportamiento a partir de la manipulación de objetos/artefactos en el desarrollo de una actividad.

Para lograr una selección de objetos se utilizan los resultados de la correlación de cada una de las imágenes de la secuencia de vídeo. Una de las ventajas de haber utilizado este tipo de filtros de correlación es que es posible reconocer varios objetos de la misma clase con un

solo filtro. Además de que es posible considerar el reconocimiento de objetos tomando en cuenta la traslación y rotación dentro de un mismo filtro. Al implementarse los filtros en paralelo utilizando GPU, es posible realizar el análisis del comportamiento de cada uno de los objetos de forma independiente, y que a su vez se puedan coordinar con los demás objetos participantes en una actividad.

Por lo que cada uno de los resultados de la correlación se compara con un umbral. Es decir, se obtienen los picos de correlación de cada una de las imágenes que son los valores más grandes como resultados del hallazgo de uno o más objetos en la imagen de la escena. Si los valores exceden el umbral se considera el reconocimiento del objeto y se interpreta como una señal que forma parte del comportamiento en el desarrollo de la actividad. Si los valores de correlación resultantes no exceden el umbral, se interpreta como una ausencia del objeto en la escena. En esta etapa se obtienen los valores máximos haciendo un recorrido de los resultados de los filtros de correlación en forma secuencial. De esta forma se discretizan los resultados de los filtros de correlación y se obtiene la representación del concepto de RB de cada uno de los filtros implementados.

Por ejemplo, en la Figura 3 considere el filtro A de la imagen. Este filtro comienza con el reconocimiento del artefacto, es decir, los resultados de la correlación con este filtro muestran un valor que se encontraba por encima del umbral. Por consiguiente, el valor es representado con un 1 desde el inicio de la secuencia de vídeo. Posteriormente en el índice 146 ocurre un cambio de 1 a 0. Es decir el artefacto sale de la escena y no es reconocido por el filtro. De este forma se discretizan los resultado de correlación y se va formando un comportamiento hasta cumplir las reglas para inferir la actividad como se menciona en [13]. En este caso, el comportamiento de este filtro debe de cumplir al menos 5 cambios de estado de 0 a 1 o viceversa. Este filtro obtuvo 6 cambios de estado (en los índices 284, 344, 410, 457, 558 y 661 de la Figura 3) y pudo inferirse la actividad.

4. Resultados

Se realizaron pruebas experimentales en laboratorio, en el cual se instaló una cámara web para realizar el reconocimiento de artefactos. La cámara genera 18 cuadros por segundo, bajo la plataforma Linux, a los cuales se les realizó el procesamiento.

En este trabajo se utilizaron etiquetas para realizar el reconocimiento de objetos como se muestra en la Figura 2. Además, es posible realizar el reconociemien-

Tabla 1. Tiempos promedios en milisegundos

	Fourier	Ley K	Correlación	Fourier Inversa
CPU	9.21	73.83	22.66	9.26
GPU	0.35	0.08	0.04	0.40

to de objetos utilizando las características físicas de los objetos. Los objetos utilizados en estas pruebas corresponden a un ambiente de cuidados de salud. Cada etiqueta representa un artefacto. El filtro A de la Figura 3 representa a una solución fisiológica relacionada con la actividad de curación; el filtro B representa el uso del papel higiénico relacionado con la actividad de higiene; el filtro C representa el objeto para tomar la presión arterial relacionada con la actividad para conocer la presión; y el último filtro representa al objeto para medir la glucosa relacionada con la actividad para medir los niveles de azúcar en la sangre. Aunque algunas de estas actividades están relacionadas con dos o más objetos, en esta demostración se asume que estas actividades se pueden inferir reconociendo sólo un objeto.

Como se mencionó anteriormente, se realizaron cuatro filtros de correlación que corresponden a cuatro objetos como se muestran en la Figura 3 (filtros A, B, C y D). Los filtros no consideraron imágenes para discriminar otro tipo de objetos. Estos filtros consideraron tres imágenes de referencia con las cuales se considera la traslación. Para todos los filtros se manejó un umbral de 0.65, por lo que todo valor por encima de este umbral es 1, y 0 para valor por debajo de umbral.

La Tabla 1 muestra los tiempos obtenidos por un filtro. Se realizaron 100 iteraciones de este filtro con 100 imágenes entrantes de la cámara de vídeo. Los filtros y las imágenes fueron de tamaño 512 × 512. Las imágenes entrantes de la cámara de vídeo se recortaron a 512 y se rellenaron con ceros para cumplir la especificación de los filtros (512), quedando las imágenes centradas. En la Tabla 1 se puede observar mejores resultados con el uso de la GPU. Esto se debe a que en esta implementación los píxeles de cada imagen son procesados por un hilo independiente, disminuyendo los tiempos de procesamiento.

La Figura 2 muestra los resultados de la implementación en paralelo mediante la GPU. Son cuatro filtros, de los cuales tres de ellos (filtros A, B, C) muestran un resultado de correlación por encima de 1.0. Se observa como el pico de los tres filtros y el filtro D con un resultado de 0.68. Todos se encuentran por encima del umbral, lo cual es requisito para discretizar la información del comportamiento de los objetos. Un detalle importante en estos resultados es que al momento de ir insertando cada uno de los objetos en el escenario, el filtro del objeto entrante inmediatamente mostraba su

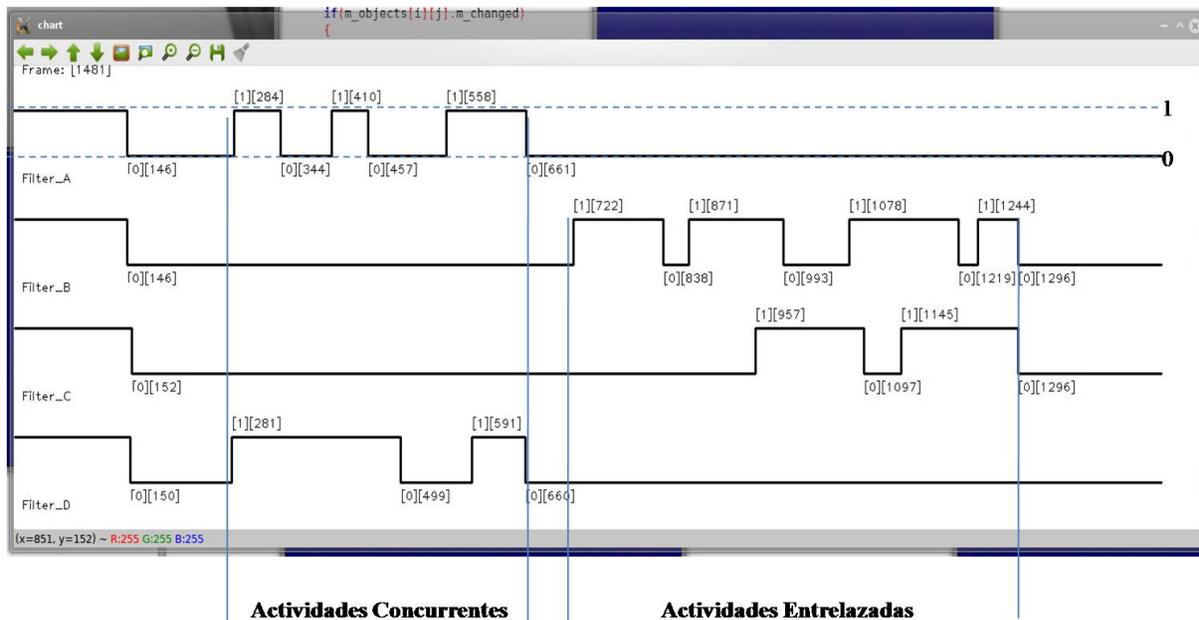


Figura 3. Ejemplo de la ejecución de actividades entrelazadas y concurrentes

pico correspondiente del reconocimiento. Por otro lado, cuando se insertaban los objetos que no correspondían al filtro, se mostraba un resultado de correlación en el rango de 0.20 y 0.30, suficiente para descartar esta información. Siendo con esto suficiente para utilizar el concepto RB e implementar la inferencia de las actividades entrelazadas y concurrentes, lo cual se describe a continuación.

4.1. Seguimiento del reconocimiento de objetos en paralelo

Este seguimiento del reconocimiento corresponde a la interpretación del comportamiento de los artefactos en el desarrollo de una actividad. Para lo cual se realizaron dos ejemplos de la ejecución de actividades concurrentes y entrelazadas. El primer ejemplo corresponde a la ejecución de dos actividades que están relacionadas con un solo objeto (Filtro A y D de la Figura 3. En este ejemplo ambos comportamientos comienzan y terminan en los mismos tiempos (281, 284 y 660 y 661 respectivamente). El segundo ejemplo corresponde a la ejecución de dos actividades pero de forma entrelazada (Filtros B y C de la Figura 3). La primera de ellas comienza en un tiempo (índice 722 del filtro B) y posteriormente se integra la segunda actividad (índice 957 del filtro C). En este ejemplo, ambas actividades terminan en el mismo tiempo (índice 1296 de los filtros B y C).

En la Figura 3, además del comportamiento de los artefactos, en cada cambio de estado de 0 a 1 ó beat, se muestran los índices de la secuencia de vídeo de la

cámara web. Con este índice se puede observar una diferencia mínima entre un reconocimiento y otro de los dos artefactos. Por ejemplo, en la actividad concurrente, el primer objeto (filtro D en la Figura 3) se reconoce en la imagen 281 mientras que el segundo objeto (filtro A de la Figura 3) se reconoce en la imagen 284. En la finalización de estas actividades el primer objeto (filtro D de la Figura 3) desaparece de escena en la imagen 660 y el segundo en la imagen 661.

Con estos resultados se corrobora la factibilidad de poder implementar las actividades entrelazadas y concurrentes a partir de utilizar el concepto RB y el utilizar un procesamiento mediante GPU. Esto, con base a que las actividades que realiza el humano son impredecibles y que es necesario realizar un procesamiento en paralelo para estar a la expectativa de todos los artefactos que sean necesarios.

5. Conclusiones y discusión

Se concluye que el uso de la GPU es una opción factible para la implementación de sistemas para el reconocimiento de actividades entrelazadas y concurrentes. Adicionalmente, es necesario tener bien definido las implicaciones para el reconocimiento de actividades. En este sentido, el utilizar el concepto RB fue de gran utilidad al marcar la pauta por el seguimiento de un comportamiento. Esto a partir del reconocimiento de los objetos con los cuales interactúa un humano en el desarrollo de una actividad.

Se utilizó una técnica de visión relacionada con filtros

de correlación compuestos. En el pasado se mostraba esta técnica para el seguimiento y detección de objetos, como costosa computacionalmente [21]. Sin embargo, con el uso de GPU, u otro tipo de tecnología, esta técnica puede llegar a ser implementada hasta en dispositivos embebidos, permitiendo el múltiple reconocimiento de objetos en segmentos de video en tiempo real.

El propósito de este artículo fue mostrar los resultados preliminares para la implementación de sistemas en el reconocimiento de actividades considerando las diversas formas que tiene el ser humano para realizarlas. Por lo que sabemos las implicaciones con las que nos enfrentamos y que los tiempos aquí mostrados pueden ser subjetivos. Nosotros consideramos estos resultados como pertinentes para continuar con nuestra investigación.

Este trabajo presenta dos de los tres componentes necesarios para un sistema de reconocimiento de actividades [2]. Por lo que nuestro trabajo futuro se enfocaría en la creación de un módulo de clasificación que se encargue de utilizar las características para inferir las actividades de una forma proactiva, inteligente y automática. Además de trabajar en un algoritmo secuencial equivalente a la paralelización mostrada en este artículo que permita medir la complejidad computacional. Adicionalmente, dirigir los esfuerzos en una optimización de los filtros de correlación compuestos para el reconocimiento de objetos similar al que se presenta en [9] utilizando GPU. Otro de los tópicos importantes en el uso de técnicas de visión es el uso de la oclusión y otras distorsiones, lo cual se plantea como trabajo futuro mediante la implementación de otro tipo de tecnología.

6. Agradecimiento

Este trabajo fue apoyado para su realización y presentación en congreso por el proyecto PROMEP del segundo autor con el número de oficio de la carta de liberación PROMEP/103.5/13/6575.

Referencias

- [1] M. Weiser, "The computer for the 21st century," *Scientific American*, Vol. 265, No. 3, pp. 94-104, 1991.
- [2] T. Choudhury *et al*, "The Mobile Sensing Platform: An Embedded Activity Recognition System," *Pervasive Computing*, pp. 32-41, 2008.
- [3] A. K. Dey and G. D. Abowd, "Towards a better understanding of context and context-awareness," *CHI 2000 workshop on the what, who, where, when, and how of context-awareness*, 2000, Vol. 4, pp. 1-6.
- [4] L. Chen, J. Hoey, C. Nugent, D. Cook, and Z. Yu, "Sensor-based activity recognition," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 42, No. 6, pp. 790-808, 2012.

- [5] I. S. Kim, H. S. Choi, K. M. Yi, J. Y. Choi, and S. G. Kong, "Intelligent visual surveillance - A survey," *International Journal of Control, Automation and Systems*, Vol. 8, No. 5, pp. 926-939, Oct. 2010.
- [6] M. Philipose *et al*, "Inferring activities from interactions with objects," *IEEE Pervasive Computing*, Vol. 3, pp. 50-57, 2004.
- [7] J. D. Owens, M. Houston, D. Luebke, S. Green, J. E. Stone, and J. C. Phillips, "GPU Computing," *Proceedings of the IEEE*, vol. 96, no. 5, pp. 879-899, May 2008.
- [8] W. Limprasert, A. Wallace, and G. Michaelson, "Real-Time People Tracking in a Camera Network," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, Vol. 3, No. 2, pp. 263-271, Jun. 2013.
- [9] D. Berjón, C. Cuevas, F. Morán, and N. García, "GPU-based Implementation of an Optimized Nonparametric Background Modeling for Real-time Moving Object Detection," *IEEE Transactions on Consumer Electronics*, Vol. 59, No. 2, pp. 361-369, 2013.
- [10] E. Kim, S. Helal, and D. Cook, "Human activity recognition and pattern discovery," *Pervasive Computing, IEEE*, Nol. 9, No. 1, pp. 48-53, 2010.
- [11] T. Gu, Z. Wu, X. Tao, H. Keng Pung, and J. Lu, "epSICAR: An Emerging Patterns based approach to sequential, interleaved and Concurrent Activity Recognition," *Pervasive Computing and Communications, 2009. PerCom 2009. IEEE International Conference on*, 2009, pp. 1-9.
- [12] T. Gu, L. Wang, Z. Wu, and X. Tao, "A pattern mining approach to sensor-based human activity recognition," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 23, No. 9, pp. 1359-1372, 2011.
- [13] F. E. Martínez-Pérez, J. Á. González-Fraga, J. C. Cuevas-Tello, and M. D. Rodríguez, "Activity Inference for Ambient Intelligence Through Handling Artifacts in a Healthcare Environment," *Sensors*, Vol. 12, No. 1, pp. 1072-1099, Jan. 2012.
- [14] F. E. Martinez-Perez, J. González-Fraga, and M. Tentori, "Automatic activity estimation based on object behaviour signature," *Proceedings of SPIE, 2010*, Vol. 7798, No. 1, p. 77980E.
- [15] *OpenCL Programming Guide for the CUDA Architecture*, Ver. 3.2. 2009, p. 61.
- [16] E. Ziegel, W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes: The Art of Scientific Computing*, Vol. 29, No. 4. 1987, p. 501.
- [17] B. Javidi and D. Painchaud, "Distortion-invariant pattern recognition with Fourier-plane nonlinear filter," *Appl. Optics*, Vol. 35, No. 2, pp. 318-331, 1996.
- [18] B. Javidi, W. Wang, and G. Zhang, "Composite Fourier-plane nonlinear filter for distortion-invariant pattern recognition," *Optical Engineering*, Vol. 36, No. 10, 1997.
- [19] A. Munshi, B. R. Gaster, T. G. Mattson, J. Fung, and D. Ginsburg, *OpenCL programming guide*. Addison-Wesley, 2011, p. 603.
- [20] F. E. Martinez-Perez, J. A. Gonzalez-Fraga, and M. Tentori, "Artifacts' Roaming Beats Recognition for Estimating Care Activities in a Nursing Home," *4th International Conference on Pervasive Computing Technologies for Healthcare 2010*, 2010.
- [21] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, Vol. 34, No. 3, pp. 334-352, 2004.

Solución Paralela en un GPU para la Ecuación de Advección-Difusión-Reacción mediante Diferencias Finitas Explícitas e Implícitas

Cesar Ivan Abrajan Barraza^a, Juan José Tapia Armenta^a

^aCentro de Investigación y Desarrollo de Tecnología Digital del Instituto Politécnico Nacional
Carr. Av. del Parque 1310 Mesa de Otay, Tijuana, B.C., México
cabrajan@citedi.mx, jtapiaa@ipn.mx

2013 Published by *DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Resumen

En este trabajo se presenta la ecuación de Advección-Difusión-Reacción (ADR) que se paraleliza en una unidad de procesamiento gráfico (GPU) para una, dos y tres dimensiones. Dicha ecuación se implementa mediante dos esquemas del método de finitas. El esquema explícito se implementa mediante la ecuación de recurrencia que se forma, sin embargo, tiene restricciones de estabilidad. Por otro lado, el esquema implícito forma un sistema de ecuaciones que se resuelve mediante el método iterativo de gradiente biconjugado en el GPU que obtiene ventaja de CUBLAS y CUSPARSE para resolver la matriz rala presente en dicho sistema de ecuaciones para cada dimensión. Finalmente se presentan los resultados donde se muestran el costo en tiempo de ejecución y convergencia que se obtiene al aplicar el gradiente biconjugado.

Palabras clave: Advección-Difusión-Reacción, Diferencias Finitas explícitas, Diferencias Finitas implícitas, GPU.

1. Introducción

La ecuación de Advección-Difusión-Reacción (ADR) es una ecuación diferencial de segundo orden que se usa ampliamente en el modelado matemático, su importancia radica en los tres fenómenos que agrupa. El término de advección indicado por los gradientes modela el desplazamiento de partículas

a una cierta velocidad, especificada por su coeficiente. El término de difusión, representado con gradientes de segundo orden, modela el flujo de una propiedad física de regiones con concentración alta a regiones con concentración baja, hasta lograr un equilibrio. La rapidez del flujo está determinado por su coeficiente y el término de la reacción es una función que aporta o retira energía al sistema; es decir, representa la fuente

del sistema. En este trabajo se modela la ecuación ADR para estudiar la evolución de su solución, a partir de una solución inicial en el tiempo $t = 0$, con las condiciones de frontera apropiadas, de acuerdo al fenómeno que se desea estudiar.

La ecuación ADR se puede aplicar para resolver problemas en una variedad de disciplinas. En particular, en el área de ecología, permite hacer estudios de medio ambiente, entre los que destacan el estudio de la interacción entre especies, el deterioro de la calidad del aire en las grandes ciudades, y la simulación de daño ambiental en fugas de plataformas petroleras. En estas aplicaciones, el término que toma mayor relevancia es el de advección.

La ecuación de calor en los sistemas térmicos modela la energía calorífica que se propaga en diversos materiales, lo que permite analizar y aprovechar dichas propiedades para la transmisión o retención de la energía. De la misma manera, en las plantas nucleares usan esta ecuación para la simulación de enfriamiento de reactores, así como también la contaminación de materiales radiactivos. La ecuación ADR se usa en medicina para modelar el crecimiento de tumores cancerígenos [1]. En el campo de estudio de la química, el análisis de las reacciones entre sustancias también puede modelarse mediante la ecuación ADR, siendo los patrones de Turing [2] los más conocidos.

Si bien existen casos donde la solución analítica o exacta es fácil de desarrollar, en la mayoría de los sistemas que son de importancia e impacto científico, es posible solo encontrar su aproximación mediante la aplicación de un método numérico que permita resolver el problema con un algoritmo computacional. Existen muchos métodos numéricos para resolver la ecuación ADR, entre los que destacan el método de diferencias finitas [3], método de elemento finito [4], método de diferencias finitas en el dominio del tiempo, métodos espectrales [5]. Un estudio detallado de la ecuación ADR se puede consultar en [6].

El modelado matemático tridimensional (3-D) mediante ecuaciones diferenciales es altamente demandante de recursos de cómputo. Hasta hace algunos años, estos problemas se resolvían en sistemas de cómputo muy costosos. Para abaratar los costos las grandes computadoras fueron sustituidas por clusters de computadoras y, en la actualidad, la opción más barata de usar cómputo de alto rendimiento es por medio de GPUs.

Es necesario también mencionar que no todos los métodos numéricos son ideales para implementarse en el GPU y es necesario saber distinguir las deficiencias de éstos [9]. En el presente trabajo se realiza el análisis numérico de la ecuación ADR mediante el método

de diferencias finitas bajo dos esquemas, el explícito que es el más intuitivo al momento de paralelizar el sistema debido a que después del análisis numérico, la ecuación resultante en cada punto a calcular no presenta dependencia espacial en el mismo tiempo [10]. Con este esquema implementado en el GPU podemos encontrar trabajos de investigación con la ecuación de calor y de onda [11], en patrones de Turing [2], [12]-[14], análisis de sistemas térmicos mediante el acoplamiento de ecuaciones de Navier-Stokes. El otro esquema es el implícito, el cual después del análisis numérico resulta en un sistema de ecuaciones que de manera directa no se puede paralelizar (funciones kernel) debido a la dependencia de datos espaciales. Sin embargo, ésta misma situación abre la oportunidad de implementar nuevas formas de interacción con el GPU como lo son los métodos iterativos y el uso de aplicaciones de interfaces de muy alto nivel (API, del inglés *Application Programming Interface*), ya sean proporcionadas de manera oficial por los diseñadores del hardware como lo son CUBLAS, CUSPARSE, o bien por terceros.

2. Fenómenos de ADR

La ecuación de Advección-Difusión-Reacción [3] con dependencia temporal de una forma general se representa como:

$$\frac{\partial u}{\partial t} = \alpha \nabla^2 u + \beta \nabla u + f(u) \quad (1)$$

donde el término de la izquierda representa el cambio de u en el tiempo. En el primer término de la derecha se representa la difusión con la segunda derivada espacial multiplicada por el coeficiente de difusión α . El segundo término representa la advección con la primera derivada respecto a la función u multiplicada por la constante del campo vectorial de velocidades β , y el tercer término representa una función de reacción.

Los fenómenos que se modelan con la ecuación ADR, pueden contener o no a todos los términos. En algunos casos, los valores que no aparecen se desprecian, tal es el caso de la ecuación del calor, donde el único término que aparece es el de difusión.

Para el caso específico de la ecuación de calor, la difusión representa el movimiento aleatorio de partículas que tienden a desplazarse de áreas con mayor energía calorífica a partes con menor intensidad. Dicho movimiento es directamente proporcional al coeficiente de difusividad del material por el cual se transporta dicha energía. Si al sistema se añade un líquido en reposo, éste tenderá a absorber o incorporar calor al sistema el cual dependerá de su temperatura. Pasado un tiempo,

tanto el líquido como el sistema tendrán la misma temperatura, debido a la difusión; por otro lado, si el líquido se encuentra en movimiento habrá un intercambio constante de energía calorífica además de un movimiento de arrastre debido a la advección.

Las condiciones de frontera intervienen directamente en el resultado de la ecuación ADR y pueden ser consideradas como funciones de reacción. En el caso de las condiciones de Dirichlet, presentan valores constantes que pueden o no cambiar en el tiempo y la ecuación que se modela tenderá a adoptar estos valores. Por otro lado, las condiciones de Neumann condicionan el flujo de la concentración que se analiza [17].

3. Aproximación numérica de la ecuación ADR mediante diferencias finitas

Para aproximar la solución de la ecuación ADR en el GPU se emplea el método de diferencias finitas, el cual consiste en remplazar las derivadas parciales por aproximaciones desarrolladas a partir de la serie de Taylor. Al ser una serie infinita, cada representación debe llevar un orden de truncamiento O (del inglés *big O*).

3.1. Aproximación numérica de la ecuación ADR con esquema explícito

Para aproximar la solución de la ecuación ADR en una dimensión, la Ec. (1) se representa como:

$$\frac{\partial u}{\partial t} = \alpha \frac{\partial^2 u}{\partial x^2} + \beta \frac{\partial u}{\partial x} + f(u) \quad (2)$$

El método de diferencias finitas bajo un esquema explícito, se inicia con una aproximación adelante en el tiempo (FT del inglés *Forward Time*) lo que resulta en:

$$\frac{\partial u}{\partial t} = \frac{u_j^{t+1} - u_j^t}{\Delta t} + O(\Delta t) \quad (3)$$

La segunda derivada espacial en x se calcula mediante el uso de diferencias finitas centradas (CS del inglés *Center Space*). El resultado que se obtiene es:

$$\frac{\partial^2 u}{\partial x^2} = \frac{u_{j+1}^t - 2u_j^t + u_{j-1}^t}{\Delta x^2} + O(\Delta x^2) \quad (4)$$

Al igual que la segunda derivada, la primera derivada respecto a x se aproxima mediante CS obteniéndose:

$$\frac{\partial u}{\partial x} = \frac{u_{j+1}^t - u_{j-1}^t}{2 \Delta x} + O(\Delta x^2) \quad (5)$$

Al sustituir las Ecs. (3)-(5) en (2), se obtiene:

$$u_j^{t+1} = u_j^t + \frac{\alpha \Delta t}{\Delta x^2} (u_{j+1}^t - 2u_j^t + u_{j-1}^t) + \frac{\beta \Delta t}{2 \Delta x} (u_{j+1}^t - u_{j-1}^t) + \Delta t f(u_j^t) \quad (6)$$

Al asignar $r = \alpha \Delta t / \Delta x^2$ y $s = \beta \Delta t / 2 \Delta x$ resulta:

$$u_j^{t+1} = u_j^t + r(u_{j+1}^t - 2u_j^t + u_{j-1}^t) + s(u_{j+1}^t - u_{j-1}^t) + \Delta t f(u_j^t) \quad (7)$$

Al agrupar términos finalmente se obtiene:

$$u_j^{t+1} = au_{j-1}^t + bu_j^t + cu_{j+1}^t + \Delta t f(u_j^t) \quad (8)$$

donde $a = r - s$, $b = 1 - 2r$ y $c = r + s$

En la Ec. (8) se aprecia como, independientemente del valor que contengan los coeficientes, la solución en el tiempo actual sólo dependerá de tres valores previamente calculados. Para el caso de dos dimensiones se sigue un proceso similar y la ecuación final es:

$$u_{i,j}^{t+1} = au_{i+1,j}^t + bu_{i,j+1}^t + cu_j^t + du_{i-1,j}^t + eu_{i,j-1}^t + \Delta t f(u_{i,j}^t) \quad (9)$$

Finalmente para tres dimensiones la ecuación es:

$$u_{i,j,k}^{t+1} = au_{i,j,k+1}^t + bu_{i+1,j,k}^t + cu_{i,j+1,k}^t + du_{i,j,k}^t + eu_{i,j-1,k}^t + fu_{i-1,j,k}^t + gu_{i,j,k-1}^t + \Delta t f(u_{i,j,k}^t) \quad (10)$$

3.2. Aproximación numérica de la ecuación ADR con esquema implícito

El método de diferencias finitas tiene, entre sus variantes, el esquema implícito Crank-Nicolson (CN) [19]. Este esquema se utiliza en el modelado matemático para encontrar la solución numérica de ecuaciones diferenciales, siendo sus principales características que el error de aproximación es de segundo orden tanto en tiempo como en espacio. Es implícito e incondicionalmente estable.

La diferencia entre los esquemas CN y FTCS radica en la aproximación de diferencias finitas que se realiza en el tiempo, ya que CN utiliza los valores en el tiempo actual y el anterior, mientras que FTCS usa los valores en el tiempo anterior. Así también, CN realiza un promedio de los valores en el espacio entre los tiempos

anteriores y actuales, en tanto que FTCS solamente necesita valores previamente calculados.

Para aproximar la solución de la ecuación ADR en una dimensión mediante CN, de la Ec. [2], se aproxima la derivada temporal mediante diferencias finitas hacia atrás (BT del inglés *Back Time*) dada por:

$$\frac{\partial u}{\partial t} = \frac{u_j^t - u_j^{t-1}}{\Delta t} + O(\Delta t^2) \quad (11)$$

Para el término de difusión donde se involucra la segunda derivada, el esquema CN aproxima la solución mediante el promedio de diferencia finitas centradas en el espacio del tiempo anterior y el tiempo actual, lo que resulta en:

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{2 \Delta x^2} (u_{j-1}^t - 2u_j^t + u_{j+1}^t + u_{j-1}^{t-1} - 2u_j^{t-1} + u_{j+1}^{t-1}) + O(\Delta x^2) \quad (12)$$

Para la primera derivada que aparece en el término de advección, el esquema CN nuevamente promedia mediante diferencias finitas centradas en el espacio del tiempo anterior y el actual, obteniéndose:

$$\frac{\partial u}{\partial x} = \frac{1}{4 \Delta x} (-u_{j-1}^t + u_{j+1}^t - u_{j-1}^{t-1} + u_{j+1}^{t-1}) + O(\Delta x^2) \quad (13)$$

Al sustituir las Ecs. [11]-[13] en Ec. [2] se obtiene el sistema de ecuaciones:

$$au_{j-1}^t + bu_j^t + cu_{j+1}^t = - au_{j-1}^t + (1 - 2r)u_j^t - cu_{j+1}^t + \Delta t f(u_j^{t-1}) \quad (14)$$

con $a = s - r$, $b = 1 + 2r$ y $c = -s - r$. A su vez $r = \alpha \Delta t / 2 \Delta x^2$ y $s = \beta \Delta t / 4 \Delta x$. Las constantes a , b y c se guardarán en una matriz de coeficientes A del sistema de ecuaciones $Au = b$.

Para dos y tres dimensiones se siguen procedimientos similares, y únicamente crece el número de términos en cada renglón en el sistema de ecuaciones. La matriz de coeficientes A presenta la peculiaridad que, al tener pocos valores diferentes de cero, se les conoce como matriz rala. Así también al aumentar el número de dimensiones en la ecuación ADR se generará un matriz de coeficiente $N^2 \times N^2$ y de $N^3 \times N^3$ para dos y tres dimensiones respectivamente.

4. Implementación de la ecuación ADR en un GPU

La programación en paralelo se utiliza para resolver programas computacionalmente demandantes, para lo cual se utilizan diferentes paradigmas de tecnologías de software que permiten expresar algoritmos paralelos para implantar aplicaciones en diferentes arquitecturas [8]. En un modelo de programación se pueden incluir distintas áreas como lo son los diferentes lenguajes de programación, los compiladores, las interfases para el desarrollo de aplicaciones, las bibliotecas de funciones, sistemas de comunicación, dispositivos de entrada y salida [7].

El principal requisito para paralelizar un problema es que no exista dependencia de datos. Esto es, que permita la descomposición del problema original en partes de tal manera que puedan ejecutarse de manera simultánea sin afectar el resultado final. Otros aspectos adicionales a considerar son asignación de tareas y la comunicación entre los procesos.

Una ventaja de utilizar plataformas en paralelo es que es posible eliminar la limitación física de la computación secuencial donde la rapidez del programa es directamente proporcional a la frecuencia del reloj. Sin embargo, más frecuencia también significa más consumo de energía lo que conlleva a saturar el sistema [1]. Ante este inconveniente, seguir con el aumento de frecuencia no es una opción y, por lo tanto, los problemas con complejidad elevada o que requieren una solución en tiempo real se ven comprometidos. En este trabajo se presenta un problema que, además de tener dependencia temporal, que habrá de resolverse un cierto número de intervalos en el tiempo, tiene la peculiaridad de que al aumentar el número de dimensiones espaciales el costo computacional se incrementa de manera drástica.

CUDA (del inglés *Compute Unified Device Architecture*), es un conjunto de tecnologías tanto en hardware como en software que proporcionan a desarrolladores los recursos necesarios para la programación de la unidad de procesamiento gráfico. Desde el punto de vista del software, CUDA es un conjunto de funciones que pueden tratarse como extensiones de lenguajes de programación tales como C, C++, Fortran, Python, entre otros [20].

En este trabajo se utiliza CUDA C para implementar los algoritmos que resuelven la ecuación ADR. De manera general, se necesitan de cuatro elementos para desarrollar programas en un GPU [7]:

1. Un GPU que soporte CUDA.
2. Un controlador de recursos para el GPU.
3. Un conjunto de herramientas de desarrollo de CUDA.

Tabla 1. Tipos de memoria en el GPU

Memoria	Cache	Acceso	Accesibilidad
Registros	-	R/W	Hilo
Local	-	R/W	Hilo
Compartida	-	R/W	Bloque
Constante	Sí	R	Hilos + CPU
Textura	Sí	R	Hilos + CPU
Global	Sí	R/W	Hilos + CPU

4. Un compilador de lenguaje C.

Al ser el GPU un procesador auxiliar al momento de ejecutar un programa, el hilo principal siempre estará en el CPU y se ejecutará de manera secuencial. Cuando se inicializa una subrutina que se habrá de ejecutar en el GPU mediante una función kernel, la ejecución se mueve al GPU, donde un determinado número de hilos son ejecutados concurrentemente para realizar el cálculo de las operaciones en paralelo. En el momento de la ejecución en el GPU se ejecutan una gran cantidad de hilos. Estos hilos son agrupados en bloques, que a su vez son agrupados en mallas. Cada hilo tiene su propia memoria, así como una memoria compartida visible para el resto de los hilos en el mismo bloque. Existe, además, la memoria principal o global la cual es visible para todos los hilos y dos memorias tipo cache; es decir, memorias de muy rápido acceso pero que están sumamente limitadas en capacidad como son la memoria de textura y la memoria constante ambas únicamente de lectura. La tabla 1 muestra un resumen de las memorias en el dispositivo y sus características principales; sin embargo, cualquier memoria que puede ser modificada por el CPU requiere un copiado de memoria.

4.1. Algoritmo para el esquema FTCS

El algoritmo 1 mostrado en la Tabla 2 se utiliza en la implementación de la aproximación de la ecuación ADR en una, dos y tres dimensiones mediante el esquema de diferencias finitas explícitas. Se parte de la reservación de la memoria tanto en el GPU como en el CPU, después se utilizan las condiciones de la ecuación diferencial para establecer la función inicial, luego se adecúa la función solución a un vector y se copia a la memoria en el GPU, en cada iteración se inicializa un kernel que valida la convergencia de la solución por medio de otro kernel que evalúa el error de norma L2 al término de la iteración. Una vez obtenida la solución se copian los resultados al CPU, y finalmente se libera la memoria tanto en CPU como GPU.

La implementación del algoritmo 1 se realiza de una manera muy intuitiva debido a que no se presenta una dependencia de datos. Los programas son divididos en

Tabla 2. **Algoritmo 1.** Algoritmo para la ecuación ADR mediante FTCS en un GPU.

```

reservar memoria en el GPU.
reservar memoria en el CPU.
condiciones iniciales.
adecuar la función solución a un vector.
copiar memoria del CPU al GPU.
for (t = 1:N) do
    inicializar kernel.
    ejecutar kernel norma L2
end for
copiar solución del GPU al CPU.
resultados.
liberar memoria.
    
```

Tabla 3. **Algoritmo 2.** Algoritmo para la ecuación ADR mediante CN en el GPU.

```

reservar memoria en el GPU.
reservar memoria en el CPU.
condiciones iniciales.
formar matriz rara.
inicializar CUBLAs.
inicializar CUSPARSE.
for (t=1:N) do
    ejecutar BiCG.
    ejecutar kernel norma L2.
    actualizar solución con llamado de un kernel.
end for
copiar solución del GPU al CPU.
resultados.
liberar memoria.
    
```

kernels, los cuales son compilados y ejecutados en el GPU de manera paralela y dependerá de la arquitectura del hardware que se utilice. Kernel es la porción de cómputo intensivo que mediante funciones en C se ejecutan N veces en N diferentes hilos [20].

4.2. Aproximación numérica de la ecuación ADR con esquema implícito

El algoritmo 2 mostrado en la Tabla 3 se utiliza en la implementación de la aproximación de la ecuación ADR en una, dos y tres dimensiones mediante el esquema de diferencias finitas explícitas. Se parte de la reservación de la memoria tanto en el GPU como en el CPU, después se utilizan las condiciones de la ecuación diferencial para establecer la función inicial, luego se adecúa la función solución a un vector y se copia a la memoria en el GPU. En cada iteración se inicializa un kernel que valida la convergencia de la solución por medio de otro kernel que evalúa el error de norma L2 al término de la iteración. Una vez obtenida la solución se copian los resultados al CPU, y finalmente se libera la memoria tanto en CPU como GPU.

En el algoritmo 2, aparecen nuevos términos como lo son CUBLAS, CUSPARSE, BiCG así como la matriz de

coeficientes A que resulta un caso de estudio [21] debido a que contiene muy pocos valores diferentes de cero. Estas matrices se conocen como matrices ralas que pueden formarse de manera aleatoria [22] y pueden ser tanto reales como complejas; así también las matrices ralas son predefinidas [21], [23]. Existen universidades que proveen base de datos con matrices ralas con características únicas en cuanto a su representación por los patrones que pueden formar, y dichos patrones resultan al aplicar diferentes métodos numéricos en la resolución de fenómenos, aunque no se descarta la creación de matrices ralas de manera artificial.

Los sistemas de ecuaciones $Au = b$ se pueden resolver mediante métodos directos e iterativos. De los primeros resaltan los métodos de eliminación directa como Gauss y sus variantes o bien descomposición LU donde L es una matriz triangular inferior y U es una matriz triangular superior. Para el caso de una implementación en el GPU se realiza mediante los métodos iterativos como el del gradiente biconjugado; por tal motivo, para implementar en paralelo la solución de un sistema de ecuaciones, se plantea como un problema de optimización puesto que se tiene que minimizar el residuo r en $Au - b = r$ [24].

Para realizar las operaciones necesarias en el algoritmo 2 al momento de realizar el gradiente biconjugado se realiza un cambio en el formato de la matriz rala. Se descompone en vectores, el caso más sencillo es guardar cada una de las posiciones con su respectivo valor. Este formato es conocido como completo o crudo [25], y es el menos recomendado debido a que los tres vectores serán muy grandes llegando incluso a superar el espacio de memoria con el que se cuenta. Por tal motivo, en este trabajo no se realizarán operaciones con este formato en la solución de la ecuación ADR.

Para la implementación de la ecuación ADR con esquema implícito se requiere el uso de CUBLAS que es una API que proporciona las subrutinas para las operaciones básicas de álgebra lineal, además de que permite el acceso, manejo y monitoreo de recursos en el GPU de una manera más eficiente. Las operaciones pueden ser de tipo flotante, dobles, complejas flotantes y complejas dobles, las cuales dependen de la capacidad de cómputo en el GPU. En este trabajo CUBLAS se utilizó para implementar el gradiente biconjugado en las operaciones con vectores densos o completos como el producto punto. Las rutinas se clasifican en:

Nivel 1 (BLAS1). Funciones para realizar operaciones tipo escalar-vector.

Nivel 2 (BLAS2). Funciones para realizar operaciones tipo matriz-vector.

Tabla 4. Comparación de los métodos explícito e implícito en una dimensión.

N	Explícito FTCS 1D		Implícito CN 1D	
	Duración	Iteración	Duración	Iteración
32	1.5	19799	5.9	19781
64	1.6	20133	6.3	20135
128	1.6	20483	6.4	20485
256	1.6	20831	6.6	20834

Nivel 3 (BLAS3). Funciones para realizar operaciones tipo matriz-matriz.

CUSPARSE es la API mediante la cual se realizan las operaciones con matrices ralas. Al igual que en CUBLAS, todas las instrucciones se invocan en el CPU y las ejecuta el GPU. CUSPARSE exige que la matriz sea convertida previamente a un formato de matriz rala y provee instrucciones para convertir de manera automática entre los formatos. En este trabajo se utilizó en todas las operaciones donde los vectores o matrices no fueran densos. Al igual que CUBLAS se pueden realizar operaciones del tipo flotante, doble, complejo flotante y complejo doble. Las rutinas se clasifican en:

Nivel 1. Operaciones entre vectores en formato de matriz rala y vectores en formato denso.

Nivel 2. Operaciones entre una matriz en formato de matriz rala y vector en formato denso.

Nivel 3. Operaciones entre una matriz y un conjunto de vectores, ambos en formato de matriz rala.

Conversión. Operaciones para convertir entre los diferentes formatos de matrices ralas.

5. Simulación

En la Tabla 4 se muestra los valores en tiempo que se necesitan para resolver la ecuación ADR en un procesador gráfico al variar la cantidad de puntos en la malla unidimensional. Se usó la función inicial $u = \text{sen}(\pi x/L)$, con condiciones en la frontera del tipo Dirichlet homogéneas. Para los diferentes valores en N se utilizó $\alpha = 10^{-5}$ y un coeficiente $\beta = 10^{-7}$. El fenómeno de reacción está dado por la función $u(x, t)$, y el incremento en el tiempo fue de 10^{-3} . La longitud de prueba se fijó a $L = 1$. Para asegurar la convergencia de la función a 10^{-8} se usó la norma L2. En las Tablas 4, 5 y 6 la columna duración indica el tiempo de ejecución en milisegundos.

En la Tabla 5 se presentan los valores encontrados tanto en tiempo como en número de iteraciones para converger al equilibrio de la ecuación ADR en dos dimensiones al usar el método de diferencias finitas explícitas como implícitas. Para llevar a cabo esta simulación

Tabla 5. Comparación de los métodos explícito e implícito en dos dimensiones.

N	Explícito FTCS 2D		Implícito CN 2D	
	Duración	Iteración	Duración	Iteración
32 ²	1.8	21389	7.2	21392
64 ²	1.8	22122	8.7	22126
128 ²	3.6	23532	12.6	23529

Tabla 6. Comparación de los métodos explícito e implícito en tres dimensiones.

N	Explícito FTCS 3D		Implícito CN 3D	
	Duración	Iteración	Duración	Iteración
8 ³	2.4	21833	8.00	21892
16 ³	2.41	22798	74.10	27533
32 ³	14.29	42303	1612.80	51759

se utilizó la función inicial de $u = \text{sen}(\pi x/L_x) + \text{sen}(\pi y/L_y)$. Los coeficientes de difusión fueron $\alpha_x = \alpha_y = 10^{-5}$, y los coeficientes de advección $\beta_x = \beta_y = 10^{-7}$. La función de reacción se define como u , el incremento en el tiempo fue de 10^{-3} , para asegurar la convergencia para 10^{-8} se utilizó la norma L2. La longitud para cada dimensión fue de $L_x = L_y = 1$, por último se establecieron condiciones de frontera de Dirichlet homogéneas.

De la Tabla 5 se puede concluir que la ecuación de ADR para dos dimensiones muestra que los tiempos de ejecución, en el método explícito, se ven afectados un mínimo respecto a los tiempos obtenidos en una dimensión. Por el contrario, el método implícito presenta una tendencia a la alza, para las iteraciones de convergencia se presentan cambios mínimos o no significativos.

En la Tabla 6 se presentan los resultados de la simulación en un espacio tridimensional de la ecuación ADR. Se usó la función inicial de $u = \text{sen}(\pi x/L_x) + \text{sen}(\pi y/L_y) + \text{sen}(\pi z/L_z)$. Los coeficientes de difusión fueron $\alpha_x = \alpha_y = \alpha_z = 10^{-5}$, y los coeficientes de advección fueron $\beta_x = \beta_y = \beta_z = 10^{-7}$. La función de reacción se define como u , el incremento en el tiempo fue de 10^{-3} , y para asegurar la convergencia de 10^{-8} se utilizó la norma L2. La longitud para cada dimensión fue de $L_x = L_y = L_z = 1$. Por último se establecieron condiciones de frontera de Dirichlet homogéneas.

De la Tabla 6 se aprecia la convergencia de la aproximación mediante la norma L_2 con condiciones de frontera de Dirichlet. En este caso en particular donde todas las dimensiones presentan condiciones de igualdad, el método explícito destaca tanto por el tiempo de ejecución como de iteración en convergencia. Al aumentar el número de dimensiones el costo por iteración en el método explícito se ve mínimamente afectado, puesto que solo aumenta el tamaño del vector que se envía al procesador gráfico. Por otro lado, el método implícito necesita resolver en cada iteración temporal una serie de iteraciones mediante el gradiente conjugado, lo cual

tiene que afectar directamente el tiempo de ejecución y su tiempo de convergencia necesita de la solución de dos normas L_2 : una en la iteración en el tiempo y una en cada iteración dentro del gradiente biconjugado, lo que se ve reflejado en la simulación con 32³ elementos.

6. Conclusiones y Trabajo Futuro

En este trabajo, se resuelve la ecuación de Advección-Difusión-Reacción con el método de diferencias finitas en un procesador gráfico. El esquema explícito se paraleliza de una manera muy intuitiva al resolver una ecuación de recurrencia en la que cada hilo de ejecución resuelve una incógnita. El esquema implícito se paraleliza al plantear un sistema de ecuaciones como un problema de optimización y se resuelve en el GPU mediante las bibliotecas de funciones CUBLAS y CUSPARSE. Para medir los tiempos de ejecución y la convergencia de los esquemas en las mismas circunstancias se emplearon parámetros similares. En la comparación para una y dos dimensiones los tiempos de ejecución del esquema explícito muestra ventaja; sin embargo, la convergencia en iteración es prácticamente la misma. Para la ecuación ADR evaluada en tres dimensiones, el costo computación aumenta dramáticamente, especialmente en el método implícito, debido a que la matriz que se tiene que resolver es de $N^3 \times N^3$ y esto se ve reflejado en el tiempo de ejecución. Además, resolver el gradiente biconjugado envuelve un error al calcular el residuo en cada intervalo de tiempo y, por lo tanto, la convergencia para tres dimensiones requiere más iteraciones. Para mejorar el método implícito es necesario mejorar el método de optimización.

Como trabajo a futuro se propone:

- Mejorar el algoritmo del gradiente biconjugado mediante el uso de matrices de pre-acondicionamiento para disminuir el tiempo de convergencia para cada iteración en el tiempo. También, usar métodos de optimización alternativos como son el método del residuo generalizado.
- Cuando la reacción domina a la difusión se tienen cambios bruscos en la solución en una región muy pequeña de la solución. Por lo que se requiere adaptación de malla. Como trabajo futuro se propone usar métodos de adaptación de malla en el GPU.
- Explotar las capacidades de la memoria de solo lectura o memoria de textura para mejorar el desempeño en la implementación en el esquema explícito.

Por otro lado, el uso de la memoria compartida podría ser una alternativa viable.

- Implementar las ecuaciones de Navier-Stokes para modelar fluidos.
- Implementar la ecuación ADR en el GPU mediante el uso de métodos numéricos alternativos como el método de elemento finito y métodos espectrales.
- Para el caso de sistemas que requieren del esquema implícito, el cálculo en tres dimensiones con una gran cantidad de datos, la memoria del GPU no es suficiente. Para este caso, se recomienda el uso de un clúster de GPUs.

Agradecimientos

Se agradece el apoyo recibido por parte del proyecto SIP-20130489.

Referencias

- [1] X. Chen, R. Summers, and J. Yao, "Kidney tumor growth prediction by coupling reaction diffusion and biomechanical model," *IEEE Trans. Biomed. Eng.*, Vol. 60, No. 1, pp. 169-173, 2013.
- [2] J. Villagrana Mancilla, L. Yamamoto, D. Miorandi, P. Collet, and W. Banzhaf, "Recovery properties of distributed cluster head using reaction-diffusion," *Swarm Intell.*, Vol 5, No. 3-4, pp 225-255. 2011.
- [3] A. R. Sanderson, M. D. Meyer, R. M. Kirby, and C. R. Johnson, "A framework for exploring numerical solutions of advection-reaction-diffusion equations using a GPU-based approach," *Comput. Vis. Sci.*, Vol. 12, No. 4, pp. 155-170, 2009.
- [4] M. Bause and K. Schwegler, "High order finite element approximation of systems of convection-diffusion-reaction equations with small diffusion," *J. Comput. Appl. Math.*, Vol. 246, pp. 52-64, July 2013.
- [5] F. Shakeri and M. Dehghan, "The finite volume spectral element method to solve Turing models in the biological pattern formation," *Comput. Math. Appl.*, Vol. 62, No. 12, pp. 4322-4336, 2011.
- [6] W. Hundsdorfer and J. Verwer, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*. Springer, 2003.
- [7] J. Sanders and E. Kandrot, *CUDA by example: An introduction to General-Purpose GPU Programming*. Addison-Wesley, 2010.
- [8] J. Dongarra, I. Foster, G. Fox, W. Gropp, K. Kennedy, L. Torczon, and A. White, eds., *Sourcebook of parallel computing*, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003.
- [9] M. H. Holmes, *Introduction to Numerical Methods in Differential Equations*. Springer, 2007.
- [10] A. Bourchtein and L. Bourchtein, "Explicit finite difference scheme with extended stability for advection equations," *J. Comput. Appl. Math.*, Vol. 236, pp. 3591-3604, September 2012.
- [11] C. A. Nahas, C. V. Krishnamurthy, K. Balasubramaniam, and P. Rajagopal, "Graphics processing unit based computation for NDE applications," *AIP Conf. Proc.*, Vol. 1430, pp. 1998-2005, 2012.
- [12] R. Acar, "An advection reaction model for flow visualization," *Pacific Visualization Symposium (PacificVis), 2010 IEEE*, pp. 137-144, Mar 2010.
- [13] D. G. Alvarado, C. Galeano, and J. Mantilla, "Turing pattern formation for reaction convection diffusion systems in fixed domains submitted to toroidal velocity fields," *App. Math. Model.*, Vol. 35, pp. 4913-4925, 2011.
- [14] F. Molnár, F. Izsák, R. Mészáros, and I. Lagzi, "Simulation of reaction diffusion processes in three dimensions using CUDA," *Chemometr. Intell. Lab.*, Vol. 108, pp. 76-85, 2011.
- [15] R. J. LeVeque, *Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems*. SIAM, 2007.
- [16] D. M. Causon, *Introductory Finite Difference Methods for PDEs*. Ventus Publishing ApS, 2010.
- [17] A. Quarteroni, *Numerical Models for Differential Problems*. Springer, 2009.
- [18] W. E. Boyce and R. C. DiPrima, *Elementary Differential Equations and Boundary Values Problems*. Laurie Rosatone, ninth ed., 2009.
- [19] J. Crank and P. Nicolson, "A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type," *Adv. Comput. Math.*, Vol. 6, pp. 207-226, 1996.
- [20] D. B. Kirk, *Programming Massively Parallel Processors: A Hands-on Approach*. Morgan Kaufmann Publishers, Elsevier, 2010.
- [21] A. Ahamed and F. Magoules, "Iterative methods for sparse linear systems on graphics processing unit," *IEEE 14th International Conference on High Performance Computing and Communication*, Liverpool, UK, pp. 836-842, June 2012.
- [22] G. Ortega, E. Garzon, F. Vazquez, and I. Garcia, "The biconjugate gradient method on GPUs," *J. Supercomput.*, Vol. 64, pp. 49-58, 2013.
- [23] R. Li and Y. Saad, "GPU-accelerated preconditioned iterative linear solvers," *J. Supercomput.*, Vol. 63, No. 2, pp. 443-466, 2013.
- [24] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 2006.
- [25] NVIDIA, CUBLAS Library. NVIDIA Corporation, 2701 San Tomas Expressway Santa Clara, CA 95050, 2012.
- [26] NVIDIA, CUSPARSE Library. NVIDIA Corporation, 2701 San Tomas Expressway Santa Clara, CA 95050, 2012.

Modelo Conceptual Inicial para Determinar la Calidad en Objetos de Aprendizaje Integrando un Enfoque a Servicios

César Velázquez Amador^a, Francisco Álvarez Rodríguez^a, Laura Garza González^a, Jaime Muñoz Arteaga^a, y Juan Pedro Cardona Salas^a

^aUniversidad Autónoma de Aguascalientes, Centro de Ciencias Básicas.
Av. Universidad 940, Col. Ciudad Universitaria, Aguascalientes, Ags., México, C.P. 20100.
vace555@hotmail.com, {fjalvar, lg, jmunozar, jpcardon}@correo.uaa.mx

2013 Published by *DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CCOMP-2013, www.cicomp.org

Resumen

Los Objetos de Aprendizaje son elementos que presentan dificultades especiales al momento de su evaluación al contar simultáneamente con las características de una aplicación de software y de un elemento instruccional. Los instrumentos para determinar la calidad en Objetos de Aprendizaje por lo general consideran la perspectiva del experto en el área, sin considerar la opinión del usuario, razón por la cual se propone un Modelo Conceptual Inicial, el cual permite determinar la calidad en estos recursos instruccionales, pero considerando la opinión del usuario por medio de la integración de la teoría de servicios.

Palabras clave: Modelo Conceptual, Objeto de Aprendizaje, Servicios.

1. Introducción

En este artículo se expone un modelo conceptual inicial para determinar la calidad en objetos de aprendizaje integrando un enfoque a servicios. Se trata de un Modelo Conceptual debido a que posee un origen en estudios relacionados encontrados en la teoría base, y se define como Inicial debido a que es la propuesta anterior a realizar una evaluación formal del modelo (en la cual se tiene contemplado realizar entre otros, un análisis exploratorio de factores y el estudio

confirmatorio), por lo que la propuesta es susceptible a presentar cambios al profundizar en la validación estadística. Una de las finalidades de generar este Modelo Conceptual Inicial es la oportunidad de poder comparar un modelo resultado de la teoría base (lo que podría considerarse debería ser) contra un modelo producto de la estadística formal (lo que realmente resulta y arroja la estadística).

El término Objeto de Aprendizaje (en inglés Learning Object) fue popularizado en 1994 por Wayne Hodgins cuando nombró al grupo de trabajo CedMA "Learning

Architectures, APIs and Learning Objects”. No existe una definición completamente aceptada del término Objeto de Aprendizaje (OA), una definición es: “Es una entidad digital o no digital, el cual puede ser usado, reusado o referenciado durante el aprendizaje soportado por la tecnología”. Existen 3 características básicas de un objeto de aprendizaje: Accesibilidad, Reusabilidad/Adaptabilidad e Interoperabilidad [1].

En la propuesta presentada se ha integrado la Teoría de Servicios, esto con la finalidad de vincular estrechamente al estudiante (usuario) en la determinación de la calidad de los OAs.

La Teoría de Servicios se refiere a todo aquello que es permanente y normal en la producción de un servicio [2]. Los servicios pueden ser definidos como la aplicación de competencias para el beneficio de otro, significando que un servicio es un tipo de acción, desempeño, o promesa que es intercambiada por valor entre el proveedor y el cliente [2]. Con relación a la calidad del servicio, esta se puede definir como la diferencia entre las expectativas del cliente sobre el servicio y el servicio percibido. Si las expectativas son mayores que el rendimiento, entonces la calidad percibida es menos que satisfactoria y por lo tanto se produce la insatisfacción del cliente [3], [4].

En el contexto del e-learning, se puede considerar a los estudiantes como el cliente final, ya que la satisfacción con un producto/servicio educativo es una de las consecuencias del intercambio entre los sistemas de e-learning y los estudiantes [5].

El empleo de un enfoque basado en servicios en la creación de un modelo que explique la calidad en OAs, se espera brinde una mayor satisfacción al usuario, esto debido a que la evaluación los tratará no simplemente como un producto, sino como un servicio. El Modelo Conceptual Inicial se vincula con la teoría de servicios, esto permite considerar el punto de vista del estudiante, como lo expresa Jim Spohrer del Centro de Investigación de IBM en Almaden “Es deseable que los estudiantes, que experimentan el servicio de primera mano usen medidas cualitativas para medir la calidad del servicio” [2].

2. Problema

Los Objetos de Aprendizaje son elementos que debido a su naturaleza presentan dificultades especiales al momento de su evaluación, ya que poseen tanto características de una aplicación de software como de un elemento instruccional [6]. La tarea de la determinación de la calidad en OA se ha abordado de distintas formas, una de las más populares es el uso de instrumentos

Tabla 1. Estudios base de la Investigación.

Estudios Base	
1	Byoung-Chan Lee, Jeong-Ok Yoon e In Lee, Learners' acceptance of e-learning in South Korea: Theories and results [10].
2	Parasuraman, Valarie A. Zeithaml y Arvind Malhotra, E-S-QUAL A Multiple-Item Scale for Assessing Electronic Service Quality [11].
3	Emmanouil Stiakakis y Christos K. Georgiadis, E-service quality: comparing the perceptions of providers and customers [12].
4	Roach, V. y Lemasters L., Satisfaction with online learning: A comparative descriptive study [13].
5	Robin Kay y Liesel Knaack, Investigating the Use of Learning Objects for Secondary School Mathematics [14].
6	DeLone W. H. y McLean E. R. T., The DeLone and McLean model of information systems success: A ten-year update [15].
7	Mercado del Collado Ricardo y López Grana-dos Mónica, Investigación institucional en el Instituto Consorcio Clavijero [16].

como el LORI [7], el cual permite realizar esta valoración desde la perspectiva del experto en el área, el problema que presenta esta evaluación es la falta de retroalimentación por parte del usuario. Otras propuestas de evaluación presentan un grupo de instrumentos y un proceso, como es el caso del trabajo de Erla Morales [8], pero al no considerar el punto de vista del estudiante, se pierde la posibilidad de obtener información de las deficiencias o errores detectados por el usuario [9].

Con el objetivo de dar solución al problema anterior, se ha propuesto el integrar la Teoría de Servicios en la Determinación de la Calidad de OAs.

3. Propuesta y Resultados

Para realizar la propuesta del Modelo Conceptual Inicial para Determinar la Calidad en Objetos de Aprendizaje Integrando un Enfoque a Servicios primeramente se hizo una extensa búsqueda de investigaciones similares en libros, bibliotecas electrónicas, memorias de congresos, trabajos de tesis y referencias de Internet, en las que se hubiera integrado la Teoría de Servicios en la Determinación de la Calidad en el e-learning de manera general, pero dando prioridad en los trabajos del área de Objetos de Aprendizaje. Los artículos más relevantes encontrados se presentan en la Tabla 1.

En los estudios presentados en la Tabla 1, se aborda

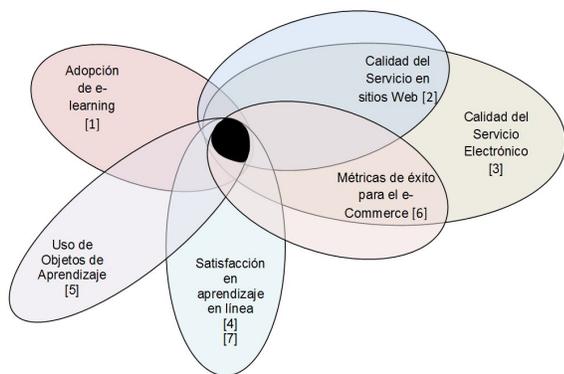


Figura 1. Áreas consultadas en la investigación.

la Calidad del Servicio en el e-learning y en los servicios electrónicos (No se encontraron investigaciones específicamente en el área de objetos de aprendizaje). Una vez terminada la búsqueda de trabajos similares, se analizaron, con la finalidad de hacer la propuesta formal del Modelo Conceptual Inicial.

El modelo tuvo su origen en la intersección de las áreas mostradas en la Figura 1. Cada una de las áreas estudiadas (Adopción de e-learning, Calidad del Servicio en sitios Web, Calidad del Servicio Electrónico, Satisfacción en aprendizaje en línea, Uso de Objetos de Aprendizaje, Métricas de éxito para el e-Commerce) cuenta con una o más investigaciones que han definido instrumentos de medición de la satisfacción (Tabla 1), estos estudios en su conjunto sirven como sustento para la propuesta del modelo propuesto en esta investigación.

El Modelo Conceptual Inicial consta de los constructores Calidad del Sistema, Calidad de Información, Calidad del Servicio y el Constructor Gozo, como parte de los elementos independientes; como parte de los elementos dependientes encontramos el Constructor Calidad Percibida del OA y el Constructor Satisfacción Obtenida (Figura 2).

El constructo calidad del sistema se conforma de las variables Tiempo de respuesta, Usabilidad, Fiabilidad, Disponibilidad y Seguridad. El constructo calidad de la información se conforma de las variables Organización, Integridad, Facilidad de comprensión, Relevancia y Elementos estéticos. El constructo calidad del servicio se conforma de las variables ayuda, interés y personalización (Figura 2).

El Modelo propuesto se encuentra diseñado para explicar la calidad en objetos de aprendizaje de tipo agregado (granularidad gruesa), en los cuales se integren distintas actividades de aprendizaje, evaluación y colaboración.

Con la finalidad de recolectar los datos para una fu-

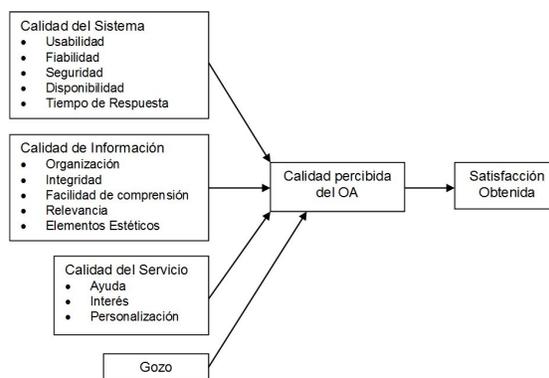


Figura 2. Modelo Conceptual Inicial para Determinar la Calidad en Objetos de Aprendizaje Integrando un Enfoque a Servicios.

tura validación estadística del modelo propuesto, se aplicó un instrumento desarrollado para determinar la importancia percibida sobre los factores que se propone pueden explicar la Calidad en Objetos de Aprendizaje considerando un Enfoque a Servicios. La aplicación del instrumento se hizo a finales del 2011 a 41 estudiantes de quinto semestre de la carrera de Ingeniero en Sistemas Computacionales y a 59 estudiantes de primero, quinto y séptimo semestre de la carrera de Licenciado en Tecnologías de Información de la Universidad Autónoma de Aguascalientes; de igual forma participaron 5 profesores de diversas instituciones (tanto nacionales como extranjeras) con amplia experiencia en el desarrollo y uso de OAs, para dar un total de 105 participantes. Para la aplicación del instrumento era necesario tener la seguridad de que el estudiante tuviera el conocimiento de que son los Objetos de Aprendizaje, para lo cual previamente se realizó una presentación sobre el tema, aún cuando el estudiante ya hubiera trabajado previamente con OAs. Para determinar la percepción de la importancia de cada factor se empleó una escala de Likert de 7 puntos, correspondiendo el 7 a completamente importante y el 1 a completamente no importante.

Los resultados se capturaron y analizaron con el SPSS ver 12.

Se aplicó un estudio de correlación sobre los datos obtenidos; del estudio realizado sobre todas las variables, se encontró una correlación positiva entre una gran cantidad de las mismas. Los resultados del estudio de correlación son parte del sustento necesario para realizar un futuro estudio EFA (Análisis Exploratorio de Factores), por medio del cual se puede comenzar con la validación estadística del modelo propuesto (es necesario encontrar una correlación entre algunas de las variables para justificar un estudio EFA). Acto seguido y fuera de una validación estadística del modelo, se aplicó

Tabla 2. Resultados del Estudio de Correlación sobre los constructores del Modelo Conceptual Inicial.

Constructor	Rango de Valores del Coeficiente de Correlación de Pearson
Calidad del Sistema	0.27 a 0.44
Calidad de Información	0.19 a 0.47
Calidad Percibida del OA	0.34 a 0.62
Calidad del Servicio	0.25 a 0.57
Gozo	0.28 a 0.44
Satisfacción Obtenida	0.54

un estudio de correlación sobre cada uno de los constructores que definen el Modelo Conceptual Inicial, esto con la finalidad de determinar los coeficientes de correlación y el nivel crítico entre las variables que integran cada constructor.

Se encontró que todas las preguntas de los constructores Calidad del Servicio, Satisfacción Obtenida y Gozo presentan una relación lineal significativa entre sí, es decir, se correlacionan significativamente.

La mayor parte de las preguntas de los constructores Calidad del Sistema, Calidad Percibida del OA y Calidad de Información presentan una relación lineal significativa, por lo que se correlacionan significativamente. Los rangos de valores en los que se encuentra el coeficiente de correlación de Pearson por constructor se presenta en la Tabla 2.

Dentro del constructor de Calidad del Sistema, la variable Seguridad no se correlacionó con las demás variables y en el constructor Calidad de Información, la variable Elementos Estéticos no se correlacionó con siete preguntas del constructor (de un total de 14), por lo que existe la posibilidad en base a futuros estudios de separar las variables Seguridad y Elementos Estéticos como constructores aparte, reacomodarlos dentro de otros o eliminarlos

Una posible explicación a la falta de correlación de la variable Seguridad puede deberse a que la Privacidad (la protección de la información personal) y la Seguridad (protección de los usuarios contra los riesgos de fraude y pérdidas financieras) han demostrado empíricamente tener un fuerte impacto en la actitud hacia el uso de servicios financieros en línea [17], pero esto pudiera ser distinto para aplicaciones educativas en las cuales pueden variar las prioridades del usuario.

4. Conclusión

En el artículo se expuso el Modelo Conceptual Inicial para Determinar la Calidad en Objetos de Aprendizaje Integrando un Enfoque a Servicios (Figura 2), así como los principales estudios que sirvieron como sustento para definirlo. La importancia de la propuesta radica en la posibilidad de integrar en la evaluación de los OAs la satisfacción del usuario, lográndose con esta retroalimentación del estudiante conocer algunos aspectos que pueden escapar a la vista del evaluador, todo con miras a mejorar estos recursos instruccionales.

Se presentaron los resultados de los estudios de correlación por constructor, que aún cuando no proporcionan una validación estadística del modelo, aporta una evidencia inicial de una redacción y agrupación adecuada de la mayoría de las preguntas extraídas de la teoría. El Modelo Conceptual Inicial propuesto, originado de la teoría base, se comparará en futuras investigaciones con el modelo resultado del estudio EFA. Tanto los resultados del estudio EFA, como el estudio confirmatorio permitirán obtener un modelo validado estadísticamente.

Por lo anterior, como estudio futuro se tiene contemplado el realizar la validación estadística del Modelo Conceptual Inicial desarrollando un estudio EFA, para luego continuar con los estudios comprobatorios.

Como estudio futuro se realizará una comparación entre los resultados de percepción de la importancia obtenidos entre los profesores con una alta experiencia en el desarrollo y uso de OAs contra la percepción de la importancia de los estudiantes, para de esta forma determinar si existe una diferencia significativa.

Se tiene planeada la aplicación de instrumentos para determinar la satisfacción esperada, la satisfacción obtenida y la calidad percibida empleando objetos de aprendizaje, el análisis de estos resultados se espera permitirá hacer también una Validación por Prueba de Concepto de Construcción de Artefacto [18].

Referencias

- [1] J. Aguilar, J. Zechinelli, y J. Muñoz, "Hacia la creación y administración de repositorios de objetos de aprendizaje," *IV Congreso Internacional de Ciencias de la Computación*, ENC, México, 2003.
- [2] J. Spohrer, P. P. Maglio, J. Bayley, and D. Gruhl, "Steps Toward a Science of Service Systems," *IEEE Computer Society*, pp. 71-77, 2007.
- [3] A. Parasuraman, V. A. Zeithaml, and L.L. Berry, "A conceptual model of service quality and its implication," *Journal of Marketing*, 49, pp. 41-50, Fall 1985.

- [4] B. R. Lewis, and V. W. Mitchell, "Defining and measuring the quality of customer service," *Marketing Intelligence & Planning*, 8(6), pp. 11-17, 1990.
- [5] L. H. Chen, and C. Lin, "Integrating Kano's model into E-learning satisfaction," *Industrial Engineering and Engineering Management*, 2007 IEEE International Conference, pp. 297-301, 2007.
- [6] C. E. Velázquez, J. Muñoz, F. J. Alvarez, y L. Garza, "La Determinación de la Calidad de Objetos de Aprendizaje," *VII Encuentro Internacional de Ciencias de la Computación ENC 2006*, pp. 346-351, 2006.
- [7] J. Nesbit, K. Belfer, and T. Leacock, "Learning Object Review Instrument (LORI), User Manual," E-Learning Research And Assessment Network, 2003.
- [8] E. Morales, F. García, A. Barrón, A. Berlanga, y C. López, "Propuesta de Evaluación de de Objetos de Aprendizaje," *II Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables (SPDECE)*, Barcelona, España, 2005.
- [9] C. E. Velázquez, J. Muñoz, F. J. Alvarez, F. J. Pinales, y L. Garza, "Estrategias de Gestión de la Calidad en el Desarrollo de Objetos de Aprendizaje," *Tercera Conferencia Latinoamericana De Tecnologías De Objetos De Aprendizaje LACLO 2008*, Aguascalientes, Ags., México, pp. 185-190, 2008.
- [10] L. Byoung-Chan, y Jeong-Ok, and L. In, "Learners' acceptance of e-learning in South Korea: Theories and results," *Computers and Education*, 53(4), pp. 1320-1329, 2009.
- [11] A. Parasuraman, V. A. Zeithaml, and A. Malhotra, "e-S-QUAL: a multiple-item scale for assessing electronic service quality," *Journal of Service Research*, 7(3), pp. 213-33, 2005.
- [12] E. Stiakakis, and C. K. Georgiadis, "E-service quality: comparing the perceptions of providers and customers," *Managing Service Quality*, Vol. 19 Iss: 4, pp. 410 -430, 2009.
- [13] V. Roach, and L. Lemasters, "Satisfaction with online learning: A comparative descriptive study", *Journal of Interactive Online Learning*, 5 (3), pp. 317-332, 2006.
- [14] R. Kay, and L. Knaack, "Investigating the Use of Learning Objects for Secondary School Mathematics," *Interdisciplinary Journal of E-Learning and Learning Objects*, Vol.4, 2008.
- [15] W. H. DeLone, and E. R. T. Mclean, "The DeLone and McLean model of information systems success: A ten-year update," *Journal Of Management Information Systems*, 19(4), pp. 9-30, 2003.
- [16] R. Mercado del Collado, y M. López Granados, "Investigación institucional en el Instituto Consorcio Clavijero," *3er Congreso Virtual Educa México, "Evaluación, Equidad y Calidad en la Educación a Distancia"*, San Cristóbal de las Casas, Chiapas, México, Marzo 2011.
- [17] M. M. Montoya-Weiss, G. B. Voss, and D. Grewal, "Determinants of Online Channel Use and Overall Satisfaction with a Relational, Multichannel Service Provider," *Journal of the Academy of Marketing Science*, 31(4), pp. 448-58, 2003.
- [18] M. Mora, "Descripción del Método de Investigación Conceptual: Tipo Conductual o Tipo Diseño, Versión 3.5," Departamento de Sistemas de Información, Universidad Autónoma de Aguascalientes, Agosto 2009.

Implementation of an Intelligent Algorithm in a PLC: A Practical Approach

Arturo Vázquez Gálvez^a, Oscar Mario Rodríguez Elias^a, Guillermo Valencia Palomo^a, Germán Alonso Ruiz Domínguez^a

^aDivisión de Estudios de Posgrado e Investigación, Instituto Tecnológico de Hermosillo
Avenida Tecnológico, El Sahuaro, Hermosillo, Sonora, México, 83170.
arturov47@yahoo.com.mx, {omrodriguez, gvalencia, gruiz}@ith.mx

2013 Published by *DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CCOMP-2013, www.cicomp.org

Abstract

This paper focuses on finding a way to implement an intelligent algorithm in an Allen Bradley PLC. It demonstrates such implementation with the purpose of improving a process, particularly water level control. The processing module followed during this research was the artificial neural networks (ANN). To achieve the programming, MATLAB was selected as the main support tool due to the powerful modules that it offers. In the paper, the steps that were necessary in order to achieve the proposed objective are described.

Keywords: Intelligent Algorithms, Neural Networks, Programmable Logic Controller, Intelligent Control

1. Introduction

Through history mankind has shared an interest on creating a machine that could imitate human behavior. It's not surprising that initial projects where conducted by the military during World War II [1]. Advances on automation technologies for industry solutions have been evolving at a great pace since the industrial revolution. Without a doubt one of the main "devices" that have had major impact on the industry is the Programmable Logic Controller (PLC). Since its beginnings on the 70's it has offered great advantages in process optimization that could not be done before. One of its main advantages is its adaptability to different

process at low cost. Even when the PLC performs an extraordinary work in diminishing the workload, industrial processes have become more complex, require to be performed at greater speeds and have better accuracy. Additionally, processes are submitted to revisions more frequently, which cause a shift in the variables involved [2]. To tackle these issues, new techniques have been implemented in order to reach better solutions; intelligent algorithms are among these new techniques.

Intelligent systems offer a great diversity of solutions; one of the most popular techniques for developing intelligent systems in industry during the last years is the artificial neural networks (ANN). ANN have been applied to several fields, like: medicine, biology, economy,

software development, etc. [3]. The conjunction of an intelligent algorithm in a PLC has found a solution to problems found every time more frequently as the processes become more complex. These problems relate to the fact that commonly PLC processor run the program based on variable values without any adaptability, even though the variables of the process have changed. Therefore, by implementing an intelligent algorithm the solution could adapt itself to changes occurring in the process.

In this paper we look for a solution that aids us to implement an intelligent algorithm, based in ANN, to a PLC that does not have a specialized module for it. The PLC model used in this research was the Compact Logix Allen Bradley #1768L43.

Following, in this paper, there is a brief explanation and definitions on different subjects that will be referred during the rest of the paper. Then the tools used on the experiment are listed and explained, to later present the steps required from the experiment that achieves the objective of the research. Finally the conclusions of the research are depicted.

2. Theoretical Framework

For choosing the intelligent algorithm to be used in this work, and to define the method for implementing it into a PLC, a literature survey was conducted following two main objectives:

1. To identify the type of intelligent algorithms being used to aid in industrial processes.
2. To identify the best way for programming the intelligent algorithm into the PLC.

Next we describe the main results of this literature survey. Particularly, we found the ANN as the best way to show the manner in which an intelligent algorithm can be programmed into a PLC for supporting complex process control. As well, we identified that ANN are not commonly programmed into PLCs, but are programmed in regular computers or in special hardware. So we required investigating and defining a method for implementing an ANN into the chosen PLC, which is described later in this paper.

2.1. Intelligent systems in industrial PLCs

The subjects that were analyzed belonged to article publications, books and academics projects that implement an intelligent algorithm in to a PLC in order to find solutions to real industry problems. The criterion that was followed in order to filter the information was search

terms, title content, abstract, etc. The results concluded that there are several ways to implement an intelligent algorithm into a PLC, which can be achieved according to the program languages that the PLC accepts.

The type of intelligent algorithms that are being implemented varies depending on the situation that is being managed, for example, genetic algorithms were employed to manage comfort variables (temperature, illumination, air quality, etc.) in the control of an intelligent building [4].

ANN were implemented to manage the control of a character recognition system [5]. Another project that implemented ANN used them for failure prevention in a conveyor belt, which only had a PLC to control the process. The processing capacity of the PLC was improved since an ANN can run its process in parallel, which caused a significant reduction in the processing time of the variables [6].

There are other types of algorithms that follow different techniques such as fuzzy logic, predictive control, etc. And the solutions are not restricted to implement only one type of algorithms; integrations of two or more intelligent algorithms can also be achieved. The benefits obtained vary depending on the areas that are being worked on, among the more relevant we can find are energy saving, time improvements, battery consumption, among others.

Since it is a relatively new method of improvement, documentation to achieve the implementation of an ANN into a PLC it is currently scarce. With the information and tools that were available at the moment in this project to achieve the objective, the ANN was selected as the intelligent algorithm that was implemented in a PLC to manage the level control in a cylindrical tank.

2.2. PLC Programming

Through its history the basic programming form of a PLC has been ladder programming; this language provides a graphic representation of the selected elements, making it easy to understand [7]. The graphic representation of the ladder programming is similar to the one that can be found in a relay circuit, this easy to understand language was made by design so its acceptance was wider. This wide acceptance helped to break the paradigm that existed at the time in process control [8]. The IEC 1131-3 norm was created by the International Electronic Commission to standardize the distinct languages that were being applied to program a PLC by several brands and have a broader compatibility between devices notwithstanding the manufacturer [9].

Among the basic types of programming one can find

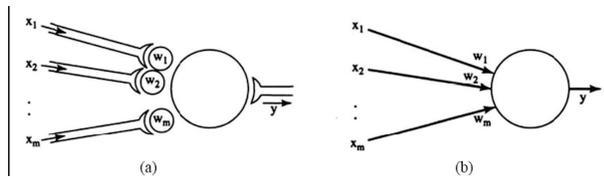


Figure 1. Neural model. a) Natural form representation. b) Abstract representation [13].

the following:

- Block Diagram, mostly used in control systems in sequences.
- Structured Text, used when the operations described in its programming are repeated constantly.
- Ladder Diagram, recommended for systems based on events.

The PLC used in this project manages RSLogix 5000 as its programming software; taking advantage of the Ethernet communication port of the PLC, and as it will be shown on chapter 4.2, the created ANN was imported as an “add-on” block to RSLogix 5000. To establish the communication between the PC and PLC, Matrikon OPC and OPC module from MATLAB were used so all the variables could be transferred. Ladder diagram language was used as the programming language to manage the variables in the control system as well as the ones inside the created ANN.

2.3. Artificial Neural Networks

The “traditional” concept of a neural network is utilized to describe the functions of the way human brain processes information that it receives through the human senses, so it can take a pertinent decision [10]. These decisions will improve according to the experience that a person gathers every time she faces the same situation. An ANN can also “learn” to take pertinent decisions by adjusting declared variables [11]. An ANN is composed of artificial neurons that can be related as nodes in an electrical circuit. These neurons are the main component of an ANN. Figure 1 presents a graphical interpretation of a single neuron model, where it can be noticed the distinct input variables (X_1, X_2, \dots, X_m) that stimulate the neurons depending on the declared weights (W_i). Inside a neuron there is a “black box” where the neuron generates a function with an output variable (Y) [12].

The following are several characteristics that ANN can offer when compared to other tools [13]:

- Applying an ANN is rather simple as its interpretation.

- Unlike statistical methods, ANN does not need to know every single data that is going to be involved in the system.
- An ANN offers the opportunity to divide its network into distinct parallel circuits, which allows having faster responses.
- An ANN can adjust the results when the data being received is updated.

It was necessary to analyze the characteristics of the particular PLC that is going to be used. In this research was the PLC 1768L43 compact Logix from Allen Bradley. According to its manual it is compatible with the IEC 1131-3 norm.

3. Methods and Materials

By following the IEC 1121-3 norm, the PLC used on this research allows its processor to be programmed by the distinct languages listed on Table 1, where some of their characteristics are mentioned.

Among the characteristics of the PLC that helped to reach the objective on this research, the following can be noticed: the PLC version and brand are compatible with MATLAB and even has a specialized option to select it in the OPC module. Additionally, Matrikon OPC also has a specialized module to configure PLC of the used brand. By using RSLogix 5000 to program the PLC, it was possible to import the ANN in to the ladder diagram.

The Compact Logix PLC allows importing add-ons blocks with additional instructions inside its programming. These modules allow for a more advanced programming like Visual Basic or C++. These tools allow for a more complex syntax that can support the program of an ANN.

In order to manage information between the PLC, with RSLogix 5000, and the PC, with MATLAB, a communication between these devices has to be established. This communication was handled by using the software Matrikon OPC, the user friendly interface of this software makes the communication easy to configure.

Other tools that were used in this project are modules contained in MATLAB. The Neural Network Toolbox was used to create the implemented ANN. SIMULINK managed the export of the ANN in to code that is compatible with RSLogix 5000.

The ANN was implemented in the PLC to control a cylindrical tank level process. The brand of these cylindrical tanks is Quanser, which contains a pump for water extraction, an amplifier and a data acquisition board. The

Table 1. Languages Allowed by the Compact Logix PLC Allen Bradley #1768L43

Language	Characteristics
Ladder Diagram	<ul style="list-style-type: none"> Graphical representation of the elements. Evaluates the networks based on the interconnected elements. Allowed elements: contacts, functions, functional blocks, among others.
Block Diagram Function	<ul style="list-style-type: none"> Graphic interface that allows the user to program the elements in a form that is similar to an electric circuit. Based on representation of interconnected functions. Similar interpretation, compared to traditional electric circuits.
Sequential Function Graphic	<ul style="list-style-type: none"> Mainly used to describe functions of sequential control. Based on GRAFCET norm (IEC848). Action properties that allows managing actions, pulses, memories, among others.
Structured Text	<ul style="list-style-type: none"> PASCAL like syntax. Variable designation, function control is managed with text succession.



Figure 2. Quanser data acquisition board [14].

data acquisition card send and receives the inputs and outputs of the ANN, this data controls the voltage that powers the pump for level control. This data acquisition card is shown in Figure 2.

4. Implementation of the ANN into the PLC

The ANN was implemented in the PLC to control a cylindrical tank level process. This was done with the objective of gather real data of the implemented ANN working on a real system. The PLC that was used in this project doesn't have a data acquisition card, therefore a constant communication with a PC was established, and so the sensor data that was sent by the Quanser tank could be read in real time. This communication was handled with MATLAB, which managed communications

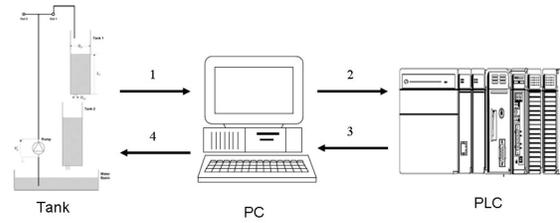


Figure 3. Data flow on the implemented system.

between the Quanser tank and the PLC.

The Quanser system is compatible with MATLAB, which makes possible to establish and analyze different configurations. However it does conflict with any external elements added on the working window. This occurs when the data acquisition card tries to convert the program from the working window into "Quanser code" for its data acquisition card. To achieve the implementation of the ANN, several adaptations where necessary in the MATLAB project so the Quanser system could communicate effectively with the rest of the devices like the Allen Bradley PLC.

Another problem we faced was that the PLC we used does not have an input/output card for managing analog data. So it was required a configuration based on the use OPC communication, as illustrated in Figure 3. First, the sensors on the tank send the data to a PC, which trough MATLAB sends the data to the PLC that contains the created ANN. Once the ANN finishes its process, the data is sent back to the PC, which sends the data back to the Quanser tank trough MATLAB.

4.1. Controlled System

The system to be controlled is a set of two coupled tanks. The cylindrical tanks are configured as shown in Figure 4. One tank is above the other so the output of the first one becomes the input of the second one. At the output of both cylinders there is a recipient that holds the liquid. The ANN created in this research controls the level of the top tank by managing the necessary voltage that is directed to the pump.

4.2. Design of the ANN

In order to create an ANN that controls the variables involved in the tanks process, first it is required to identify the variables as the desired outputs. Additionally, the appropriate ANN type to manage this process needs to be identified.

When implementing the ANN it has to be considered that this algorithm can handle a wide variety of ways to manage the variables. Once the inputs values are



Figure 4. Tank configuration.

defined, the value of the outputs has to be declared and by doing so, the activation level of the pump can be controlled.

As it was previously explained, the creation and manage of an ANN is easier to handle with the different tools that the modules of MATLAB offer. The ANN was created with the graphical interface, from where it is exported to the command window; then it is exported one more time to SIMULINK where it can be coded to RSLogix. The ANN designed is a multilayer perceptron network containing 4 neurons as input and 4 neurons as output which represents a number in binary code. The values were represented in binary code so the variable management in the PLC would be easier to handle. Additionally a hidden layer with 4 neurons is declared. Since the PLC was not made with the idea of handling ANN, it is possible to find errors when importing a complex ANN into RSLogix. In order to simplify the networks and achieve the control for distinct levels, different neural networks were designed (one for each level). As it is shown in Figure 5, when selecting a level on the system, a variable will select the proper ANN from the ones implemented on the PLC.

It is important to have a reliable database of possible values that can be obtained in the system. This data has to represent the behavior of the system so it can be properly simulated when creating the ANN. In this system, the water level can fluctuate between 0 to 30 centimeters for the level which is measured by a sensor with outputs values that vary from 0 to 5 V DC. The levels that are going to be managed in this project vary from 0 to 4 V DC. Regarding the output values of the ANN, they will be managed as voltages with a value of 9 V; the pump will raise the water level at an acceptable speed without overheating the pump motor; a value of 4 V will be declared for the minimum value, by doing so, the water level will drop at a controlled pace and the signal adjustment will be minimum.

In Table 2, the input and output values are listed for

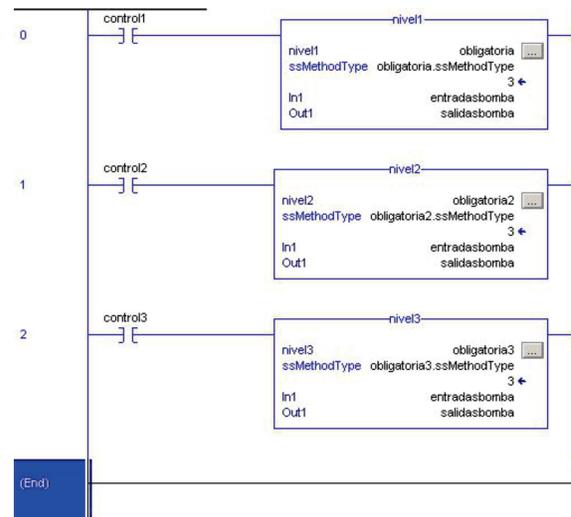


Figure 5. RSLogix code that shows three ANN imported as individual add-ons.

Table 2. Input/Output values, inferior level.

Level	Table of voltage values	
	Input	Output
Inferior Level	0	9
	1	9
	2	5
	3	4
	4	4

the inferior level so it shows the logic that is being followed in the other levels.

It is common to wonder if by adding more neurons in the hidden layer it will optimize the results but this is not always the case; when adding more neurons, it is possible that the system will require more learning cycles. In a similar way, if the learning curve is diminishing in an unnecessary manner, it can reflect in unwanted results [15].

The training parameters will remain the ones that MATLAB has as default values for this type of ANN. The weights for the layers are also managed by the neural network toolbox of MATLAB.

Once the ANN is completed, a diagram that shows a preview of it is displayed, as shown in Figure 6, where the layers are displayed along with their respective neurons.

4.3. Coding the ANN in RSLogix

Once the ANN was created, the next step was to translate it into a code recognizable by RSLogix 5000. To achieve this, the ANN needs to be exported to the MATLAB command window where it can be exported to SIMULINK module with the following command:

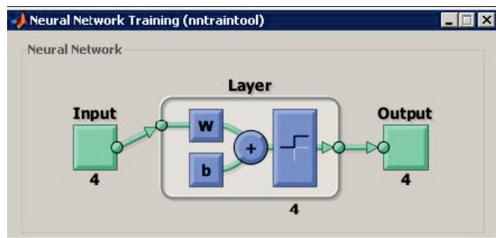


Figure 6. ANN preview, displayed by Neural Network Toolbox in MATLAB.

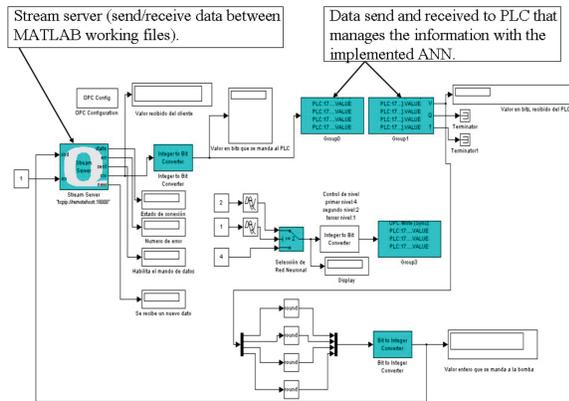


Figure 7. MATLAB working file that handles data between the PC and the PLC.

“gensim(name of the network)”. In SIMULINK, the imported ANN needs to be compressed into a subsystem and then, by configuring the available SIMULINK, to export the subsystem into a code compatible with RSLogix 5000.

The created ANN can be imported to RSLogix 5000 as an add-on and can be added to a ladder logic program. Once the ANN has been imported, all the different variables that are involved in the process need to be configured in RSLogix 5000. It is worth noticing that the variables that represent the inputs and outputs of the ANN need to be declared as REAL with a $1 \times n$ dimension so they work as a matrix, where n is the number of input or output neurons.

Next, the working file which handles the project that controls the tank level in MATLAB needs to be configured, Figure 7 shows this file project in MATLAB. Since the ANN was trained using binary code representing the real values, it is necessary to add a block that converts the decimal values into binary code and vice versa. These values represent the inputs and outputs of the ANN, so they will need to be managed as a 1×4 matrix, so it continues with the logic that was used when the ANN was created.

Additionally, it is necessary to configure the OPC communication blocks that connect the working files, the PLC, and the tanks process system.

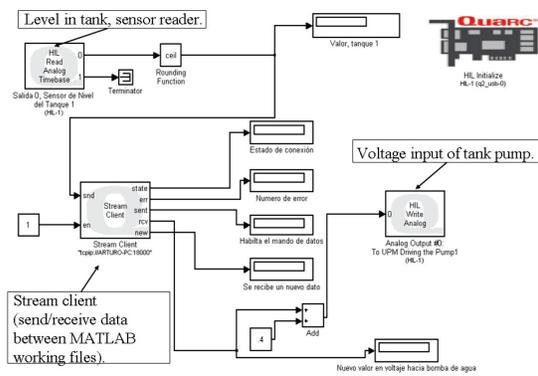


Figure 8. MATLAB working file that handles data between PC and Quanser tank.

All the blocks that are involved on the working window with the system need to be adjusted at the same working simulation time, in our case it was 0.05, so they synchronize. If they are not synchronized, the system returns unwanted results.

Once all the programming has been completed, the MATLAB working file that manages data between PC and the Quanser tank has to be downloaded into the Quanser data acquisition board, Figure 8 shows the MATLAB working file that handles data between PC and the Quanser tank as displayed in the PC window.

To download the file into the Quanser data acquisition board, the specific board that is being used has to be selected and the simulation parameters need to be configured for the correct system that the PC is running, for example its bits properties (32 or 64 bits). The programmed ANN will be controlling the levels for three different phases, the highest level will be selected (at twenty centimeters) first and remain at it for sixty seconds, then the medium level will be selected (at fifteen centimeters) and also remain at it for sixty seconds, finally the last inferior level is selected (at ten centimeters); to do this, a delay will be added into the MATLAB working file.

4.4. Discussion of results

It is worth noticing that the main objective of this research “implementation of an intelligent algorithm in a Compact Logix PLC #1768L43 by Allen Bradley” was achieved.

By analyzing the PLC user’s manual it was possible to note the advantages and disadvantages that this particular PLC possesses, specifically that it follows the IEC 1131-3 norm and communication with other devices can be established via Ethernet.

The Quanser tank system that was controlled had its restrictions regarding the use of external devices but

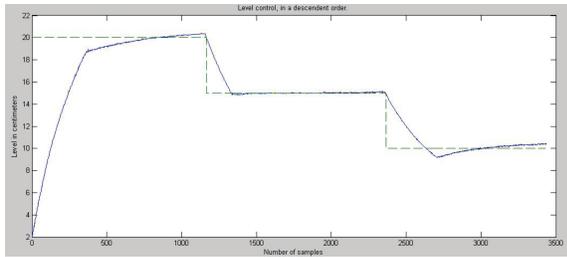


Figure 9. Controlled system behavior in 3 minutes.

a solution was found using support software (Matrikon OPC) and modules that MATLAB includes (OPC toolbox).

Besides noting the benefits that can be obtained by implementing an intelligent algorithm in a PLC, the systematic review of the applications that are being used in the industry helped to confirm that for each type of problem, there is a specific solution; where several factors intervene like the type of algorithm, PLC characteristics, support tools, among others. Therefore, it's not commendable to use a single type of algorithm to optimize every type of problem.

To define the type of intelligent algorithm to be used, different techniques were analyzed, like genetic algorithms, fuzzy logic, among others and due to its characteristics, described on chapter 2.3, the ANN was selected as the one to be implemented. The algorithm was programmed using the neural network toolbox included in MATLAB due to the advantages that its graphical interface offers. Another module included in MATLAB that was important is SIMULINK. With this module, it was possible to export the created ANN into code that is compatible with RSLogix 5000.

Three ANN were implemented into the PLC in order to control three different levels of the coupled tanks system. This solution was employed because, when importing a complex ANN into RSLogix 5000, error messages of invalid expressions in the programming appeared when downloading the project to the PLC, therefore, to avoid this situation, three ANN were managed, one for every single level to be controlled.

The use of all different programs, Quanser, MATLAB, RSLogix, and Matrikon OPC can be difficult for the PC to handle. Unexpected delays occurred when running the process, which caused significant errors. Initially, the PC that was intended to be used on this research had an Intel Dual Core with 1.6 GHz and 1.5 GB of RAM. The PC that was used to run the control process without problems had an Intel Core i5 2.4 GHz processor with 6 GB of RAM.

Figure 9, shows the behavior of the controlled system with a graph that is obtained at real time with MATLAB

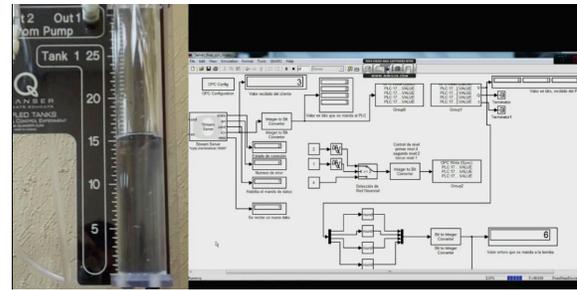


Figure 10. Controlled system at middle level (15 centimeters), showing the cylindrical tank and MATLAB working file.

scopes when the process was running. The Y axis represents the water level in centimeters and X axis represents the number of samples that were taken in a 3 minutes period. As it was previously mentioned in Section 4.2 the voltage level to control the system vary from 9 V as maximum to 4 V as minimum. The transfer from level to level is at the minute mark that passes in the total running clock, until the inferior level is reached. The dotted line represents the desired level and the stream line represents the current level of water in the tank.

Figure 10 shows an image where the process is currently controlled at 15 centimeters. The left side shows the actual tank and the right side shows the MATLAB working file.

5. Conclusions

After analyzing the results of the systematic review it can be concluded that the implementation of an intelligent algorithm into a PLC in order to optimize process like the control of hydraulic systems, solar panels or robotic arms, etc, can be achieved, in this research it was implemented to control the level in a cylindrical tank.

To make the above possible, there are different ways to achieve it, among which we followed the languages allowed by the PLC, specifically the ladder diagram. It is important to notice that MATLAB comes as a relevant support tool that offers many possibilities to reach the desired objective, thanks to all the modules it supports.

During the research, some problems emerged for which their solution caused to a certain point to adjust the parameters of the research itself. Among these issues, we can find the distinct limitations that need to be considered when programming in ladder diagram language for the Allen Bradley PLC, mainly the import of the ANN code that it is created in MATLAB. Other issues were more simple, like a delay when reading the water level sensors, which were caused by the limited processing power of the PC that was being used. To

solve this it was just a matter of using a more powerful PC.

It was also relevant to consider the limitations of the tanks system by Quanser since it is not very “friendly” when adding distinct process to what they are programmed for, specifically when transferring the program to the data acquisition card. It was necessary to create additional process so the system could adapt to the Quanser requirements.

It is also important to notice the wide spectrum of solution that MATLAB offers along all its modules. Using these modules made possible to create the ANN, adapting it to the Quanser tanks system and importing the code to RSLogix into the PLC and all of it in a practical manner due to its different graphical interfaces.

Acknowledgements

Thanks to the support from scholarships system by DGEST and the “Divisin de Estudios de Posgrado e Investigacin from Instituto Tecnolgico de Hermosillo”. This project was partially financed by PROMEP (Oficio PROMEP/103.5/12/4633).

References

- [1] H. Moravec, and Mind Children, The future of Robot and Human Intelligence. Harvard University Press, p. 222, 1988.
- [2] F. Lewis, J. Campos, and R. Selmic, Neuro-Fuzzy Control of Industrial Systems with Actuator Nonlinearities, Philadelphia: SIAM, p. 244, 2002.
- [3] K. Suzuki, Artificial Neural Networks Industrial and Control Engineering, India: InTech, p. 490, 2011.
- [4] D. Kolokotsa, G. S. Stavrakakis, K. Kalaitzakis, and D. Agoris, “Genetic algorithms optimized fuzzy controller for the indoor environmental management in buildings implemented using PLC and local operating networks,” *Eng. Appl. Artif. Intell.*, Vol. 15, No. 5, pp. 417428, Sep. 2002.
- [5] I. Topalova, and A. Tzokev, “Automated Texture Classification of Marble Shades with Real-Time PLC Neural Network Implementation,” *Neural Netwokrs (JCNN)*, 2010 Int. Jt. Conf., pp. 18, 2010.
- [6] I. S. Jung, B. Mulman, D. Thapa, L. Koo, J. H. Bae, S. Hong, S. Yeo, C. M. Park, S. C. Park, and G. N. Wang, “PLC Control Logic Error Monitoring and Prediction Using Neural Network,” *Fourth Int. Conf. Nat. Comput.*, pp. 484488, 2008.
- [7] C. Peshek, and M. Mellish, “Recent Developments and Future Trends in PLC Programming Languages and Programming Tools for Real-Time Control,” *Forth-Fifth Annual Conference of Electrical Engineering Problems in the Rubber and Plastics Industries*, pp. 219230, 1993.
- [8] A. A. Ghandakly, M. E. Shields, and A. M. Farhoud, “Enhancement of Exiting PLC’s with an Adaptive Control Technique,” *Industry Applications Conference, Thirtieth IAS Annual Meeting, IAS '95*, pp. 16341640, 1995.
- [9] E. M. Perez, J. M. Acevedo, C. Silva, y J. Armesto, *Automatas Programables y Sistemas de Automatizacion*, 2nd ed. Mxico, D.F., Grupo Alfa Omega, p. 1120, 2010.
- [10] N. K. Kasabov, *Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering*, 2nd ed., Cambridge, Massachusetts: MIT Press, p 581, 1998.
- [11] J. Jantzen, *Foundations of Fuzzy Control*, Chichester, UK: John Wiley & Sons, Ltd, p. 220, 2007.
- [12] B.M. del Brio, y A. Sanz Molina, *Redes Neuronales y Sistemas Borrosos*, 3rd ed. Madrid: Alfa Omega, p. 404, 2007.
- [13] T. Munakata, *Fundamentals of the New Artificial Intelligence Neural, Evolutionary, Fuzzy and More*, 2nd Ed., Cleveland, OH: Springer, p. 266, 2008.
- [14] Q. Inc., “Q2-USB User Manual.” Canada, p. 14, 2010.
- [15] P. I. Viñuela, y I. M. G. León, *Redes Neuronales Artificiales Un Enfoque Practico*, Madrid: Prentice Hall, p. 229, 2008.

Análisis de Factores que Afectan el QoS que Ofrecen las WSN Aplicado a Entornos de Salud

Raymundo Buenrostro Mariscal^{a,b}, Juan Iván Nieto Hipólito^a, Juan Antonio Guerrero Ibáñez^b, María Cosío León^a, Mabel Vázquez Briseño^a, y Juan de Dios Sánchez López^a

^aUniversidad Autónoma de Baja California, Facultad de Ingeniería, Arquitectura y Diseño.
Carr. Ensenada-Tijuana 3917, Col. Playitas, Ensenada, B.C., México, 22860.

{rbuenrostro, jnieto, maria.cosio, mabel.vazquez, jdios}@uabc.edu.mx

^bUniversidad de Colima, Facultad de Telemática.

CAv. Universidad 333, Col. Las Víboras, Colima, Col. México, CP. 28040.

antonio_guerrero@ucol.mx

2013 Published by *DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Resumen

Actualmente se han considerado las Redes Inalámbricas de Sensores (WSNs) para la recolección de signos vitales de personas. Sin embargo, por el tipo de datos se considera una aplicación crítica, lo cual exige una comunicación eficiente. Dentro de los protocolos de comunicación de red, la capa de transporte y la sub-capa de acceso al medio juegan un papel importante en el desempeño de la comunicación. En este trabajo se analizó el protocolo TCP y el estándar IEEE 802.15.4 bajo un escenario de aplicación médica, específicamente, los factores que afectan las métricas de tasa de entrega de datos y el retardo de transmisión. Los resultados demostraron que el tamaño del paquete es un factor crucial en el diseño de redes bajo IEEE 802.15.4. Además, TCP fue la causa principal del bajo desempeño de las métricas, debido a la necesidad de una sesión de datos para transmitir los paquetes del nodo sensor.

Palabras clave: IEEE 802.15.4, QoS, TCP, WSN.

1. Introducción

Las Redes Inalámbricas de Sensores (WSNs, Wireless Sensor Networks) son redes compuestas por dispositivos interconectados entre sí, que trabajan de forma cooperativa para recolectar datos de interés y transmitirlos de forma inalámbrica a otro punto. En los últimos años se ha incrementado los trabajos que exploran su uso en aplicaciones médicas, por ejemplo,

la supervisión de la salud de forma remota [1], [2]. Las WSN pueden recuperar signos vitales de personas en hospitales o en casa, como temperatura corporal, saturación de oxígeno, ritmo cardíaco, electrocardiograma (ECG), entre otros, los cuales son transmitidos hasta un punto central para ser analizadas por personal médico. Bajo este escenario y por la naturaleza crítica de los datos, éste tipo de aplicaciones demandan una comunicación más eficiente. Por lo tanto, existen diferentes

requerimientos de transmisión de datos que deben ser cubiertos por las WSNs: métricas de calidad de servicio (QoS, Quality of Service) tales como retardo de transmisión del paquete y la tasa de entrega de datos; junto con la métrica de "Equidad" en el uso de los recursos (referida como la oportunidad que tiene cada nodo a enviar sus datos). Sin embargo, las WSNs tienen limitaciones importantes para cumplir con estos requerimientos [3], debido a su naturaleza inalámbrica y a su medio compartido, donde todos los nodos compiten por el acceso al medio para transmitir su datos. Como resultado, la transferencia de datos está sujeta a diferentes problemas que afectan la calidad de la comunicación (colisiones, degradación de la señal por interferencias, congestión del enlace, entre otras). Además, la WSN puede estar formada por múltiples nodos sensores que mantienen flujos de datos la mayor parte del tiempo, provocando que se acentúen éstos problemas. Lo anterior conduce a la pérdida de paquetes y al incremento del tiempo de transmisión de los datos; lo cual tiene un impacto directo en el desempeño de la red y por ende en los niveles de QoS y Equidad que pueden ofrecer las WSNs.

Dentro de los protocolos de comunicación de las WSNs, la capa de Transporte y la subcapa de acceso al medio (MAC, Medium Access Control) juegan un rol importante en el desempeño en la entrega de datos [4]. Por lo tanto, es importante analizar los factores que afectan el desempeño de las WSNs, considerando Transporte y MAC, con el fin de alcanzar los mejores niveles de desempeño y habilitar este tipo de redes para aplicaciones médicas.

Existen varios trabajos que han analizado el desempeño de las WSNs. En [4] se evalúa el desempeño de diferentes protocolos de transporte que utilizan los protocolos IEEE 802.11 e IEEE 802.15.4 por medio de simulación. Sin embargo, el escenario de experimentación fue limitado, ya que la cantidad de nodos sensores (4 nodos) y el tamaño del paquete fueron siempre los mismos. El máximo desempeño y el mínimo retardo de transmisión alcanzado por el estándar IEEE 802.15.4 son investigados en [5]. Sin embargo, sólo se consideran un solo nodo como fuente de tráfico. En [6] se analizaron los efectos de variar la cantidad de nodos en la red, el tamaño del paquete, la frecuencia de muestreo, y el ciclo de transmisión en una red basada en IEEE 802.15.4 que consiste de sensores de monitorización de ECG. Sin embargo, no incluye en su análisis el efecto de la capa de transporte. Ninguna de las propuestas revisadas presenta resultados sobre la métrica de Equidad. Además, excepto en [6], ninguna de ellas considera aplicaciones médicas para modelar la fuente

de tráfico. Por lo tanto, el objetivo de este trabajo es analizar los factores que afectan el QoS y la Equidad que ofrecen las WSNs en aplicaciones médicas, que utilizan el protocolo de transporte TCP [7] y el estándar IEEE 802.15.4 [8].

El resto de este trabajo está organizado como sigue. La Sección 2 describe aspectos técnicos relacionados con una WSN aplicada a la salud. El escenario de trabajo se presenta en la Sección 3. En la Sección 4 se detallan las métricas de evaluación y los casos de estudio. El análisis de resultados se explica en la Sección 5. Finalmente, en la Sección 6 se dan las conclusiones y el trabajo futuro.

2. WSN para aplicaciones médicas

Cuando se proponen las WSNs para la monitorización de pacientes de forma remota, el desempeño en la transmisión de los datos es mucho más importante que en otras aplicaciones. Ya que los datos son utilizados para la atención médica personal y esto lo vuelve una aplicación crítica. Por lo tanto, es necesario que el diseño de la WSN considere los protocolos de la capa de transporte y la subcapa MAC como se indico en la introducción. Los protocolos de transporte, aseguran la comunicación de los datos extremo a extremo y pueden ofrecer un trato especial a los datos de la aplicación, ya sea para programarlos y/o priorizarlos. Por su parte, los protocolos MAC complementan el proceso de comunicación entre los nodos de la red, administrando el acceso al canal común. Además, los dispositivos que forman las WSNs deben trabajar en el entorno de las personas. Lo cual exige otros requerimientos importantes en el diseño de la WSNs; tales como: los dispositivos que forman la WSN deben ser de tamaño reducido, de bajo consumo de energía, con mínimos niveles de mantenimiento, de bajo costo y que permitan la movilidad de las personas.

2.1. Protocolo de Control de Transmisión (TCP)

TCP es un protocolo orientado a conexión que opera en la capa de transporte y ofrece servicios de comunicación extremo a extremo entre dos nodos de la red. Además, incluye servicios de confiabilidad en la entrega de datos, control de flujo, establecimiento de sesión, entre otros; con el objetivo de proveer una transmisión confiable [7].

En TCP cuando dos procesos, que se ejecutan en nodos diferentes, desean comunicarse, necesitan establecer una conexión antes de iniciar el envío de los

datos. TCP utiliza un mecanismo de sincronía de “tres-pasos”: solicitud de conexión, respuesta a la solicitud de conexión y un mensaje de confirmación por parte del nodo solicitante. Sólo si la respuesta fue exitosa, el nodo emisor inicia con la transmisión de sus datos. Para las funciones de confiabilidad, TCP necesita recibir un mensaje de acuse de recibo, del nodo receptor, por cada paquete que se envía, para asegurar una entrega exitosa. Este mensaje se le conoce como mensaje de Ack (del inglés Acknowledgment). Si no lo recibe en un determinado tiempo, debe retransmitir dicho paquete. En suma TCP está diseñado para ofrecer una comunicación confiable proporcionando control de congestión y un mecanismo de confiabilidad en la entrega de datos.

2.2. Estándar IEEE 802.15.4

Una de las tecnologías utilizadas para desarrollar las WSNs es el estándar IEEE 802.15.4. Este estándar fue diseñado para dispositivos que necesiten una transmisión de datos de corto alcance con una velocidad de datos moderada (250 kbps a 2.4 GHz), con muy bajo consumo de energía, de baja complejidad, portable y de bajo costo de implementación [8]. IEEE 802.15.4 no fue diseñado para competir con otras tecnologías inalámbricas, si no que complementa la gama de tecnologías inalámbricas disponibles [9].

IEEE 802.15.4 comprende las especificaciones de la capa Física y la subcapa MAC para redes inalámbricas de área personal. Para el acceso al canal de comunicaciones, el estándar utiliza un mecanismo de acceso múltiple por detección de portadora con prevención de colisiones (CSMA/CA, del inglés Carrier Sense Multiple Access with Collision Avoidance). Cada vez que un dispositivo de la red necesita transmitir un paquete de datos, éste espera un periodo de tiempo aleatorio o de Backoff, antes de verificar el canal. Si el canal está libre, el dispositivo transmite inmediatamente su paquete; de lo contrario, éste espera por otro periodo de Backoff antes de intentarlo de nuevo. Este proceso se repite hasta que se alcanza el máximo número de intentos de acceso al canal; si esto ocurre, se declara una falla de acceso al canal.

El estándar IEEE 802.15.4 ofrece servicios de confiabilidad, similar a TCP, entre dos nodos vecinos que están en la misma área de cobertura (a un salto como máximo). Este servicio puede ser activado o desactivado de acuerdo a las necesidades de la aplicación desde capas superiores (por ejemplo la capa de transporte). Puede ser activado para incrementar los niveles de confiabilidad en la entrega de datos, o eliminarla para reducir el retardo de transmisión extremos a extremo.

Tabla 1. Mediciones de señales biomédicas [10]

Señal	Tasa muestras/seg	Resolución bits/muestra	Datos bits/s
ECG	1250	12	15000
Nivel de glucosa	50	16	800
Ritmo cardíaco	25	24	600
Temp. corporal	5	16	80

Por lo mencionado anteriormente, el estándar IEEE 802.15.4 se concibe como una opción adecuada para crear WSNs; ya que se alinean a las necesidades principales de las aplicaciones médicas. Sin embargo es necesario su análisis al respecto del desempeño de las métricas de QoS y Equidad para confirmar su factibilidad.

2.3. Caracterización del tráfico

Conocer el tipo de tráfico que generan las aplicaciones es importante, para determinar el tipo de atención que los protocolos de comunicación deben ofrecerle para su transporte dentro de la WSN. Para lograr esto, existen varios factores que deben ser considerados, tales como: *tipo*, *cantidad* y *frecuencia* de los datos. Para el caso de la supervisión de la salud, el tráfico es generado por los sensores que recolectan las señales fisiológicas de los pacientes. Dichas señales se necesitan estudiar desde el punto de vista eléctrico y desde el punto de vista de la aplicación médica. Para éste fin, la señal ECG es seleccionada; con la intención de que éste estudio pueda ser aplicable a otras señales médicas.

La señal ECG es el resultado de las señales eléctricas que produce el corazón en el rango de frecuencia de 0.01 a 250 Hz (Figura 1). Estas señales eléctricas son generadas continuamente, y por lo tanto, deben ser periódicamente muestreadas para digitalizarlas. La frecuencia de muestreo y los métodos de digitalización juegan un papel importante en determinar las características del tráfico. Por ejemplo, si la señales eléctricas del corazón se muestrean a 500 muestras por segundo (2 veces la frecuencia máxima) y cada muestra tiene un tamaño de 10 bits; entonces, se espera que el flujo de datos de la aplicación ECG genere un tráfico de 5000 bps (bits/segundo). La Tabla 1 muestra una propuesta de requerimientos de tráfico, para algunas señales biomédicas [10]. Observe que la señal con más requerimientos de comunicación es el ECG. Por el contrario la temperatura corporal es la que menos bits por segundo genera.

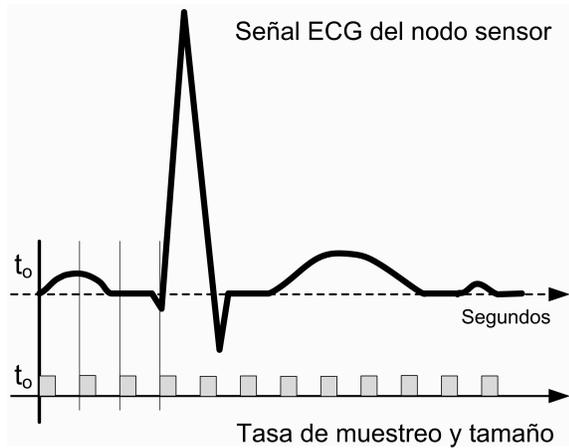


Figura 1. Señal ECG y digitalización.

Desde el punto de vista de la aplicación, el tipo tráfico se debe generar de forma periódica durante todo el tiempo que incurra la supervisión médica. A éste tipo de tráfico se le considera de carácter crítico, con flujos de datos de tipo continuo. Además, el usuario de la aplicación espera recibir los datos lo más rápido posible. Por lo tanto, el retardo de transmisión de los datos es un importante parámetro. En [11] se establece que el valor de retardo de transmisión de un ECG de tres electrodos debe ser menor a 500 ms y una tasa mínima de 4 Kbps. Por su parte en [12] indica que una señal ECG genera una tasa de datos de 1 a 20 Kbps y el retardo de transmisión debe ser menor a 1 segundo. Por lo tanto, consideramos para nuestro análisis un requerimiento QoS de: al menos 5 Kbps para la tasa de datos y retardo de transmisión de paquete no mayor a 500 ms.

Estos requerimientos de comunicación y el tipo de tráfico de las señales ECG son un dato importante al momento de diseñar los protocolos de comunicaciones para garantizar los niveles QoS que la aplicación médica necesita.

3. Escenario de trabajo

Nuestro escenario de trabajo involucra una WSN con una arquitectura de red que incluye las capas de aplicación, de Transporte y la sub-capa MAC (Figura 2). La capa de aplicación involucra los requerimientos de la aplicación médica, detallados en la sección anterior. Cabe resaltar que el objetivo de la aplicación médica es para la supervisión de la salud del paciente y no con fines de diagnóstico. Además, se propone que cada nodo sensor envíe una señal ECG (de tres electrodos) de forma constante y permanente. La cantidad de tráfico que cada nodo puede generar se establece en 18 kbps durante todo el tiempo de experimentación, mayores

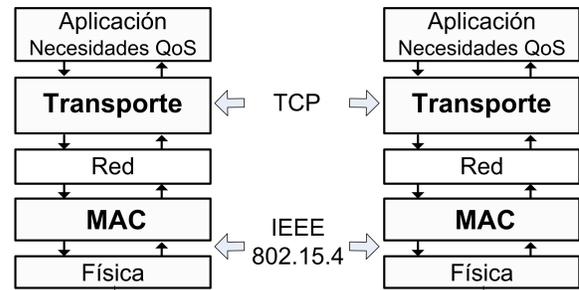


Figura 2. Arquitectura del escenario de trabajo.

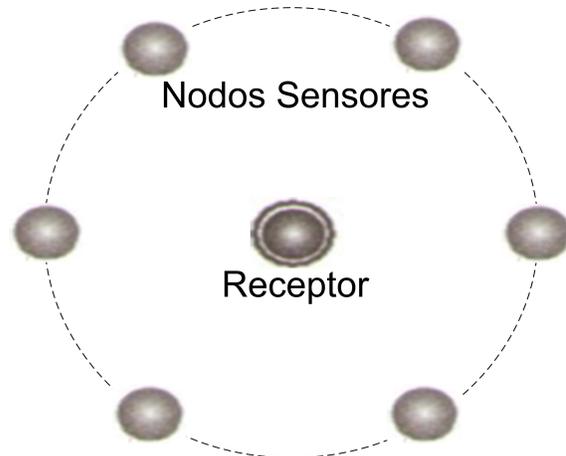


Figura 3. Topología de la red WSN.

a los establecidos en la Tabla 1, con el fin de considerar casos irregulares de pacientes con problemas de arritmias, como una taquicardia. En lo referente a la capa de Transporte se propone el protocolo TCP, como medio para establecer la comunicación a nivel de proceso, debido a sus prestaciones de control de flujo y confiabilidad en la entrega de paquetes.

Par la subcapa MAC y capa Física se utiliza el estándar IEEE 802.15.4 en su versión de 2.4 GHz a una tasa de 250 Kbps, que cubre las especificaciones de ambas entidades. IEEE 802.15.4 en modo No-Beacon es elegido para la operación de la red, el cual ofrece el máximo desempeño y el mínimo retardo de transmisión [5], [6]. En todos los casos se utiliza el mecanismo CSMA/CA para usar el canal inalámbrico.

En cuanto a la topología de la red esta se basa en una configuración en estrella. Los nodos sensores están distribuidos alrededor del nodo receptor a la misma distancia, como se aprecia en la Figura 3. La red se propone “totalmente conectada”, esto significa que todos los nodos de la red puedan escuchar sus transmisiones entre sí; para evitar el problema del nodo oculto presente en las redes inalámbricas.

4. Métricas de desempeño y casos de estudio

Para conocer si el protocolo TCP y el estándar IEEE 802.15.4 cubren las necesidades de transmisión del tráfico ECG se propone dos métricas de evaluación:

- Métricas QoS:** tasa de entrega de datos (Td) y retardo de transmisión de paquete (Rt).
- Equidad:** Se refiere a las oportunidades de transmisión de paquetes de cada nodo hacia el destino o Sink.

Las métricas fueron seleccionadas por que integran los principales parámetros de evaluación del desempeño de una red de comunicaciones [13]. Además, se consideran los requerimientos de QoS para la señal ECG de: 500 ms para el retardo de transmisión y de 5 Kbps para la tasa de entrega de datos.

La tasa de entrega de datos (Td) se define como la cantidad de bits recibidos en el nodo destino por unidad de tiempo, desde cualquier nodo sensor. Este valor es conocido como desempeño de la red o Throughput (por su significado en inglés). Td está dada en bits por segundo (bps) y se calcula como:

$$Td = \frac{\text{bits recibidos}}{\text{Ventana de tiempo}} \text{ bps} \quad (1)$$

El retardo de transmisión de paquete (Rt) es el intervalo de tiempo desde que el paquete i es enviado por el nodo sensor hasta el momento en el cual el paquete i es recibido en el nodo receptor:

$$Rt(P_i) = \text{tiempo de llegada} - \text{tiempo de salida} \quad (2)$$

La *Equidad* de la red es medida como la cantidad de paquetes recibidos en el nodo receptor de cada nodo sensor por unidad de tiempo. A través de esta métrica podemos conocer que señal médica se está reportando con más frecuencia hacia el nodo destino.

La evaluación de todas las métricas será efectuada en el nodo receptor y serán calculadas tanto para TCP como para IEEE 802.15.4.

4.1. Casos de estudio

Con la intención de conocer los parámetros que proporcionan el mejor desempeño de las métricas bajo el escenario planteado, se proponen dos casos de estudio:

- Tamaño del paquete de datos:** carga útil dentro del paquete que corresponde a los datos de la señal ECG muestreada.
- Cantidad de nodos sensores:** número de nodos que forman la WSN, los cuales generan datos de manera continua.

Estos casos están propuestos debido a que afectan la operación del estándar IEEE 802.15.4 y del protocolo TCP. En el caso de IEEE 802.15.4 nos interesa conocer el desempeño del mecanismo de acceso al medio CSMA/CA. El desempeño de este mecanismo se ve seriamente afectado por el la cantidad de nodos que compiten por el uso del canal inalámbrico; en este sentido, la cantidad de nodos y el tamaño de paquete juegan un rol crítico. Por un lado, más nodos sensores intentando transmitir se traduce en un incremento del nivel de contienda en la red. Por otro lado, un mismo nodo sensor puede causar mayor o menor contienda de acuerdo a la cantidad de paquetes por unidad de tiempo que desee transmitir.

En el caso de TCP, queremos conocer cuál es el desempeño alcanzado en el transporte de los datos. Por ejemplo, se requiere establecer una sesión para cada nodo que necesiten transmitir; por ello, mayor número de nodos mayor es la demanda hacia el mismo nodo receptor, lo cual puede incrementar el retardo de transmisión o que no se logre la sesión. TCP espera un paquete de Ack por cada paquete enviado. En consecuencia, mayor número de paquetes provoca una mayor cantidad de procesos y tráfico por la red. Este tráfico además de afectar las métricas de retardo incrementan el nivel de contienda del a nivel MAC.

Para cada caso de estudio se establecieron los siguientes parámetros de configuración:

- Para simular el tráfico ECG se utilizará un generador de tráfico constante tipo CBR (Constant Bit Rate) a una tasa de 18 Kbps, en la capa de aplicación.
- Cada nodo inicia el envío de tráfico de forma aleatoria una vez que la red está totalmente construida (asociación de los nodos de la red).
- Se utilizó la versión básica de TCP con las configuraciones por omisión.
- En lo referente a IEEE 802.15.4, se utilizó: con y sin servicios de confiabilidad (Ack), valores por omisión para CSMA/CA (minBE=3, maxBE=5 y CSMABackoff=4).

Para el primer caso, se consideraron 6 nodos como fuente de tráfico y un tamaño de paquete, a nivel TCP, desde 10 hasta 80 bytes; resultando en 57 y 127 bytes respectivamente para la capa MAC (debido a los 40 y 7 bytes de sobre carga de control que TCP y MAC le agregan al paquete respectivamente). Para el caso 2, se fijó el tamaño del paquete al valor encontrado en el primer caso (con el cual se tuvo el mejor equilibrio

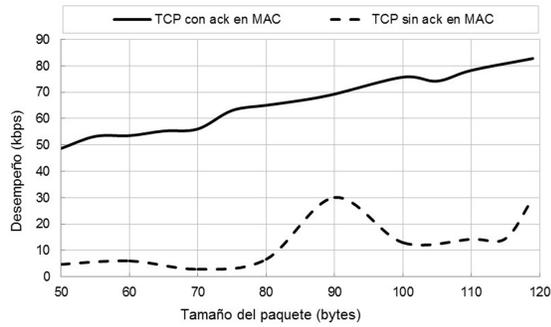


Figura 4. Tasa de entrega de datos en TCP.

entre T_d y R_t) y la variable de estudio fue la cantidad de nodos en la red (1 hasta 6 nodos).

La experimentación se realiza con el simulador de redes NS2, herramienta orientada a eventos que cuenta con módulos para el registro detallado de la operación de la red bajo el escenario propuesto. Cada resultado medido en el simulador fue el promedio de 200 eventos y cada experimento se repitió 10 veces para obtener el promedio del promedio.

5. Análisis de resultados

La Figura 4 presenta el T_d alcanzado por la red a nivel TCP (mismo valor para nivel MAC). Observe que en general conforme el tamaño de paquete se incrementa también lo hace el T_d . Por ejemplo, para el caso de servicio con confiabilidad a nivel MAC, para un tamaño de 50 bytes se alcanzó un desempeño de 48.65 Kbps y cuando el paquete MAC fue de 110 bytes el valor del desempeño fue de 78.24 Kbps; lo que representó un incremento del 62.18 %. Esto es consecuencia del mecanismo CSMA/CA; ya que la red necesita transmitir más del doble de paquetes para transportar la misma cantidad de datos, cuando se utiliza paquetes de 50 bytes. Lo cual se traduce en más periodos de contienda en un corto intervalo de tiempo, incrementando la probabilidad de colisiones y la probabilidad de “falta de acceso al canal”.

En ambos casos el paquete se pierde, primero por el daño del paquete, y segundo por el descarte del paquete debido al algoritmo CSMA/CA. Por lo tanto, es importante manejar paquetes de mayor tamaño para provocar que los nodos transmitan paquetes a “intervalos de tiempo mayores” y en consecuencia, la probabilidad de éxito de acceso al canal sea mayor. Otra conclusión importante de los resultados de la Figura 4 fue, que el valor de T_d alcanzado fue mejor cuando se usa el mecanismo Ack en la operación de IEEE 802.15.4. Esto se explica, porque el mecanismo de confiabilidad de MAC recu-

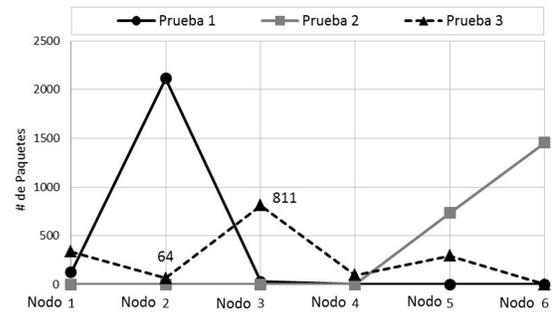


Figura 5. Paquetes recibidos en el receptor de cada nodo sensor.

pera, con un cierto grado de eficiencia, los paquetes que no fueron recibidos. Por lo tanto, se incrementa la probabilidad de éxito en la entrega de paquetes. El beneficio del mecanismo se maximiza cuando el paquete que recupera corresponde a un paquete de solicitud o de confirmación de sesión de la capa de transporte. Puesto que, evita que el 100 por ciento de los paquetes de ese flujo se pierdan, debido a al mecanismo de TCP para enviar sus datos. Esto puede observarse en los resultados de la Figura 5. Por ejemplo, en la prueba 1 y 2 sólo se recibieron paquetes de dos nodos, de los 6 posibles. Revisando el archivo de traza, que genera el simulador NS2, se confirma que la causa principal por la que el resto de los nodos no enviaron sus datos fue la falla para iniciar sesión; incluso a pesar, que la capa MAC operaba con el mecanismo Ack.

En lo referente a la métrica de “Equidad”, nos apoyaremos de la Figura 5. Observe los resultados de la prueba 3, el nodo que más paquetes logró enviar al nodo destino fue el nodo tres, con un total de 811 paquetes, en contraste con los 64 paquetes enviados por el nodo 2. Esto nos permite concluir que TCP e IEEE 802.15.4 no cuentan con mecanismos adecuados para proporcionar equidad a los nodos. Este problema impide que se pueda garantizar un cierto nivel de QoS por aplicación específica. Esto deja un espacio importante de investigación para implementar mecanismos de priorización de paquetes y/o nodos.

El retardo de transmisión es una métrica importante en las aplicaciones médicas de tiempo crítico. Los resultados del retardo promedio en TCP, para diferentes tamaños de paquetes, se muestran en la Figura 6. La primera conclusión fue, que la prueba con el mecanismo Ack de MAC presenta mayores retardos para cualquier tamaño de paquete. Lo cual es comprensible, ya que el nodo transmisor debe esperar por un paquete de confirmación antes de enviar el siguiente paquete; lo cual incorpora un valor extra de tiempo a cada comunicación. Otro resultado importante fue, que el valor del retardo promedio disminuye conforme se incrementa el

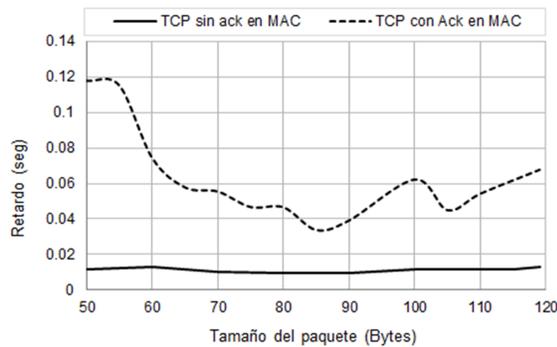


Figura 6. Retardo promedio de transmisión de paquete a nivel TCP.

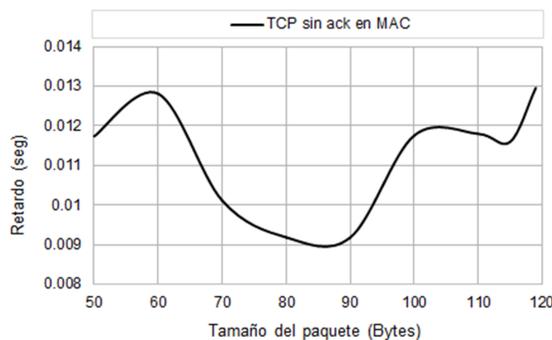


Figura 7. Retardo promedio de transmisión de paquete en TCP sin Ack en MAC.

tamaño del paquete hasta que el paquete alcanza un cierto tamaño, entonces el retardo promedio comienza a incrementarse.

Observe la Figura 6, para el caso de TCP con Ack en MAC, el retardo disminuye de 117 ms (para un tamaño de 50 bytes) hasta los 33.7 ms, cuando el paquete tiene un tamaño de 85 bytes. Sin embargo, con forme el tamaño del paquete se incrementa el valor de retardo comienza a crecer.

En el caso de TCP sin Ack en MAC, parece que el efecto no fue tan evidente, sin embargo, esto es debido a la escala del eje “Y”. Por ello se presenta la Figura 7, con el mismo caso. Observe que el comportamiento antes descrito es muy evidente. Además, al igual que en el caso anterior, los mejores valores de retardo se obtienen para un tamaño de paquete MAC entre 80 y 90 bytes.

Este comportamiento se explica debido a los procesos involucrados en el envío de los datos: el acceso al canal (retardo de Backoff) y la transmisión del paquete (retardo de transmisión). Cuando el tamaño del paquete es pequeño la principal contribución al retardo lo aporta el tiempo de acceso al canal del mecanismo CSMA/CA, ya que más paquetes a transmitir por cada nodo ocasionan mayor demanda de acceso, y por lo tanto mayores periodos de Backoff. Por otro lado, conforme el tamaño del paquete rebasa los 85 bytes, la principal contribu-

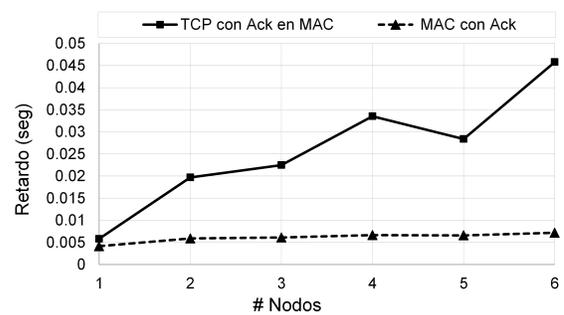


Figura 8. Retardo promedio de transmisión de paquete a nivel TCP y MAC.

ción al retardo de transmisión la da el tiempo que le toma al emisor enviar el paquete hasta el nodo destino, una vez que obtiene el acceso al medio. Ya que más bits por paquete requieren periodos mayores de transmisión incrementando el retardo total.

Recordando que uno de los objetivos de la experimentación es obtener el tamaño de paquete que nos diera el mejor valor de la métrica de Retardo y Td. Se concluye, a partir del estudio anterior que se tiene mayor Td cuando se utiliza el mecanismo de Ack en MAC y se tienen los más bajos valores de retardo cuando se utiliza un tamaño de paquete en MAC de alrededor de 80 bytes (y un tamaño de 33 bytes en TCP por los encabezados que se adicionan, 47 bytes).

Ahora que se tiene definido el tamaño de paquete (80 bytes), con el cual se tiene un adecuado equilibrio entre el valor de Retardo de paquete y Ancho de Banda alcanzado por la red, nuestro siguiente paso es conocer el efecto del número de flujos o nodos en la red en las métricas definidas. Con ello, establecer el número máximo de flujos ECG que una red WSN puede soportar bajo los requisitos QoS establecidos en la sección 2.3. La Figura 8 muestra los resultados de las pruebas para el retardo de transmisión para un tamaño de paquete fijo a 80 bytes y la activación del Ack en capa MAC. Los resultados muestran que el retardo promedio para TCP fue mayor que en MAC, sobre todo conforme el número de nodos se incrementa. Es claro que el retardo en TCP debe ser mayor que MAC, ya que éste incluye el tiempo de MAC; sin embargo, TCP aporta la mayor parte del valor final del retardo. Por ejemplo, para el caso de 3 nodos, el retardo de TCP fue de 22.48 ms y para MAC fue de 6.1 ms; que representa un incremento de 72.84 % para TCP. Lo cual evidencia una oportunidad de mejora al cambiar de protocolo de transporte para el envío de los datos.

Para el análisis de la cantidad de nodos que la WSN puede soportar consideramos que un ECG genera 5 Kbps o 625 bytes por segundo, y considerando que TCP

debe incorporar 33 bytes como máximo para obtener los 80 bytes en capa MAC, se tendrían 18.94 paquetes en TCP por cada señal ECG (o 19 paquetes). Por lo tanto, el retardo de transmisión de los 19 paquetes no debe ser mayor a 500 ms. En este sentido y de los resultados de la Figura 8, la WSN puede soportar hasta 3 nodos si utilizamos TCP, ya que el retardo promedio por paquete fue de 22.487 ms por 19 paquetes de la señal ECG, se alcanza un retardo total de 427.25 ms. Sin embargo para 4 nodos, con un retardo de 33.528 ms, se alcanza 637.03 ms, rebasando los 500 ms que requiere la aplicación médica del ECG. Caso totalmente diferente si consideramos sólo el retardo de transmisión de capa MAC, que puede soportar los 6 nodos sin problema, ya que el retardo promedio con 6 nodos es de 136.352 ms (retardo unitario de 7.176 ms).

Para el caso de la métrica de desempeño (Td), la Figura 9 muestra los resultados obtenidos en la simulación. Como se observó en los resultados del primer caso de estudio, el desempeño es significativamente mayor cuando se activa el mecanismo de confiabilidad de MAC. Sin embargo, para un solo nodo sensor, el desempeño que alcanza la red es prácticamente el mismo en ambos casos. Esto es explicable, ya que no existen otros nodos que provoquen colisión o incrementen la congestión de la red. Sin embargo, cuando la cantidad de nodos crece, la cantidad de contienda por acceso al canal también lo hace; provocando problemas en la operación de TCP y del mecanismo CSMA/CA que afectan el desempeño de la red. Otra observación importante de la Figura 9 y del caso TCP con Ack en MAC fue que, excepto para uno y dos nodos sensores, el resto de los casos no alcanzo el desempeño esperado, ya que el crecimiento se mantiene alrededor de los 62.5 Kbps. La causa principal fue la falla en el inicio de sesión de TCP, lo cual provocó que algunos nodos no iniciaran la transmisión de sus datos. Por lo tanto, aunque TCP e IEEE 802.15.4 tiene mecanismos para recuperar paquetes (Ack), sin embargo, no fue suficiente.

6. Conclusiones

En este trabajo se analizaron algunos factores que afectan la operación de TCP e IEEE 802.15.4 para cumplir los requerimientos QoS (tasa de entrega de datos y retardo de transmisión) y de Equidad para la supervisión de la salud. Estos factores fueron el tamaño de paquete, la cantidad de nodos en la red y las características del tráfico de la aplicación ECG. Los resultados muestran que el desempeño alcanzado en las métricas por IEEE 802.15.4 dependen del nivel de contienda por el acceso al medio debido al CSMA/CA, por lo tanto

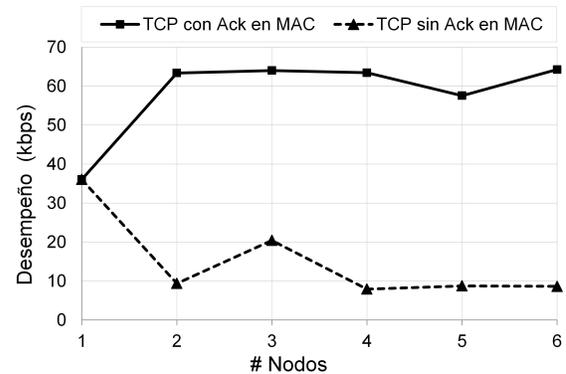


Figura 9. Desempeño de TCP contra el número de nodos.

cuando se trabaja un tamaño de paquete a nivel MAC entre 80 y 90 bytes se obtiene los valores más bajos de retardo y un valor aceptable para la tasa de entrega de datos. Otra conclusión importante es que TCP es la causa principal del retardo de transmisión total, debido a sus mecanismos de entrega y recepción de sus mensajes. Al respecto del número de nodos sensores ECG, el IEEE 802.15.4 puede soportar los 6 nodos del escenario sin alcanzar los 500 ms de retardo máximo permitido; sin embargo TCP sólo puede soportar tres nodos.

Por último, se concluye que TCP e IEEE 802.15.4 no ofrecen la misma oportunidad de transmisión a los nodos de la red y sólo garantiza el “mejor esfuerzo”. Por lo tanto, es necesario administrar el tráfico que pasa en cada nodo; con el fin de garantizar los mejores valores de QoS, de acuerdo a los niveles de prioridad de cada aplicación.

Como trabajo futuro consideramos la evaluación del mecanismo de acceso al medio CSMA/CA administrando los tiempos de Backoff e incorporando un protocolo de transporte que considere las áreas de oportunidad aquí descritas.

Referencias

- [1] M. A. Ameen, A. Nessa, and K. S. Kwak, “QoS Issues with Focus on Wireless Body Area Networks,” *Convergence and Hybrid Information Technology, ICCIT '08. Third International Conference on*, Vol. 1, pp. 801 -807, 2008.
- [2] B. Latré, B. Braem, I. Moerman, C. Blondia, and P. Demeester, “A survey on wireless body area networks,” *Wirel. Netw.*, Vol. 17, No. 1, pp. 1-18, Jan. 2011.
- [3] M. A. Yigitel, O. D. Incel, and C. Ersoy, “QoS-aware MAC protocols for wireless sensor networks: A survey,” *Comput. Netw.*, Vol. 55, No. 8, pp. 1982-2004, Jun. 2011.
- [4] A. Sharif, V. Potdar, and A. J. D. Rathnayaka, “Performance evaluation of different transport layer protocols on the IEEE 802.11 and IEEE 802.15.4 MAC/PHY layers for WSN,” *Proceedings of the 7th International Conference on Advances in*

- Mobile Computing and Multimedia*, pp. 300-310, New York, NY, USA, 2009.
- [5] B. Latré, P. De Mil, I. Moerman, N. Van Dierdonck, B. Dhoedt, and P. Demeester, "Maximum throughput and minimum delay in IEEE 802.15.4," *Proceedings of the First international conference on Mobile Ad-hoc and Sensor Networks*, Berlin, Heidelberg, pp. 866-876, 2005.
- [6] X. Liang and I. Balasingham, "Performance Analysis of the IEEE 802.15.4 based ECG Monitoring Network," *Wireless and Optical Communications*, Quebec, Canada, pp. 99-104, 2007.
- [7] DARPA. (Julio, 2013), "TCP-Transmission Control Protocol," <http://www.ibiblio.org/pub/docs/rfc/rfc793.txt>.
- [8] IEEE. (2006), "IEEE Standard for Information technology- Local and metropolitan area networks- Specific requirements- Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (WPANs)," *IEEE Std 802154-2006 Revis. IEEE Std 802154-2003*, pp. 1-320.
- [9] J. A. Gutiérrez, Callaway, and R. L. Barrett, "Low-rate wireless personal area networks: enabling wireless sensors with IEEE 802.15.4," *IEEE*, New York, 2007.
- [10] N. Chevrollier and N. Golmie, "On the use of wireless network technologies in healthcare environments," *Proc. 5th IEEE workshop on Applications and Services in Wireless Networks*, ASWN 2005, Paris, France, pp. 147-152, 2005.
- [11] J. R. Gallego, A. Hernandez-Solana, M. Canales, J. Lafuente, A. Valdovinos, and J. Fernandez-Navajas, "Performance analysis of multiplexed medical data transmission for mobile emergency care over the UMTS channel," *IEEE Trans. Inf. Technol. Biomed.*, Vol. 9, No. 1, pp. 13-22, 2005.
- [12] S. Arnon, D. Bhastekar, D. Kedar, and A. Tauber, "A comparative study of wireless communication network configurations for medical applications," *IEEE Wirel. Commun.*, Vol. 10, No. 1, pp. 56-61, 2003.
- [13] C. Wang, B. Li, K. Sohraby, M. Daneshmand, and Y. Hu, "Upstream congestion control in wireless sensor networks through cross-layer optimization," *IEEE J. Sel. Areas Commun.*, Vol. 25, No. 4, pp. 786-795, May 2007.

Arquitectura de Agentes ARSEC-AMS para Diseño de Sistemas de Seguridad

María de Guadalupe Cota Ortiz^a, Juan Pablo Soto Barrera^a, y Pedro Flores Pérez^a

^aUniversidad de Sonora.

Blvd. Luis Encinas y Rosales S/N, Colonia Centro, Hermosillo, Sonora, México, 83000.

{lcota, jpsoto, pflores}@gauss.mat.uson.mx

2013 Published by *DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>

Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Resumen

En este artículo se presenta la arquitectura híbrida para diseño y desarrollo de sistemas de seguridad computacional basados en tecnología de agentes, denominada ARSEC-AMS (Security Architect Multiagent System), la cual permite implementar mecanismos de interacción y razonamiento de agentes para detectar situaciones anormales que pueden representar amenazas no registradas en la base de conocimientos del sistema. El enfoque difiere de otras técnicas y metodologías que se vienen implementando en herramientas de protección, que aunque exitosas, no logran controlar eventos relacionados con el tema denominado “Zero day”, problemática que consiste en el surgimiento constante de nuevas formas de intrusión o programas que no son reconocidos por software de protección y toman el control de equipos de cómputo para vulnerar o comprometer sus recursos. La arquitectura cuenta con los módulos intérprete, pizarra, deliberativo-cognitivo, seguridad y reactivo, que en forma coordinada permiten al sistema de agentes alcanzar los objetivos de seguridad previamente establecidos.

Palabras clave: Arquitectura, Agente, Seguridad computacional.

1. Introducción

La tecnología de agentes es un campo innovador en el desarrollo de aplicaciones, donde los agentes son identificados como entidades software programadas para realizar una serie de operaciones con cierto grado de independencia, utilizando la información que proviene de su entorno y el conocimiento proporcionado por los diseñadores de este tipo de sistemas para dotarles de formas de razonamiento que les permitan percibir sus capacidades y el alcance funcional

para la toma de decisiones [1].

Por otra parte, la evolución tecnológica en materia computacional proporciona a todo tipo de organizaciones y particulares la posibilidad de modernizar sus servicios y aumentar el nivel de ingresos y productividad, pero esto también implica el tener que afrontar problemas de seguridad que cada vez son más complejos y difíciles de resolver, los cuales pueden ocasionar pérdidas millonarias en materia de información o recursos financieros [2]. En la actualidad, los países más desarrollados son los que hacen uso

de este tipo de tecnología, y por lo tanto, son los más susceptibles a delitos cibernéticos que se cometen con el fin de corromper, destruir o modificar los datos que se almacenan en sistemas de cómputo o se transmiten a través de redes de computadoras; situación que los obliga a adoptar medidas de seguridad que permitan proteger las operaciones que se realizan por este medio, y así, poder garantizar un nivel óptimo de confidencialidad, integridad y disponibilidad de la información [3, 4, 5, 6, 7].

Una de las amenazas más graves en seguridad computacional es llamada *amenazas del día cero* o *zero day*, y se identifica como el período de tiempo que inicia cuando se activa *malware* que no puede ser detectado por software de protección como sistemas antivirus, detectores de intrusos, etc., por no tener registrados los patrones de comportamiento en sus bases de datos [6]. Como ejemplo, está el caso de los *gusanos* o *worms*, que generalmente tienen el propósito de comprometer recursos de los equipos que han infectado, exponiéndolos a intrusiones o ataques que aprovechan esta situación, la cual perdura hasta que el problema es detectado y controlado por personal de seguridad [4] en horas, meses, años, o simplemente quedar sin solución por tiempo indeterminado.

Para controlar problemas relacionados con *Amenazas del día cero* o *Zero day* con un enfoque distinto a los actuales, una vez que se revisaron especificaciones de arquitecturas de agentes existentes, se llegó a la conclusión de que, en forma individual, adolecen de elementos básicos para diseñar sistemas de seguridad complejos. Retomando ideas de arquitecturas como BDI [8], GAIA [9], [10] y PRS [11], se ha diseñado la arquitectura híbrida orientada al área de seguridad computacional que ha sido denominada ARSEC-AMS, la cual permite abordar problemas de seguridad con tecnología de agentes, implementar mecanismos de comunicación y razonamiento de agentes, establecer criterios de seguridad y protección de información, aplicar comportamientos reactivos y dotar a los agentes de capacidades para la toma de decisiones e implementación de acciones que repercutan y se reflejen en su entorno. Dicha arquitectura está organizada en los siguientes módulos: el intérprete que se encarga de filtrar los mensajes entrantes y salientes del sistema de agentes, y de la canalización de transacciones que pueden estar dirigidas al control de la pizarra como elemento de comunicación, al módulo deliberativo-cognitivo que permite implementar formas de representación del conocimiento, al módulo de seguridad donde se pueden establecer criterios y

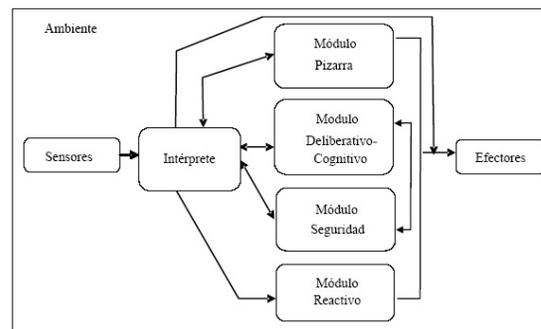


Figura 1. Arquitectura ARSEC-AMS.

niveles de seguridad, registrar alertas, definir instrucciones, etc., o al módulo reactivo que se encarga de implementar acciones programadas que respondan a cierto tipo de eventos previamente definidos por el administrador o por el sistema de agentes.

Cabe mencionar que bajo el modelo de la arquitectura ARSEC-AMS y sus componentes, se han concluido en el último año cuatro tesis en la Licenciatura en Ciencias de la Computación de la Universidad de Sonora, con el desarrollo de sistemas basados en tecnología de agentes para el diseño y desarrollo de herramientas software en materia de seguridad computacional, obteniendo una de ellas mención honorífica.

Para efectos de organización, para el desarrollo de este documento se utilizará el mismo orden en que se han mencionado los módulos de la arquitectura, agregando al final la descripción de elementos complementarios para el diseño y desarrollo de sistemas de seguridad basados en agentes tomando como base la arquitectura ARSEC-AMS.

2. Arquitectura de agentes ARSEC-AMS

La arquitectura ARSEC-AMS (ver Figura 1), se compone de módulos que en forma integral, permiten interpretar los eventos del ambiente donde se desenvuelven los agentes, diseñar mecanismos de razonamiento de tipo deliberativo-cognitivo, establecer criterios y niveles de seguridad, implementar tareas de tipo reactivas y mantener el control de la pizarra que se utiliza como elemento de apoyo en el sistema de comunicación.

2.1. Roles de agentes

Los roles que han sido diseñados para establecer las distintas funcionalidades de los tipos de agentes tienen

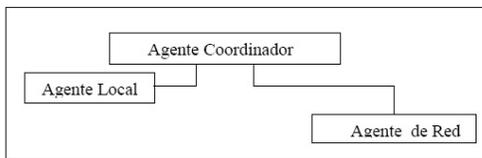


Figura 2. Jerarquía de Agentes.

correspondencia con un nivel jerárquico que define el grado de autoridad, ventajas y limitaciones que pueden asumir los agentes, y se asocian con conceptos relacionados con aspectos de seguridad como lo es el de responsabilidad, permiso y privilegio.

En la jerarquía diseñada para la arquitectura ARSEC-AMS (ver Figura 2), se establecen tres perfiles: Agente Coordinador (AC), Agente Local (AL) y Agente de Red (AR), los cuales se implementan atendiendo los siguientes criterios:

1. Conforme disminuye el nivel jerárquico, el agente tiene menor grado de autoridad para acceder a los recursos del sistema.
2. Cada rol se asocia con un nivel jerárquico, de tal forma que independientemente del grado de jerarquía que tenga un agente, no podrá realizar las funciones de los roles que corresponden a otro nivel.

En este contexto:

- a) El Agente AC es el encargado de administrar y coordinar el trabajo especializado del conjunto de agentes y realizar actividades especializadas y complejas.
- b) El Agente AL administra el funcionamiento interno del equipo donde se crea y realiza actividades generalizadas para vigilar el cumplimiento de políticas de seguridad, revisar el comportamiento de los recursos del sistema, y monitorear eventos locales en su entorno.
- c) El Agente AR supervisa el comportamiento de red del equipo donde se crea, realiza actividades de monitoreo de paquetes de red, detecta eventos que pueden considerarse como anormales e implementa acciones establecidas previamente por el administrador del sistema o por un agente de tipo Coordinador.

2.2. Esquema funcional de agentes

Con el fin de llevar un seguimiento de transacciones e intercambio de mensajes del sistema de agentes, los roles han sido diseñados bajo el esquema funcional que se muestra en la Figura 3, estableciendo un control interno para intercambio de mensajes basado en el protocolo *request_protocol* propuesto por FIPA [12] (ver

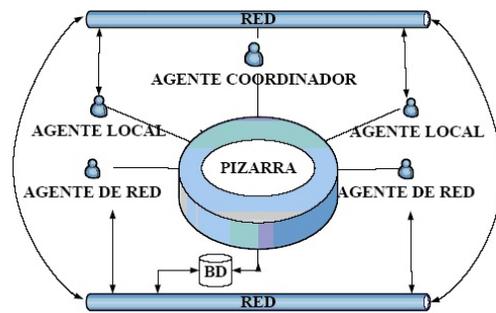


Figura 3. Esquema funcional de agentes.

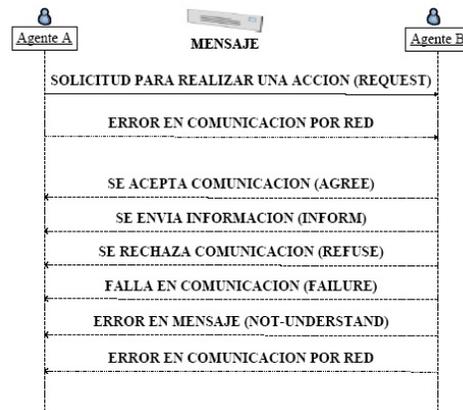


Figura 4. Esquema básico para intercambio de mensajes.

Figura 4). El procedimiento para lograr este objetivo, consiste en establecer un protocolo de comunicación que sirva como base para implementar los mensajes relacionados con las siguientes opciones:

- **Solicitud para realizar una acción (request).** Un agente puede solicitar a otro la realización de una acción.
- **Error en comunicación por red.** Además de los casos contemplados en el protocolo *request_protocol* de FIPA, en esta propuesta se incluye una opción adicional que permite al agente emisor, en base a un tiempo de espera previamente definido, detectar si el mensaje enviado ha llegado a su destino. Cuando se presenta este tipo de error, se recomienda reprogramar la transacción para el envío del mensaje.
- **Aceptar comunicación (agree).** Este mensaje, además de informar sobre la aceptación de realizar la acción solicitada, sirve como un aviso de que el mensaje transmitido ha llegado en forma correcta a su destino.
- **Envío de información (inform).** Consiste en enviar un mensaje que transmita cierta información sin requerir respuesta.

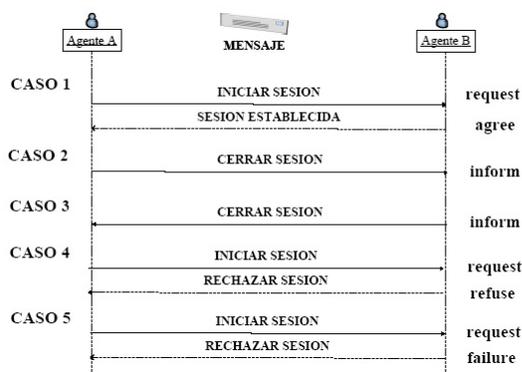


Figura 5. Esquema para control interno de sesiones.

- **Rechazo de comunicación (refuse).** Cuando por alguna razón, la solicitud de un agente no puede ser atendida, se envía un mensaje de rechazo, y opcionalmente, se agrega información adicional que indica el motivo por el cual no se puede realizar la acción solicitada o no puede aceptar establecer una sesión de trabajo.
- **Falla en comunicación (failure).** Error en secuencia de comunicación. Cuando una comunicación ha sido interrumpida antes de llegar a una confirmación de finalización de la misma por alguna de las partes, pueden emitirse mensajes que informen sobre las fallas de comunicación, y opcionalmente, pueden incluirse reportes con indicios que permitan detectar el origen del problema. En este caso existe la posibilidad de reiniciar completamente la transacción, o continuarla desde el punto donde fue interrumpida.
- **Error en el mensaje (not-understand).** Si un mensaje no puede comprenderse pero se tienen los datos de referencia del emisor, puede enviarse un mensaje informado la situación para que se re programe el envío del mensaje original.

Por otra parte, tomando en cuenta que la arquitectura ARSEC-AMS está orientada a problemas de seguridad computacional, en el caso de control de agentes alojados en múltiples equipos, opcionalmente, para efectos de evitar intrusiones en el sistema de agentes, se recomienda utilizar un esquema de control interno de sesiones de trabajo implementado por agentes AC (ver Figura 5), apoyándose para esta tarea en información interna del sistema de agentes que debe ser centralizada en un sistema remoto de bases de datos.

En base a lo descrito anteriormente, y para efectos de mantener un control interno de las sesiones de agentes y el envío/recepción de mensajes interactivos, se ha diseñado una transacción específica identificada como

sesión, a través de la cual se implementan los siguientes casos (ver Figura 5):

- **Iniciar sesión (caso 1).** Para que un agente pueda solicitar a otro que se realice una acción o se identifiquen como válidos los mensajes recibidos, es necesario que exista una confirmación de sesión establecida, y que este hecho sea registrada en un contenedor de datos, que básicamente contenga el nombre de los agente y el IP o dirección de red. En caso de que una sesión se finalice por cualquier causa, el registro de ésta debe ser eliminado de la lista en forma inmediata.
- **Sesión establecida (caso 1).** Cuando un agente recibe una solicitud para iniciar una sesión, puede responder con un mensaje de *Sesión establecida*, con lo cual acepta mantener un canal de comunicación para envío y recepción de mensajes.
- **Cerrar sesión (Caso 2 y 3).** Si una transacción ha llegado a su fin o la sesión de trabajo debe finalizarse, el mensaje que debe enviarse es el que corresponde a *Cerrar sesión*.
- **Rechazar petición de sesión (Caso 4 y 5).** Si el agente que recibe una solicitud para iniciar una sesión de trabajo, debe rechazar la petición, y opcionalmente, puede agregar información que detalle o explique el motivo de este tipo de respuestas.
- **Sin respuesta (failure).** Atendiendo a lo estipulado en el esquema básico, esta es una opción que permite verificar si los mensajes enviados han llegado a su destino o se han presentado errores en la transmisión de los datos y reprogramar la transacción o abortarla.

En el control interno planteado para manejo de sesiones, en los casos 1, 4 y 5 (ver Figura 5), el agente que envía el mensaje *Iniciar sesión*, debe establecer un tiempo límite para obtener respuesta del receptor, de tal forma sea capaz de detectar fallos en el envío/recepción de mensajes, con el fin de reprogramar las peticiones de sesión que no hayan sido atendidas y/o registrar alertas para el administrador que permitan detectar nodos inestables en los equipos donde se encuentra instalado el sistema.

2.3. Módulos de la arquitectura ARSEC-AMS

Los comportamientos de agentes de tipo general asociados a los módulos de la arquitectura ARSEC-AMS, pueden presentar variaciones debidas a la identificación

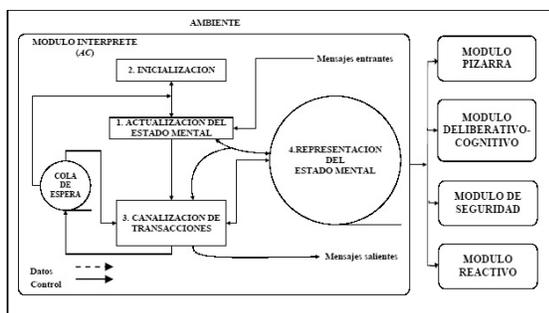


Figura 6. Flujo de mensajes en el módulo intérprete del Agente tipo Coordinador.

de funciones que corresponden al rol y nivel jerárquico del agente que se trate, o por cambios contextuales que ocurren en su medio ambiente.

2.3.1. Módulo intérprete

El módulo que funge como *intérprete* se encarga de filtrar los eventos registrados en el sistema y dar seguimiento en forma coordinada al conjunto de acciones a realizar para alcanzar los objetivos definidos para el sistema de agentes. Tiene como función interpretar los mensajes relacionados con los eventos filtrados por *sensores* del sistema, o redireccionarlos hacia el sistema de salida a través de *efectores* (ver Figura 6).

Las tareas básicas que sirven de soporte a este módulo para agentes AC son las siguientes:

- a) Obtener datos relacionados con los agentes del sistema y coordinar el trabajo conjunto.
- b) Crear sesiones de trabajo y mantener control del estado en que se encuentran.
- c) Controlar las instancias de pizarra.
- d) Aplicar lineamientos o criterios establecidos por el administrador del sistema en la toma de decisiones relacionados con períodos de revisión, instrucciones, políticas, niveles de seguridad, alertas, acciones a tomar, etc.
- e) Emitir alertas al administrador cuando se presenten casos en los cuales existe sospecha de un problema de seguridad, y aplicar medidas precautorias para resguardar los recursos involucrados en tanto se definen acciones para el tratamiento de eventos del mismo tipo, o se recibe la instrucción para desactivar en forma permanente la protección preventiva que se haya activado.
- f) Mantener un sistema de control en el envío de mensajes para efectos de reprogramarlos en una cola de espera cuando éstos no lleguen con éxito a su destino. Para realizar esta tarea es necesario imple-

mentar un mecanismo que permita la detección de la recepción de los mensajes.

La única diferencia en el comportamiento de los agentes en este módulo, es que para los agentes AL y AR el proceso “3 Canalización de transacciones” cambia a “3 Ejecución de instrucciones”.

Flujo de control del módulo intérprete

En los puntos marcados como 1 y 2 en la Figura 6 se realizan acciones de inicialización y actualización del estado mental de los agentes para identificar creencias, capacidades y el alcance de funciones de cada rol. En el punto 3 se controla la canalización de transacciones y mensajes salientes, se ejecutan instrucciones, y se mantiene control sobre una lista de espera de transacciones que deben ser reprogramadas. Por otra parte, el punto 4 de la Figura 6 identifica al módulo que proporciona información al sistema de agentes relacionada con la representación del estado mental del agente (creencias), la cual puede ser actualizada con la recolección de datos relacionados con características básicas de agentes, eventos, o información relevante que es extraída de mensajes entrantes, planes a seguir (intenciones), registro de metas u objetivos (deseos), y registrar el estado de avance actual de los objetivos registrados.

2.3.2. Módulo Pizarra

Si bien es cierto que uno de los recursos más importantes de un sistema de agentes es la información que se define para efectos de organización y funcionamiento, también lo es la que se genera en forma dinámica durante períodos de actividad, cuyo significado e interpretación proporciona un medio de identificación de patrones en los sucesos que se presentan en el medio ambiente, así como el elemento que se utilice para registrarla. En este sentido, la arquitectura de pizarra sirve como medio de comunicación en un sistema de agentes a través de sus componentes y se constituye en un conjunto de fuentes de conocimiento (KSs, Knowledge Sources), y un mecanismo de control [13].

La arquitectura de pizarra denominada BLACKBOARD-ECRE (Blackboard-Element Control for Register of Events) que se muestra en la Figura 7, ha sido diseñada para implementar el módulo de pizarra en la arquitectura ARSEC-AMS, para efectos de registrar eventos relevantes que puedan ser analizados con el fin de detectar patrones de comportamiento que puedan constituir un problema de seguridad. Se implementa a través del esquema de base de datos que se muestra en la Figura 8 y funciona como elemento de

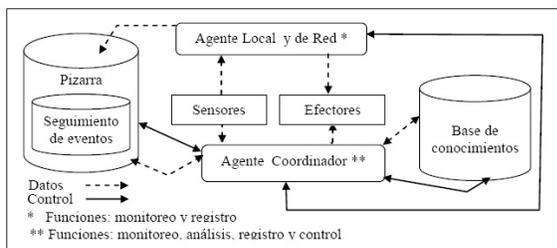


Figura 7. Arquitectura de Pizarra BLACKBOARD-ECRE.

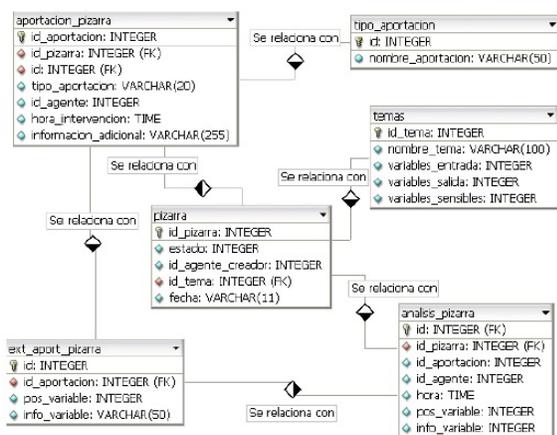


Figura 8. Esquema de Base de Datos.

comunicación en el sistema de agentes en base al ciclo de control que se presenta en la Figura 9.

Ejemplo de uso del módulo pizarra

En el diagrama de actividad de agentes que se muestra en la Figura 10, se muestra como ejemplo, el intercambio de mensajes para monitoreo de la pizarra con el tema *puertos*, con el objetivo de detectar puertos abiertos que registren más de 100 intentos de conexión por un mismo ordenador en un período de 10 minutos.

2.3.3. Módulo Deliberativo-Cognitivo

Un aspecto importante en seguridad computacional que se toma en cuenta al momento de diseñar sistemas o herramientas de protección o de prevención, consiste en incluir componentes que permitan monitorear eventos, registrar en contenedores la información que se considera relevante en base a criterios previamente establecidos, aplicar técnicas de análisis para clasificar o reconocer patrones que permitan identificar comportamientos que puedan representar una amenaza, y aplicar procedimientos correctivos o preventivos cuando se requiere. Esta metodología, aunque exitosa, no ha logrado controlar la problemática “Zero day”, y por lo tanto se asume que es necesario buscar alternativas que permitan automatizar y hacer más flexibles los procedimientos



Figura 9. Ciclo de Control de Pizarra.

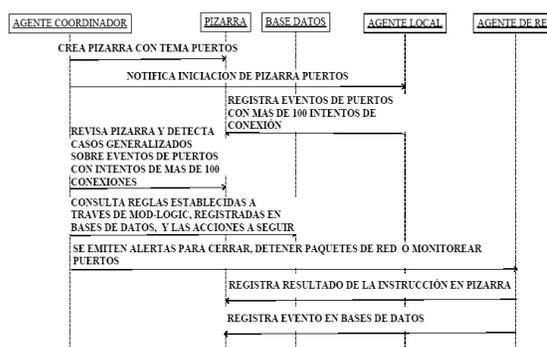


Figura 10. Diagrama de actividad para monitoreo de puertos.

actuales, de tal forma que puedan implementarse y modificarse estrategias en tiempo de ejecución para atacar problemas de este tipo.

Tomando en cuenta lo antes expuesto y que los sistemas basados en agentes, entre otras cosas, pueden utilizar formas de razonamiento que les permitan interactuar y asumir comportamientos que definan el nivel y alcance de sus funciones, en la arquitectura ARSEC-AMS se incluye el módulo deliberativo-cognitivo, a través del cual pueden implementarse técnicas de reconocimiento de patrones, sistemas expertos, reglas basadas en lógica, etc.

Ejemplo de uso del módulo deliberativo-cognitivo

Siendo la programación lógica una de las formas más eficientes y parecidas a las que el ser humano utiliza para representar el conocimiento, a continuación se presenta un ejemplo de un sistema de reglas que puede ser aplicado en este módulo donde se visualiza un posible escenario relacionado con el tema *puertos* y los siguientes supuestos semánticos:

Supuesto 1 Se ha detectado que existe un *gusano*, que para efectos de ejemplificar, ha sido denominado *scanworm*, que se compone de la parte *cliente* y

la parte *servidora*, siendo ésta última la que se descarga en ordenadores para infectarlos y seguir propagándose a otros equipos que se encuentran cercanos al punto donde se encuentran localizados.

Supuesto 2 La parte *cliente* es manejada por el autor del *gusano*, y se dedica a escanear la red buscando equipos que tengan abierto el puerto 2127, el cual es abierto en *listening* por la parte *servidora* que una vez activa se encuentra en espera de que la parte *cliente* se conecte.

Supuesto 3 El objetivo de este *gusano* es establecer conexión entre ambas partes (*cliente* y *servidor*) con el fin de obtener información confidencial que está guardada en el equipo infectado.

Supuesto 4 Se detectan anomalías en los equipos infectados provocados por este *gusano*, pero aún no se cuenta con la herramienta capaz de detectarlo, detenerlo y eliminarlo del sistema.

Supuesto 5 Se ha propagado una alerta administrativa que permite registrar una regla en el módulo lógico, con el fin de cerrar el nodo de conexión, que es el puerto, y evitar la pérdida o robo de información hasta que se solucione el problema.

Supuesto 6 Se registra la regla *scanworm* y se obtiene la siguiente plantilla:

semántica:

```
scanworm(n_puerto, t_operador, n_conexion,
t_accion).
num_puerto(n_puerto).
num_conexion(n_conexion).
tipo_accion(t_accion).
```

reglas:

```
scanworm(n_puerto, t_operador, n_conexion,
t_accion):
num_puerto(n_puerto),
num_conexion(n_conexion),
tipo_accion(t_accion),
```

datos:

```
num_puerto('2127').
num_conexion('0').
tipo_accion('cerrar').
tipo_accion('revisar').
```

Interpretación de la regla 'scanworm': Si el número del puerto es '2127' y el número de conexiones es '0', entonces se trata de 'scanworm' y se debe tomar la acción de 'cerrar' el puerto y 'revisar' el estado del mismo en períodos de tiempo predeterminados.

2.3.4. Módulo de Seguridad y Reactivo

Módulo de seguridad. El módulo de seguridad trabaja conjuntamente con el módulo deliberativo-cognitivo, ya que permite establecer criterios y niveles de seguridad que pueden aplicar los agentes en determinados casos, e incluir datos relacionados con alertas, instrucciones y acciones que sirvan como elementos complementarios en la toma de decisiones.

Módulo reactivo. En el caso del módulo reactivo pueden programarse acciones que respondan a cierto tipo de eventos previamente fijados por el administrador o diseñador del sistema, siendo los agentes AL y AR quienes pueden asumir este rol. Como ejemplo de un agente reactivo, se puede tomar el caso de un agente que atiende la actividad de monitoreo de paquetes de red aplicando reglas derivadas de un árbol de decisión para descartar paquetes de red que puedan ser dañinos, o aplicando técnicas de reconocimiento de patrones para identificar texto que se incluye en las cabeceras de paquete de red que es considerado como un patrón que permite identificar a cierto tipo de *malware*.

3. Elementos básicos complementarios a la arquitectura ARSEC-AMS

Para efecto de complementar la arquitectura ARSEC-AMS, como elemento de comunicación, se ha diseñado la gramática *Content Language Security System Agent for Windows* (CLASS-W) [14], la cual ha sido diseñada para generar lenguajes de contenido y ser utilizada en el desarrollo de sistemas de seguridad basados en agentes, de acuerdo a las especificaciones fijadas por FIPA [12].

Por otra parte, se ha desarrollado la librería denominada Mod-Logic, Mod-Logic, traductor de predicados tipo Prolog a una base de datos relacional e intérprete de consultas con plantillas diseñadas con SQL [15], con la cual se puede guardar y recuperar predicados tipo Prolog en base de datos, con la misma funcionalidad que lo haría un compilador de prolog pero con la ventaja de que puede utilizarse en forma independiente en cualquier tipo de aplicaciones que permitan utilizar librerías de enlace dinámico.

Existen otro tipo de elementos diseñados para ser implementados en la arquitectura ARSEC-AMS, como por ejemplo, el procedimiento de cifrado de información a través de intercambio de clave pública y privada, pero que por cuestiones de espacio, al igual que ejemplificar todos los módulos no pueden ser descritos en este documento.

Por último, cabe mencionar que el contenido de este

artículo forma parte del trabajo que se realiza dentro de un proyecto de investigación en materia de seguridad computacional. Con la implementación de la arquitectura ARSEC-AMS se han desarrollado herramientas que han sido motivo de cuatro tesis de licenciatura.

4. Conclusiones

La seguridad en las redes de computadoras es un tema que cada día cobra más importancia debido a la gestión masiva de datos y a la generalización del uso de tecnología, lo cual ha permitido a todo tipo de organismos aumentar su productividad y riqueza pero también se ven obligados a buscar medios de protección que eviten pérdidas de información que generalmente tienen repercusiones financieras significativas.

A pesar del avance que existe en alternativas de protección en materia de seguridad computacional, como lo es el desarrollo de herramientas software como antivirus, sistemas detectores de intrusos, etc., la problemática denominada “amenazas de día cero” o “zero day” es la que más preocupa a los expertos en seguridad, ya que la naturaleza de este problema consiste precisamente en el surgimiento constante de nuevas amenazas que el software de protección no puede detectar porque no han sido diseñadas para inferir eficientemente nuevas amenazas y desconocen patrones de comportamiento ajenos a los ya registrados.

Con un enfoque distinto a los actuales, la arquitectura ARSEC-AMS, como herramienta de diseño, permite a los expertos, a un administrador o desarrollador de sistemas de seguridad, administrar o diseñar y desarrollar sistemas basados en tecnología de agentes aprovechando la experiencia adquirida en materia de seguridad.

Por último, hay que agregar que si bien existen diferentes tipos de gramáticas para desarrollo de sistemas de agentes, no se encontró evidencia de alguna gramática que en forma rápida y eficiente permita el diseño y desarrollo de sistemas de agentes para resolver problemas complejos en materia de seguridad computacional. Por ello, se concluye que los elementos creados para implementarse en la arquitectura ARSEC-AMS, como lo es la gramática CLASS W y la librería Mod-Logic son componentes básicos, pero no necesariamente son elementos obligatorios para el uso de la arquitectura, ya que pueden utilizarse gramáticas de agentes existentes y métodos que permitan controlar el sistema de razonamiento de los sistemas basados en tecnología de agentes.

Referencias

- [1] J.M. Bradshaw. (1997), *Software Agents*, <http://agents.umbc.edu/introduction/01-Bradshaw.pdf>.
- [2] La Flecha. (Enero, 2006), *Diario de Ciencia y Tecnología, El coste de los virus informáticos asciende a 92,000 millones de euros en 10 años (110,000 millones de dólares)*, Agencia EFE, <http://laflecha.net/canales/seguridad/200601301>.
- [3] McAfee Proven Security. (Diciembre, 2006), *Microsoft Word 0-Day Vulnerability III*, http://vil.nai.com/vil/Content/v_vul27264.htm.
- [4] R. D. Pethia. (Septiembre, 2003), “Viruses and Worms: What Can We Do About Them?,” Software Engineering Institute, Carnegie Mellon University, CERT@Coordination Center, Pittsburgh, Pennsylvania, USA, <http://archive.cert.uni-stuttgart.de/isn/2003/09/msg00092.html>.
- [5] El Universal.com.mx. (Junio, 2007), “Se propaga Gusano en YouTube,” <http://www.el-universal.com.mx/articulos/40642.html>.
- [6] R. Clarke. (Julio, 2004), “Cómo afrontar los cambios radicales de la seguridad en el ciberespacio,” http://www.symantec.com/region/mx/enterprisesecurity/content/expert/LAM_4198.html.
- [7] B. Schneier. (2004), “SIMS: Solution, or Part of the Problem?,” *IEEE Security and Privacy*, Vol. 2, No. 5, pp. 88, ISSN: 1540-7993, <http://doi.ieeecomputersociety.org/10.1109/MSP.2004.83>.
- [8] D. Kinny, M. Georgeff and A. Rao. (1997), “A Methodology and Modelling Technique for Systems of BDI Agents,” Report, <http://www.dia.fi.upm.es/~phernan/AgentesInteligentes/referencias/kinny96methodology.pdf>.
- [9] F. Zambonelli, N. Jennings and M. Wooldridge, “Developing multiagent systems: the Gaia Methodology,” *ACM Trans on Software Engineering and Methodology*, Vol. 12, No. 3, pp. 317-370, 2003.
- [10] A. Garro, “GAIA.Modeling Notation Source, Version 03-03-1217:25EST,” <http://www.auml.org/auml/documents/Gaia.doc>.
- [11] F. Ingrand and M. Georgeff. (1990), *Managing Deliberation and Reasoning in Real-Time AI Systems*, In Proceedings of the Workshop Innovative Approaches to Planning, Scheduling and Control, pp. 284-291, <http://books.google.com.mx>.
- [12] FIPA, “Foundation for Intelligent Physical Agents,” <http://www.fipa.org>.
- [13] N. Carver and V. Lesser, “The evolution of blackboard control architectures,” Technical report, University of Massachusetts Amherst, pp. 19-23, 1992, <http://www.cs.siu.edu/~carver/ps-files/tr92-71.ps.gz>.
- [14] G. Cota, P. Flores and J. Suárez, “CLASS-W, a Grammar for Security System Development based on Environment Equipment Agents with Windows-Technology NT,” *Research in Computing Sciences. Advances in Computer Science and Engineering*, ISSN: 1870-4069, pp. 65-76, 2009, <http://prometeo.cic.ipn.mx/2009/index.html>.
- [15] G. Cota, P. Flores, I. López and M. Morales, “Mod-Logic: Un Traductor de Predicados Tipo Prolog a una Base de Datos Relacional e Intérprete de Consultas con Plantillas Diseñadas con SQL,” *Revista Iberoamericana de Sistemas, Cibernética e Informática*, ISSN: 1690-8627, Vol. 5, No. 2, pp. 25-29, 2008, http://www.iiis.org/iiis/best_paper.asp?year=2008.

Reconocimiento Biométrico de la Palma de la Mano Basado en el Promediado de Filtros de Correlación Simples

Everardo Santiago Ramírez^a, José Ángel González Fraga^a, Everardo Gutiérrez López^b, y Sergio Omar Infante Prieto^c

^aUniversidad Autónoma de Baja California, Facultad de Ciencias.
Carr. Ensenada-Tijuana 3917, Col. Playitas, Ensenada, B.C., México, 22860.
 {everardo.santiagoramirez,angel_fraga}@uabc.edu.mx

^bUbilogix, S. de R.L. de C.V.
Álvaro Obregón Núm. 110-4, Ensenada, Baja California, C. P. 22800.
everardo@ubilogix.com

^cUniversidad Autónoma de Baja California, Facultad de Ingeniería, Arquitectura y Diseño.
Carr. Ensenada-Tijuana 3917, Col. Playitas, Ensenada, B.C., México, 22860.
sinfante@uabc.edu.mx

2013 Published by *DI²U_{100ci}* @ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CCOMP-2013, www.cicomp.org

Resumen

Este artículo presenta un algoritmo biométrico para reconocer personas en base a la información contenida en la palma de la mano usando la técnica de correlación. Cada imagen de entrenamiento se segmenta en cuatro regiones iguales y donde cada una permanece en su posición original. Mediante el promediado de estas regiones se construye un filtro de correlación para cada persona, la cual se almacena en una base de datos. Dada una imagen de la palma de la mano a reconocer, ésta se correlaciona con cada uno de los filtros de correlación. Si se encuentra alguna coincidencia, entonces se devuelve la identidad asociada a dicho filtro. El algoritmo propuesto fue evaluado con un conjunto de datos que contiene 500 clases diferentes de imágenes de la palma de la mano, alcanzando una tasa de identificación y detección igual a 98.48 %.

Palabras clave: Filtros de correlación, Reconocimiento biométrico, Imágenes palmares.

1. Introducción

El reconocimiento de la palma de la mano es una modalidad biométrica que usa la estructura física de la palma de la mano de una persona para

finde identificación [1]. La palma de la mano es la parte que va de la muñeca hasta el inicio de los dedos. Esta parte muestra protuberancias largas y estrechas conocidas como crestas, las cuales son separadas por espacios conocidos como valles. También presenta sur-

cos y arrugas. Este conjunto de características es único y permanente para cada persona, por lo que permite realizar una identificación de forma confiable.

Existen varios enfoques para el reconocimiento de la palma de la mano. Los más comunes se basan en la huella palmar, la cual se representa a través de la información de la impresión de surcos de fricción [2, 3]. En una imagen binaria de la huella palmar, los surcos de fricción (líneas negras) se producen por las crestas. Mientras que los espacios en blanco son causados por los valles y surcos. Wen-xin et al. [4] presentan un método basado en la extracción de características mediante la aplicación de la transformada de Fourier (TF) sobre las imágenes de la palma de la mano. Estas características se usan como índices para las plantillas en la base de datos, y también para determinar el grado de acoplamiento con una imagen a reconocer. El método fue evaluado con una base de datos de 500 clases, 12 imágenes de la palma de la mano por clase, alcanzando una exactitud de identificación de 92.50 %. Otro método basado en el dominio de las frecuencias, es el que presentan Imtiaz y Anowarul-Fattah [5]. La imagen se segmenta en varias bandas espaciales de pequeña anchura. El reconocimiento se realiza extrayendo las características espectrales dominantes de cada una de esas bandas usando la transformada de Coseno discreto. Este método fue evaluado sobre dos bases de datos, obteniendo una exactitud de reconocimiento de 99.97 % y 99.92 %, respectivamente. En la referencia [6], los autores emplean filtros de correlación del tipo OTSDF, donde la estrategia consiste en diseñar el filtro con subregiones de imágenes de entrenamiento. La estrategia logró una exactitud de reconocimiento de 99.94 % en el mejor de los casos. En [7], el autor presenta una estrategia basada en la combinación de imágenes multispectrales, y se emplean los algoritmos de análisis de componente principales y análisis lineal discriminante para realizar el reconocimiento. Dicho algoritmo alcanzó tasas de reconocimiento que van del 98.20 % al 99.57 %.

Los enfoques previamente mencionados se basan en imágenes de la huella palmar, por lo que necesitan de una buena estrategia de extracción de características para obtener desempeños aceptables. Sin embargo, no aprovechan de manera conjunta la información proporcionada por las crestas, valles, arrugas y la textura de la superficie de la palma de la mano. De ahora en adelante, se referirá a la imagen de la palma de la mano como imagen palmar. La propuesta presentada en este trabajo se basa en la premisa de que el uso en conjunto de toda la información contenida en la imagen palmar, proporcionará a los algoritmos mayor robustez para dis-

criminar una clase auténtica de clases impostoras, aún y cuando las crestas no sean distinguibles completamente. Este artículo presenta un algoritmo basado en filtros de correlación. Cada imagen palmar de entrenamiento se segmenta en cuatro regiones, donde cada región permanece en su posición original. Mediante el promediado de estas regiones se construye un filtro de correlación para cada persona. A la señal promedio se le aplica la TF. Luego un factor de no linealidad, la Ley-k, como se describe en [8]. Por último, el filtro se almacena en una base de datos. Una imagen palmar de prueba se compara con la base de datos, y en caso de que se encuentre alguna coincidencia se devuelve la identidad asociado a dicho filtro. El resto de este artículo está organizado de la siguiente manera. En la Sección 2 se presentan conceptos básicos del filtrado de retina y del proceso de correlación. El algoritmo propuesto se presenta en la Sección 3. Mientras que en la Sección 4 se presentan los resultados experimentales, así como la discusión de los mismos. Finalmente, en la Sección 5 se presenta la conclusión de este trabajo.

2. Conceptos básicos

Una ventaja de los filtros de correlación es que pueden ser entrenados con muestras que representan las distorsiones que se espera presente el objeto objetivo. Sin embargo, una de las distorsiones más difíciles de representar es la causada por los cambios en iluminación, pues esta es muy variable y difícil de predecir. Varios trabajos de investigación han abordado este problema cuyo reto principal es mejorar la imagen sin pérdida de detalles.

En esta sección se describe brevemente el filtrado de retina para la mejora de imágenes con variaciones en iluminación. También se describe brevemente el proceso de filtrado de correlación.

2.1. Filtrado de retina

Se implementó el filtrado de retina descrito en [9] con el fin de lidiar con problemas de iluminación tanto en las imágenes de entrenamiento como en las imágenes de prueba sin perder detalles de las crestas.

El filtrado de retina imita el funcionamiento de la retina humana mediante la combinación de dos funciones no lineales adaptativas. Sea $I(x, y)$ una imagen palmar normalizada. La primera función no lineal es un filtro pasa bajas dado por

$$F_1(x, y) = I(x, y) \otimes G_1(x, y) + \frac{\bar{I}_{en}}{0.1} \quad (1)$$

donde $F_1(x, y)$ es el factor de adaptación, \otimes denota la operación de convolución, \bar{I}_{en} es la media de la imagen de entrada y $G_1(x, y)$ es un filtro pasa bajo Gaussiano:

$$G_1(x, y) = \frac{1}{2\pi\sigma_1^2} \exp\left(\frac{-x_2 + y_2}{2\sigma_1^2}\right) \quad (2)$$

Después, $I(x, y)$ es procesada de acuerdo al factor de adaptación $F_1(x, y)$:

$$I_{la1}(x, y) = \text{máx}(I(x, y)) + F_1(x, y) \frac{I(x, y)}{I(x, y) + F_1(x, y)} \quad (3)$$

donde $\text{máx}(I(x, y)) + F_1(x, y)$ es el factor de normalización. $\text{máx}(I(x, y))$ es el valor máximo de intensidad en la imagen de entrada.

La segunda función no lineal trabaja de forma similar. La imagen adaptada a la luz está dada por:

$$I_{la2}(x, y) = \text{máx}(I_{la1}(x, y)) + F_2(x, y) \frac{I_{la1}(x, y)}{I_{la1}(x, y) + F_2(x, y)} \quad (4)$$

con

$$F_2(x, y) = I_{la1}(x, y) \otimes G_2(x, y) + \frac{\bar{I}_{la1}}{0.1} \quad (5)$$

y

$$G_2(x, y) = \frac{1}{2\pi\sigma_2^2} \exp\left(\frac{-x_2 + y_2}{2\sigma_2^2}\right) \quad (6)$$

La constante 0.1 en las ecuaciones (1) y (5) difieren del trabajo original. El valor se determinó experimentalmente, ya que fue el valor que mejor recuperó la información de los bordes en la imagen.

2.2. Filtrado por correlación

El reconocimiento de patrones por correlación está basado en la selección o creación de una señal de referencia $h(x, y)$, denominado filtro de correlación, para luego determinar el grado de similitud entre el filtro y la señal de prueba $s(x, y)$ [10].

La aplicación de un filtro de correlación a una imagen de prueba produce un plano de correlación $g(x, y)$, al aplicar las siguientes operaciones:

$$g(x, y) = s(x, y) \otimes h(x, y) = F^{-1} \{S(k, l)H^*(k, l)\} \quad (7)$$

donde $S(k, l)$ y $H(k, l)$, corresponden a las transformadas de Fourier de $s(x, y)$ y $h(x, y)$, $*$ y F^{-1} son el complejo conjugado y la transformada inversa de Fourier. Si son similares, $g(x, y)$ mostrará un pico agudo y alto. En caso contrario, el pico será muy pequeño o simplemente no se producirá.

Existen varios tipos de filtros de correlación a los cuales se les puede incorporar invariencias como desplazamiento, tolerancia a ruido y a ligeras variaciones

en iluminación, rotación, pose y escala. El filtro de correlación más simple es el filtro de acoplamiento [11], el cual está dado por:

$$H^*(k, l) = F \{f(x, y)\} \quad (8)$$

donde $f(x, y)$ es la imagen de referencia. El filtro de acoplamiento es robusto para reconocer imágenes afectadas por ruido blanco aditivo. Sin embargo, es muy sensible a distorsiones causadas por la iluminación.

En la literatura han surgido varios enfoques de filtros de correlación que buscan una mayor robustez a alguna distorsión en específico y/o producir un pico alto y agudo.

3. Algoritmo de reconocimiento biométrico de la palma de la mano

El algoritmo propuesto consta de dos etapas principales. La primera consiste en el registro de la plantilla biométrica del usuario en la galería. Considerando el argot biométrico, las plantillas biométricas corresponden a los filtros correlación. Al conjunto de todas las plantillas se le denomina galería. Esta primera etapa se realiza solo en el registro del usuario o cuando se desea realizar una actualización de las plantillas biométricas. La segunda etapa consiste en el reconocimiento de alguna imagen palmar.

La primera etapa supone que las imágenes palmares fueron detectadas y extraídas previamente. Cada plantilla biométrica se construye con el siguiente algoritmo:

Algoritmo 1. Construcción de plantilla biométrica.

Entradas: Conjunto de imágenes palmares de la misma clase $p_1(x, y), p_2(x, y), \dots, p_n(x, y)$.

Salida: Plantilla biométrica $H^k(k, l)$ almacenado en la galería G .

1. Inicializar $h(x, y)$.
2. Para $i = 1$ hasta n .
3. $p'_i(x, y) \rightarrow$ Aplicar filtrado de retina a $p_i(x, y)$ como se describió en la subsección 2.1.
4. $s_j(x, y) \rightarrow$ Segmentar $p'_i(x, y)$.
5. Sumar a $h(x, y)$ cada segmento $s_j(x, y)$.
6. Promediar $\bar{h}(x, y) = \frac{1}{n}h(x, y)$.
7. Aplicar la TF $H(k, l) = F \{\bar{h}(x, y)\}$.
8. Aplicar la no linealidad $H^k(k, l) = |H(k, l)|^k \exp(-i * \phi_{H(k, l)})$.
9. Almacenamos $H^k(k, l)$ en G .

La palma de la mano es un objeto no rígido, sujeto a variaciones en forma. Con la segmentación de la imagen se obtiene un conjunto de clasificadores débiles. De esta forma, es posible reconocer una imagen que presenta ligeros cambios de forma, ya que se espera que las regiones no deformadas se acoplen con su correspondiente clasificador débil. Por otra parte, el promediado enfatiza aquellas características comunes entre los clasificadores débiles.

Una imagen palmar puede ser identificada mediante el siguiente algoritmo:

Algoritmo 2. Identificación de la palma de la mano.

Entradas:

- Imagen palmar a identificar $p_r(x, y)$.
- Umbral de reconocimiento r_{th} .

Salida: Id o “Desconocido”.

1. Se inicializa vector de puntuaciones v_p .
2. Aplicamos el filtrado de retina a p_r .
3. Aplicamos la TF a la imagen de prueba $p_r(x, y)$, tal que $P_R(k, l) = F \{p_r(x, y)\}$.
4. Aplicamos el factor de no linealidad: $P_R^k(k, l) = |P_R(k, l)|^k \exp(-i * \phi_{P_R(k, l)})$.
5. Para cada $H_i^k(k, l) \in G$.
6. $g(x, y) = F^{-1} \{P_R^k(k, l)H_i^{k*}(k, l)\}$.
7. Se examina $g(x, y)$ en busca de un pico: $pico = \max\{g(x, y)\}$.
8. Se calcula agudeza del pico: $pico_{psr} = PSR(g(x, y))$.
9. Si $pico_{psr} > r_{th}$:
10. Se agrega $pico_{psr}$ a v_p .
11. Si v_p no está vacío entonces $Id \in \max\{v_p\}$.
12. En caso contrario, se devuelve “Desconocido”.

La imagen de prueba se compara con cada una de las plantillas biométricas. Aquellas comparaciones que producen un $pico_{psr}$ mayor o igual que el umbral de reconocimiento r_{th} se agregan a un vector de puntuaciones v_p . Al final, la identidad del individuo corresponde a la plantilla cuyo valor del $pico_{psr}$ es el más alto en v_p .

El valor PSR [10] representa la agudeza del pico de correlación. Este valor indica a cuántas desviaciones estándar sobre la media está dicho pico:

$$pico_{psr} = \frac{pico - \mu_{r(x,y)}}{\sigma_{r(x,y)}} \quad (9)$$

donde $\mu_{r(x,y)}$, $\sigma_{r(x,y)}$ son la media y la desviación estándar de la región $r(x, y)$ en la salida de correlación $g(x, y)$ pero excluyendo la región del pico.

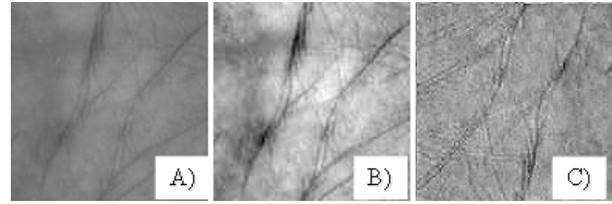


Figura 1. A) Imagen original, B) imagen con iluminación mejorada mediante el filtrado de retina y C) Plantilla biométrica sintetizada.

4. Resultados experimentales

El algoritmo propuesto fue evaluado usando la base de datos de imágenes palmares espectrales de la Hong Kong Polytechnic University (PolyU) [12]. Este conjunto de datos fue recolectado de 250 personas. A cada una se le recolectó 12 imágenes de cada mano, considerando cada mano como una clase, se obtuvieron un total de 500 clases diferentes. Solo se emplearon las imágenes en los espectros rojo (R: Red), verde (G: Green) y azul (B: Blue) para construir imágenes RGB manteniendo la dimensión original de píxeles. Cada imagen RGB construida se transformó a escala de grises.

La galería G para la tarea de identificación fue creada usando el Algoritmo 1. Para cada plantilla biométrica se usaron cuatro imágenes diferentes de la misma clase. A cada imagen se le aplicó el filtrado de retina configurando las ecuaciones (2) y (6) con $\sigma_1 = 0.5$ y $\sigma_2 = 1$, respectivamente. En la Figura 1A) se muestra una imagen que contiene valores de intensidad bajos. Esto dificulta aprovechar las diferentes características de la palma de la mano. En la Figura 1B) se muestra la misma imagen cuya iluminación mejoró sin mucha pérdida de características mediante el filtrado de retina. Luego, cada imagen fue segmentada en cuatro partes iguales. Finalmente, al filtro de correlación (sintetizado y mostrado en la Figura 1C)) resultante del promediado de las señales se le aplicó la TF y la no linealidad con un factor $k = 0.5$.

Mediante el uso de un valor PSR umbral $r_{th} = 10$ se puede realizar una perfecta discriminación entre las dos clases. De esta manera, las imágenes de prueba que corresponden a la misma clase que la plantilla biométrica producen valores PSR mayores o iguales que r_{th} (Línea segmentada roja). Mientras que las imágenes de las clases falsas producen valores PSR menores que r_{th} (Líneas continuas verdes).

En un primer experimento se correlacionaron, mediante el Algoritmo 2, 6000 imágenes palmares de prueba con la plantilla biométrica de la clase 1. Las Figuras 3 y 4 presentan los resultados de las clases 200 y 500, respectivamente. Como se puede observar en dichas figuras (al igual que en la Figura 2), las primeras dos

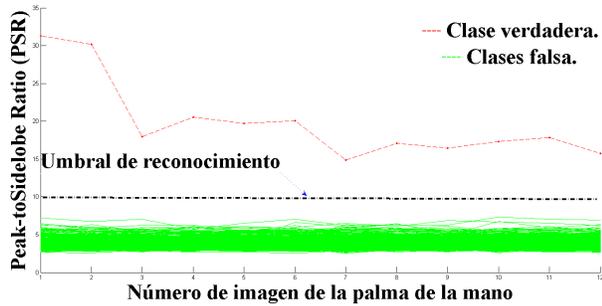


Figura 2. Resultado en valores *PSR* de la correlación de la plantilla biométrica de la clase 1 con el conjunto de prueba.

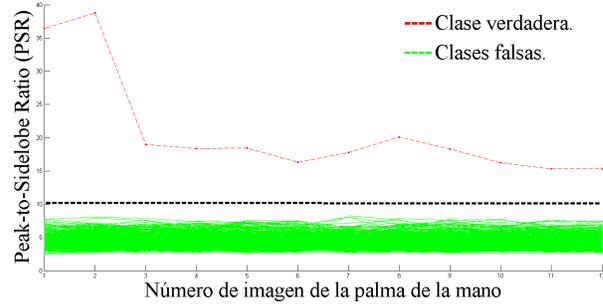


Figura 4. Resultado en valores *PSR* de la correlación de la plantilla biométrica de la clase 500 con el conjunto de prueba.

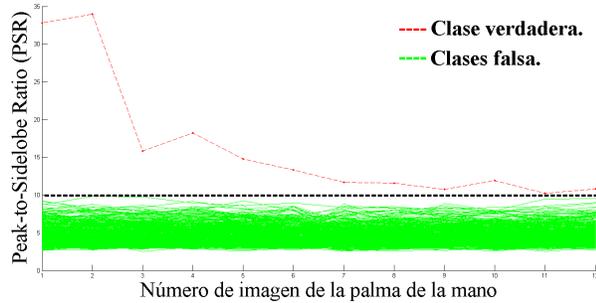


Figura 3. Resultado en valores *PSR* la plantilla biométrica de la clase 200 con el conjunto de prueba.

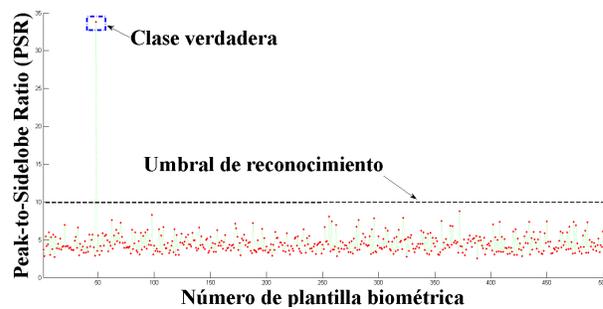


Figura 5. Resultado en valores *PSR* de la correlación de una imagen de la palma de la mano perteneciente a la clase 50 con la galería *G*.

imágenes de la palma de la mano producen valores *PSR* altos debido a que forman parte de la plantilla biométrica. Para el resto de las imágenes, su valor *PSR* es menor pero sigue siendo mayor o igual que r_{th} .

En el segundo experimento se hizo una evaluación intensiva usando la métrica conocida como Tasa de Identificación y Detección (TID) del protocolo de evaluación FERET [13]:

$$TID(r_{th}, 1) = \frac{|p_j(x, y) : rank\{p_j(x, y)\} = \Upsilon picop_{sr} \geq r_{th}|}{|P_G|} \quad (10)$$

donde $p_j(x, y)$ es la imagen de prueba, $rank$ corresponde al paso 11 del Algoritmo 1 y P_G que es el conjunto de prueba que contiene imágenes palmares de personas registradas en la galería G . Una imagen de prueba es detectada e identificada cuando se identifica correctamente y su valor $picop_{sr}$ es mayor o igual que el umbral de reconocimiento r_{th} .

El conjunto P_G contiene 500 clases \times 12 imágenes/clase = 6000 imágenes palmares a identificar. Cada imagen fue correlacionada con 500 plantillas biométricas almacenadas en la galería. La Figura 5 muestra el resultado de correlacionar una imagen de prueba (de la misma clase que la plantilla biométrica en el registro 50) con la galería. Como se puede observar, se produce un valor *PSR* mayor que el del resto con la plantilla en el registro 50. Esto indica que la plantilla biométrica contiene la suficiente información para reconocer correctamente

imágenes de su misma clase.

El algoritmo propuesto es capaz de identificar y detectar correctamente una imagen palmar en el 98.48 % de las veces.

5. Conclusiones

Se ha propuesto un algoritmo para el reconocimiento biométrico de la palma de la mano mediante filtros de correlación. Se concluye lo siguiente: a) Se puede identificar a una persona, empleando únicamente cuatro imágenes palmares en el diseño del filtro, b) La segmentación de las imágenes de entrenamiento y el promediado de dichos segmentos, proporciona mayor robustez al algoritmo frente superficies de la palma de la mano deformadas, c) A base de experimentos se determinó que el factor de no linealidad $k=0.5$ hace que el algoritmo discrimine mejor la clase verdadera de las clases falsas, y d) Las imágenes con variaciones en iluminación mejoran manteniendo la información de bordes aplicando el filtrado de retina.

Como trabajo futuro se propone el desarrollo de un algoritmo que mediante una imagen obtenida de una cámara, detecte y extraiga la región palmar para la creación de una plantilla biométrica o realizar el reconocimiento en tiempo real.

Agradecimientos

Este trabajo fue parcialmente apoyado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) mediante la beca escolar No. 344833/239152 para el primer autor.

Referencias

- [1] Subcommittee on Biometrics and Identity Management, "Biometrics glossary," National Science and Technology Council (NSTC), Technical report, September 2006.
- [2] Subcommittee on Biometrics and Identity Management, "Palm print recognition," National Science and Technology Council (NSTC), Technical report, August 2006.
- [3] D. Zhang, W. Zuo and F. Yue, "A Comparative Study of Palmprint Recognition Algorithms," *ACM Comput. Surv.*, Vol. 44(1), pp. 2:1-2:37, January 2012.
- [4] L. Wen-Xin, D. Zhang and X. Zhuo-qun, "Palmprint Recognition Based on Fourier Transform," *Journal of Software*, Vol. 13-5, pp. 879-886, January 2002.
- [5] H. Imtiaz and S. Fattah, "A DCT-based feature extraction algorithm for palm-print recognition," *IEEE International Conference on Communication Control and Computing Technologies*, pp. 657-660, October 2010.
- [6] A. Kong, D. Zhang and M. Kamel, "A survey of palmprint recognition," *Pattern Recognition*, Vol. 42(7), pp. 1408 - 1418, July 2009.
- [7] J. Cui, "Multispectral fusion for palmprint recognition," *Optik - International Journal for Light and Electron Optics*, Vol. 124(17), pp. 3067 - 3071, September 2013.
- [8] B. Javidi, W. Wang and G. Zhang, "Composite Fourier-plane nonlinear filter for distortion-invariant pattern recognition," *Optical Engineering*, Vol. 36, pp. 2690-2696, May 1997.
- [9] N.-S. Vu and A. Caplier, "Illumination-robust face recognition using retina modeling," *IEEE international conference on Image processing*, pp. 3253-3256, November 2009.
- [10] B. V. K. Vijaya-Kumar, A. Mahalanobis and R. Juday, *Correlation pattern recognition*. Pennsylvania: Cambridge University Press, 2005, p. 130-147.
- [11] A. VanderLugt, "Signal detection by complex spatial filtering," *IEEE Transactions on Information Theory*, Vol. 10, p.p. 139-145, April 1964.
- [12] D. Zhang, G. Zhenhua, L. Guangming, Z. Lei and Z. Wang-meng, "An Online System of Multi-spectral Palmprint Verification," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, pp. 480-490, February 2010.
- [13] P.-J. Phillips, P. Grother and R. Micheals, *Evaluation methods in face recognition*, New York: Springer, 2005, p. 329-348.

Medidor Modular Inalámbrico de la Tensión Hídrica del Suelo

Javier Villagrana Mancilla^a, Ricardo Francisco Núñez Pérez^b, Manuel Moisés Miranda Velasco^a, Miguel Enrique Martínez Rosas^a, y Horacio Luis Martínez Reyes^a

^aUniversidad Autónoma de Baja California, Facultad de Ingeniería, Arquitectura y Diseño.
Carr. Ensenada-Tijuana 3917, Col. Playitas, Ensenada, B.C., México, 22860.
<http://fiad.ens.uabc.mx/>

^bCICESE, Laboratorio de Control, Departamento de Electrónica y Telecomunicaciones.
Carr. Ensenada-Tijuana 3918, Zona Playitas, Ensenada, B.C., México, 22860.
<http://www.cicese.edu.mx/>

2013 Published by *DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Resumen

En este trabajo se presenta el diseño e implementación de un instrumento para la monitorización en forma continua de la tensión hídrica y temperatura del suelo, cuyo objetivo es incrementar la eficiencia de los cultivos agrícolas. El instrumento recarga su batería de manera autónoma y además incluye una interfaz gráfica para la visualización de los datos medidos, la cual opera desde un servidor Web dedicado montado sobre un sistema operativo Linux. El instrumento, que puede ser configurado de acuerdo a las necesidades del usuario final está diseñado para operar en condiciones ambientales reales y la selección tanto del sensor como de los componentes utilizados en la construcción del instrumento fue respaldada por pruebas de laboratorio y campo.

Palabras clave: Control de riego agrícola, Monitorización de variables ambientales, Redes de sensores, Tensión hídrica.

1. Introducción

El volumen de agua que se precipita sobre la tierra sin incluir los océanos es del orden de 110,000 km^3 por año. De esta cantidad, casi dos tercios se evaporan, o son transpirados por la vegetación. Los restantes 40,000 km^3 por año, se convierten en aguas superficiales y subterráneas (acuíferos). Parte de esta agua, se remueve (extrae) mediante infraestructura instalada por el hombre. A nivel mundial, las proporciones de extracción en función de su utilización son: 70 % agropecuaria, 11 % municipal y 19 % industrial [1].

Puesto que la producción agrícola depende de la cantidad de agua disponible para riego, es necesario hacer un uso racional de este preciado recurso. Para el caso particular del municipio de Ensenada, Baja California, en México, no se cuenta con ríos, y sólo existen arroyos que llevan agua cuando el volumen pluvial en invierno o verano es considerable [2]. Según datos de la Comisión Nacional del Agua (CONAGUA), el 82.6 % del agua que se extrae en la Zona Costa de B.C. proviene de acuíferos en equilibrio y sobreexplotados [3]. El problema en dicho municipio y en general, en México, radica en que, en la mayoría de las explotaciones agrícolas, el

riego se realiza de forma empírica y manual, es decir, la calendarización y la observación son los métodos de riego más utilizados.

Debido a las razones mencionadas, se deduce que es necesario utilizar el agua racionalmente, de modo que se garantice su propia existencia y el cuidado del medio ambiente, para ello se debe emplear tecnología que permita evaluar en forma continua la cantidad de agua utilizada, de manera que los usuarios puedan tomar decisiones (basadas en información confiable) en sus políticas de riego que den lugar a un aprovechamiento eficiente del recurso hídrico.

En este trabajo se describe el desarrollo de un sistema de monitorización continua que permita medir las condiciones de humedad y temperatura de suelo agrícola. Los datos obtenidos de las condiciones del suelo, en conjunto con la información del agua de riego aplicada, y la evolución de los frutos le permitirían al agricultor hacer una estimación de la cantidad mínima de agua requerida para mantener las condiciones de su cultivo y por lo tanto podría modificar tanto la cantidad como los periodos de riego, es decir se tendría un manejo eficiente de los recursos. La falta de instrumentación necesaria para medir la humedad del suelo, así como de una metodología o asesoramiento sobre la programación de riegos que indique cuando regar un cultivo y cuanta agua utilizar, se cuentan entre los principales factores que afectan el uso eficiente del agua. En general, los expertos sobre este tema están de acuerdo en que el factor que limitará la producción de los alimentos en el mundo no será la tierra o la tecnología sino el agua [4].

Este escenario requiere la introducción de tecnologías modernas, ya que el agua es el factor vital en la producción de alimentos, el mal uso, la contaminación y el desperdicio del agua, no es sólo irresponsable sino un atentado contra la vida misma (particularmente en ésta época de producción intensiva).

2. Desarrollo del instrumento

El diseño del instrumento se puede describir a partir de los módulos que se muestran en el diagrama a bloques de la figura 1.

En los siguientes apartados se presenta la descripción de cada uno de los módulos, incluyendo la circuitería y sus componentes principales.

2.1. Sistema de administración de energía

El sistema de administración de energía, también conocido como BMS (*Battery Management System*),

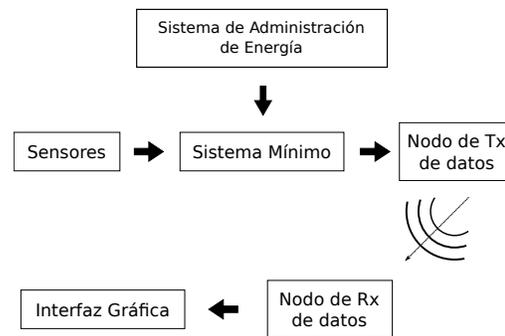


Figura 1. Diagrama a bloques del instrumento.

está compuesto por los siguientes elementos:

- **Batería.** En base al rendimiento y densidad de energía de las diferentes tecnologías descritas en el Manual de Baterías de Linden [5], se decidió utilizar baterías secundarias de iones de Litio (Li-ion), puesto que el instrumento requiere funcionar las 24 horas al día, durante todo el año.
- **Circuito controlador de energía.** Se seleccionó un controlador de carga apropiado (circuito MCP73833 de Microchip) para baterías de iones de Litio, el cual cumple con las funciones básicas para conformar el sistema de administración de energía.
- **Panel solar.** Se optó por utilizar celdas solares para alimentar al circuito controlador de energía, ya que es uno de los métodos que presenta una mayor capacidad de generación de energía con respecto a otras formas colectoras de energía [6].

Este sistema en particular cumple con las funciones de proporcionar la alimentación al instrumento, así como de proteger y prolongar la vida de las baterías.

2.2. Sistema mínimo

Para el control de periféricos y de las etapas de comunicación del instrumento se utilizó un dispositivo microcontrolador Atmega 328P del fabricante Atmel. Se decidió emplear éste dispositivo porque sus características de desempeño cubren las necesidades del proyecto.

Entre las características de mayor interés para este proyecto se encuentran las interfaces de comunicación proporcionadas por el microcontrolador, tales como: USART (*Universal Synchronous/Asynchronous Receiver/Transmitter* - Receptor/Transmisor Universal Síncrono/Asíncrono), SPI (*Serial Peripheral Interface* - Interfaz Serial de Periféricos) e I²C (*Inter-Integrated Circuit* - Inter-Circuitos Integrados). El módulo USART del

microcontrolador es fundamental para la configuración del instrumento, ya permite la comunicación con una computadora. Para nuestro caso, la conexión se realizó mediante un cable del tipo FTDI (marca de fabricante del circuito utilizado en conjunto con el cable), que es un conversor serie-USB (5 V, nivel TTL).

La utilización del protocolo SPI, que proporciona el microcontrolador, se utiliza para establecer comunicación con el convertidor analógico-digital, extraer los valores de los sensores, comunicarse con la memoria del tipo EEPROM (*Electrically Erasable Programmable Read Only Memory* - Memoria de solo escritura programable y borrable eléctricamente) y almacenar los datos adquiridos. La interfaz I²C permite la comunicación con el circuito temporizador de tiempo real, gracias al cual se puede realizar la escritura y lectura de la hora, fecha, alarmas, interrupción y temporizador.

El protocolo I²C es fundamental para establecer la sincronía, que el circuito de tiempo real proporciona, con todos los dispositivos electrónicos utilizados, así como también, para ajustar el periodo de muestreo de las variables de interés. Otra de sus actividades que realiza, es la conversión de diferentes datos tales como: hora, fecha y datos de las variables de interés que proporcionan los diversos dispositivos electrónicos, de tal manera que se permita una visualización adecuada para los usuarios.

2.3. Sensores

De acuerdo a [7], hay dos objetivos principales en la medición del contenido de agua en el suelo de los cultivos agrícolas.

1. Determinar el contenido de humedad del suelo. Esto se refiere a la cantidad de agua contenida en una unidad de masa o volumen de suelo. Con este dato se puede calcular el agua necesaria para restaurar la humedad del suelo a la capacidad de campo del cultivo.
2. Determinar la magnitud de la tensión del agua en el suelo o tensión hídrica. Representa la cantidad de trabajo necesario para remover una cantidad equivalente a una unidad de agua retenida más flojamente, medida en centibares (ó *kPa*) de tensión de agua del suelo.

Desde el punto de vista de la ingeniería, muchas de las propiedades del suelo están más íntimamente relacionadas con la medición de la magnitud de la tensión del agua en el suelo y no en su contenido de humedad, por el hecho de que la cantidad de agua no es tan im-

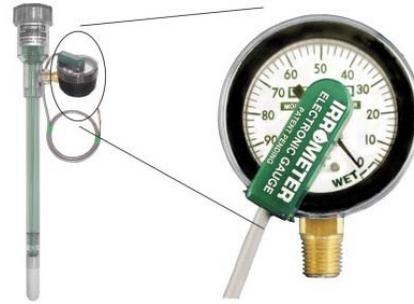


Figura 2. Tensiómetro Irrrometer modelo SR [4].

portante como lo difícil que es para la planta extraerla de la tierra.

Para llevar a cabo la medición de este parámetro se utilizó un tensiómetro como el que se muestra en la figura 2, que es un tubo lleno de agua cerrado herméticamente por uno de sus extremos y teniendo en el otro una cápsula de material cerámico poroso que se pone en contacto con el suelo. Junto al extremo cerrado se coloca un medidor de presión o vacuómetro, para medir presiones inferiores a la atmosférica (*i.e.*, depresiones). Esta cápsula porosa permite que la succión del suelo se transmita al agua del tensiómetro y la depresión se lee por el vacuómetro. La medida del vacuómetro se da en centibares (1 *cbar* = 0.10 *m* de columna de agua) y su escala esta graduada de 0 a 100 centibares. Una lectura cero indica que el suelo no ejerce ninguna retención sobre el agua por lo que está saturado o sobresaturado [8].

Procesos tales como la asimilación, la respiración, la transpiración, la fotosíntesis y el crecimiento dependen de la temperatura. La mayoría de las reacciones químicas que tienen lugar en la planta y en el suelo ocurren con mayor velocidad a temperaturas altas. Los procesos físicos de difusión y flujo viscoso dependen de la temperatura. Las plantas disponen de más agua en un suelo caliente que en uno frío, porque el suelo caliente retiene menos agua, cualquiera que sea su potencial. En sistemas equilibrados, la dependencia de la temperatura del potencial de agua indica un aprovechamiento inadecuado de energía [7].

La sonda de temperatura seleccionada (modelo 6470 de Davis Instruments), está diseñada para medir la temperatura del suelo. El encapsulado es de acero inoxidable y el sensor es un termistor de precisión y coeficiente negativo de temperatura (NTC por sus siglas en inglés) que produce un cambio de resistencia proporcional a la temperatura. El termistor es un dispositivo extremadamente no lineal y su curva característica varía dependiendo del fabricante; sin embargo, la curva de comportamiento de un termistor puede obtenerse

de forma bastante aproximada por medio de la Ec. (1) conocida como ecuación de Steinhart-Hart.

$$\frac{1}{T} = A + B [\ln(R_{term})] + C [\ln(R_{term})]^3 \quad (1)$$

donde:

T : Temperatura en $^{\circ}K$.

R_{term} : Resistencia del termistor en $k\Omega$.

A, B, C : Coeficientes de la curva de aproximación.

Los coeficientes A, B , y C pueden calcularse al seleccionar tres puntos de la tabla o curva que acompaña el termistor, sustituyendo los valores de temperatura y resistencia de cada uno de los tres puntos seleccionados en la Ec. (1) se forma un sistema de 3 ecuaciones simultáneas con tres incógnitas, al resolver dicho sistema de ecuaciones simultáneas se obtienen los valores de A, B , y C .

2.4. Nodos para la transmisión inalámbrica

Para agregar conectividad inalámbrica al instrumento se eligieron los módulos XbeePro de MaxStream S1. Los módulos Xbee pueden ser configurados desde una computadora utilizando el programa X-CTU, el cual permite realizar estas operaciones de manera más intuitiva, fácil y rápida que utilizando la Hyperterminal de Windows o bien desde el propio microcontrolador.

La Serie 1 incluye el firmware 802.15.4 que permite redes punto a punto o de topología en estrella. Este firmware ofrece conversión analógica-digital (ADC por sus siglas en inglés), así como entradas y salidas digitales. El sistema basado en el estándar 802.15.4 es mucho más rápido que ZigBee.

2.5. Interfaz gráfica

Para la visualización de los datos de tensión hídrica y temperatura del suelo, se optó por instalar un servidor Web en una computadora con sistema operativo Linux, en particular con una distribución Ubuntu versión LTS 10.04 con ayuda de la paquetería Apache 2 y PHP 5. De esta forma, se puede tener acceso a la información en línea y sin necesidad de instalar algún programa en el dispositivo de despliegue (computadora, smarthphone, tablet o cualquier otro dispositivo con acceso a internet).

A grandes rasgos, en la computadora que mantiene en operación el servidor corre un programa que lee los datos a través del puerto serie universal (USB por sus siglas en inglés) y genera un archivo de texto, en el que se almacenan de forma ordenada los datos recibidos. Por otro lado, el paquete Apache se utiliza en conjunto con el paquete Php, el cual lee el archivo de texto

para posteriormente mostrar los datos en formato html y javascript, de tal manera que se puede generar una gráfica de la variable medida (que en realidad es una imagen) por medio de la librería Jpgraph.

3. Resultados experimentales

El tensiómetro provee un voltaje de salida en relación con el voltaje de alimentación, que está restringido al alcance de 2.7 V_{cd} mínimo y 5.5 V_{cd} máximo. Para poder convertir el voltaje medido a centibares el fabricante proporciona la Ec. (2), así como los valores correspondientes cuando se alimenta con 5 V_{cd} .

$$cb = \frac{(V_{sm} - V_m)}{((265V_g)/(V_{mam} - V_{mim}))} \quad (2)$$

donde:

V_{sm} : Voltaje máximo de salida calculado del medidor.

V_m : Voltaje medido.

V_g : Volts por grado.

V_{mam} : Valor máximo en centibares del medidor.

V_{mim} : Valor mínimo en centibares del medidor.

Para corroborar la información proporcionada por el tensiómetro, se realizó el montaje de la figura 3. El sensor se excitó con 5 V_{cd} y los datos fueron registrados empleando el módulo de adquisición de datos NI USB-6216 de National Instruments, el cual cuenta con 16 entradas analógicas y una resolución de 16 bits. El módulo se controló mediante el uso de rutinas escritas en lenguaje Matlab, las cuales permiten efectuar mediciones del sensor cada 15 minutos y generar un archivo por día para los datos obtenidos, los cuales además incluyen el día y hora de la medición.

Los resultados obtenidos de la caracterización del tensiómetro se pueden observar en la figura 4, en donde se aprecia como al principio, el medidor arroja una lectura de aproximadamente 61 cb y aumenta a 65 después de los 1000 minutos de estar midiendo. Esto nos indica que el suelo se estaba secando. Poco antes de los 1500 minutos, se agregó agua (simulando el riego) al contenedor con tierra en donde se insertó el tensiómetro, lo que provocó que la humedad del suelo llegara a 12 cb lo que se interpreta como un suelo adecuadamente mojado.

La simulación de un pequeño riego en el montaje es semejante a la dinámica que tienen los suelos de cultivo, para los que se busca mantener la tensión hídrica del suelo en cierto intervalo (ni seco ni saturado), así que este proceso de secado natural y regado se mantendrá mientras el cultivo lo requiera dentro de los márgenes óptimos para el mismo.

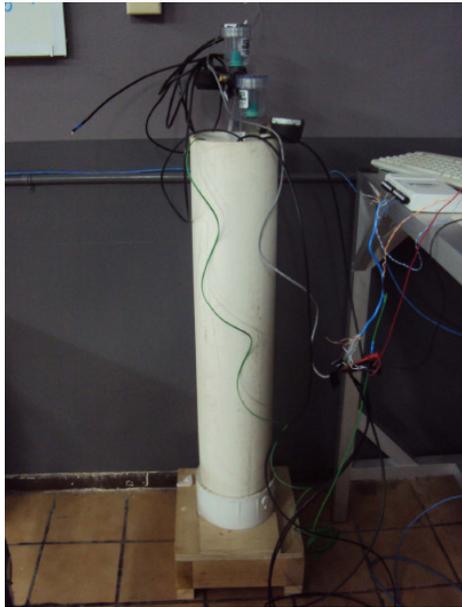


Figura 3. Montaje para la caracterización del tensiómetro.

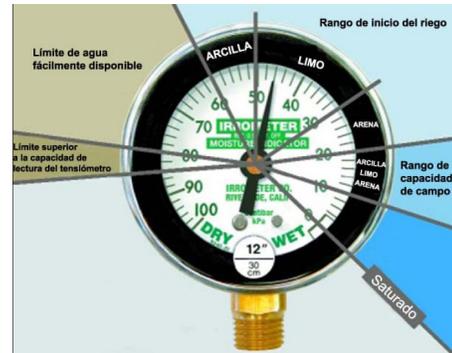


Figura 5. Interpretación de lecturas del tensiómetro (tipo de suelo).

Tabla 1. Interpretación de lecturas del tensiómetro [9].

Lectura (cb)	Interpretación de lecturas
0-10	Suelo saturado
10-30	Suelo adecuadamente humedecido
30-60	Alcance usual para irrigación
60-100	Alcance usual para irrigación en suelos pesados
100-200	El suelo se está secando peligrosamente, la producción puede afectarse severamente

Aunque los valores óptimos de tensión hídrica dependen tanto del tipo de suelo (figura 5), como del tipo de cultivo [4], los valores de las lecturas de la tabla 1 se pueden utilizar como una guía general, teniendo en cuenta que la capacidad de lectura del tensiómetro para una lectura confiable de acuerdo al fabricante es de alrededor de 80 centibares.

3.1. Integración del instrumento

Después de las pruebas preliminares, tanto en el laboratorio como en el campo, de los diversos módulos que componen el instrumento, se llegó al diseño final de las tarjetas de circuito impreso.

Para instalar y proteger estas tarjetas se escogió una caja de montaje de policarbonato, que cumple con las normas Nemas 4, 4X, 12, 13, i.e., está sellada contra agua y polvo y es resistente a la corrosión. Por lo que es ideal para uso en interiores y exteriores, protegiendo el equipo contra salpicaduras de agua, filtraciones de agua, derrames de agua sobre la caja y condensación externa severa.

El prototipo final del instrumento se muestra en la figura 6.

Para poder conectar los sensores a la caja, fue necesario agregar cuatro conectores especiales que cumplan con la norma Nema 4x; esto asegura que aún con estas perforaciones la caja sigue manteniendo toda la electrónica protegida. En la tabla 2, se muestran las

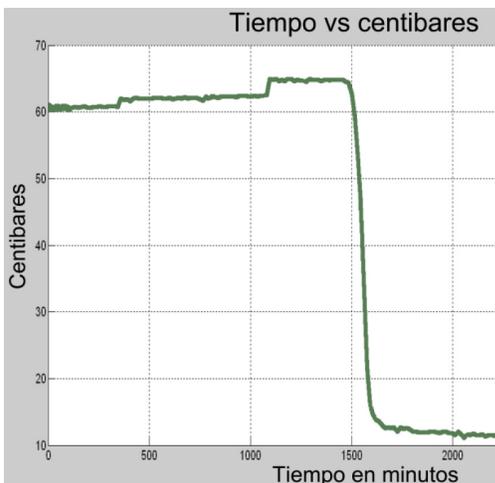


Figura 4. Caracterización del tensiómetro Irrometer modelo SR.



Figura 6. Prototipo final del instrumento.

Tabla 2. Especificaciones del instrumento.

Tipo	Valor	Unidades
Alcance de tensión hídrica.	0-75	Centibares (<i>KPa</i>)
Alcance de temperatura.	25 ± 15	°C.
Precisión tensión hídrica.	± 3	Por ciento.
Precisión temperatura.	± 0.5	Por ciento.
Voltaje de alimentación.	3.3	V
Resolución.	0.8	mV
Consumo energético.	49.5	mWh.

principales especificaciones con las que cumple el instrumento desarrollado

El instrumento se mantuvo en funcionamiento en campo, ininterrumpidamente, durante 4 meses, en un cultivo de fresa en San Quintín, B.C. México. En la figura 7 se puede observar el instrumento operando en el campo de cultivo.

3.2. Despliegue y visualización de la información

La visualización de los datos que llegan del medidor al servidor Web, vía inalámbrica, con ayuda de los radios



Figura 7. Medidor instalado en un cultivo de fresas.



Figura 8. Página principal para el usuario del instrumento.

Xbee, es de la forma mostrada en la figura 8. Para facilitar la lectura de datos, la gráfica cuenta con un soporte interactivo en donde se utilizan menús para selección, por ejemplo para el caso en que se desea ver las mediciones de un mes particular, al seleccionar el mes deseado se despliegan los días que cuentan con datos en el servidor.

Después de haber seleccionado el día, la información se mostrará en forma gráfica (figura 9). En la parte inferior, se indica la figura y color que representa cada uno de los 4 sensores con los que cuenta el instrumento. El tensiómetro 1 y el termistor 1, se encuentran enterrados a una profundidad de 15 centímetros, por otro lado el tensiómetro 2 y el termistor 2 se encuentran a 30 centímetros de profundidad.

Claramente se ve que los sensores que se encuentran más cerca de la superficie, presentan una dinámica mayor, al encontrarse más expuestos a la humedad superficial; por el contrario, los sensores que se encuentran a mayor profundidad entregan datos que permanecen casi constantes.

En las abscisas, se despliega la hora en que se adquirió la medición y en las ordenadas la tensión hídrica o la temperatura (en *cb* ó °C, respectivamente) dependiendo del sensor que se esté consultando.

4. Conclusiones

Se diseñó e implementó un instrumento para medir la tensión hídrica y temperatura del suelo, el cual incluye un suministro de energía que le permite funcionar de manera continua e ininterrumpida durante una temporada completa de cultivo. El instrumento permite verificar

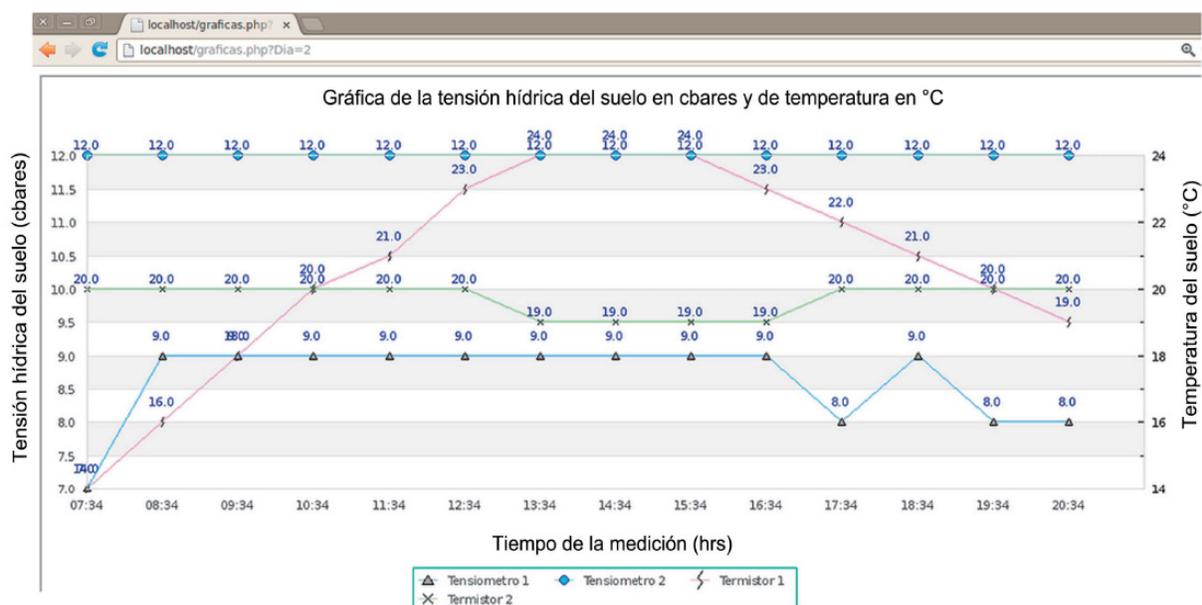


Figura 9. Despliegue de los datos obtenidos del instrumento.

el estado de las variables medidas de forma remota, ya que incluye un módulo de transmisión inalámbrica de datos que se enlaza con una computadora personal.

El prototipo es totalmente configurable a las necesidades de presentación de información del agricultor y puede ser utilizado prácticamente en cualquier tipo de cultivo sobre suelo agrícola. La programación del instrumento se realizó con el objetivo de que fuera adaptable a las necesidades específicas del usuario final.

Durante el desarrollo del instrumento, se efectuaron varias pruebas en el laboratorio y en los campos de cultivo. Los resultados obtenidos, corroboran su adecuado funcionamiento, en condiciones reales de operación, por lo que se puede mencionar que se tiene un sistema capaz de competir con otros equipos comerciales más costosos.

Por otro lado, cuando se colocan dos sensores a diferentes profundidades, se puede conocer el gradiente de infiltración del agua en el suelo, con todo el potencial que esto representa.

El usuario puede encontrar en éste instrumento una herramienta que le permita tomar decisiones basadas en información para controlar la programación de riegos, lo cual implicaría las siguientes ventajas: ahorro de agua, optimización de aplicación de fertilizantes, reducciones de costo por mano de obra y en general un mejor control del desarrollo de los cultivos. Este instrumento, tiene además la característica de que al ser un sistema abierto puede ser modificado para cubrir necesidades particulares, lo que no sucede con otros instrumentos comerciales que además por lo regular involucran modificaciones tecnológicas periódicas para mantener en

operación los equipos y por lo tanto incrementan el costo para el agricultor.

Agradecimientos

Los autores desean agradecer el apoyo brindado por la FIAD-UABC, CONACYT y CICESE para la realización de este trabajo.

Referencias

- [1] FAO. "AQUASTAT - Sistema de Información global sobre el agua y la agricultura de la FAO", 2013.
- [2] J. Villagrana Mancilla. "Medidor modular inalámbrico de la tensión hídrica del suelo". Tesis de Maestría, Universidad Autónoma de Baja California, Ensenada, B.C., 2013.
- [3] Comisión Nacional del Agua. "Atlas del agua en México 2012", 2012.
- [4] Irrrometer. "Moisture Sensor Agricultural Irrigation Design Manual". Technical report, Irrrometer Company, Riverside, CA, 2005.
- [5] T. Reddy. *Linden's Handbook of Batteries*. McGraw-Hill Professional, 4th edition, 2010.
- [6] J. M. Conrad. "A survey of energy harvesting sources for embedded systems". *IEEE SoutheastCon 2008*, pp. 442-447, April 2008.
- [7] A. G. Sapat *Física de los suelos: principios y aplicaciones*. Limusa-Wiley, México, D.F., 1976.
- [8] C. Cadahia López. *Fertirrigación. Cultivos hortícolas, frutales y ornamentales*. Ediciones Paraninfo S.A, Madrid, 2005.
- [9] Irrrometer. "Soil Moisture Basics", 2013.

Uso de Matlab para la Síntesis en Lenguaje VHDL de Decodificadores Viterbi

Luis Alberto Luna Espinosa^a, Juan de Dios López Sánchez^a, Juan Iván Nieto Hipólito^a, y Mabel Vázquez Briseño^a

^aUniversidad Autónoma de Baja California, Facultad de Ingeniería, Arquitectura y Diseño.
Carr. Ensenada-Tijuana 3917, Col. Playitas, Ensenada, B.C., México, 22860.
kel.night.eyes@hotmail.com, {[jddios](mailto:jddios@uabc.edu.mx), [jnieto](mailto:jnieto@uabc.edu.mx), [mabel.vazquez](mailto:mabel.vazquez@uabc.edu.mx)}@uabc.edu.mx

2013 Published by *DI^{TU}100ci* @ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Resumen

En el presente trabajo se muestra una forma más sencilla de programar decodificadores de Viterbi utilizando dispositivos programables FPGA (Field Programmable Gate Array). Estos decodificadores son creados a partir de los puntos de suma que describen a las salidas de un codificador convolucional no retroalimentado a una tasa de codificación de $\frac{1}{2}$. Esta relación ha sido utilizada para implementar un programa en Matlab, el cual genera decodificadores en lenguaje VHDL para un dispositivo FPGA a partir de un conjunto de entidades básicas utilizados para formar células ACS.

Palabras clave: Células de suma-compara-selección, Código Convolucional, Decodificador Viterbi.

1. Introducción

La codificación de canal es una herramienta ampliamente utilizada en los sistemas de comunicaciones debido a las mejoras que esta puede aportar al sistema, tales como mejor desempeño del Bit Error Rate (BER), reducción de potencia, incluso incrementar la tasa de transmisión manteniendo la misma calidad del enlace. Además, este tipo de técnica supone una mejora considerable sin implicar altos costos de hardware especializado, siendo adaptable a todo tipo de esquemas de comunicación debido a sus diversas propiedades y capacidades correctivas, por lo que para cada canal existen una o varias opciones viables [1].

Un caso en particular de este tipo de codificaciones que a pesar de que ha transcurrido un largo tiempo desde su aparición, son los códigos convolucionales, mismos que en la actualidad siguen siendo ampliamente utilizados mediante diversas variantes, y que dependiendo de la necesidad de la aplicación tiene varias ventajas sobre otros tipos de codificaciones [2].

En la industria de las comunicaciones es común la utilización de arreglos lógicos programables como lo FPGA para la implementación de funciones u operaciones que requieren tiempo real tales como la sincronización, filtrado, ecualización, etc. debido a su estructura que permite implementar de funciones y operaciones lógicas de forma eficiente en términos de tiempo de operacio-

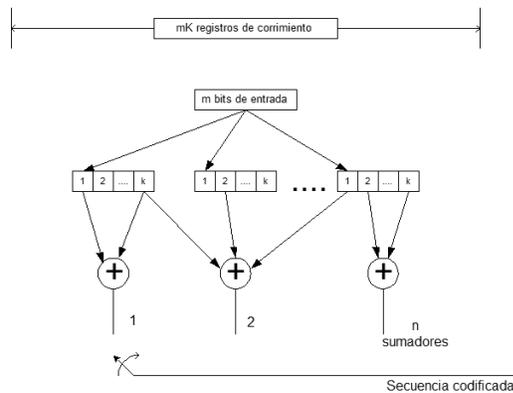


Figura 1. Diagrama de bloques de un codificador convolucional.

nes permitiendo realizar operaciones en pocos ciclos de reloj reduciendo sustancialmente el retardo [3]. Además, al ser programables permite actualización y la reconfiguración del receptor [4]. En el presente trabajo se trata el diseño e implementación de código de Matlab, el cual, a partir de los vectores que describen a los puntos de suma de un código convolucional, puede generar decodificadores convolucionales utilizando un conjunto de bloques básicos implementados como Entidades en VHDL. Dichos bloques son utilizados para crear células de Suma-Compara-Selección (ACS: Add-Compare-Select), mismas que serán interconectadas para dar origen al decodificador en cuestión. La complejidad de los decodificadores también es descrita.

2. Teoría de funcionamiento

2.1. Codificador Convolucional

Un código convolucional es generado cuando una secuencia de datos atraviesa una serie de registros de estados finitos, los cuales forman una secuencia nueva al hacer una combinación lineal de los datos en dichos registros. Los registros consisten en mK registros de corrimiento y n generadores de funciones lineales (puntos de suma), las conexiones entre los registros y los puntos de suma son generalmente descritos por n vectores V de longitud K . El codificador es alimentado con m bits a la vez, mientras que la salida está dada por los n generadores, de modo que la tasa del codificador puede definirse aproximadamente como m/n . El parámetro K es llamado constante de restricción y define la cantidad de registros a través de los cuales pasa un bit de información influyendo la salida del codificador antes de abandonar el mismo [5]. Un codificador convolucional genérico es ilustrado en la Figura 1.

Este diagrama puede ser también utilizado para conocer la salida generada por un codificador a partir de

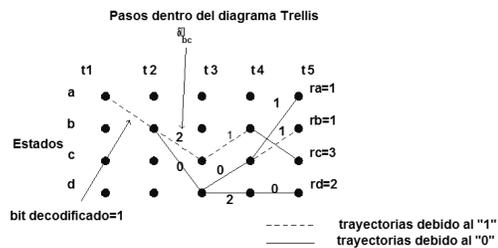


Figura 2. Pasos dentro del diagrama de enrejado.

la secuencia de datos entrantes, mediante observar los trayectos entre estados generados por la entrada.

2.2. Algoritmo del decodificador Viterbi

El algoritmo de Viterbi [6] se basa en un decodificador de máxima verosimilitud y en la recursividad que presenta el diagrama de Trellis. El algoritmo de esta decodificación consiste en hacer una comparación de la secuencia recibida en los instantes t_i (con $i = 1, 2, 3, \dots$) con todos los caminos de Trellis entrando a los diversos estados en t_i , dicha comparación deja unas marcas sobre los caminos denominadas Distancia de Hamming, $\delta_{x,y}$, donde x es el estado de origen y y es el estado al que se llega, posteriormente se calcula la Distancia de Hamming Acumulada de cada estado para ese instante t_i que es la suma de los $\delta_{x,y}$ desde el tiempo t_1 hasta t_i . Una vez obtenida la Distancia de Hamming Acumulada para cada estado, se comparan las pertenecientes a un mismo estado, es decir se compara $\delta_{x,y}$ con $\delta_{z,y}$ para cada uno de los estados, eliminando aquellas Distancias de Hamming Acumuladas más grandes, de modo que solo quede un único camino entrando a cada estado, convirtiéndose la distancia de Hamming Acumulada en la Medida del Estado Γ_y , lo cual facilita el cálculo de las posteriores Distancias de Hamming Acumuladas. El proceso es repetido para los tiempos t_i posteriores y después de un número de comparaciones y avances dentro del diagrama de Trellis (en práctica 4 o 5 veces K), los caminos van convergiendo en un mismo origen del cual es posible extraer los datos codificados, como se muestra en la Figura 2 [6].

3. Implementación del decodificador Viterbi

Las células ACS son una implementación del algoritmo de Viterbi las cuales se encargan de obtener la diferencia entre la secuencia recibida y el código generado por cada transición entre los diversos estados, calcula las métricas de estado, y decide que trayectos almacenar dada la menor métrica de cada estado para cada momento t_i . Esta implementación permite proce-

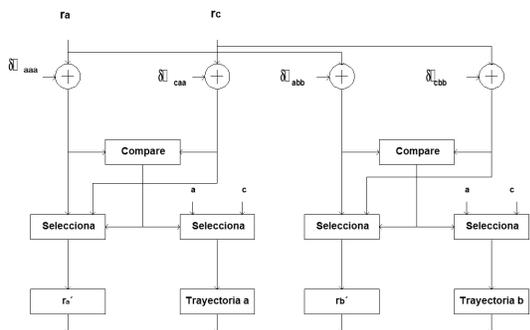


Figura 3. Diagrama de bloques de una implementación de células ACS en el decodificador Viterbi con $K = 3$.

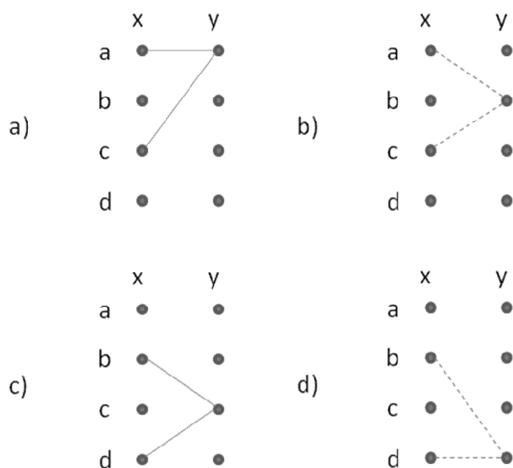


Figura 4. Relación entre los estados origen (x) y los estados (y) para los estados a, b, c y d respectivamente.

sar dos estados a la vez cuyos estados precedentes sean comunes a los estados procesados. Actualmente existen variantes de esta implementación en función del dispositivo en el cual se implementa, dichas variaciones permite mejorar ciertos aspectos tales como la latencia, minimizar la memoria requerida para almacenar las trayectorias de los estados, a cambio de algún parámetro no significativamente relevante para determinada aplicación [2],[7]. Un diagrama de bloques que ilustra el funcionamiento de las células ACS se muestra en la Figura 3.

Las células ACS son capaces de procesar dos estados consecutivos (estados destino), cuyos estados anteriores son comunes (estados origen). La relación entre estados origen y estados destino resulta de observar el diagrama de trellis y está dada por la cantidad de registros utilizados en el codificador, de modo que para un codificador con $K = 3$ los estados origen de cada estado destino quedan ilustrados en la Figura 4.

Una relación más precisa puede ser obtenida de generalizar la cantidad de registros que el codificador pueda tener. Para un determinado cantidad de registros K ,

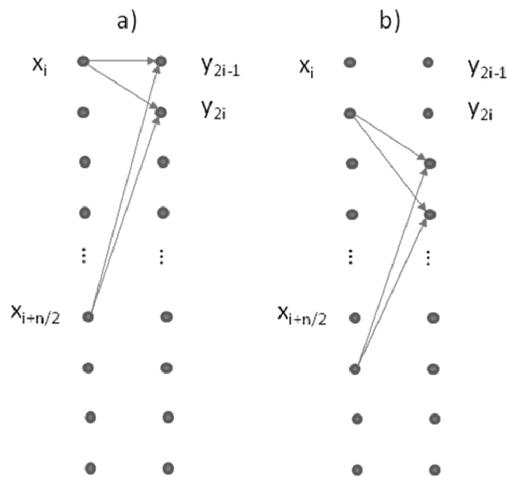


Figura 5. Relación genérica entre los estados x y los estados y para a) $i = 1$, b) $i = 2$.

se tienen 2^{K-1} estados. Dado que la transición entre estados está únicamente en función de la entrada tenemos para cada estado destino y_{2i-1} y y_{2i} le corresponden los estados origen x_i y $x_{i+\frac{n}{2}}$, con $n = 2^{K-1}$ e $i = 1, 2, \dots, \frac{n}{2}$. En la Figura 5 observamos la representación de esta relación entre estados.

4. Implementación en el FPGA

Un primer decodificador fue hecho a partir de un código $K = 3$, $n = 2$, mismo que sirvió para definir las entidades de los bloques básicos requeridos para la implementación de las células ACS. Dichas entidades son:

DeltaXY Este componente se encarga de comparar la secuencia codificada contra la secuencia generada por los diversos trayectos en el diagrama de Trellis, dando como salida la cantidad de bits en los que difieren.

PathHistory Es una memoria que almacena los el camino con menor métrica de estado entrando a un determinado estado destino, mientras que el bit decodificado es el ultimo bit de la memoria asociada al estado con menor métrica de todos.

ACS Es el componente central del decodificador, ya que para cada estado realiza el cálculo de la trayectoria con menor métrica, esto en función de la información recibida del componente *DeltaXY*.

A partir de las entidades que describen a los bloques básicos y la relación general descrita anteriormente, se elaboró un programa (script) en Matlab, el cual, a partir de seleccionar los vectores V que describen a

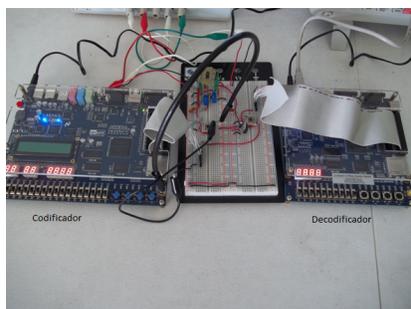


Figura 6. Implementación física.

los puntos de suma, genera código en VHDL con las instancias e interconexiones necesarias para la descripción del decodificador especificado, siendo este código perfectamente portable a cualquier dispositivo FPGA, con el único requerimiento de las señales de reloj que indiquen los tiempos de bits.

5. Resultados

El programa en código Matlab realizado fue utilizado para generar un conjunto de codificadores óptimos [6]-[8] en cuanto a sus capacidades correctivas para diversos valores de K con la finalidad de evaluar la complejidad del mismo. Estos códigos fueron compilados en el IDE QUARTUS II™ para un dispositivo FPGA Cyclone II™. El consumo de unidades lógicas con respecto a K para valores de $K = 3, 4, 5, 6, 7, 8$, de la cual se observa que la complejidad, relacionada a la cantidad de unidades lógicas requeridas, incrementa exponencialmente de acuerdo a [1]. Para la verificación del correcto funcionamiento se implementó el código obtenido en los dispositivos de FPGAs, tanto para la codificación convolucional como la etapa de decodificación Viterbi. La implementación física del codificador y el decodificador se muestran en la Figura 6.

En la Figura 7 se muestran formas de onda No Retorno a Cero (NRZ) que se utilizaron para la transmisión por un canal físico entre el transmisor (codificador) y el receptor (decodificador).

6. Conclusiones

Se ha presentado una relación que permite predecir la forma en la cual estarán interconectados los estados para la implementación de un decodificador utilizando el algoritmo de Viterbi y células ACS. La utilización de un programa en código de Matlab que haga uso de estas entidades permite la generación de diversos tipos de decodificadores sin presentar ninguna dificultad en el



Figura 7. Formas de ondas requeridas para la transmisión de los datos codificados.

desarrollo del mismo, lo cual permite realizar pruebas experimentales con rapidez, bajo diversos escenarios como variantes de canal, modulaciones, o sistemas en general, facilitando en diseño de los decodificadores. Este programa permite que el usuario no tenga la necesidad de implementar los decodificadores en código VHDL, sino de utilizar un script o código en Matlab que permite implementarlo rápido y fácilmente.

Agradecimientos

Los autores desean agradecer a la Administración de la Facultad de Ingeniería, Arquitectura y Diseño (FIAD) de la Universidad Autónoma de Baja California (UABC) y al CONACYT.

Referencias

- [1] A. Neubauer, *Coding Theory: Algorithms, Architectures and applications*, England Wiley, 2010.
- [2] S. V. Maiya, Daniel J. Costello and T. E. Fuja, "Low Latency Coding: Convolutional Codes vs. LDPC Codes," *IEEE Transactions on Communications*, Vol. 60, No. 5, May, 2012.
- [3] S. Karris. *Digital Circuit Analysis and Design with Simulink Modeling and Introduction to CPLDs and FPGAs*. Orchard Publications, USA, 2007.
- [4] W. Tuttlebee. *Software Defined Radio*. John Wiley and Sons, Sussex, Inglaterra, 2002.
- [5] A. Viterbi. *Communication systems engineering*. Prentice Hall, Upper Saddle River, New Jersey, 2001.
- [6] A. Viterbi. "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, Vol. 13, pp. 260-269, Abril 1967.
- [7] B. He, R. Schober. "Bit-Interleaved Coded Modulation for Hybrid RF/FSO Systems," *IEEE Transactions on Communications*, Vol. 57, pp. 3753-3763, Diciembre 2009.
- [8] B. Sklar. *Digital Communications*. Prentice Hall, New Jersey, p.408, 2001.

Sistema Inmerso de Bajo Costo para la Localización y Segmentación de Uvas

Eduardo Antonio Murillo Bracamontes^{a,b}, Miguel Enrique Martínez Rosas^b, Manuel Moisés Miranda Velasco^b, Horacio Luis Martínez Reyes^b, Jesús Raúl Martínez Sandoval^b

^aUniversidad Autónoma de Baja California, Centro de Ingeniería y Tecnología.
Blvd. Universitario #1000, Valle de las Palmas, Tijuana, B. C., México, 21500.
emurillo@uabc.edu.mx

^bUniversidad Autónoma de Baja California, Facultad de Ingeniería, Arquitectura y Diseño.
Carr. Ensenada-Tijuana 3917, Col. Playitas, Ensenada, B.C., México, 22860.
{[emartine](mailto:emartine@uabc.edu.mx),[mmiranda](mailto:mmiranda@uabc.edu.mx),[hmartine](mailto:hmartine@uabc.edu.mx),[rmartinez33](mailto:rmartinez33@uabc.edu.mx)}@uabc.edu.mx

2013 Published by *IEEE U100ci* @ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Resumen

En la determinación óptima del estado de madurez de frutas y vegetales, el color es uno de los parámetros más importantes para los consumidores. Recientemente, el uso de la tecnología en el área de agricultura ha sido cada vez más importante en la evaluación de la madurez de los frutos, a través del empleo de sistemas de visión por computadora, reduciendo considerablemente el margen de error. Este artículo presenta el diseño y desarrollo de un sistema inmerso (o empotrado) basado en una plataforma ARM de bajo costo y el software de código abierto OpenCV, mediante el cual se utiliza visión por computadora para localizar uvas en racimos, así como bayas de uva que se utilizan en muestreos de evaluación. Mediante el empleo de algoritmos de segmentación se obtienen sub-imágenes de cada uva para extraer la información de los parámetros de color a^* y b^* en el espacio CIELAB.

Palabras clave: Madurez de la uva, Morfología Digital, Sistema de Visión, Transformada de Hough.

1. Introducción

El color es la primera sensación agradable o desagradable que los consumidores perciben de las frutas, lo cual también afecta el resto de las sensaciones percibidas. Por esta razón, es esencial el poder cuantificar esta primer sensación. Puesto que el color constituye la primer relación establecida entre el producto agrícola y el consumidor mediante la cual se juzga el estado de madurez, es entonces uno de los

parámetros más importantes para tomar la decisión de compra [1]. Por esta razón, monitorear y controlar el estado de madurez antes de la cosecha es una tarea muy importante en la agricultura, debido a que las diferentes etapas que le siguen a la cosecha, tales como el almacenamiento y distribución de mercado determinan la calidad del producto final en términos de satisfacción al cliente [2].

El empleo de sistemas de visión por computadora para analizar imágenes, tiene mucho potencial en tareas

agrícolas automatizadas y su uso se ha incrementado recientemente[3]. Los investigadores en ésta área indican la factibilidad particularmente en el uso de sistemas de visión por computadora para mejorar la calidad de los productos y con ello liberar el recurso humano de las tareas de selección manual de frutos en la industria agrícola [4], lo cual brinda ahorro significativo de mano de obra [5]. Un sistema básico de visión por computadora está integrado por las siguientes etapas: adquisición de imagen, pre-procesamiento, segmentación, extracción de parámetros, clasificación e inspección [6].

La industria vitivinícola es una de las más interesadas en el empleo de sistemas de visión para aumentar la calidad de sus cultivos [7]. Un número cada vez más creciente de productores de vino reconocen las ventajas de entender las características biofísicas y el rendimiento de sus viñedos que conlleven a una mejor administración de sus recursos y toma de decisiones [8].

Los productores de vino, comúnmente tienen una meta para el estado de madurez que desean alcanzar para el vino que producen. Tal meta puede variar, aún dentro de la misma variedad de uva, dependiendo del tipo o estilo de vino deseado [9]. Para la caracterización de la madurez de la uva, se realizan análisis fisicoquímicos, tales como: contenido de azúcar, acidez y pH [10], siendo el contenido de azúcar uno de los parámetros más importantes [11]. Por esta razón, el desarrollo de instrumentos de caracterización que sean más prácticos y no invasivos es muy importante, razones mencionadas anteriormente dieron origen al estudio del cual se deriva el presente artículo.

2. Morfología Digital

El término “morfología” se refiere a la forma y estructura de un objeto. La morfología digital es la manera en que se describe y analiza la forma de un objeto digital [12]. La idea fundamental de la morfología digital es que las imágenes consisten en un conjunto de elementos (píxeles), los cuales se agrupan en dos dimensiones para formar estructuras (formas). Se pueden utilizar operaciones matemáticas sobre conjuntos de píxeles para resaltar aspectos específicos de formas, las cuales pueden ser por ejemplo, contadas o reconocidas [12]. En términos más técnicos, la forma de un objeto es la información acerca de los aspectos geométricos de la superficie de un objeto.

Las formas en nuestro alrededor, pueden ser divididas principalmente en dos categorías: formas de objetos manufacturados y formas de objetos naturales. La razón de esta subdivisión se debe a que las técnicas matemáticas que se utilizan para la descripción de formas

de objetos manufacturados son inadecuadas para las formas de objetos naturales [2].

La descripción de formas es uno de los campos de estudio más utilizados en las áreas de visión por computadora, robótica, reconocimiento y análisis de patrones, procesamiento de imágenes y diseño asistido por computadora. Desafortunadamente, en la mayoría de los modernos desarrollos, las aproximaciones a la descripción de formas son ad hoc a necesidades prácticas [2].

Nuestro concepto de forma implica necesariamente la existencia de un contorno. Al hablar de objetos comunes como una pelota de fútbol o una sandía, estamos implícitamente refiriéndonos al contorno de ese objeto. Típicamente, la segmentación de una imagen involucra métodos automáticos para identificar contornos de objetos y regiones de interés en una imagen.

3. Sistema de Vision Inmerso

Un sistema inmerso es un sistema basado en computadora dentro de un equipo o componente. Por consecuencia, un sistema de este tipo está usualmente diseñado para realizar una tarea específica, o un conjunto pequeño de tareas específicas, usualmente en tiempo real [13]. El crecimiento acelerado en los últimos años en potencia computacional y en capacidad de las tarjetas de memoria flash, asociado con la reducción en tamaño, han hecho de los sistemas inmersos, una opción atractiva para aplicaciones de procesamiento de imágenes [13].

Los sistemas inmersos de visión son dispositivos que no solamente capturan imágenes, sino que además las procesan de acuerdo a los requerimientos de la aplicación. Un requerimiento para la mayoría de estos sistemas es que deben tener un tamaño pequeño y ser de peso ligero. Algunos funcionan con baterías y por lo tanto requieren operar con baja potencia y aún aquellos que no son operados con baterías usualmente tienen un consumo de potencia limitado [13]. La Figura 1 muestra un diagrama a bloques sobre los niveles de procesamiento que se llevan a cabo en un sistema de visión inmerso, de acuerdo a su nivel de complejidad.

Con el empleo de sistemas inmersos de visión en las áreas de agricultura, se obtienen muchas ventajas, tales como clasificación por morfología, detección de defectos y clasificación por variedad y calidad [1]. La Figura 2 muestra el diagrama a bloques de un sistema de visión convencional y un sistema de visión inmerso [14].

Una de las principales restricciones de los sistemas inmersos, es que tienen memoria limitada. Por esta

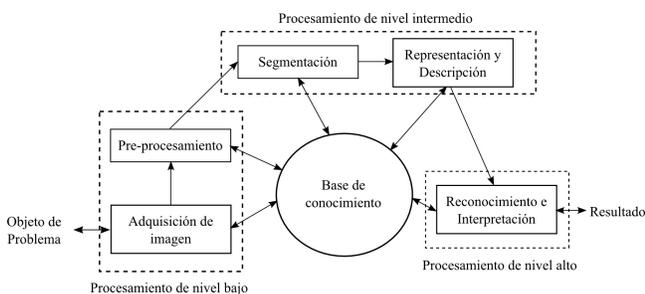


Figura 1. Diagrama a bloques de los diferentes niveles de procesamiento.

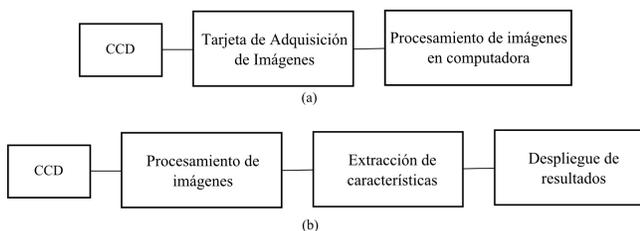


Figura 2. (a) Diagrama a bloques de un sistema de visión convencional. (b) Diagrama a bloques de un sistema de visión inmerso.

razón, los ingenieros de software inmerso han desarrollado habilidades para programar en ambientes con memoria limitada [15]. Usualmente los sistemas en tiempo real, deben ser sistemas inmersos para incrementar la robustez, confiabilidad y reducir costos. Consecuentemente, plataformas basadas en el sistema operativo Linux se han convertido en una opción práctica para los sistemas de visión inmersos [16].

4. Desarrollo del Sistema

El sistema descrito en este trabajo, detecta objetos (uvas) por su morfología mediante la Transformada de Hough. La Transformada de Hough es una herramienta muy robusta que se emplea para extraer bordes rectos, círculos o elipses, en general, cualquier polígono que pueda ser descrito mediante su ecuación paramétrica. La principal utilidad de esta herramienta es localizar, extraer y cuantificar formas aún cuando los bordes estén interrumpidos, incompletos o degradados por ruido en la imagen.

Debido a que la forma de las uvas que se utilizan para elaborar vino es semicircular, se puede aprovechar la Transformada de Hough para círculos. Si un círculo en una imagen puede ser descrito mediante la Ecuación 1, donde (a,b) son las coordenadas del centro del círculo y r es su radio, entonces un punto arbitrario (x_i,y_i) será transformado en el espacio Hough a un cono circular en el espacio de parámetros (a,b,r). Si todos los puntos

pertenecen a un círculo entonces los conos se intersectarán en un simple punto (a,b,r) correspondientes a los parámetros del círculo.

$$(x - a)^2 + (y - b)^2 = r^2 \tag{1}$$

4.1. Hardware

Las aplicaciones inmersas de visión que utilizan microprocesadores o procesadores inmersos, en la mayoría de las veces, adoptan la tecnología ARM (Advanced Risk Machine). Esto tiene muchas ventajas, entre las que se pueden mencionar: bajo costo, bajo consumo de energía, así como suficiente número de periféricos e información abundante sobre diversas aplicaciones de los usuarios. Una de las posibles desventajas en cuanto a la información que se encuentra en Internet, es que la mayoría de las aplicaciones utilizan tecnología ARM para realizar tareas de control, más que para realizar procesamiento de imágenes. Los sistemas inmersos de visión basados en tecnología ARM son usualmente diseñados para propósitos muy particulares, cuyas tareas abarcan la adquisición de imágenes, procesamiento, segmentación, detección de bordes, seguimiento de objetos y posicionamiento [17]. El sistema que se desarrolló en este trabajo, está basado en la plataforma de bajo costo y bajo consumo de potencia Pandaboard ES, con arquitectura OMAP 4460, el cual tiene dos procesadores tipo ARM Cortex-A9 a 1.2 GHz. Cuenta también con 2 entradas de USB para periféricos y salida para monitor HDMI y DVI, además de conexión Ethernet y WiFi, con lo que expande el tipo de aplicaciones. La Figura 3 muestra los periféricos y características con que cuenta la plataforma Pandaboard [18].

4.2. Software

La elección del sistema operativo, así como del software para los sistemas inmersos de visión es un factor crítico para el rendimiento total del sistema. Para este desarrollo se optó por utilizar el sistema operativo Linux, debido que es muy robusto, estable, rápido y de distribución libre, lo cual es muy importante para la reducción de costos de desarrollo.

En conjunto con el sistema operativo Linux, se instalaron las librerías de OpenCV en el sistema, el cual ofrece un amplio rango de funcionalidades para la adquisición y procesamiento de imágenes. OpenCV tiene la ventaja de ser multiplataforma, lo que permite que sus librerías se puedan acceder utilizando aplicaciones con lenguajes basados en C o Python [19]. La Figura 4 muestra de forma esquemática la metodología utilizada

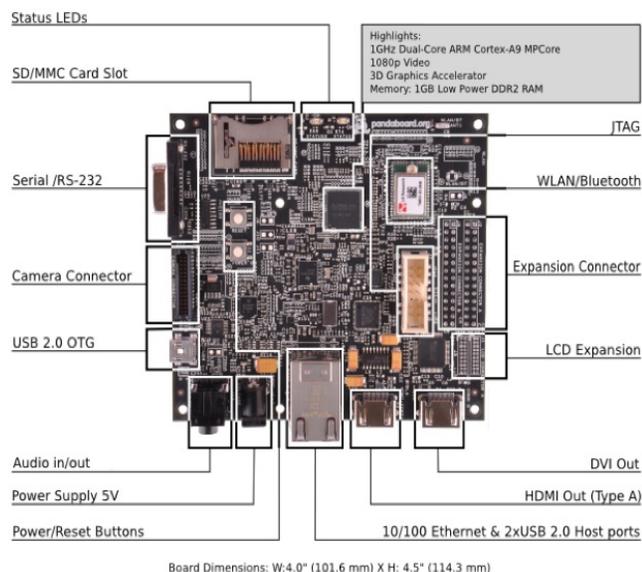


Figura 3. Diagrama a bloques del sistema utilizado.

en el algoritmo de procesamiento de imágenes para el sistema desarrollado.

5. Resultados

Con el sistema descrito, se adquirieron imágenes tanto de racimos completos como de uvas individuales que se utilizan para muestreos de calidad en el laboratorio. La figura 5 muestra la detección de un racimo de uvas, en primer lugar se aplicó un filtrado de tipo Canny (imagen del lado izquierdo) y posteriormente se obtuvieron los centros y radios de los círculos de cada uva aplicando la Transformada de Hough (imagen del lado derecho).

El mismo análisis se aplicó con uvas que fueron muestreadas para pruebas de laboratorio. El procesamiento es más rápido en este caso, debido a la menor cantidad de uvas y su separación bien definida. Finalmente, se utilizó una referencia para poder medir el diámetro de cada uva de forma automática, tal y como se muestra en la Figura 6.

Una vez detectadas las uvas en la imagen, se realiza el proceso de segmentación para cada uva y se extrae el promedio en color de los parámetros a y b para llevar a cabo su análisis. Finalmente, se efectúa el análisis de color también por medio del sistema desarrollado y se almacenan los datos en memoria. Los resultados que despliega el sistema se muestran en la Figura 7, de acuerdo a la cantidad de uvas analizadas.

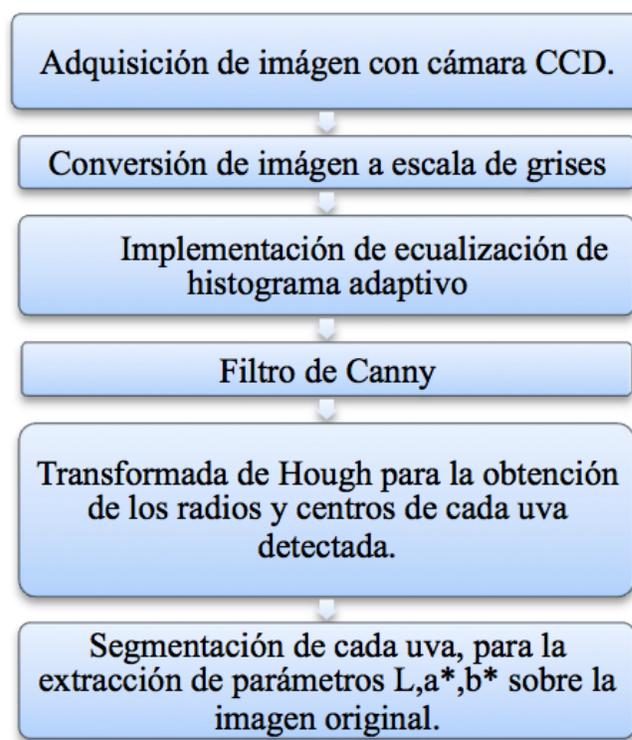


Figura 4. Metodología utilizada.

6. Conclusiones

Se presentó el desarrollo de un sistema inmerso de bajo costo para analizar color en uvas, tanto en racimos como en frutos individuales en el espacio de color CIELAB L, a, b mediante el uso del sistema operativo Linux, librerías de distribución libre como OpenCV y el lenguaje de programación Python, que es de distribución libre también, aún para aplicaciones comerciales. En este sistema, se realizaron los procesos de adquisición de imágenes, segmentación, procesamiento, almacenamiento y despliegue de datos. En ocasiones, entendemos por bajo costo, algún producto con pocas funcionalidades, pero cuando se utilizan recursos libres y robustos se pueden hacer desarrollos muy completos y al alcance de muchas personas que requieren tecnología para acelerar sus procesos productivos.

Referencias

- [1] C. Garrido-Novell, D. Pérez-Marín, J. M. Amigo, J. Fernández-Novales, J. E. Guerrero, and A. Garrido-Varo, "Grading and color evolution of apples using RGB and hyperspectral imaging vision cameras", *Journal of Food Engineering*, vol. 113, no. 2, pp. 281-288, Noviembre 2012.
- [2] K. Deguchi and P. K. Ghosh, *Mathematics of Shape Description?: A Morphological Approach to Image Processing and Computer Graphics*. Wiley, 2009, p. 4.

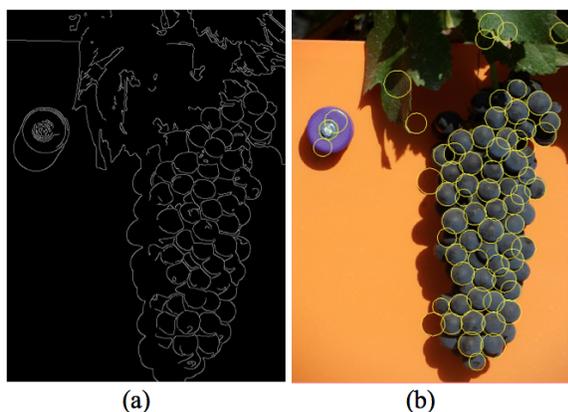


Figura 5. (a) Resultado de aplicar el Filtro de Canny a la imagen del racimo. (b) Círculos detectados con la Transformada de Hough.

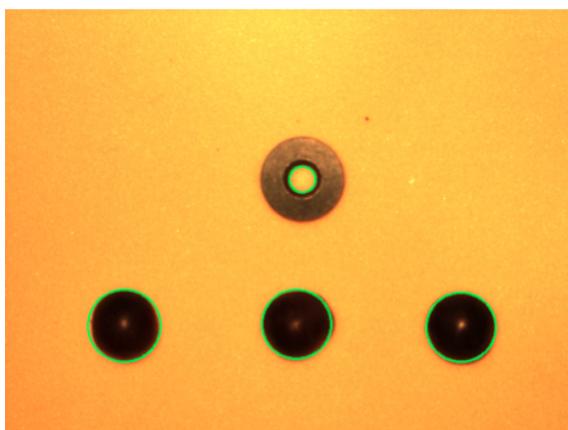


Figura 6. Uvas detectadas en la imagen para posterior segmentación.

Racimo	a_mean			b_mean		
	Uva 1	Uva 2	Uva 3	Uva 1	Uva 2	Uva 3
1	8.15	10.33	5.89	5.96	7.16	5.64
2	6.12	9.58	4.05	5.46	8.02	4.54
3	16.10	7.27	8.03	8.04	6.26	6.13
4	5.72	6.55	5.70	5.35	6.23	5.65
5	6.72	7.71	6.20	5.77	7.06	6.08
6	8.85	6.40	7.70	5.49	5.59	6.24
7	12.79	8.04	6.51	7.22	6.55	5.63
8	7.40	8.82	7.36	6.12	6.76	6.79
9	10.04	12.57	8.95	6.67	7.37	7.57
10	7.29	7.18	7.19	6.97	6.65	6.64
11	5.87	6.61	6.40	5.56	5.86	5.96
12	8.47	14.46	8.37	5.77	9.40	6.64
13	6.42	7.23	5.24	5.55	6.27	5.54
14	7.64	7.24	7.49	6.05	6.26	6.81
15	6.25	7.02	7.23	5.70	6.59	5.99

Figura 7. Resultados de color.

[3] Y. Al Ohali, Computer vision based date fruit grading system: Design and implementation, *Journal of King Saud University - Computer and Information Sciences*, vol. 23, no. 1, pp. 29-36, 2011.

[4] M. Khojastehnazhand, M. Omid, and A. Tabatabaefar, Development of a lemon sorting system based on color and size, *Plant Science*, vol. 4, no. April, pp. 122-127, 2010.

[5] H. Alimohamadi and A. Ahmadyfard, Detecting Skin Defect of Fruits Using Optimal Gabor Wavelet Filter, 2009 International Conference on Digital Image Processing, pp. 402-406, Marzo 2009.

[6] F. Leta, F. Feliciano, I. de Souza, and E. Cataldo, Discussing Accuracy in an Automatic Measurement System using Computer Vision Techniques, *ABCM Symposium in Mechatronics*, vol. 2, pp. 645-652, 2005.

[7] R. Chamelat, E. Rosso, A. Choksuriwong, C. Rosenberger, H. Laurent, and P. Bro, Grape Detection By Image Processing, 32nd Annual Conference on IEEE Industrial Electronics, Noviembre 2006.

[8] T. Proffitt and R. Bramley, Further developments in Precision Viticulture and the use of spatial information in Australian vineyards, 15 Romeo Bragato Conference, Napier, pp. 20-22, Agosto 2009.

[9] E. Hellman, How to judge grape ripeness before harvest, in

Southwest Regional Vine & Wine Conference, Albuquerque, NM, USA, Febrero 2004.

[10] L. Draghici, G. Rapeanu, and T. Hopulele, Evolution of polyphenolic compounds during maturation of Cabernet Sauvignon grapes from Dealu Mare vineyard, 2011 Ovidius University Press, vol. 22, no. 1, pp. 15-20, 2011.

[11] I. Dami, Determining Time of Harvest?: The Constant Dilemma, *Ohio Grape Wine Electronic Newsletter*, pp. 2-5, Septiembre 2008.

[12] J. R. Parker, *Algorithms for Image Processing and Computer Vision*, Segunda Edición, New Jersey: Wiley, 2010, pp. 111-113.

[13] D. Bailey, *Design for Embedded Image Processing on FPGAs*. New Jersey: Wiley, 2011, p. 12.

[14] N. Ohlsson. (Junio 2013), A Model-Based Approach to Computer Vision and Automatic Control using Matlab Simulink for an Autonomous Indoor Multirotor UAV, Tesis de maestría de la Universidad Chalmers de Suecia. Septiembre 2013. <http://publications.lib.chalmers.se/records/fulltext/179662/179662.pdf>

[15] C. Walls, *Embedded Software: The Works*. Newnes, 2006, p. 3.

[16] I. Johansen and M. Hayes, *Java Software and Embedded Systems*. Nova Science Publishers, Inc., 2010, pp. 213-214.

[17] D. Xu, *EMBEDDED VISUAL SYSTEM AND ITS APPLICATIONS ON ROBOTS*. Bentham Science, 2010, p. 2.

[18] Pandaboard references, Pandaboard Inc, Septiembre de 2013, <http://pandaboard.org/content/resources/references>.

[19] J. Coombs and R. Prabhu, *OpenCV on TIs DSP+ ARM@ platforms: Mitigating the challenges of porting OpenCV to embedded platforms*, Texas Instruments, 2011.

Determinación de Heurísticas de diseño en Objetos de Aprendizaje como elemento en la definición de un Modelo de Preferencias de Usuario

Pedro Cardona Salas^a, Ángel Muñoz Zavala^a, Cesar Velázquez Amador^a, Jaime Muñoz Arteaga^a, Francisco Álvarez Rodríguez^a

^aUniversidad Autónoma de Aguascalientes, Centro de Ciencias Básicas.
Av. Universidad 940, Col. Ciudad Universitaria, C.P. 20100, Aguascalientes, Ags., México.
{jpcardon, aemz, fjalvar, jmuozar}@correo.uaa.mx

2013 Published by *IEEE U100ci* @ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CCOMP-2013, www.cicomp.org

Resumen

La construcción de un modelo de preferencias de usuario debe considerar varios factores, proponemos describir y elicitar el diseño de composición de Objetos de Aprendizaje desde la perspectiva de las Heurísticas. Las Heurísticas son procesos cognitivos que tienden a maximizar el proceso de decisión mediante la omisión de información redundante pero por otro lado pueden generar errores más graves que los que ocasionan las decisiones basadas en modelos estadísticos del modelo racional de decisión, se considera que la mayoría de las decisiones que toman las personas se basan fuertemente en Heurísticas, dada esa situación proponemos integrar un modelo de preferencias de usuario basado en Heurísticas al diseño de Objetos de Aprendizaje del contexto de e-learning; como primer paso proponemos la determinación o identificación de Heurísticas de Diseño al decidir o elegir interfaces de Objetos de Aprendizaje, para posteriormente determinar qué Heurísticas usan las personas y en qué situaciones.

Palabras clave: Heurísticas, Objetos de Aprendizaje, Modelo de Preferencias de Usuario.

1. Introducción

Un reto en el diseño de objetos de “aprendizaje” es la aceptación por el usuario; el uso y aceptación de un objeto de aprendizaje depende en gran medida de la aceptación de la interfaz gráfica y de la usabilidad percibida. La usabilidad percibida perdura aun con el uso de la interface [1]. En este estudio se considera la aceptación o no aceptación como una deci-

sión, las decisiones que tomamos son influidas por las preferencias y las preferencias son influidas por diversos factores como regionales o culturales.

Este estudio descriptivo se enfoca a determinar las preferencias de los usuarios respecto a interfaces de objetos de aprendizaje desde el enfoque descriptivo de la teoría de la decisión y más específicamente de las reglas no compensatorias conocidas como heurísticas que recientemente han tenido un papel sobresaliente

en la determinación del proceso de toma de decisiones y preferencias [2] (ver 1.2 Teoría de decisión).

1.1. Objetos de Aprendizaje

Las definiciones de OA en la literatura son numerosas, pero la idea central referente a OA es proveer una experiencia de aprendizaje y la facultad de reusarse, y reusarse en diferentes contextos, aunque haya conceptos en la literatura como objeto de aprendizaje, unidad de aprendizaje, unidad informativa, objeto informativo, objeto de aprendizaje reusable y otras, todas comparten la misma preocupación de reuso y alcanzar aprendizajes [3].

Las estrategias para alcanzar el reuso de OA mayormente se enfocan a la interoperabilidad, metadata, granularidad; el presente trabajo se enfoca a seleccionar los OA más preferidos por la usabilidad de la interface apoyándose en la experiencia de otros usuarios.

También el diseño de los OA se basó en una de las definiciones más aceptadas de OA: es una unidad de aprendizaje independiente y autocontenida para ser reutilizada en múltiples contextos educativos [4]. El objeto de aprendizaje como tecnología emergente es una opción viable de integración de las tecnologías de computación e información y los avances en teorías de aprendizaje.

La tecnología de OA según la escala de Rogers [5] de adopción de tecnología, pasó del nivel 1 de tecnología con potencial al nivel 2 como una tecnología que continua siendo nueva y hay dificultad para encontrar personal para implementar o dar soporte, mientras continúa la división entre entusiastas y detractores, se realizan investigaciones para determinar la mejor forma de aplicar, entender y explotar su potencial. [6].

1.2. Teoría de la Decisión

La teoría de decisión se divide en dos principales ramas: la primera es cómo debería de hacerse el proceso de toma de decisiones a fin de hacerlo de manera racional y metódica (teoría prescriptiva o racional); la segunda es cómo se hace el proceso de toma de decisiones en realidad (teoría descriptiva o positiva).

Las condiciones deseables de la teoría prescriptiva es tener la información completa, se pueden hacer los cálculos con exactitud y el tiempo no es una limitación, en esas condiciones las personas que deciden, si así lo disponen, pueden integrar toda la información necesaria, combinando y ponderando, usando modelos algebraicos - típicamente lineales-[7], [8].

La teoría descriptiva establece que las personas usualmente no basan sus decisiones en reglas axiomáticas sino que frecuentemente se hacen violaciones a la optimización o errores estructurales en el proceso de toma de decisiones. Biases, teóricos señalan que algunas causas de esta conducta son las limitaciones de la capacidad cognitiva humana Bounded Rationality [9], también se señala que el ser humano realiza procesos prescriptivos de decisión modelos lineales pero de manera reducida heurísticas, los mismos estudios señalan que las personas pueden hacer estas prácticas altamente efectivas [10], este campo de estudio también es conocido como teoría de decisión conductual.

La situación de las heurísticas respecto a las dos ramas de la teoría de la decisión es la pertenencia a la teoría descriptiva pero las bases de cálculo y razonamiento están mayormente asociadas a la teoría prescriptiva, otra característica es que el grado y forma de aplicación de las heurísticas es altamente dependiente del contexto donde se aplique, además los teóricos señalan que una condición de las heurísticas es que no buscan la optimización al 100 % sino una respuesta satisfactoria en tiempo y resultados [11], por lo que se considera a las heurísticas como mayormente adaptativas.

Respecto a la clasificación de reglas de decisión basadas en atributos compensatorias y no compensatorias, las heurísticas se ubican dentro de las reglas no compensatorias (teoría descriptiva, como ejemplo: regla conjuntiva, disyuntiva, eliminación de aspectos y heurísticas lexicográficas), reglas compensatorias (teoría prescriptiva, como ejemplo modelos lineales).

1.3. Definición y Detección de Heurísticas

Heurística se refiere a métodos basados en la experiencia que nos ayudan a resolver problemas como prueba y error, en Inteligencia Artificial, Ciencias Computacionales y Optimización Matemática, son técnicas que se utilizan cuando los métodos clásicos prescriptivos no hallan la solución óptima o se tardan demasiado en encontrar las soluciones, la psicología desde los años 70 estudian la heurística como una fuente de explicación de errores de razonamiento humano (biases) [1], aunque básicamente tienen el mismo objetivo que las heurísticas en otras áreas, encontrar soluciones aceptables en condiciones de incertidumbre.

Estudios señalan que la gente usa heurísticas lexicográficas en su proceso de decisión cuando la tarea de cálculo es considerablemente alta [12], se ha comprobado que la gente usa lexicográficos aun cuando los cálculos sean mínimos, [13].

Uno de los programas con mayor influencia en la teo-

ría de toma de decisiones es el de Tversky y Kahneman [14], el tema es sobre heurísticas y biases (errores), este programa ha compilado evidencia empírica en donde el proceso de toma de decisiones se realiza con heurísticas, este trabajo forma una importante base teórica del proceso psicológico de la toma de decisiones [15].

Las heurísticas son utilizadas por personas en una amplia variedad de tareas complejas como toma de decisiones, preferencias [16].

Modelos de heurísticas se han documentado y comprobado mediante pruebas empíricas pero no hay una teoría sustentada completa, el primer paso hacia la solidificación de la teoría es determinar los bloques de donde son construidas las heurísticas a modo de principio organizador, esto permitirá reducir el gran número de heurísticas a un número menor de componentes, similar a los elementos químicos en la tabla periódica, [1].

La propuesta de los bloques básicos que en mayor o menor medida esta presente en todas las heurísticas son: 1) regla de búsqueda, validación de atributos, 2) regla de stop, cuándo un atributo discrimina convincentemente a las alternativas, 3) regla de decisión, escoger alternativa favorecida por el atributo [17].

1.4. Heurísticas de Diseño

Los principales avances en heurísticas se han realizado en áreas como Inteligencia Artificial, Ciencias Computacionales, Optimización Matemática y recientemente en las bases psicológicas del proceso de toma de decisiones.

Las mismas condiciones se dan en el diseño y composición gráfica, donde hay técnicas especializadas aplicadas por los expertos y técnicas similares pero reducidas (heurísticas) que utilizan las personas para diseñar o elegir algún diseño (en este caso un diseño de interface gráfica de objetos de aprendizaje).

Entre las técnicas especializadas de diseño más conocidas están: en artes y matemáticas la proporción aurea o golden ratio, donde 2 elementos (a y b) son armónicos si el volumen de a respecto $b = a$ respecto $a + b$.

Otra técnica de composición es la ley de tercios, una composición es mas armónica cuando se divide en tres, tanto de manera vertical como horizontal, se le considera una extensión o variante de la proporción aurea.

También esta la técnica de cropping donde se resalta el objeto de interés de la composición por contraste de tamaño o de color, así como despejar los elementos distractores, las técnicas mencionadas son reglas generales que pueden integrarse en una sola composición.

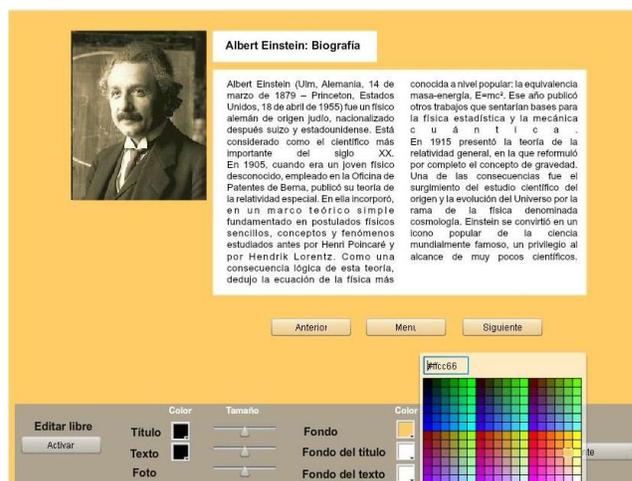


Figura 1. Interface modificable que almacena los atributos modificados (colores).

2. Problemática

Una situación problemática en el desarrollo de objetos de aprendizaje es el costo de desarrollo y que una vez desarrollado el costo de mantenimiento o mejoras también resulta oneroso, por lo tanto se buscan diseños de objetos de aprendizaje que satisfagan las preferencias de los usuarios. La propuesta es definir y refinar modelos de preferencias de usuarios, mejorar el espacio del problema al definir con mayor certeza el modelo de preferencia de las interfaces de objetos de aprendizaje.

Se trata de conocer mejor el proceso de toma de decisiones del usuario al determinar las heurísticas de diseño que forman parte de las preferencias del usuario respecto a las interfaces de objetos de aprendizaje.

3. Materiales y Métodos

Se realizó un experimento con un grupo de 33 alumnos (15 alumnos de Ing. en Computación Inteligente de 2do semestre con edades entre 19 y 21 años y 18 alumnos de la carrera de Matemáticas Aplicadas de 2do semestre con edades entre 19 y 21 años) de la Universidad Autónoma de Aguascalientes, sobre las preferencias de interfaces gráficas de objetos de aprendizaje.

El experimento consistió en que los estudiantes manipularon una interfaz gráfica modificable hasta llegar a su interfaz preferida, la interfaz almacenó los atributos modificados (colores), la interfaz se muestra a continuación en la Figura 1.

Las variables son:

- FG: Color del Fondo Global (color mostaza)
- FL: Color del Fondo del Texto (color blanco)
- FT: Color del Fondo del Título (color blanco)
- L: Color de la Fuente del Texto (color negro)

El criterio que se siguió para comparar los tonos de cada color fue categorizar los tonos mediante una herramienta de contraste de tonos que utilizan los diseñadores de sitios web -esta herramienta está disponible en [18].

Esta herramienta fue compatible con la paleta de colores utilizada el recuadro con 6 colores (3 verde azul y 3 rojo amarillo) en la parte inferior derecha de la figura 1-, esta paleta facilitó el trabajo por a) es una paleta reducida de 216 colores, b) los tonos de la paleta fueron múltiplos de 10 hasta llegar a 100 con lo que se facilita el estudio y análisis de correlación.

La base de colores utilizada es el código de color HTML, cada código HTML contiene el símbolo # y 6 letras o números que son números del sistema hexadecimal, el primer par de código representa la intensidad del rojo -00 menos intenso y FF mas intenso-, el segundo par de códigos representa la intensidad del verde, el tercer par de códigos representa la intensidad del azul, al combinar la intensidad de rojo, verde y azul se puede representar casi cualquier color, el numero de combinaciones de los 6 códigos (cada código tiene el rango de 0 a 16) nos da 16,777,216 colores diferentes pero con el uso de la paleta de Flash, la herramienta con que fue creada la interface modificable reduce eficientemente la cantidad de colores y tonos a utilizar.

4. Resultados

El primer resultado es una matriz de correlaciones entre las covariables, no son basados en una escala sólo son relaciones comparativas al considerar las 4 variables en el modelo, como se muestra a continuación.

- FG: Color del Fondo Global
- FL: Color del Fondo del Texto
- FT: Color del Fondo del Título
- L: Color de la Fuente del Texto

Matriz de Correlaciones				
	FG	FL	FT	L
FG	1.0000000	0.5769788	0.4824865	-0.2381358
FL	0.5769788	1.0000000	0.8384333	-0.4290934
FT	0.4824865	0.8384333	1.0000000	-0.5733016
L	-0.2381358	-0.4290934	-0.5733016	1.0000000

Se probaron diferentes modelos estadísticos que reflejaran mejor una tendencia y el resultado se muestra

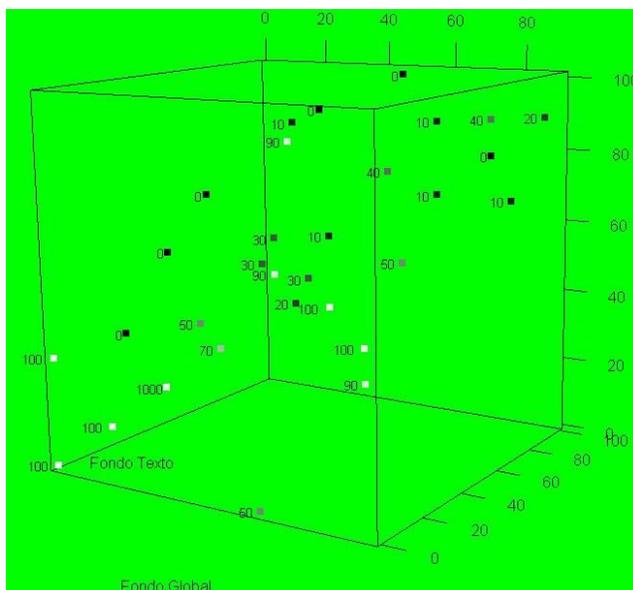


Figura 2. Diagrama de Correlación de color Fondo Global y color Fondo de Texto.

en la figura 2.

Los resultados obtenidos de los 33 usuarios reflejan una fuerte correlación positiva entre el color que se aplica al fondo del título y el color que se aplica al fondo del texto. Es decir, si el color del fondo del título es oscuro (o claro), el color del fondo del texto también lo será la mayor parte de las ocasiones.

Sin embargo, la correlación entre el color del fondo del título y el color del fondo del texto con respecto al color de la fuente del texto es negativa en ambos casos, -0.57 y -0.43, respectivamente. Contrario a los casos anteriores, cuando el color de la fuente del texto es oscuro, el color del fondo del título y del fondo del texto serán claros, y viceversa.

Para ilustrar lo explicado anteriormente, en la Figura 2 se presentan los resultados de los diseños finales de los 33 usuarios. Cada eje, X, Y y Z, representa el grado de claridad (0 a 100) en el color del fondo global, fondo del texto y fondo del título, respectivamente. Utilizando una escala de grises, blanco=100 a negro=0, ilustramos por medio de cuadros etiquetados el grado de claridad elegido por el usuario en el color de la fuente del texto.

5. Conclusiones

Se propuso detectar heurísticas de diseño en la elección de interfaces gráficas de objetos de aprendizaje, se esperaba una mayor correlación entre Fondo Global y Fondo del Texto por ser los elementos de mayor tamaño de la interface y que el usuario buscara un mayor efecto

de *cropping*, pero no fue así.

En cambio la mayor correlación fue entre el Título y el Texto, lo cual coincide con los resultados de eye-tracking realizados en laboratorios de usabilidad [19], donde el usuario sigue un patrón en forma de F donde la atención se va desvaneciendo siendo el título y los siguientes 2 párrafos como los elementos que reciben más atención.

El diseño de la interface no facilitó la evaluación de la técnica de diseño *golden ratio*.

En general las correlaciones son débiles por lo que es necesario realizar en un futuro el experimento con una muestra mayor.

El principal beneficio esperado al determinar la existencia de las heurísticas y las áreas donde se aplican es transformar estas intenciones en acciones concretas al diseñar interfaces inteligentes (reactivas), en este caso el primer beneficio es que al localizar las áreas de aplicación de las heurísticas se está reduciendo el espacio donde se aplicarán las acciones de diseño.

Referencias

- [1] M. Kurosu and K. Kashimura, "Apparent Usability vs. inherent Usability: experimental analysis on the determinants of the apparent Usability", *Conference companion on Human factors in computing systems*, p.292-293, May 07-11, 1995.
- [2] G. Gigerenzer, and W. Gaissmaier, Heuristic decision making. *Annual Review of Psychology*, 62, 451-482 (2011).
- [3] C. Cechinel, y S. Da Silva Camargo, . *Localization of Learning Objects*. Cap. 3 Analysis of State-of-the-Art Solutions for Personalised Learning Support 2011, documento del proyecto IGUAL - Innovation for Equality in Latin American University Page, DCI-ALA/19.09.01/10 /21526/ 245-315/ALFAHI (2010)123 Alfa III programme.
- [4] P.R. Polsani, (2003) Use and Abuse of Reusable Learning Objects, *Journal of Digital Information*, volumen 3, número 4, artículo No. 164. Disponible en: <http://jodi.ecs.soton.ac.uk/Articles/v03/i04/Polsani/> 2003.
- [5] E. Roger, *Diffusion of innovations* Glencoe, Free Press 1962.
- [6] L.I. Santacruz-Valencia, I. Aedo y C. Delgado, Designing Learning Objects with the ELO-Tool *The 3rd IEEE International Conference on Advanced Learning Technologies (ICALT'03)*. ISBN: 0-76951967-9. pp: 372-373. 9-11 de Julio. Atenas-Grecia. 2003.
- [7] N.H. Anderson, *Foundations of information integration theory*. New York: Academic Press 1981.
- [8] B. Brehmer, The psychology of linear judgement models. *Acta Psychologica*, 87, 137-154. 1994.
- [9]
- [10] D. Kahneman, and S. Frederick, *Representativeness revisited: attribute substitution in intuitive judgment*. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, ed. T Gilovich, D Griffin, D Kahneman, pp. 49-81. New York: Cambridge Univ. 2002.
- [11] G. Gigerenzer, P.M. Todd,. and the ABC Research Group. *Simple heuristics that make us smart*. New York: Oxford University Press 1999.
- [12] A. Drolet, and M.F. Luce, The rationalizing effects of cognitive load on response to emotional tradeoff difficulty. *Journal of Consumer Research*, 31(1), 63-77. 2004.
- [13] A.Tversky, S. Sattath, and P. Slovic Contingent weighting in judgment and choice, *Psychological Review*, 95:371-384 1988.
- [14] A. Tversky and D. Kahneman, *Judgements under uncertainty and heuristics and biases*, Science New Series Vol 185, No. 4157 (1974) 2011.
- [15] E.M. Pothos and J.R. Busemeyer, *Formalizing heuristics in decision-making: a quantum probability perspective*, Frontiers in psychology 2011.
- [16] J.W. Payne, "Heuristic Search Processes in Decision Making", in NA - Advances in Consumer Research Volume 03, eds. Beverlee B. Anderson, Cincinnati, OH : Association for Consumer Research, Pages: 321-327 1976.
- [17] G. Gigerenzer, y D.G. Goldstein, Reasoning the fast and frugal way: Models of bounded rationality, in: *Psychological Review*, Vol. 103, pp. 650-669 1996.
- [18] Oto255.com, Designed and Developed By shaunchapman, consultado 5 julio 2013
- [19] Nielsen Norman Group , creador de las 10 heurísticas de diseño de Nielsen, consultado 1 Julio 2013 <http://www.nngroup.com/topic/eyetracking>.

Control de Asistencia y Tiempo de Permanencia de Personas en un Área Laboral Mediante la Detección de Smartphones

Nicolás Quiroz Hernández^{a,b}, Mauro Martín Letras Luna^b, Mario Bustillo Díaz^b, Apolonio Ata Pérez^b, Alejandro Rangel Huerta^b

^aBenemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación.
Av. San Claudio y 14 Sur, Ciudad Universitaria, Puebla, México 72570
nquiroz@cs.buap.mx

^bBenemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Electrónica.
Av. San Claudio y 18 Sur, Ciudad Universitaria, Puebla, México 72570
[romauro_11,bustillo1956}@hotmail.com](mailto:{romauro_11,bustillo1956}@hotmail.com), [apolonio,arangel}@solarium.cs.buap.mx](mailto:{apolonio,arangel}@solarium.cs.buap.mx)

2013 Published by \mathcal{UT} 100Ci@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Resumen

Se presenta un sistema desarrollado para la detección de personas dentro de un área laboral haciendo uso de teléfonos inteligentes (smartphones) a través del módulo Bluetooth integrado. Utilizando una serie de módulos (equipados con un transceptor Bluetooth-Ethernet) se puedan detectar los smartphones de un grupo de trabajadores para así llevar un control de la asistencia y tiempo de permanencia en el lugar trabajo asignado. Debido al impacto que los smartphones han tenido en la sociedad moderna, es muy común que la mayoría de las personas tengan uno de estos dispositivos, los cuales tienen un transceptor Bluetooth integrado.

Palabras clave: Bluetooth, Control de Asistencia, Detección de smartphone, Ethernet.

1. Introducción

Un trabajador en un ambiente laboral tiene la obligación de asistir, llegar puntual y permanecer en el trabajo, estos aspectos repercuten directamente en el rendimiento de la empresa en la que trabaja, esto puede representar ganancias o pérdidas (si el empleado es impuntual y falta mucho). En base al cumplimiento de los puntos mencionados anteriormen-

te a un trabajador se le puede premiar (con algún tipo de bono de puntualidad y/o asistencia) o castigar (con algún descuento en el salario). Es aquí donde surge la necesidad de llevar un registro de la asistencia y puntualidad de un grupo de trabajadores, y proporcionar un mecanismo que permita garantizar la permanencia de los mismos en su lugar de trabajo durante un horario asignado. La solución propuesta consiste en el uso de tecnología Bluetooth. El objetivo radica en la



Figura 1. Comparativa del uso de telefonía celular y el servicio de internet, (Comisión Federal de Telecomunicaciones).

localización de smartphones utilizando el transceptor Bluetooth que tienen integrado, se estará supervisando constantemente un grupo de smartphones que esta ligados con un grupo de trabajadores. Para garantizar la asistencia se asignará un smartphone únicamente con un trabajador, aprovechando que los dispositivos Bluetooth tienen asignada una dirección única llamada dirección Bluetooth (muy similar a la dirección MAC que también es única para cada tarjeta de red Ethernet).

Uno de los motivos para usar smartphones es el gran impacto que han tenido en la sociedad moderna. Actualmente hay seis billones de usuarios de tecnología celular a nivel mundial, lo cual representa tres cuartas partes de la población mundial. Además hay un incremento acelerado del uso de telefonía celular en países en vías de desarrollo en comparación con países desarrollados [1, 2]. En lo que respecta a México, la población registrada en 2010 por INEGI (Instituto Nacional de Estadística y Geografía) fue de 112 millones 336 mil 538, de esta población según cifras de la COFETEL (Comisión Federal de Telecomunicaciones) hay 97.6 millones usuarios de telefonía celular [3]. Este indicador quiere decir que de cada 100 mexicanos 86.9 tienen acceso a un equipo celular (Figura 1).

Además realizando una consulta de los equipos que ofrece la más grande empresa de telefonía celular en el país, hay 216 equipos disponibles en el mercado de los cuales 180 cuentan con un transceptor Bluetooth.

Estas razones fueron el motivo por el cual usar Bluetooth, ya que no es necesario comprar algún tipo

de transceptor adicional, usando la tecnología disponible y debido a la gran difusión de los smartphones se

puede dar solución a este problema.

2. Antecedentes

El control de asistencia consiste en verificar que una persona va a estar en un lugar por determinado tiempo realizando una tarea en específico (trabajo, escuela, etc.). No solo es necesario indicar que la persona asistió a un sitio a una hora determinada sino saber cuánto tiempo permaneció en su área de trabajo, cuánto tiempo estuvo ausente de la misma y la hora en la cual se retiró del sitio. Por eso un sistema que se encargue de controlar la asistencia debe de garantizar, por lo mínimo los siguientes puntos.

- Registrar hora de entrada, para comprobar puntualidad.
- Registrar permanencia en el área de trabajo.
- Registrar salida.

La manera más común de comprobar la asistencia de un grupo de personas es a través de pase de lista, esto se observa dentro de los salones de clases, donde el profesor menciona el nombre de cada alumno para así saber si se encuentra presente. La desventaja principal de esta propuesta, es que no se puede garantizar la permanencia de las personas, debido que una vez pasada la lista, la persona puede abandonar el lugar sin que se vea afectada su asistencia. También existen sistemas basados en biométricos que sirven para control de acceso y asistencia, algunos de estos sistemas usan reconocimiento de huella digital, geometría de la mano, reconocimiento de iris, escáner de retina [4], etc.

Aplicados a sistemas de acceso y control de asistencia, los sistemas biométricos están orientados para identificar a la persona que requiere el acceso.

Los sistemas biométricos operan en dos modos. En modo de reconocimiento, el sistema compara una muestra obtenida en tiempo real con todas las plantillas en la base de datos y determina la mejor coincidencia. En modo verificación, el sistema requiere que la persona se identifique primero, con una clave que le fue asignada con anterioridad, después el sistema busca las plantillas correspondientes de la base de datos y acepta o rechaza la asistencia en base a las coincidencias.

Incluso para bases de datos pequeñas, el reconocimiento es muy demandante en comparación con la verificación. Las circunstancias en las cuales usar uno u otro enfoque dependen en gran medida a los requerimientos del proyecto. Pero es notable que al usar un

enfoque de verificación el tiempo de respuesta del sistema sea muy rápido pero la desventaja es que requiere intervención directa de la persona a la cual se toma asistencia. En la propuesta que se presenta, el pase de asistencia se hace de manera no intrusiva. Por lo que las personas no notaran cuando se toma su asistencia y se guarda la información dentro de una base de datos para poder generar reportes estadísticos sobre la respuesta de los trabajadores. Una de las aplicaciones es la monitorización de asistencia de profesores en una institución de educación. Donde se requiere saber el tiempo que asiste un profesor al salón de clases, y no solo saber que asistió. El trabajo se enfoca a esta aplicación en particular.

3. Desarrollo

La implementación del sistema cuenta con los siguientes módulos:

- Modulo Bluetooth-Ethernet.
- Base de Datos.
- Aplicación para Administración de Datos.
- Página Web.
- Proceso de Pase de Lista.

Se desarrolló una red LAN (figura 2), donde cada uno de los nodos de esta red es un módulo Bluetooth-Ethernet, que serán los encargados de detectar los smartphones de los trabajadores. El proceso "Pase de lista", es un proceso que se ejecuta en el servidor y su tarea principal es obtener toda la información de todos los módulos Bluetooth-Ethernet de la red para posteriormente almacenarlos en una base de datos. La información estará disponible en una página web, para ver en tiempo real los resultados obtenidos por el sistema.

4. Módulo Bluetooth-Ethernet

La figura 3 muestra un diagrama a bloques del módulo Bluetooth-Ethernet. Se utilizó un microcontrolador PIC18F4550 para controlar los transceptores de Bluetooth (BC4 V2.0) y Ethernet (ENC28J60), estos tres elementos forman el módulo Bluetooth-Ethernet (figura 4). El microcontrolador se comunica con el módulo Bluetooth por RS-232 utilizando comandos AT y con el Ethernet por SPI. El microcontrolador controla los datos que se envían y reciben por los módulos, mediante el algoritmo mostrado en la figura 5.

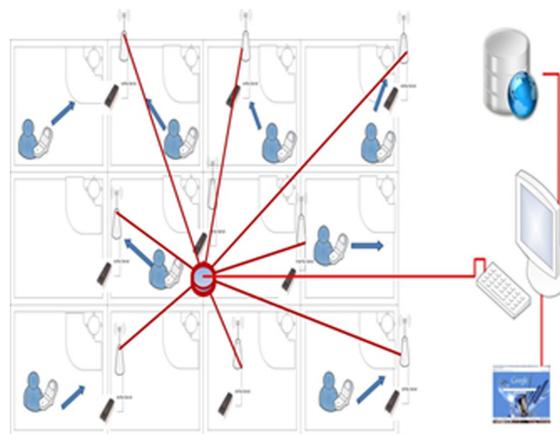


Figura 2. Creación de una red LAN, a través del uso de transceptores Ethernet.

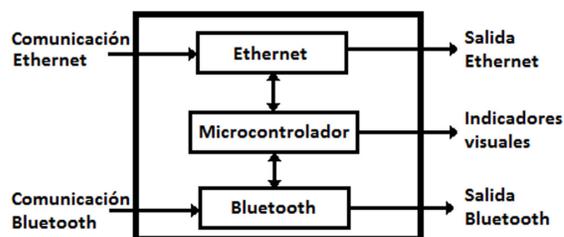


Figura 3. Diagrama Módulo Bluetooth-Ethernet.

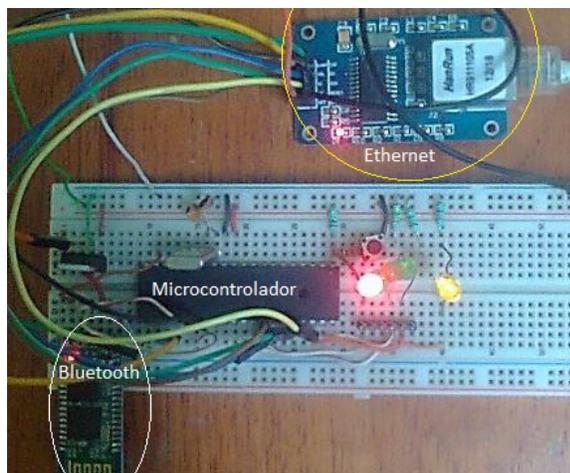


Figura 4. Módulo Bluetooth-Ethernet.

```

Cadena comando? Leer_Bluetooth();
Si (comando es igual a Activar Sondeo)
  sondeoActivo? verdadero
Fin Si
Si (AgoTó Tiempo? sondeoActivo)
  Resultado? sondearDireccion(direccionSP)
Fin Si
Si (comando es Enviar Resultado)
  Escribir_Ethernet(Resultado);
Fin Si
Si (comando es Recibir Direccion Monitoreo)
  direccion_SP? Leer_Ethernet();
Fin Si
Si (comando es igual a Detener Sondeo)
  sondeoActivo? falso
Fin Si
    
```

Figura 5. Algoritmo módulo Bluetooth-Ethernet.

El algoritmo verifica si está activa la monitorización de los Smartphone (SP) y el tiempo asignado para dicho proceso, si encontró el SP envía la información por Ethernet. Puede verificar un nuevo SP, o terminar el proceso de monitorización. La programación se realizó en lenguaje C.

El módulo de Bluetooth-Ethernet cuenta con tres leds indicadores para mostrar el estado del módulo. La comunicación con el Bluetooth se realiza a una tasa de transferencia de 38,400 bps, con dicha frecuencia se puede pasar al modo AT [5, 6].

5. Base de Datos

La base de datos del sistema sirve para almacenar la información personal de los trabajadores, el edificio sobre el cual se instaló el sistema, los horarios y los smartphones a checar (figura 6).

La base de datos permite llevar un registro de toda la información para posteriormente realizar reportes de la asistencia y permanencia de los trabajadores. Dicha información es necesaria para la página web. La base de datos se implementó SQL Server para tener una mejor integración con las demás aplicaciones que se desarrollaron con tecnología .Net.

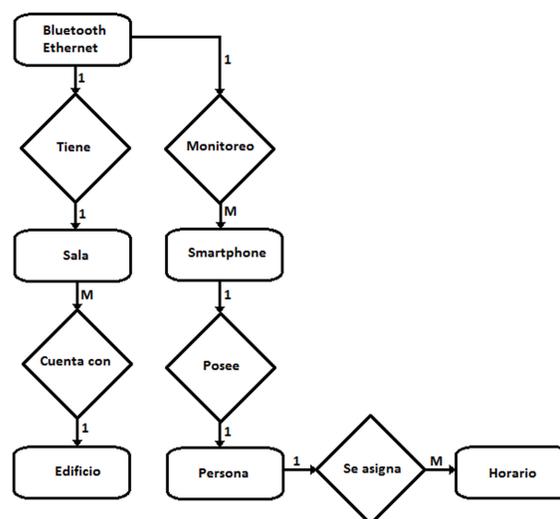


Figura 6. Diagrama Entidad Relación.

6. Proceso de Monitorización

El proceso de “pase de lista” se encarga de recabar la información de la monitorización de todos los módulos que conforman la red. La intención de este proceso es que se ejecute permanentemente en el servidor para que de manera automática y dependiendo de los horarios establecidos se envíe a cada módulo la instrucción de realizar la búsqueda. La manera de comunicación hacia cada módulo de la red se realiza a través de sockets TCP/IP [7]. El tiempo de monitorización se puede programar desde segundos hasta minutos, según se desea saber. El sistema puede reportar el tiempo que el trabajador permaneció en su espacio de trabajo, y el tiempo que se ausentó.

7. Aplicación para Administración de Datos y Página WEB

En la figura 7, se muestra la pantalla principal de la aplicación para la administración de datos que aparece en forma de menú indicándonos cuatro opciones que son: usuarios, teléfonos, módulos y red.

La finalidad de esta aplicación, es para ingresar, modificar y/o dar de alta o baja información en la base de datos.

La página web desarrollada para este proyecto sólo servirá para mostrar los resultados de la monitorización a smartphones en tiempo real. Toda la información que se mostrara será obtenida de la base de datos que previamente fue llenada por el proceso de “Pase de lista”.



Figura 7. Interfaz de la aplicación.

Tabla 1. Horario utilizado en las pruebas del sistema.

HORA	Smartphone	Dirección
7:00 - 8:00	Smartphone1	3c:c2:43:60:f1:08
8:00 - 9:00	Smartphone2	9c:02:98:96:a8:32
9:00 - 10:00	Smartphone1	3c:c2:43:60:f1:08
10:00 - 11:00	Smartphone1	3c:c2:43:60:f1:08
11:00 - 12:00	Smartphone2	9c:02:98:96:a8:32

8. Pruebas y Resultados

Una vez que se realizó la integración del sistema, se prosiguió a realizar las pruebas del sistema en su totalidad para esto se realizaron mediciones en base a un horario que se muestra en la Tabla 1. Este ejemplo se aplicó en un espacio educativo para verificar la asistencia de profesores a su salón de clases. Y determinar el tiempo que permanecen. Para esta prueba se determinó tomar datos cada 20 segundos y puede variar dependiendo del tiempo que el módulo se tarde en detectar el Smartphone. Solo se presenta una pequeña muestra de los datos obtenidos.

Esta prueba se realizó para dos salas donde se colocó cada uno de los módulos Bluetooth-Ethernet. En esta primera prueba, se hizo que el smartphone1 si acudiera a su lugar de trabajo en todas las horas programadas, mientras que el smarphone2 se le pidió que no asistiera de 8:00 a 9:00 horas, pero si en las demás horas. En la Tabla 2 se identifica como asistencia un 1 y como falta un 0, en caso de no haber comunicación se registra un -1. Se puede observar la correspondencia entre el resultado obtenido y los resultados esperados. Estos datos presentados solo son un subconjunto de los datos obtenidos, se puede observar peculiaridades como la que ocurrió a las 7:59, donde el sistema no pudo conectarse con el módulo pero a la siguiente prueba se pudo enlazar la comunicación marcándolo con -1 por el sistema.

A partir de la Tabla 2 los datos muestran el resultado esperado, se realizaron varias pruebas mostrando la eficacia del sistema. El sistema puede mostrar reportes

Tabla 2. Resultados obtenidos de la prueba del sistema.

Smartphone	Hora Registrada	Resultado Sistema	Resultado Esperado
Smartphone1	7:39:31	1	1
Smartphone1	7:43:32	1	1
Smartphone1	7:53:01	1	1
Smartphone1	7:59:47	-1	1
Smartphone2	8:18:06	0	1
Smartphone2	8:23:23	0	1
Smartphone2	8:28:24	0	1
Smartphone2	8:33:36	0	1
Smartphone2	8:59:15	0	1
Smartphone1	9:05:16	1	1
Smartphone1	9:11:17	1	1
Smartphone1	9:18:17	1	1
Smartphone1	9:23:18	1	1
Smartphone1	9:31:13	1	1
Smartphone1	9:40:41	1	1
Smartphone1	9:48:48	1	1
Smartphone1	9:54:13	1	1
Smartphone1	10:14:06	1	1
Smartphone1	10:40:33	1	1
Smartphone1	10:47:05	1	1
Smartphone2	11:38:41	1	1
Smartphone2	11:42:42	1	1
Smartphone2	11:48:42	1	1
Smartphone2	11:54:58	1	1
Smartphone2	12:00:00	0	0

del tiempo de permanencia de una persona en su área de trabajo y no solo que asistió.

9. Conclusiones

Después de realizar este proyecto y haber cubierto cada una de las etapas consideradas en la definición del problema, se puede concluir que a través de la detección de smartphones se puede llevar el control de asistencia y el tiempo de permanencia en un área laboral. Este sistema podría utilizarse en una universidad para obtener información de los alumnos y del profesor si asisten a clase y porque tiempo lo hacen. Las aplicaciones son diversas de sistema, se puede aplicar donde se tiene personal que debe permanecer por un cierto tiempo en área menor a diez metros. La distancia es debido al alcance de los transeptores Bluetooth. Además debido al alcance de los módulos Bluetooth resulta una ventaja su corto alcance más que una desventaja. A través de los datos obtenidos podemos decir que el sistema cumple con los requisitos de garantizar la permanencia y asistencia de un grupo de personas en su área trabajo. El sistema puede presentar algunas fallas que no se contemplan como puede ser la deshonestidad, dejando el smartphone sin estar presente la persona. Estos problemas también se pueden presentar con los sistemas tradicionales donde las personas checan su entrada y se quedan a laborar.

Referencias

- [1] [1] World Bank (Enero 20, 2013), <http://www.worldbank.org/ict/IC4D2012>. A. Serrano Santoyo and J. Organista-Sandoval, Implications of 4G Connectivity Related to M-Learning Contexts, *Journal of the Research Center for Educational Technology*, Vol. 6, No. 1, pp. 129-135, 2010.
- [2] T. Jamil, Automatic Attendance Recording System Using Mobile Telephone, Telecommunications Forum, 2011 19th. pp. 1297-1299, Serbia: IEEE. 2011.
- [3] Comision Federal de Telecomunicaciones. (Enero 17 2013). Cofetel, <http://www.cft.gob.mx:8080/portal/>.
- [4] Tsai-Cheng Li ; Huan-Wen Wu ; Tiz-Shiang Wu, The Study of Biometric Technology Applied in Attendance Management Systems. Digital Manufacturing and
- [5] Automation (ICDMA), 2012 Third International Conference on Digital Object Identifier, pp. 943-947, 2012. B. Millert and C. Bisdikian, Bluetooth Revealed, Prentice Hall, 2001.
- [6] R. Mettala (Enero 25, 2013), Bluetooth Protocol Architecture. SIG., 1999. Bluetooth Special Interest Group, http://srohit.tripod.com/bluetooth_sec.pdf
- [7] C. Webb, WAP, Bluetooth and 3G Programming: Cracking the Code, New York, Hungry Minds Inc, 2002.

Aplicación Móvil como Asistente en la Comunicación de Personas Discapacitadas y con Limitaciones del Habla

Jesús Hernández Cosío^a, Mónica Adriana Carreño León^a, Jesús Andrés Sandoval Bringas^a, Italia Estrada Cota^a, Elvia Esthela Aispuro Félix^a, Jaime Suárez Villavicencio^a

^aUniversidad Autónoma de Baja California Sur, Departamento Académico de Ciencias Computacionales.
Carretera al Sur km. 5.5. Col. El Mezquitito, La Paz, B.C.S., Mexico 23080.

<http://www.uabcs.mx{jhernandez, mcarreno, sandoval, iestrada, aispuro, jsuarez}@uabcs.mx>

2013 Published by \mathcal{UT} 100ci@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CCOMP-2013, www.cicomp.org

Resumen

Los grupos de atención especial son porciones poblacionales de importantes factores sociales y económicos, entre ellos se identifica al grupo de personas con un espectro de discapacidades, particularmente sensible y con requerimientos urgentes e inobjetable. Como parte de ese conjunto de necesidades fundamentales, la comunicación como actividad natural del ser humano, se posiciona entre las más altas prioridades a considerar. Las personas con discapacidades y con limitaciones del habla, enfrentan variados y difíciles obstáculos para lograr su interlocución con los demás, por lo que cualquier asistencia puede revelar una gran ayuda o mejora para erradicar sus barreras. Esto es sumamente deseable, ya que se crea la posibilidad de mejorar la vida de estas personas. El presente trabajo describe la propuesta de una aplicación para dispositivos móviles en beneficio de personas con alguna limitación, o bien con cualquier manifestación de dificultad en su comunicación diaria.

Palabras clave: Aplicación móvil, Asistente de comunicación, Discapacidad, Tecnología móvil.

1. Introducción

En las comunidades actuales, la realidad social de cada uno de sus integrantes debe ser obligadamente de igualdad de condiciones. El avance científico aunado al desarrollo tecnológico ha traído notables beneficios dentro de la salud y del entorno social, cambiando la calidad, los estilos y esperanzas de vida de la mayoría de las personas [1]. Por su par-

te los gobiernos han manifestado constantemente sus intenciones por favorecer la formación de un entorno social de imparcialidad y una adecuación a las distintas y múltiples circunstancias que pudiesen presentarse en el panorama específico de cada individuo, integrante de los diferentes sectores. Sin embargo, a pesar de la implementación de estas políticas sociales, aún no se ha llegado a un nivel de consolidación de aceptables condiciones, esto debido a la coexistencia de distintos

factores tales como la edad, género, condición económica, nivel educativo, etc. En México, la Secretaría de Desarrollo Social (SEDESOL), a través de la Dirección General de Atención a Grupos Prioritarios brinda apoyo a dichos grupos mediante programas especiales, y además promueve la participación de la sociedad, organizaciones, instituciones y organismos internacionales, con el propósito de que contribuyan al desarrollo social y humano de los grupos prioritarios en las diversas regiones del país [2]. Es necesario indicar que integran a este cúmulo de grupos, también conocidos como grupos de atención especial, las personas con discapacidad, madres solteras, jefas de familia, grupos indígenas, grupos vulnerables (como adultos mayores o niños de la calle), inmigrantes, entre otros. Para efectos de la actual propuesta se delimitó a lo concerniente al grupo de atención especial conformado por las personas discapacitadas residentes en el estado de Baja California Sur (BCS). Derivado de lo anterior, es posible indicar como información cuantitativa, que el Instituto Nacional de Estadística y Geografía (INEGI) informó que en México existen 5 millones 739 mil 270 personas con alguna discapacidad, de las cuales 26 mil 816 se encuentran en B.C.S.[3], esto de acuerdo al último censo poblacional del año 2010. A partir de estas cifras se calcula que el 4.23 POR CIENTO de los 633 mil 854 habitantes del estado, presentan alguna discapacidad identificada [3]. En la figura 1 aparecen dichas discapacidades, clasificadas de acuerdo al tipo de limitante presente en el individuo. Sin lugar a dudas estos datos resultan sumamente interesantes, pues permiten descubrir con facilidad el importante sector poblacional bajo estudio. También clarifica las limitantes en los individuos, observadas en la gráfica de la figura anterior. De acuerdo a su valor porcentual, las limitantes del habla no se visualizan como un punto relevante, sin embargo, la situación real se conduce bajo circunstancias de discapacidades múltiples. Las discapacidades múltiples se consideran aquellas situaciones que incluyen combinaciones de dos o más discapacidades, tales como sensoriales, comunicativas, motrices y mentales [1]. Esto significa que existe un número importante de personas con problemas relativos a dos o más limitantes, lo cual propicia la proliferación de obstáculos para el entorno de los individuos bajo esta situación. Es importante señalar que no es propósito de este artículo analizar en profundidad estos complejos temas sociales, es claro que se requiere la especialización y la entrega disciplinaria, como cualquier otra área de conocimiento para investigarlos adecuadamente. No obstante, se desea establecer un breve contexto introductorio y marco conceptual para abordar la temática, y así poder resaltar la oportunidad de la propuesta tec-



Figura 1. Gráfica con la distribución porcentual de la población con discapacidad por tipo de limitante en B.C.S. [3].

nológica descrita, identificada como una alternativa de apoyo directo a este grupo de atención especial.

Las universidades comparten una imperiosa obligación social como instituciones educativas, consistente en concebir proyectos como alternativas viables de solución a las necesidades y problemas existentes en las comunidades, ciudades, o países a los que pertenecen. Por lo que, mediante esta propuesta la Universidad Autónoma de Baja California Sur (UABCS), a través del Departamento Académico de Sistemas Computacionales (DASC) desea concretar la voluntad de estructurar una alternativa tecnológica que proporcione un apoyo significativo, y además, entregue una gama de servicios configurables de comunicación a personas en condiciones de alguna discapacidad, ya sea múltiple, o bien para cualquier individuo que tenga afectaciones en el habla o cierta parálisis cerebral que le imposibilite realizar adecuadamente su proceso de comunicación.

2. Justificación del Proyecto

Una verdadera incorporación de las personas con discapacidad al ambiente familiar y social, es un anhelo permanentemente presente bajo el cual existen múltiples y complicados obstáculos de distintas circunstancias, los cuales no favorecen la inclusión de estas personas a sus entornos. A lo largo de sus vidas deben considerar aspectos relevantes, como los que se indican en la figura 2, los cuales exigen interactuar para poder informarse, aprender, decidir, conocer, relacionarse, etc. Por lo tanto, es menester una adecuada comunicación para alcanzar mejores niveles de comprensión e involucramiento para todos los seres humanos [4]. La tecnología puede ser un sustituto, como un asistente de comunicación que proporciona una salida vocal para una persona que no puede comunicarse mediante su voz. Esto significa que seleccionando un botón, un



Figura 2. Aspectos más significativos del desarrollo de una persona con discapacidad.

niño podría pedir una manzana a su mamá, o podría hablar con sus amigos. También puede significar que un trabajador puede conversar con otras personas en su oficina. Bajo estas circunstancias la UABCS entiende que con su labor científica, tecnológica y de investigación, puede ofrecer un apoyo de valía social, ya que mediante el desarrollo de proyectos ha logrado incorporar atinadamente, la preparación de sus alumnos junto a la experiencia y conocimiento de sus docentes, convirtiéndose en una trascendencia participativa y plena para la comunidad.

3. Contexto Conceptual

3.1. Objetivo

El objetivo principal del proyecto es desarrollar la aplicación, para crear una herramienta de software para ayudar a las personas con discapacidad que requieran apoyo complementario para su interlocución. En relación al objetivo, es necesario considerar que la discapacidad es un término genérico que comprende las deficiencias en las estructuras y funciones del cuerpo humano, las limitaciones en la capacidad personal para llevar a cabo tareas básicas de la vida diaria y las restricciones en la participación social que experimenta el individuo al involucrarse en situaciones del entorno donde vive [3].

3.2. Modelo social y de interlocución

El modelo ilustrado en la figura 3, muestra tres grupos principales, la familia, los grupos de apoyo y la sociedad en general, comúnmente presentes en la relación con las personas en cuestión. A pesar de que se visualizan espacios semejantes ocupados por cada grupo, esto

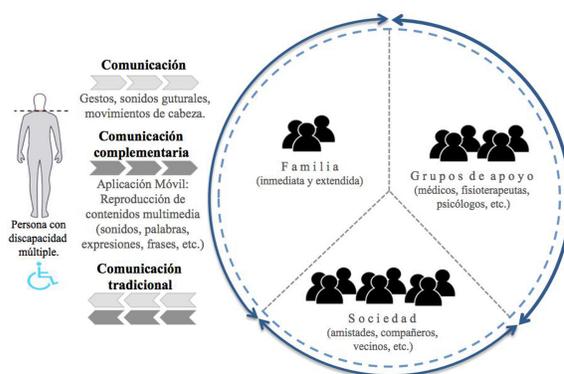


Figura 3. Modelo de interlocución propuesto para la interacción de la persona discapacitada y los grupos con los que se relaciona.

no significa que tengan que existir los tres, pues esto dependerá del contexto único de cada persona, además es factible considerar que la familia generalmente abarcará un mayor espacio en las relaciones interpersonales, aunque puede variar (disminuir) conforme la persona crezca o se independice. El modelo plantea que debe existir alguna modalidad de comunicación implementada por la persona y los grupos. Empleando gestos faciales, sonidos guturales y movimientos de cabeza, por parte de la persona discapacitada, y una comunicación tradicional, conformada por la voz, gestos, señales, posturas, etc., del lado de los grupos. En esta parte es donde se ubica a la aplicación tecnológica, convirtiéndose en habilitador de nuevas modalidades dentro de una comunicación complementaria para la persona discapacitada, entregándole a su vez la posibilidad de utilizar sonidos y una voz artificial para su comunicación. Este producto de software deberá ser fácil de usar, accesible, flexible y de eficiente interacción. Todas estas características serán la combinación perfecta para la interacción persona-computadora (dispositivo móvil).

3.3. Tecnología Móvil

No por el sólo hecho de desarrollar una aplicación móvil significa que será exitosa, esta debe proveer un valor a los usuarios. Una aplicación móvil es una pieza de software especialmente diseñada para ejecutarse en dispositivos móviles, tal como un teléfono inteligente (smartphone) o una tableta (tablet)[5]. Los dispositivos móviles son considerados en la actualidad como pequeñas unidades de cómputo, esto debido a que la tecnología móvil continua avanzando, además, los usuarios pueden realizar gran parte de las tareas que se realizaban comúnmente en una computadora personal [6].

3.4. Tecnologías del Desarrollo

Sistema operativo de dispositivos móviles: Android como sistema operativo de plataforma abierta, permite una compatibilidad cruzada, ya que puede correr en dispositivos de diferentes tamaños y resoluciones de pantalla, incluyendo teléfonos y tabletas [6]. Además, ofrece muchas herramientas para el desarrollo de aplicaciones. Lenguaje de programación: Debido a que las aplicaciones para Android son codificadas en Java, este será el lenguaje de programación a utilizar. Como parte de las librerías del lenguaje, se incluirá el subconjunto acorde para los dispositivos móviles. Específicamente se investigará exhaustivamente sobre el uso y control de la cámara, la cual será un medio de entrada de instrucciones o comandos. Gracias a experiencias previas se sabe de algunos componentes de librerías serán ampliamente útiles para los propósitos del proyecto

4. Aspectos del Diseño

4.1. Fases

Las fases generales del diseño seguirán las siguientes etapas:

- UML Diagrama de clases.
- diagrama de secuencia.
- Casos de uso e historias de usuarios.
- Pruebas de código y a nivel de aplicación.

4.2. Menus de Opciones

La aplicación pondrá a disposición de la persona una serie de menús contenedores de opciones básicas en formato gráfico (iconos), agrupadas en base a necesidades, las cuales al ser seleccionadas reproducirán un sonido o grabación de voz para comunicar la acción deseada, como se muestra en la figura 4.

Al inclinar la cabeza hacia el lado derecho, se desplaza la selección a la opción que se encuentre inmediatamente a la derecha. Al realizar un movimiento de inclinación a la izquierda, se desplazará la selección a la opción disponible inmediatamente a la izquierda. Finalmente, para activar la opción deseada se deberá mover la cabeza hacia arriba e inmediatamente se emitirá un sonido o se reproducirá la grabación de voz acorde a la opción elegida.

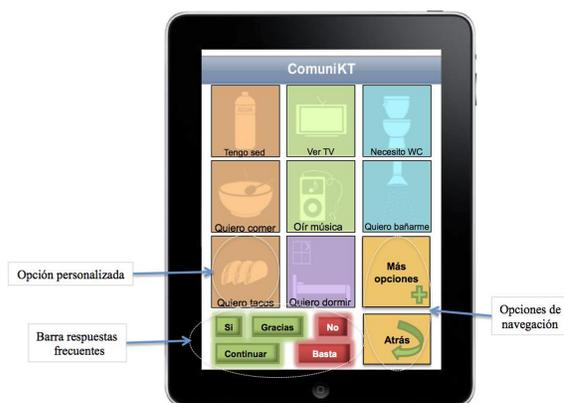


Figura 4. Interface con el menú de opciones.

4.3. Movimientos para la Interacción

La operación de la aplicación será mediante movimientos de la cabeza, para lo cual se espera que la persona tenga un adecuado control motriz. Los movimientos determinarán las posibles operaciones de navegación a través de las pantallas de la aplicación. Los movimientos válidos se ejemplifican con las posturas mostradas en la figura 5.

Al inclinar la cabeza hacia el lado derecho, se desplaza la selección a la opción que se encuentre inmediatamente a la derecha. Al realizar un movimiento de inclinación a la izquierda, se desplazará la selección a la opción disponible inmediatamente a la izquierda. Finalmente, para activar la opción deseada se deberá mover la cabeza hacia arriba e inmediatamente se emitirá un sonido o se reproducirá la grabación de voz acorde a la opción elegida.

4.4. Personalizaciones

Se configurarán opciones personalizables, tales como la voz, pues se ofrecerán varios tipos de voz, acordes al sexo y edad de las personas. También dentro de las opciones de personalización, será posible definir colores para los botones de acciones, tipo y tamaño de letra. Asimismo, se podrán crear plantillas de contenido para las opciones más frecuentes que faciliten la operación. La aplicación está basada en las nuevas tecnologías móviles (celulares y tabletas). Estos dispositivos al ser relativamente pequeños y livianos hacen fácil su portabilidad por lo que no representa un peso extra. Esto favorece a que las personas lleguen a considerarlas como objetos familiares y a sentirse identificadas con ellos [5].

	Posición de Reposo
	Posición Izquierda
	Posición Derecha
	Aceptación del Usuario

Figura 5. Movimientos de cabeza captados por la cámara del dispositivo asociados con operaciones de desplazamiento dentro de la aplicación.

4.5. Usuarios

Los perfiles de usuarios se encuentran plenamente identificados. Entre las características de los usuarios de la aplicación están:

- Usuarios de cualquier edad.
- Experiencia mínima utilizando tecnología de dispositivos móviles.
- No será necesario conocer términos técnicos.

Las personas que utilicen la aplicación deberán tener la capacidad visual y auditiva para tener un manejo adecuado de la aplicación. La aplicación utilizará la cámara frontal del dispositivo y con movimientos sutiles de la cabeza se llegará a utilizar la aplicación.

5. Requerimientos

5.1. Dispositivos de Entrada

El dispositivo debe contar con una cámara frontal para captar el rostro del usuario y por medio de los movimientos de cabeza que realice, la aplicación identificará el desplazamiento y se producirá la respuesta.

5.2. Dispositivos de Salida

Bocinas del dispositivo para emitir los sonidos y reproducir los audios. Cuando el usuario seleccione una de las opciones del menú, el dispositivo responderá con una pantalla, mostrando algunas imágenes. Podrá navegar entre ellas, y al seleccionar una de las imágenes, el dispositivo generará un audio o sonido correspondiente a la imagen.

6. Apoyo Interdisciplinario

Asumiendo la presencia de los responsables del proyecto obviamente pertinentes al desarrollo de aplicaciones móviles, como parte del plan también se identifican las siguientes áreas como necesarias para conjuntar conocimientos hacia el proceso de elaboración de la herramienta de software [7].

6.1. Diseñador Gráfico

Esta disciplina será útil en la creación de las imágenes idóneas para los distintos perfiles de usuarios, como son niños, jóvenes y adultos. En general, permitirá conjugar colores, tipos de letras, tamaños y formas de los componentes de la interfaz de usuario, en un modo interesante para cada perfil. Además, esto ayudará a que la interacción fluya correctamente.

6.2. Pedagogo

La aplicación debe servir de apoyo para construir esquemas de aprendizaje orientados a la enseñanza de auto suficiencia. De tal forma que esta disciplina aportará grandes e importantes aspectos.

6.3. Psicólogo

El permitir que las personas con discapacidad no se sientan inferiores y motivarlos hacia la superación personal y desarrollo humano, son temas significativos que podrán ser cubiertos con la ayuda de un profesional de esta área.

6.4. Fisioterapeuta

Valorar los esfuerzos físicos generados por la operación de la aplicación para definir las mejores posturas y flexiones derivadas de los movimientos de la cabeza, entre otros temas será abordado en esta disciplina.

7. Conclusiones

Los momentos actuales fluyen con la convergencia de distintas tecnologías, sobresaliendo entre ellas, la tecnología de dispositivos móviles. Cada vez crece el número de usuarios de estos dispositivos, a cada minuto son parte de la vida cotidiana de una gran cantidad de personas. Esta gran disponibilidad ha permitido que sean empleados para diversas actividades que van desde el entretenimiento y ocio, hasta el desarrollo de labores profesionales de trabajo. Paralelamente a esto, en las sociedades existen y surgen problemas que requieren ser atendidos con prioridad. Uno de estos problemas es la dificultad de comunicación que existe entre una persona con discapacidad y su entorno. Cabe mencionar que en el mercado de aplicaciones móviles existe una buena oferta de alternativas para personas con discapacidad, específicamente centradas en ayudar la comunicación. Sin embargo, gran parte de estas herramientas están desarrolladas con las leyendas, etiquetas y audios en inglés, lo cual resulta inapropiado. Otro aspecto a considerar es el costo envuelto en la adquisición, aunque existen opciones aceptablemente buenas y libres de pago, las mejores y más completas resultan ser versiones comerciales que exigen el pago de una licencia. Por todo esto, la aplicación propuesta será absolutamente gratuita y buscará atender prioritariamente a las poblaciones locales de forma inicial. Se considera mejorar la vida de las personas especiales, brindando una aplicación móvil que sin lugar a dudas resultará útil y valiosa. Como parte de la formación de los alumnos, se encuentra el sentido social, por lo que el proyecto será un enlace para satisfacer necesidades de personas que sufren marginaciones o exclusión social.

8. Resultados Esperados

Como resultados esperados se plantean los siguientes puntos:

- Fomentar que las personas con discapacidad reciban la oportunidad al derecho a su independencia y participación en todos los ambientes.
- Atender una necesidad social, prioritaria para el grupo de atención especial, discapacitados.
- Permitir que la utilización sea con plena facilidad y propiciar una verdadera familiaridad de operación.
- Servir de apoyo en las instituciones dedicadas al tratamiento de personas con discapacidad.

Referencias

- [1] M.J. Scherer, *Living in the State of Stuck: How Assistive Technology Impacts the Lives of People With Disabilities*, Brookline Books, 4ta. Edición, E.U.A., Enero de 2005.
- [2] Sitio Web de la Secretaría de Desarrollo Social. [http://www.sedesol.gob.mx/Es/SEDESOL/Direccion General de Atencion a Grupos Prioritarios](http://www.sedesol.gob.mx/Es/SEDESOL/Direccion%20General%20de%20Atencion%20a%20Grupos%20Prioritarios).
- [3] Sitio Web del INEGI. <http://www.inegi.org.mx> Consultado el 24 de junio del 2013.
- [4] Sitio Web de United Cerebral Palsy, <http://www.ucp.org>. Consultado el 4 de mayo del 2013.
- [5] J. McWherter and S. Gowell, *Professional Mobile Application Development*. Wrox, 1ra. Edición, E.U.A., Septiembre de 2012.
- [6] [6] P.A. Salz and J. Moranz, *The Everything Guide to Mobile Apps: A Practical Guide to Affordable Mobile App Development for Your Business*, Adams Media, E.U.A., Marzo de 2013.
- [7] [7] D. Pilone and R. Miles, *Head First Software Development*, O'Reilly Media, 1ra. Edición, E.U.A., Enero de 2008.

Análisis Térmico para un Circuito SoC de Disparo de un Rectificador Controlado de Alta Eficiencia

José Treviño Martínez^a, José de Jesús López Villalobos^a, José Valderrama Chaires^a

^aInstituto Tecnológico de Nuevo León.

Av. Eloy Cavazos No. 2001, Col. Tolteca, Guadalupe, N.L., México, 67170.

jtrevm2009@yahoo.com.mx, xe2n@yahoo.com.mx, jvc68@hotmail.com 2013 Published by *DI²U*_{100ci}@

<http://www2.uaz.edu.mx/web/www/publicaciones>

Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Resumen

Este trabajo establece una propuesta de Encapsulado para un control específicamente en la fase de disparo, que mediante agentes (máquinas finitas) para el diseño de Cruce por cero y los pulsos necesarios para el disparo de SCR's (Silicon-Controlled Rectifier) de un motor de Corriente Directa (CD), utilizando VHDL, simulando en Ise Simulator (ISim) y utilizando FPGA (Field Programmable Gate Array) para las pruebas de laboratorio y el análisis térmico de los componentes de un encapsulado usando el Evaluador de Potencia de Xilinx (XPE).

Palabras clave: Análisis térmico, Automatas, FPGA, SCR's ,VHDL.

1. Introducción

Durante las últimas tres décadas ha existido la necesidad de controlar la potencia eléctrica de los sistemas de tracción y de los controles industriales impulsados por motores eléctricos; esto ha llevado a investigar en el entorno de sistemas de Control con el objeto de obtener un voltaje variable para el control de los motores. Las líneas de investigación actuales buscan la integración de dispositivos de potencia y control en un único chip (SoC, por sus siglas en inglés, System on a Chip), reduciendo costos y multiplicando sus potenciales aplicaciones. No obstante existen dificultades a salvar como el aislamiento entre zonas trabajando a altas tensiones y circuitería de control, así como la disipación de la potencia perdida. En el presente traba-

jo se establece la propuesta de una metodología para elección del Encapsulado para un circuito de control específicamente en la fase de disparo, de un motor de Corriente Directa (CD) por medio del ajuste en los impulsos de disparo de Rectificadores Controlados de Silicio (SCR's por sus siglas en inglés), para el control de los impulsos se utilizan agentes (máquinas finitas) que ajustan los retardos a partir de los cruces por cero y dichos agentes se implementan (programan) en dispositivos FPGA (Field Programmable Gate Array), la programación o implementación de los circuitos físicos en los dispositivos FPGA se realiza con ayuda del lenguaje VHDL y los ambientes de software integrado (ISE -Integrated Software Environment- para el caso del fabricante XILINX), simulando en ISE Simulator (ISim),

además permiten la síntesis, el análisis y por tanto la optimización de los circuitos diseñados por medio de herramientas tales como simuladores de desempeño térmico y evaluación de potencia de los circuitos implementados (XPE -Evaluador de Potencia de Xilinx) y utilizando la metodología Top-Down donde se encapsula los procesos representativos del circuito para el disparo de los SCR's.

2. Manejo Térmico

La carga extraída por la naturaleza para el procesamiento de la información es el calor residual que se genera. Como la energía se acumula, aumenta la temperatura de unión en la soldadura. La diferencia de temperatura entre el dispositivo y sus alrededores permite conducir el flujo de calor fuera del sistema. Cuanto mayor sea la diferencia de temperatura, mayor es el flujo de energía. La temperatura máxima de una unión de un dispositivo se alcanza cuando el flujo de calor hacia el ambiente es capaz de mantenerse con la energía que se produce por la unión. La Gestión de ingeniería térmica determina qué cantidad será esta temperatura [1]. Más allá de los problemas relacionados, existe una fiabilidad y pérdida de velocidad asociado con el aumento temperatura de la unión. En general, todos los mecanismos de falla tienen una energía de activación asociada con ellos que son acelerados por la temperatura creciente. Para una energía de activación para el mecanismo de falla de 0,5 eV, típica de afectación a través de los fallos, por ejemplo, cada 10 °C de aumento en la unión de soldadura los resultados de temperatura en la vida útil de un dispositivo se acortan más o menos a la mitad (Figura 1). Debido a que el canal resistencia aumenta con la temperatura, la demora de conmutación para una puerta también se incrementa aproximadamente dos por ciento por cada 10 °C de aumento en la temperatura. Por estas razones, es necesario evitar que en la unión, la temperatura se eleve demasiado. La utilización de los FPGA (Field Programmable Gate Array) ha ido creciendo de manera impresionante debido a que ha resultado una herramienta para generar prototipos en corto tiempo y con bajos costos. El consumo de potencia y la disipación térmica ha sido un factor de estudio por mucho tiempo en la industria y en la investigación que ha generado metodologías y estándares que pueden ser aplicados para un mejor diseño de los dispositivos o sistemas de componentes que son enfocados a solucionar un problema en específico. No es desconocido que actualmente la disipación de calor dentro de encapsulado sobrepasa la capacidad térmica del paquete sobre todo en aquellos donde no se lleva

un buen diseño y manejo de la energía [2].

Con los niveles de disipación de calor de los componentes que llegan a 500 W/cm² y en ocasiones sobrepasa este número, los sistemas convencionales de enfriamiento de aire son inadecuados para eliminar el exceso de calor. La investigación se ha intensificado hacia el desarrollo de técnicas más innovadoras de refrigeración de chips. El objetivo final es reducir la resistencia térmica de la unión del chip a la temperatura ambiente, y mantener la temperatura de la unión de soldadura del chip tan bajo como sea posible [3]. En general, la temperatura de la unión debe mantenerse por debajo de 115 °C durante un período de vida razonable de un dispositivo. IBM utiliza el 85 °C como su directriz de diseño. Intel utiliza 90 °C como la especificación de unión temperatura máxima para el procesador Pentium. Con un ambiente que podría ser de 40 °C(113 °F) dentro de un recinto, la caída de la temperatura máxima entre la unión y el ambiente es típicamente del orden de 50 °C. El propósito de una estrategia de gestión térmica es proporcionar un método de extracción de cualquier tipo para que la energía térmica que se genera se disipe en el medio ambiente, mientras se mantiene la unión a una temperatura por debajo de 90 °C, utilizando como máximo un diferencial de temperatura de 50 °C. Al mismo tiempo, a menudo hay una especificación de ruido acústico que se deben de cuidar al momento de buscar la estrategia del manejo de la temperatura [1].

Elementos Finitos

Los elementos finitos son cada vez más utilizados para poder aproximar los resultados de un prototipo de un circuito integrado o un dispositivo electrónico inclusive sobre su comportamiento térmico [4]. El método de los elementos finitos (MEF) ha adquirido una gran importancia en la solución de problemas de ingeniería, de física, de aproximaciones esféricas, ya que permite resolver casos que hasta hace poco tiempo eran prácticamente imposibles de resolver por métodos matemáticos tradicionales. Esta circunstancia obligaba a realizar prototipos, realizar varias pruebas e ir realizando mejoras de forma continua, lo que traía consigo un elevado coste tanto económico como en tiempo de desarrollo y re trabajo [5].

Estimados de Potencia de Xilinx (XPE)

El Estimador XPower (XPE) es una herramienta de estimación de potencia típicamente usados en el pre-diseño y pre-implementación de fases de un proyecto. El XPE ayuda con la evaluación de la arquitectura, selección de dispositivos, componentes

adecuados de suministro de energía, manejo térmico específicos para una aplicación determinada. El XPE considera el uso de recursos del diseño, los factores de cambio, la carga de E/S, y muchos otros factores que se combinan con los modelos de dispositivos para calcular la distribución de potencia estimada. Los modelos de los dispositivos se extraen de las mediciones, la simulación, y/o extrapolación. La precisión del XPE depende de dos conjuntos principales de entradas: Dispositivos utilizados, configuración del componente, el reloj, lo activo, y los factores de cambio, y otra información que se introduce en la herramienta. Los modelos de los dispositivos de datos integradas en la herramienta, para realizar las estimaciones precisas de la aplicación, se debe introducir información realista que sea lo más completa como sea posible. Actualmente es utilizado en la Industria e Investigación [6].

Introducción a los paquetes de Xilinx

Los paquetes Electrónicos son capsulas interconectadas para dispositivos de semiconductores [7]. La mejor función de los paquetes Electrónicos es proveer de interconexiones eléctricas entre los Circuitos Integrados (CI) y la tarjeta y para remover eficientemente el calor generado por el dispositivo. Xilinx provee de un alto rango de soluciones con plomo y arreglos empaquetados para los proyectos avanzados. Las soluciones avanzadas de Xilinx incluyen arreglos de redes de bolas de plástico sobre moldeadas (PBGA, por sus siglas en inglés, overmolded plastic ball grid arrays), Paquetes de Chips escala pequeña (small form factor Chip Scale Packages), Cavidad Baja BGA's ("Cavity-Down" BGAs), Matrices de Arreglos de Columnas con chip invertido cerámicos (Ceramic column grid arrays flip chip (CCGA) por su siglas en inglés), así como el nuevo forma de paquetes como los paquetes Cuadrados planos sin plomo (Quad Flat No-Lead (QFN)) para ofrecer varias cantidades de pines y requerimientos de densidad. Los paquetes de Xilinx son diseñados, optimizados, y preparados para soportar los requerimientos mecánicos de larga vida y proponer el funcionamiento más avanzado de requerimientos térmicos y eléctricos de los proyectos avanzados FPGA de alta velocidad. Xilinx también desarrolla soluciones de paquetes que son seguras para el ambiente. Actualmente, los paquetes estándar de Xilinx no contienen sustancias que son identificadas como peligrosas para el ambiente incluyendo cadmio, cromo hexavalente, mercurio, Polibromobifenilos PBB, y éteres polibromodifenil PBDE. Las soluciones libres de plomo (Pbfree) es un avance sobre las antecesoras. Esta propuesta de Xilinx cumple con la reducción

de Sustancias peligrosas. Los Paquetes Pb-free de Xilinx también consideran el cumplimiento de la norma JEDEC J-STD-020, lo que significa que los paquetes son hechos más robustos que los hace capaz de resistir altos reflujos de temperatura.

Los métodos de caracterización

Xilinx utiliza varios métodos para obtener las características térmicas de rendimiento de paquetes de circuitos integrados. Los métodos incluyen la simulación térmica utilizando herramientas de software de elementos finitos, y un método indirecto que utiliza un diodo eléctrico aislado en una matriz de prueba especial térmica o incluso en una FPGA Xilinx ubicado en el paquete de su interés. La mayoría de los datos comunicados por Xilinx se basa en el método indirecto de diodo. Las herramientas de simulación, calibrado con datos de mediciones reales, se utilizan para complementar la garantía térmica de generación de datos. La mayoría de los datos publicados del modelo térmico se basa en ese esfuerzo.

Máquinas de Estado

Una máquina de estados es frecuentemente usada para el diseño de un sistema digital que representa los pasos de un procedimiento o algoritmo [8]. Los diagramas de estados o los grafos de estados, los círculos representan los estados y los arcos representan la transición que tradicionalmente ha sido usada para especificar la operación de la máquina de estados controladora. Como una alternativa para usar grafos de estados, un tipo especial de flujo de datos, llamado flujo de maquinas de estados (SM chart, State Machine chart), que puede utilizar para representar la conducta de las maquinas de estados. Estas representaciones también son llamados algoritmos gráficos de máquinas de estados (ASM charts, Algorithmic State Machine charts). SM charts son frecuentemente para diseño unidades de control para sistemas digitales. El diseño se puede simular e implementar sobre el mismo FPGA para reducir los riesgos de lo esperado con respecto al modelo y el circuito final [9]. La generación del listado de estados permite realizar la lógica para plantear los ajustes necesarios como eliminar estados repetidos o reutilizando salidas. El concepto de estados equivalentes es importante para el diseño y pruebas de circuitos digitales. Identificar los equivalentes ayuda a reducir el consumo de hardware por ende evitar el SoC sea más grande mas allá de lo necesario abarcando parte del circuito y disminuir la energía de consumo y por tanto la temperatura.

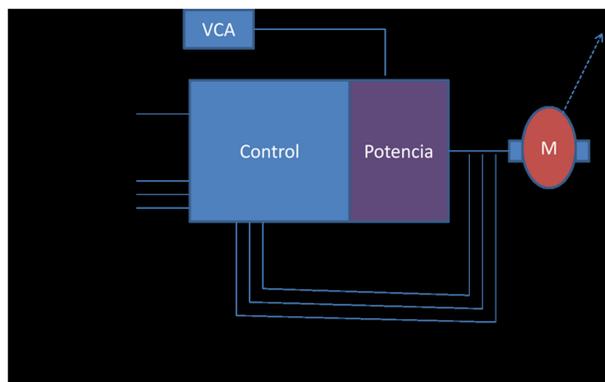


Figura 1. Modelo de Control Propuesto

VHDL y Agentes

Para el diseño se aplican en técnicas de optimización que están incorporadas en técnicas de diseño asistido por computadora (CAD, Computer Asisted Design). La importancia de minimizar estados ha ido disminuyendo en los recientes años por la cantidad abundante de transistores que se utilizan, sin embargo es necesario que al menos se reduzcan por eliminación obvia y así reducir el área del circuito y la energía [10]. Por ésta razón se hace necesario realizar un diseño óptimo, aprovechando los recursos que estén disponibles dividiendo el problema en varios módulos menos complejos, utilizando como se mencionó anteriormente un diseño TOP-DOWN y dónde herramientas como el VHDL genera la sincronización y concurrencia con máquinas pequeñas donde cada uno de estos subsistemas se realiza la interconexión de los mismos donde cada parte se comporta como un pequeño sistema digital. Mediante la descripción de cada uno de los módulos se llega a la solución total deseada. [10].

3. Implementación

El circuito a utilizar es el presentado en [11], donde se tiene el alcance del circuito de cruce por cero y el de disparo de los SCRÁ's. Las compuertas lógicas necesarias para representar el circuito se encapsulan dentro un integrado y mediante la herramienta ISE, se genera el Soc. Los resultados generados se llevan a la herramienta de simulación XPE, para el cálculo de la temperatura que va a estar sometido el integrado. Este se muestra en la Figure 1.

El encapsulado resultante se muestra en la fig. 2. Dentro de las familias de FPGA's de Xilinx, se seleccionó Spartan 3E FPGA para poder hacer la implementación y simulación del comportamiento del prototipo. La última columna son los bits configurados.

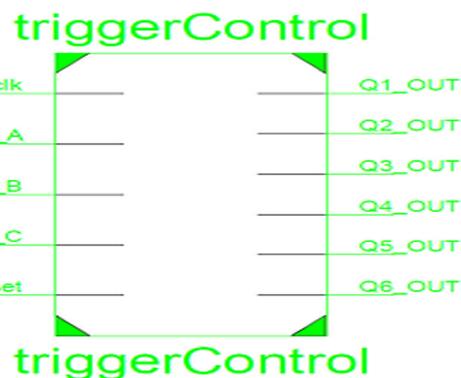


Figura 2. Características del Paquete

Spartan-3E FPGA	XC3S100E	581,344
	XC3S250E	1,353,728
	XC3S500E	2,270,208
	XC3S1200E	3,841,184
	XC3S1600E	5,969,696

Figura 3. Tomado del manual ug332 Xilinx. (Xilinx U., 2009)

En este caso es la que soporta nuestra tarjeta de pruebas y de acuerdo a la configuración del manual de Xilinx (Ug320), se seleccionó el XC3S500E.

Dentro de la parte seleccionada de la tabla anterior se seleccionó fg320. El paquete fgg320 la diferencia de éste es que propone materiales más nobles con el ambiente, pero en funcionamiento tienen las mismas características que el fg320 [6]. El paquete fg320 es del tipo de matriz de arreglo de bolas de Soldadura [12]. También se realiza una prueba exhaustiva con nueve veces el tamaño con un 70 % de utilización con el peor caso de análisis con 39 entradas y 117 salidas con reloj a 50 MHz. Lo siguiente es generado con el Simulador XPE de Xilinx [6] que internamente utiliza Elementos Finitos y datos de Experimentación en Laboratorio.

4. Análisis de Resultados

Los autómatas desarrollados con VHDL y la herramienta isim de xilinx presentaron excelentes resultados dejando un diseño de bajo consumo que nos permite un mejor manejo de la temperatura. Lo que nos da los siguientes consumos de energía, utilización de componentes y propiedades térmicas cercanas al real de Silicio. La curva térmica es como se muestra en la siguiente Figura 6 Como Extra por si se quiere obtener un prototipo y aun que no se recomienda el uso comercial se puede considerar el Cuadrado Plano por si se quiere realizar más pruebas de laboratorio. Para el prototipo Cuadrado Plano con la misma cantidad de componentes incrementado nueve veces, en este caso ocupando

Información Térmica	
Temp Ambiente (°C)	50.0
Flujo de Aire (LFM)	0
ΘJA (°C/W)	25.9
Configurable ΘJA	
Max Ambiente (°C)	80.1
Unión Temp(°C)	54.9

Figura 4. Especificación Xilinx

Encapsulado	Potencia(W)	Usado	Disponible	Utilización
Relojes	0.002	9	---	---
Lógico	0	449	9312	4.8
Señal	0	485	---	---
Entrada/Salida	0	11	232	4.7
Filtración	0.081			
Total	0.084			
Propiedades Térmicas		TJA Efectivo (C/W)	Max Ambiente (C)	Temp Junction (C)
		26.1	82.8	27.2

Figura 5. Resultado Térmico de SoC de Disparo

el 98 % del paquete. (Con autómatas implementados sobre una tarjeta FPGA para uno eficiente de Energía), se llega a la conclusión que el paquete seleccionado son una ventilación en la herramienta XPE) aún que se comportamiento Térmico de los materiales y las uniones del encapsulado así como las técnicas de arreglos de bolas de soldadura, se obtiene excelentes respuestas (Con Técnicas de Elemento Finito).

En este caso fue necesario utilizar un ventilador con 750 Líneas por minuto (LFM). Esto nos permite mantener en las uniones una temperatura de 55.4 a una temperatura de 50 °C pero con un 98.7% de utilización. Con Encapsulado cuadrado y con 3 cm aproximadamente con 1 cm de diferencia con respecto al fg320, nos permite construirlos manualmente muy útil para prototipos y aun teniendo resultados aceptables.

Conclusiones

Mediante una configuración para la tarjeta Spartan 3E con fg320, es decir un encapsulado en matriz de arreglos de bolas de soldadura, se obtiene excelentes respuestas (Con Técnicas de Elemento Finito de la herramienta XPE) amplifica el tamaño del circuito, lo que indica que se espera un buen comportamiento del dise-

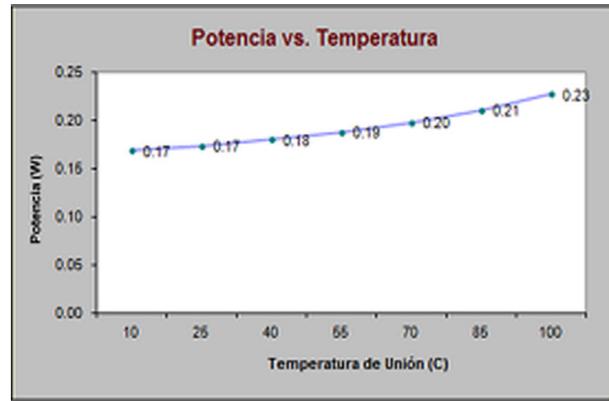


Figura 6. Temperatura de un fg320

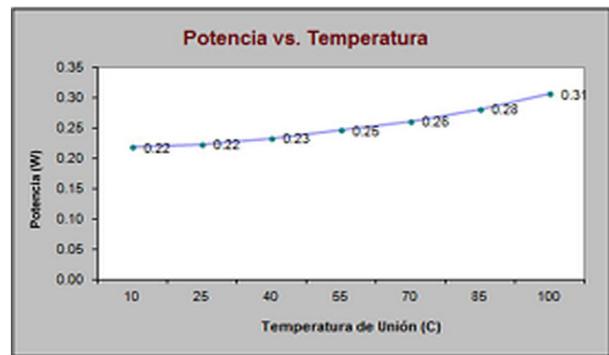


Figura 7. Temperatura de prototipo Cuadrado Plano

ño propuesto alternativa para encapsular el circuito de disparo en una pastilla SOC, de manera comercial con fg320 o si desea utilizar un prototipo si se desea realizar con el PQ208 también presenta buenos resultados por el uso eficiente de la Energía

Referencias

- [1] T. Yamada. Roadmaps Of Packaging Technology. Thailand: IC Engineering Corporation. 1994.
- [2] S. Nevado. Implementación de Técnicas de Bajo Consumo en FPGAs. International Technology Roadmap for Semiconductors. 2009.
- [3] B. Tavassoli. Heat removal with microchannel heat sinks. 2011.
- [4] C. D. Chaudhuri. Thermal Effects in Design of Integrated CMOS MEMS High Resolution Pressure Sensor. International Journal On Smart Sensing And Intelligent Systems September, VOL. 2, NO. 3. 2009.
- [5] A. Carnicero. Introducción a los Elementos finitos. España: Universidad Tecnológica Nacional de Córdoba. 2009.
- [6] Xilinx, U. XPower Estimator User Guide. New York: Xilinx. 2012.
- [7] Xilinx, U. Device Package User Guide. Obtenido de http://www.xilinx.com/support/documentation/user_guides/ug112.pdf, 22 de Septiembre de 2010.
- [8] C. H. Roth y L. K. Jonh. Digital Systems Design Using VHDL (Second Edition Ed.). Stamford, CT: CENGAGE Learning. 2008.

- [9] L. A. Barragán, I. Urriza, D. Navarro, J. I. Artigas, J. Acero, y J. M. Burdio. Comparing simulation alternatives of FPGA-based controllers for switching converters. Dept. Ingeniería Electrónica y Comunicaciones, Universidad de Zaragoza, 419-424. 2007.
- [10] L. D. Murillo. Redes de Petri: Modelado e implementación de algoritmos para autómatas programables. *Tecnología en Marcha*, 21 (4), 102-125. 2008.
- [11] J. Treviño, J. Villalobos y J. Valderrama. Diseño de un circuito SoC para Disparo de SCR. *Cicomp 2012*.
- [12] Xilinx, F. Fine-Pitch BGA (FG320/FGG320) Package. New York: Xilinx. 2005.

Implementación de una Comunidad Virtual de Aprendizaje en los Subsistemas de los Tecnológicos Descentralizados del Estado de Tabasco para el Fortalecimiento Académico

Roberto Custodio Martínez^a, Miguel Alberto López Guzmán^a, y Ana Lilia Sosa López^a

^aInstituto Tecnológico Superior de Centla.
Calle Ejido s/n, Colonia Siglo XXI, Frontera Centla, Tabasco, México, C.P. 86751
rcustodiom@itscentla.edu.mx, mlopezg@itscentla.edu.mx, asosal@itscentla.edu.mx
<http://www.itscentla.edu.mx>

2013 Published by *DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CCOMP-2013, www.cicomp.org

Resumen

Los cambios tecnológicos, de métodos y los recursos disponibles para el proceso de enseñanza-aprendizaje sea un éxito y que cumpla con las metas establecidas tanto por los profesores como por los estudiantes, ha provocado la aparición de nuevos conceptos dentro del campo educativo como las comunidades virtuales de aprendizaje. Este concepto es de interés para los docentes como para los estudiantes, ya que ha tomado auge en los últimos años debido a la aparición e incursión de la tecnología en el campo educativo, este proyecto tratara de integrar las actividades académicas, proyectos de trabajo, capacitación, talleres y eventos que organizan cada una de las instituciones educativas de los Institutos Tecnológicos Descentralizados del Estado de Tabasco es el objetivo del proyecto de investigación “Implementación de una Comunidad Virtual de Aprendizaje en los subsistemas de los Tecnológicos Descentralizados del Estado de Tabasco para el fortalecimiento académico”.

Palabras clave: Comunidad Virtual de Aprendizaje, Fortalecimiento Académico, Red de Aprendizaje.

1. Introducción

Debido a la creciente innovación tecnológica que se presenta día a día y con la oportunidad de

aprovecharlas en un cien por ciento, este proyecto se origina de la incertidumbre e inquietud que se encontró en el estado de Tabasco en detonar el desarrollo académico, cultural, de investigación y de capaci-

tación al personal administrativo, docente y comunidad estudiantil, debido a que no existe un aprovechamiento integral en el desarrollo productivo de proyectos académicos en los diferentes Institutos Tecnológicos de nuestro estado, es por eso la relevancia de desarrollar una Aplicación Web que permita asesorías en línea de materias específicas y en tiempo real, por medio de un video/chat para facilitar la comprensión de los estudiantes sobre las dudas que se le presenten en su formación profesional. Así mismo, los profesores de los diferentes Tecnológicos, podrán interactuar con el resto de la plantilla docente que conformaran la comunidad virtual de aprendizaje, entre otras cosas, para intercambiar métodos y técnicas de enseñanza, colaborar en proyectos de investigación entre otras actividades, los cuales los estudiantes se vean beneficiados, siendo este nuestro principal objetivo. En el estado de Tabasco falta trabajar en proyectos integrales en donde las universidades del estado trabajen para el beneficio de Tabasco.

2. Estado del Arte

2.1. Comunidades Virtuales

Las Comunidades Virtuales son espacios en Internet destinados a facilitar la comunicación entre los miembros del grupo al que pertenecen y que se encuentran en distintos puntos geográficos. Este concepto no es nuevo, en realidad la comunidad ha existido siempre a lo largo de la historia, la única diferencia radica en que antes el territorio de la comunidad era geográfico y hoy este es virtual, es decir, la comunidad no ocupa un espacio en el mundo físico sino en la web. Así, el ser humano traslada a este espacio virtual, sus modos de vida en sociedad, su cultura y configura ese espacio para que responda a sus necesidades.

Una de las características que tiene una comunidad virtual es que pertenecen a la misma red de trabajo, es decir, los actores pueden ser personas, organizaciones y todos deben coincidir en su misma relación social-laboral. Es evidente que si el manejo de estos nuevos medios de información y comunicación es encaminada desde la perspectiva del aprendizaje, los profesores tendrán en sus manos una herramienta muy completa que les servirá de apoyo a su práctica docente ya que a través del uso de Internet no solamente tendrán acceso a la información sino que además podrán trabajar bajo un esquema de comunicación multidireccional que les permitirá trabajar de manera conjunta, exponiendo sus ideas, intercambiando experiencias y de esta forma se retroalimentarán y aprenderán unos de otros, rompiendo así con el esquema tradicional de enseñanza donde el

papel del estudiante dejara de ser un simple receptor pasivo.

2.2. Naturaleza y Características de las Comunidades Virtuales de Aprendizaje (CVA)

Para Linda Harasim (2008) existen cinco rasgos que distinguen la comunicación cuando se trabaja en redes de aprendizaje:

- a) La comunicación tiene lugar en grupo.
- b) Es independiente del lugar.
- c) Es asincrónica, y por lo mismo es temporalmente independiente.
- d) Se basa en el hipertexto y cada vez más en entornos multimedia. Se envían mensajes por la computadora.

Lo anterior nos permite acercarnos a un concepto más acabado de comunidades virtuales de aprendizaje y de grupos virtuales de aprendizaje colaborativo. Una comunidad virtual de aprendizaje consiste en un espacio de encuentro en la virtualidad en donde concurren sus diferentes miembros para hacer uso de los distintos tipos de estructuras carpetas, módulos, etcétera creadas para albergar y dar forma a la acción comunicativa de carácter interactivo, en forma ordenada y en función de los objetivos comunes de aprendizaje y metas a lograr. Los estudiantes pasan a ser un grupo cuando, además de la interacción que se establece entre sí, participan en el desarrollo de actividades comunes y acuerdan, por negociación entre docentes y estudiantes, las reglas de operación del grupo. De este modo se van integrando grupos vivos que mediante la red deciden aprender a través de la comunicación y la colaboración mutua.

2.3. Las primeras CVA en México.

Las universidades han invertido grandes cantidades de dinero en crear laboratorios de cómputo que acerquen las TICs a los investigadores, docentes y alumnos, pero invertir en los espacios de cómputo es sólo el inicio de lo que puede ser un gran problema, hay que garantizar la capacitación de los usuarios para tener un nivel aceptable de la tecnología y sacarle provecho a inversión [5]. La UNAM FES Iztacala iniciaba sus grandes procesos de inversión en equipos de cómputo para los laboratorios escolares, 70 equipos en tres laboratorios, «cantidad considerable para una facultad que se dedica a la enseñanza de disciplinas del área de las ciencias biológicas y de la salud», la inversión estaba hecha, los equipos instalados y las promociones al uso de los

laboratorios tenían poco impacto. Dentro de la universidad se implementó una metodología de enseñanza que le llamaron “Laboratorio en Línea de Enseñanza de Cómputo (LLEC)” y que permitía enseñar los mismos contenidos en cursos de 20 a 40 horas. El LLEC descansaba entre nociones básicas: aprendizaje situado, tutelaje cognoscitivo y el aprendizaje colaborativo, así como rescatábamos las nociones de experto novato y comunidades virtuales de aprendizaje. Los estudiantes podían elegir por cursar el LLEC en una línea estructurada (se enfrentaban a una secuencia de contenidos preorganizada) o de forma exploratoria (los participantes decidían que necesitaban aprender). En cualquiera de las dos opciones se desarrollaran habilidades de interacción con los diferentes componentes funcionales de la computadora [2]. Observa la figura 1 en donde se muestra la Arquitectura del proyecto LLEC.

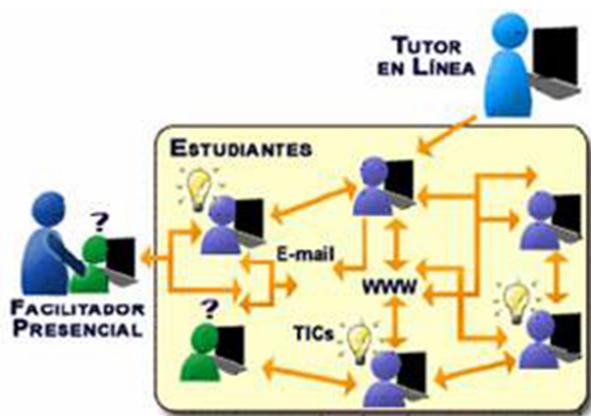


Figura 1. Muestra la interacción entre la herramienta, alumnos y facilitadores en el Laboratorio en Línea de Enseñanza de Cómputo

El LLEC fue aplicado cuatro años de forma ininterrumpida tiempo en que se almaceno una considerable cantidad de mensajes y páginas web, elementos que eran usados por los nuevos estudiantes de manera recurrente para realizar actividades sugeridas, resolver dudas y hacer comentarios. “El éxito del LLEC radicaba principalmente en la idea de situar el aprendizaje, bajo el supuesto de que el cómputo como muchas más actividades, se aprende haciendo, si vinculamos esta práctica con el soporte de una tutoría presencial y en línea que modela un discurso, formas de uso y actitudes hacia la tecnología el resultado era una comunidad de alumnos, docentes e investigadores sensibilizados de manera eficiente en los usos del cómputo aplicados a la educación” [1]. Otros ejemplos correspondientes a la implementación de una comunidad virtual de aprendizaje se llevaron a cabo en la Ibero On-Line, donde se demuestra como el modelo pedagógico es el eje rector para el diseño y operación de las comunidades. “Se dan

evidencias de los trabajos que realizan los participantes dentro de los foros de discusión de las comunidades, a través de cuadros comentados donde queda patente como los conceptos teóricos se convierten en quehacer cotidiano de las comunidades virtuales de Ibero On-Line”[2].

3. Metodología o Técnica usada

Este trabajo propone el implementar un espacio a través del cual se lleven a cabo las actividades académicas, proyectos de trabajo, capacitación, talleres y eventos que organizan cada una de las instituciones educativas de los Institutos Tecnológicos Descentralizados del Estado de Tabasco con el fin de integrarlos y trabajar en el mismo fin académico. De tal forma que su uso permite a los participantes de la comunidad virtual de aprendizaje retroalimentar las actividades de manera rápida y sencilla.

3.1. Recopilación y Análisis de la Información

Al integrar esta comunidad virtual de Aprendizaje en los Institutos Tecnológicos Descentralizados en el estado de Tabasco con el fin de transferir los conocimientos y habilidades, contribuyendo en la formación académica de los estudiantes y profesores institutos mediante la implementación de una Aplicación Web la cual en estos momentos está en desarrollo con un 40 % de avance. Para el desarrollo del proyecto se realizó una investigación previa para generar un análisis y un estudio de factibilidad para la realización del Proyecto CVA, se llevó a cabo una encuesta donde se evaluó el conocimiento que tiene cada estudiante de una comunidad virtual, ver figura 2.



Figura 2. Conocimiento de los estudiantes de una CVA.

Esto significa que el 62% de los estudiantes podrán interactuar fácilmente con la Comunidad Virtual de Aprendizaje y que antes de culminar el proyecto se debe de impartir cursos de capacitación para los usuarios que no han interactuado o no conocen un CVA. El objetivo

de la recopilación de la información fue de obtener la información necesaria para formar los módulos reales que se necesitan para la Comunidad Virtual de aprendizaje, el proceso de Recopilación de la Información se concluyó satisfactoriamente y se están realizando las etapas análisis y diseño del CVA.

3.2. Arquitectura de la CVA

A través de este diagrama podemos darnos cuenta como estará distribuido la comunidad virtual de aprendizaje, ver figura 3.

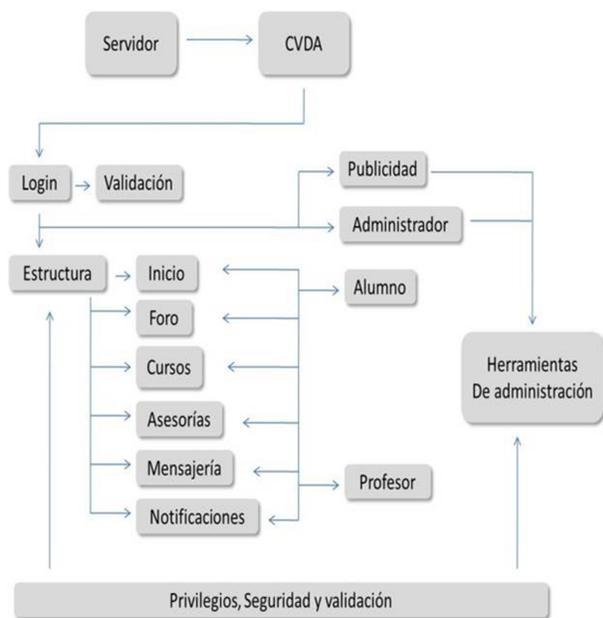


Figura 3. Diagrama de Navegación del CVA

En la Etapa del Análisis se trabajó en la Descripción de los Requisitos obteniendo el siguiente resultado del análisis. En el Diagrama se presentan los Tecnológicos Descentralizados que participaran en el uso del CVA ver figura 4. Los Tecnológicos involucrados en el son:

- Instituto Tecnológico Superior de Centla (ITS-Ce)
- Instituto Tecnológico Superior de Macuspana (ITSM)
- Instituto Tecnológico Superior de Comalcalco (ITSC)
- Instituto Tecnológico de Villahermosa (ITV)
- Instituto Tecnológico Superior de los Ríos (ITS Ríos)
- Instituto Tecnológico Superior de la Sierra (ITSS)
- Instituto Tecnológico Superior de la Venta (ITSV).

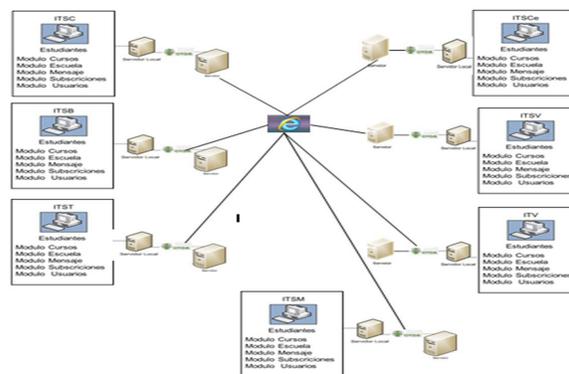


Figura 4. Diagrama del contexto de la Comunidad Virtual de Aprendizaje (CVA)

4. Resultados Experimentales

Las comunidades virtuales de aprendizaje tienen beneficios en el aspecto profesional, como estudiante y personal a los integrantes de la comunidad. En el aspecto profesional, a través de las comunidades virtuales de aprendizaje, se comparten conocimientos e información que ayuda a aumentar el conocimiento en su materia de estudio y/o su área de interés [3].



Figura 5. Pantalla Principal de la Comunidad Virtual de Aprendizaje (CVA).

En la Fig. 5 Se pudo observar el prototipo de la pantalla de inicio de la comunidad virtual de aprendizaje, en esta primera pantalla los usuarios de la CVA deberán autenticarse a través de su matrícula o clave que se le asigne y una contraseña la cual es generada por el usuario, después de haber ingresados los datos correspondiente para su autenticación pulsaremos el botón de ingresar y mostrara la siguiente pantalla la cual se muestra en la Figura 6.

A continuación se describen cada uno de los módulos que contiene la CVDA:

Módulo Inicio: mostrara los anuncios y eventos que

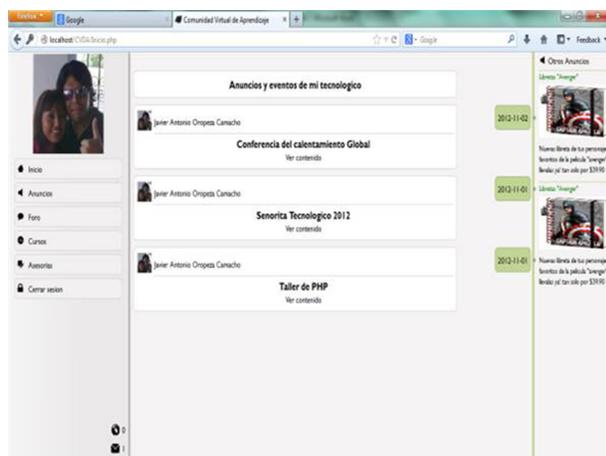


Figura 6. Pantalla Principal de la Comunidad Virtual de Aprendizaje (CVA).

cada uno de los tecnológicos participantes publiquen en la comunidad virtual la cual será visible para cada uno de los alumnos que tengan acceso a la comunidad virtual.

Módulo Foro: en ella se anunciarán los foros de discusión o de temas innovadores que ayuden en el aprendizaje del estudiante.

Módulo Cursos: en este módulo se publicarán los cursos con que cuenta cada tecnológico en la cual los estudiantes estarán enterados y así poder inscribirse al curso que mejor les convenga para su actualización constante para su perfil académico.

Módulo Asesorías: de la misma manera se publicarán las asesorías disponibles en la cual los estudiantes interesados podrán tener en línea.

La comunicación que asumirán los estudiantes con los Profesores Académicos o Tutores en la Comunidad Virtual de Aprendizaje se establecerá en los módulos de Cursos, Asesoría y Foros, algunos de los módulos mencionados contará con un chat en línea para lograr la interacción entre el profesor o Tutor con el estudiante y de igual manera se podrán dejar mensajes o avisos a los estudiantes y profesores según sea el caso.

Al integrar esta Comunidad Virtual de Aprendizaje en los Institutos Tecnológicos Descentralizados en el estado de Tabasco con el fin de transferir los conocimientos y habilidades formando con ello una Red de Aprendizaje para el mejoramiento académico en los estudiantes de las instituciones participantes y así, contribuir a la formación académica de los estudiantes y profesores

de los institutos mediante la implementación de una Aplicación Web la cual está en desarrollo con un 40 % de avance hasta al momento.

5. Conclusiones y Trabajos Futuros de Investigación

El CVDA ayudará a los estudiantes a mejorar su nivel educativo mejorando las competencias comunicativas orales y digitales. Las actividades deben estar bien diseñadas para alcanzar nuestros objetivos, y el resultado que se obtendrá en los profesores es la aceptación del método comunicativo; el desafío ahora es saber cómo estructurar lecciones y tareas significativas en el marco de la comunidad virtual. La generación del aprendizaje a través del CVDA es la razón primordial en esta investigación, se busca formar un espacio común entre estudiantes y profesores de distintos Tecnológicos en donde los estudiantes puedan acudir a asesorías, cursos o foros con otros profesores diferentes a su institución para profundizar y conocer opiniones distintas de acuerdo a las temáticas que se trate. En la actualidad la dinámica y movilidad del aprendizaje es vital en el desarrollo de las regiones, el CVDA busca motivar en los estudiantes una manera diferente de aprender y conocer opiniones de diferentes profesionistas de su carrera con el afán de progresar, formando redes de investigación y de aprendizaje en un futuro en los Institutos Tecnológicos del país, a través de los estudiantes. Lo que se requiere concluir al final de esta investigación es que los Tecnológicos Descentralizados del Estado de Tabasco puedan trabajar proyectos integrales para el beneficio del estado de Tabasco. Los trabajos futuros del proyecto es evaluar el rendimiento del CVDA en las Instituciones Tecnológicas ya mencionadas en encontrar una nueva forma de trabajo colectivo para el bienestar social, económico y académico del estado de Tabasco. Más adelante está la idea de proponerlo ante la Dirección General de Educación Superior Tecnológica (DGEST) de implementarlo a nivel nacional para integrar proyectos académicos y de investigación para el beneficio del país.

Referencias

- [1] M. Díaz, G. Alejandro. (2004, Noviembre 10). Artículo de los Ambientes virtuales de Aprendizaje a las Comunidades de Aprendizaje en Línea. Revista Digital Universitaria. Consultado el 14 de Septiembre del 2012, http://www.revista.unam.mx/vol.5/num10/art62/nov_art62.pdf.
- [2] A. Lemus, O. Rodríguez, R. Gaytán. Diseño y Operación de las Comunidades Virtuales de Aprendizaje de Ibero On-line. IV

- Encuentro Internacional de Educativo Superior. UNAM 2005.
Del 20 al 24 de Junio 2005. México D.F
- [3] STAIR, Reynolds, (2005) Principios de Sistemas de Información. Editorial Pearson 7ª Edición, México.
 - [4] C. Pérez, (2005) Arquitectura de computadoras, Editorial McGraw Hill 2ª Edición, México.
 - [5] A. Torres Velandia. 2004. RED UAM: Grupos Virtuales de Aprendizaje Colaborativo. Reencuentro: Análisis de problemas universitarios. Universidad Autónoma Metropolitana-Xochimilco. ISSN 0188-168X. P 38-48. Distrito Federal.

Selección de Características para Análisis de Sentimientos Basado en Computación Evolutiva: Resultados Preliminares

Neftalí David Watkinson Medina^a, Carlos Alberto Brizuela Rodríguez^a

^aCentro de Investigación Científica y Educación Superior de Ensenada
Carr. Ensenada-Tijuana 3918, Zona Playitas, Ensenada, B.C., México, 22860.
nwatkins@cicese.edu.mx, cbrizuela@cicese.edu.mx

2013 Published by *DIFU*_{100ci}@ <http://www2.uaz.edu.mx/web/www/publicaciones>
Selection and peer-review under responsibility of the Organizing Committee of the CCOMP-2013, www.cicomp.org

Resumen

En soluciones recientes para el análisis de sentimientos automatizado se han utilizado diversas herramientas para extraer características de un texto que van desde la representación de cada palabra hasta características que combinan información semántica y léxica del texto. Sin embargo, es técnicamente imposible para un clasificador automatizado hacer diferencia entre aquellas características que otorgan información sobre la polaridad del texto y aquellas que no, para lo cual se han diseñado diferentes métodos para realizar selección de características utilizando información estadística. Puesto que la mayoría de estos métodos son el tipo voraz, en muchas ocasiones fallan en eliminar características ruidosas y descartan otras que pueden servir para la clasificación, por esto se han diseñado métodos alternativos que optimicen la tarea de selección, En este trabajo se propone la implementación de un algoritmo basado en Evolución Diferencial, experimentos preliminares muestran que los resultados son competitivos con otras propuestas del estado del arte.

Palabras clave: Evolución diferencial, análisis de sentimientos, selección de características.

1. Introducción

El análisis de sentimientos (también conocido como minería de opiniones) es un problema de clasificación de textos, donde dado un conjunto de clases se busca determinar a cual clase pertenece un texto dado. Sin embargo, a diferencia de otros problemas de clasificación de textos, el análisis de sen-

timientos intenta identificar la información subjetiva que indique la polaridad del documento, que en el caso de clasificación binaria puede ser positiva o negativa.

El análisis de sentimientos dista de ser un problema trivial, en un estudio realizado por [1] se reporta que expertos en clasificación llegan a un 80 % acuerdo entre ellos o precisión (es decir, el 80 % de los documentos fueron clasificados de la misma

manera, esto quiere decir que para un experto el otro clasificó con ese porcentaje de precisión), por lo tanto cualquier sistema automatizado que pueda lograr empatar o sobrepasar este número se considera como bueno.

Existen diversos métodos para la clasificación automatizada, [2] y [3] presentan una lista detallada de los utilizados hoy en día, entre estos se encuentran el uso de clustering, Entropía Máxima, redes bayesianas, SVM (Máquina de Vectores de Soporte), redes neuronales, árboles de decisión, regresión lineal. De estos SVM y redes bayesianas naif (NB) son de los que obtienen mejores resultados de precisión sin sacrificar el costo computacional.

Los clasificadores automatizados no pueden entender el texto directamente como lo hace una persona, por lo cual es necesario representar la información explícita e implícita en un formato entendible para la máquina. A esta tarea se le llama extracción de características, donde se obtiene cada elemento que será utilizado para la clasificación (palabras, patrones de frase, signos de puntuación, etcétera) y dependiendo del clasificador se traducen a un formato con el cual este pueda trabajar (ej. en el caso de SVM los textos se presentan como un vector de índices los cuales apuntan a las características extraídas). En [4] se presenta una lista de las diferentes características que pueden ser extraídas de un texto. En este trabajo nos enfocamos en aquellas que se pueden obtener con herramientas que no dependan de intervención humana como lo son los diccionarios morfo-sintácticos, sino de herramientas de extracción directa de características como lo son n-gramas, signos de puntuación, largo de oraciones y etiquetado morfológico.

La alta dimensionalidad del espacio de características, hace deseable la implementación de una selección previa de las mismas que no sacrifiquen la precisión de clasificación, por lo tanto, una vez obtenidas las características se realiza un proceso de selección previa a la clasificación para eliminar aquellas características de ruido que no solo distan de ser representativas de una clase sino que también afectan negativamente a la clasificación en tiempo y precisión. Las técnicas más comunes de selección de características son aquellas que están basadas en frecuencia de aparición de las características (ej. tf-idf), en información mutua (ej. ganancia de información), y en pruebas de estadística (ej. χ^2 o Chi Cuadrada). La primera se utiliza principalmente para pre-procesar el texto eliminando las características

que por frecuencia de aparición son muy comunes en ambas clases como para determinar la pertenencia a una de ellas (ej. el uso de pronombres y artículos generalmente no es indicativo de que un texto sea positivo o negativo), la segunda utiliza información individual de cada característica y su valor de entropía en un árbol de decisión para determinar si es un buen divisor de clases, y la última aplica pruebas estándar de estadística de valor esperado y valor obtenido para determinar la dependencia que hay entre la presencia de una característica y la pertenencia de un texto a una clase. En [5] se establece una correlación entre estos tres métodos, ya que bajo ciertas condiciones arrojan resultados similares.

En la Sección 2 de este documento se presentan los antecedentes que inspiraron el método que aquí se presenta, en la Sección 3 se describe el método basado en Evolución Diferencial y en la Sección 3.3 se exponen los resultados. Finalmente, la Sección 4 presenta las conclusiones obtenidas a partir de los resultados observados y trabajo futuro para el desarrollo del algoritmo.

2. Trabajo relacionado

El análisis de sentimientos comienza a principios de los años 90 como una tarea manual de clasificación. Sin embargo, a partir del año 2000 se considera el nacimiento de la versión moderna del análisis de sentimientos con el surgimiento de los primeros trabajos que utilizan clasificadores y herramientas automatizadas para la clasificación, trabajos como el de [6], [7] y [8] entre otros vienen a formar la base para el trabajo presente y han inspirado en gran parte los estándares de comparación de resultados (ej. En [7] se define por primera vez el corpus de prueba utilizado por un gran número de trabajos recientes). En esta sección se presenta de manera superficial aquellos trabajos que tienen una relación directa con el Algoritmo de Selección basado en Evolución Diferencial (ASED).

2.1. Corpus de texto

Como ya se menciona al inicio de esta sección, en [3] y [7] se presenta un conjunto de documentos (corpus de texto) que consiste en 2000 críticas de cine en inglés extraídas de la página de internet IMDB y etiquetadas 1000 como positivo y 1000 como negativo (versión 2.0 del corpus). Este corpus ha sido utilizado en sus diferentes versiones en varios trabajos de análisis de sentimientos y facilita la comparación de resultados con

sistemas que no se encuentran disponibles pero que han sido probados con este corpus.

2.2. Selección de características basada en Algoritmos Genéticos

En [4] se presenta el algoritmo EWGA (Algoritmo Genético con Peso de Entropía por sus siglas en inglés). Aunque no es la primera vez que se utiliza el cómputo evolutivo en la clasificación de textos (ver [9]), este es de los primeros trabajos publicados acerca del uso de los mismos para la selección de características para clasificación supervisada.

El EWGA utiliza una representación binaria del individuo donde la posición de los valores en el vector indican el índice de la característica en cuestión y el valor 1 dentro del vector indica que la característica es utilizada y el 0 que esta se descarta. La población es construida al azar y uno de los individuos es generado con el método de ganancia de información con un umbral de 0.0025 y procede a realizar los siguientes pasos:

1. Obtener los pesos de las características utilizando la función de ganancia de información (IG)
2. Incluir las características elegidas de acuerdo a un umbral de IG para formar el primer individuo de la población y los $n-1$ individuos restantes generarlos de manera aleatoria.
3. Evaluar y seleccionar las soluciones basándose en la función de aptitud
4. Cruzar los pares de soluciones de tal forma que se maximice la diferencia total de IG entre las dos soluciones
5. Mutar las soluciones basándose en el peso IG de la característica para definir la probabilidad de mutación dónde el valor de IG es la probabilidad de que mute de 0 a 1 y $1-IG$ de que mute de 1 a 0
6. Repetir los pasos 3-5 hasta alcanzar el criterio de paro

La aptitud del individuo es la precisión de clasificación obtenida utilizando el método de 10 pliegues (10-fold) con un clasificador SVM con un kernel lineal, en [7] y en [2] se halla más información con respecto al papel que tienen los diferentes tipos de kernel en un SVM. Cabe notar que con este método se realizan $g \times n$ clasificaciones donde g es el número de generaciones y n el número de individuos en la población. Además, el paso 4 requiere del cálculo de la diferencia de IG cuyo costo será en el orden del largo del vector. Sin embargo, la selección de características se considera un problema fuera de línea, por lo que el costo computacional no es

tan relevante siempre y cuando sea razonable a consideración de quien lo esté implementando. Otro detalle del EWGA es que en el paso 5 se van a descartar todas las características cuyo valor de IG es 0. Aunque no hay una justificación formal del porque se deban considerar las características que no tienen ganancia de información, tampoco lo hay para no hacerlo.

2.3. Comparación de métodos del estado del arte

En la Tabla 1 se muestran los resultados comparativos de distintos métodos de selección de características, donde Base se refiere a la clasificación sin selección de características, IG es utilizando el método de Ganancia de Información con un umbral de 0.0025, GA son los resultados del Algoritmo Genético de [4] sin utilizar Ganancia de Información, SVMW es utilizando el método de Máquina de Vectores de Soporte con pesos (véase [1]) y EWGA es el algoritmo descrito por [4] con una población de 50 individuos por 200 generaciones. Los resultados de [8], con grupos de apreciación, y de [7] se obtuvieron sobre una versión previa del corpus de 1300 críticas de cine. Los resultados de EWGA parecen ser muy favorables ya que no solo tiene buena precisión sino que también disminuye de manera efectiva el número de características totales sobre el cual se realiza la clasificación. Sin embargo, EWGA aún no ha sido probado con texto nuevo que no haya formado parte del entrenamiento, cabe notar que el uso de SVM de 10 pliegues ayuda a determinar el desempeño de un modelo ante texto que no formó parte del entrenamiento, pero por el diseño del algoritmo en el que la clasificación se realiza un número de veces y esta rige la aptitud del individuo, es imposible conocer si el algoritmo realmente está mejorando la precisión para un caso general o solo para los documentos que están siendo evaluados dentro del mismo. Es decir, es necesario probar la precisión del método con texto que no formó parte de las evaluaciones de aptitud.

Técnica	Precisión	No. Características
Base	87.95 %	26,870
IG	92.50 %	2,316
GA	92.55 %	2,017
SVMW	92.86 %	2,000
EWGA	95.55 %	1,752
Whitelaw [8]	90.20 %	–
Pang [4]	87.20 %	–

Tabla 1. Comparación de métodos

3. Selección de características basada en Evolución Diferencial

En [10] se describe la heurística de Evolución Diferencial, la cual es una versión modificada de Algoritmos Genéticos. La diferencia principal consiste en que la mutación se realiza antes del cruce, y utiliza una mutación definida por:

$$u_i(t) = x_{i_1}(t) + \beta(x_{i_2}(t) - x_{i_3}(t)) \quad (1)$$

Esta consiste en agregar al valor de un individuo tomado al azar ($x_{i_1}(t)$), la diferencia de los valores de otros dos individuos ($x_{i_2}(t) - x_{i_3}(t)$) multiplicado por un factor de escalamiento (β) para generar un vector mutado llamado vector de prueba ($u_i(t)$), que será posteriormente utilizado para cruzarlo y generar el individuo hijo. La variable t indica el índice dentro del vector del individuo.

Originalmente fue diseñado para problemas de optimización continuos, sin embargo, en [9] se presentan distintas técnicas para aplicar Evolución Diferencial a problemas de permutación, una de esas técnicas es el de modificar ligeramente la fórmula de mutación para obtener solo valores enteros y que sean válidos para la permutación al no repetir resultados dentro de un mismo vector. Puesto que en el algoritmo para selección de características, solo importa mantener los valores dentro de un intervalo válido que va de 1 al número total de características, y que los valores sean enteros. La repetición de índices se puede lidiar de manera externa eliminando las repeticiones de un índice previo a la clasificación.

3.1. Representación del individuo

Una de las diferencias entre este algoritmo y el propuesto por [1] es la manera en la que se representa el cromosoma del individuo. Una de las desventajas observadas en la representación binaria para este problema en específico es que mientras más características son descartadas, estas serán sustituidas por un 0 binario, lo cual producirá cromosomas con la mayoría de los genes apuntando a características que no se van a utilizar. Para tratar de solucionar este problema, se ha propuesto una representación de números enteros con longitud variable, cada vector o cromosoma incluirá los índices directos de cada característica (Figura 1). No es la primera vez que se utilizan vectores de longitud variable en algoritmos genéticos o evolución diferencial (ver [12] y [13]), además esto no solo produce vectores de menor tamaño sino que puesto a que la posición del gen no



Figura 1. Representación del individuo

influye en el índice de la característica, permite realizar ordenamiento aleatorio (esto es cambiar el orden en que aparecen los índices dentro del vector) y otras operaciones que permiten prevenir la convergencia prematura del algoritmo. Existe la probabilidad de que dos o más individuos tengan los mismos índices pero en diferente orden, se espera compensar esto con la capacidad de exploración del cruce, donde dos individuos iguales puedan moverse a puntos distintos, además se incluyeron mecanismos que procuran mantener la diversidad en la población agregando características aleatorias a individuos similares.

3.2. Algoritmo basado en Evolución Diferencial

El algoritmo ASSED realiza los siguientes pasos:

1. Inicializar la población de manera aleatoria y añadir un individuo generado con ganancia de información.
2. Evaluar la aptitud de cada individuo
3. Realizar la mutación de los individuos
4. Realizar el cruzamiento entre los individuos y su respectivo vector de prueba
5. Evaluar la aptitud de los individuos hijo
6. Elegir entre el hijo y el padre aquél que tiene una aptitud más alta para la siguiente generación
7. Repetir hasta alcanzar el criterio de parada o número de generaciones

En el Algoritmo 1 se observa el pseudocódigo del método propuesto. En este algoritmo los parámetros de entrada son el número de individuos en la población (n), número de generaciones (Gen), probabilidad de cruzamiento (p_c), el valor del factor de escalamiento (β), y el umbral de Ganancia de Información ($delta$).

1 Algoritmo: DifferentialEvolution($n, Gen, p_c, \beta, \delta$)

Entrada: Número de individuos (n), número de generaciones (Gen), probabilidad de cruce (p_c), valor de escalamiento (β) y umbral de Ganancia de Información (δ)

Salida: Individuo con mejor aptitud de clasificación

```

t ← 0;
InitializePopulation(n, C(0));
f(C(0)) ← EvaluatePopulationFitness(C(0))
while t < Gen do
  for  $x_i \in C(t)$  do
     $u_i \leftarrow$  Mutate(C(t),  $x_i, \beta$ );
     $x'_i \leftarrow$  Crossover( $u_i, x_i, p_c$ );
     $f(x'_i) \leftarrow$  EvaluateFitness( $x'_i$ );
    if  $f(x'_i) > f(x_i)$  then
       $C(t+1) \leftarrow C(t+1) \cup \{x'_i\}$ 
    else
       $C(t+1) \leftarrow C(t+1) \cup \{x_i\}$ 
    end if
  end for
end while
return BestFitness(f(C(t)))
    
```

En la línea 2, la población es inicializada tomando índices al azar del conjunto de características. Esto puede tener como consecuencia la generación de individuos de distintas longitudes. En la línea 3 se añade un individuo extra que es construido utilizando un filtro basado en ganancia de información con un umbral δ . El método EvaluatePopulationFitness() (línea 4) obtiene la aptitud de cada individuo de la generación 0 haciendo pruebas de clasificación. En las líneas 5 a 16 se realiza la mutación (línea 7) y cruzamiento (línea 8). Además, el método EvaluateFitness() (línea 9) obtiene la aptitud de un solo individuo, mientras que el método BestFitness() (línea 17) obtiene el individuo con la mejor aptitud obtenida en una generación y todo se repite hasta terminar el número determinado de generaciones.

3.3. Mutación

Para la mutación se utiliza el procedimiento descrito en esta misma sección. Para enfrentar el problema de individuos de diferentes longitudes, el factor de escalamiento (β) es multiplicado o dividido por 10 según sea el caso para lidiar con los índices que rebasan al individuo de menor longitud. En la Figura 2, se muestra un ejemplo gráfico de lo que sucede con el vector de prueba, donde **a** incluye los valores que fueron obtenidos sin modificación a la ecuación, **b** se obtuvieron multiplicando por diez el factor de escalamiento para compensar la falta del primer valor y **c** son los valores que fueron

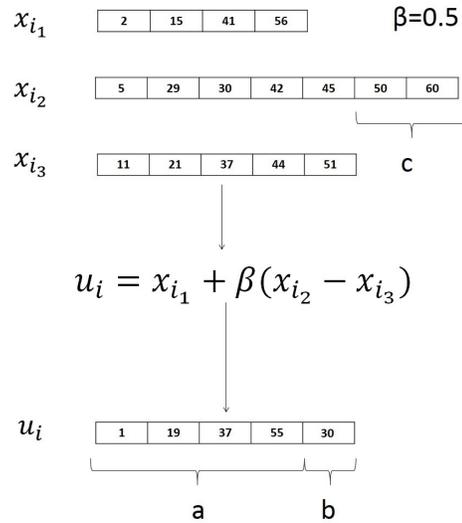


Figura 2. Mutación

descartados como excedente.

3.4. Cruzamiento

El cruzamiento se realiza en base al valor de ganancia de información. En la Figura 3, se observa un ejemplo de cruzamiento con probabilidad de 0.7 de cruce. El elemento con mayor ganancia de información tiene una probabilidad p_c de ser añadido al individuo contra $1-p_c$ del que tiene menor ganancia. Si ambos valores son iguales tienen una probabilidad de 0.5, y si hay diferencia de longitud entre el vector de prueba y el vector padre, la probabilidad de agregar los elementos sobrantes es de 0.5.

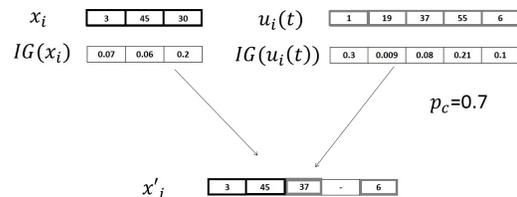


Figura 3. Cruzamiento

4. Evaluación

Los detalles para la experimentación son los siguientes:

- Lenguaje de programación: Java JDK (versión 7.21)
- Plataforma de desarrollo: Eclipse Juno (4.2.2)
- Sistema operativo: Windows 7 46 bits SP1
- Procesador: Intel Core i5 2.67GHz
- Memoria RAM: 6 GB
- Modelo: Asus U43F

4.1. Experimentos

Debido a que el algoritmo requiere de un gran número de clasificaciones para obtener el valor de aptitud, y el tiempo de clasificación aumenta exponencialmente con el tamaño del corpus se extrajeron distintos conjuntos del corpus de críticas de cine para realizar las pruebas. En todas las pruebas se utilizaron n-gramas hasta grado 3 con umbral de aparición de 5 para las características. Primero se realizó una prueba con 300 críticas de cine (150 positivas, 150 negativas) para encontrar el valor óptimo de β . Para mantener los parámetros lo más cercano posible a los empleados en [1], las pruebas se realizaron con un umbral de ganancia de información $\delta=0.0025$, con una población de 20 individuos con un límite de 200 generaciones. En la primer prueba (Figura 4) se hicieron 20 corridas variando el valor de β de 0.1 a 2 y el valor de precisión más alto obtenido fue de 98 % cuando la precisión base fue de 67 % y ganancia de información obtenía una precisión de 95.6 %. La mejoría es consistente cuando $1 < \beta < 2$. El número base de características es de 20726 entre unigramas, bigramas y trigramas. Ganancia de información obtiene 14000 y el algoritmo basado en evolución diferencial obtiene 426 características.

La tercera prueba se realizó con 500 críticas de cine (250 positivas, 250 negativas), precisión base de 82 %, el resultado obtenido fue de 99.8 % de precisión con 3566 (de 43,521 de base) características con $\beta=1.5$.

Método	Precisión
Base	73.33 %
IG	73.33 %
ASED	80 %

Tabla 2. Comparación preliminar

En la Tabla 2 se presentan los resultados de una cuarta prueba con el corpus de 300 documentos donde se extrajeron 30 de estos para que no formaran parte de la selección, al finalizar el algoritmo se clasificaron estos

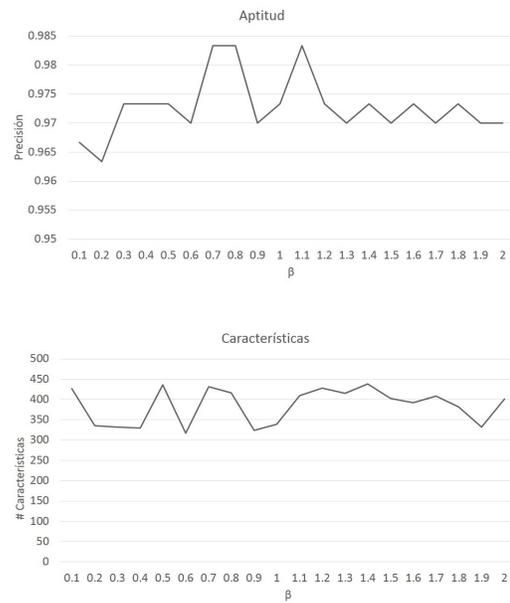


Figura 4. Resultados de primera prueba

30 documentos utilizando el modelo construido por los 270 restantes. Cuando no hay selección ninguna (Base) se obtiene una precisión de 73.33 % utilizando el modelo de 270 documentos para clasificar a 30 restantes. El filtro basado en ganancia de información obtiene una solución similar, mientras que la lista de características obtenida por el algoritmo logra el 80 % de precisión.

4.2. Comparaciones

Con el fin de comparar este método con otros, la quinta y última prueba se realizó con todo el corpus (2000 críticas de cine), 53.425 características de base con precisión de 81.3 % y al final se obtuvieron 2447 características con un 93.25 % de precisión. Aunque aún es necesario realizar más pruebas, si comparamos este resultado con el de la Tabla 1, el método aquí descrito no dista mucho del propuesto por [1] y utiliza una población de menor tamaño lo que se traduce en menor número de clasificaciones. Sin embargo, este resultado no es directamente comparable puesto que no se ha obtenido el mismo número de características base ni de porcentaje de precisión reportado en la Tabla 1.

5. Conclusiones y trabajo futuro

Los resultados obtenidos muestran que el sistema logra mejorar la precisión obtenida por el método es-

estadístico que en este caso es el basado en ganancia de información. Además, requiere de un menor número de iteraciones y de individuos en la población que el algoritmo EWGA propuesto por [1]. Aún es necesario realizar más pruebas con el mismo corpus de texto y el mismo número de características para poder concluir si el método basado en evolución diferencial es mejor que el basado en algoritmos genéticos. Puesto que los resultados dependen fuertemente de la precisión inicial de ganancia de información, hace falta probar este método integrando otros del tipo estadístico como lo es χ^2 . La representación del individuo como un vector de números enteros de longitud variable combinado con las operaciones de mutación y cruzamiento y el uso de ordenamiento aleatorio permiten explorar índices que no están presentes en la población, aun cuando los individuos son muy similares, lo cual previene la convergencia prematura. Aún es necesario probar este método y otros del estado del arte con documentos desconocidos por el clasificador.

Referencias

- [1] T. Wilson, J. Wiebe, & P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis." In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05. Morristown, NJ, USA: Association for Computational Linguistics, 2005, pp. 347-354
- [2] P. Manning, C. D. Raghavan, & H. Sch  tze An Introduction to Information Retrieval 1st Edition. Cambridge University Press, New York, NY, 2008.
- [3] A. Westerski, Sentiment Analysis: Introduction and the State of the Art overview. Universidad Politecnica de Madrid, Espa  a, 2007, pp 211-218
- [4] A. Abbasi, H. Chen, & A. Salem, "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums." In ACM Transactions on Information Systems, 2008, pp. 1-34.
- [5] A. Ginsca, E. Boros, A. Iftene, D. Trandabat, M. Toader, M. Corici, D. Cristea, "Sentimatrix: multilingual sentiment analysis service." In Proceedings of the second Workshop ACL - WAS-SA, 2011, pp. 189-195.
- [6] B. Pang, & L. Lee, "Opinion Mining and Sentiment Analysis". Foundations and Trends in Information Retrieval, 2006, pp 91-231.
- [7] A. Go, R. Bhayani, & L. Huang, "Twitter sentiment classification using distant supervision." CS224N Project Report, Stanford. Stanford, California, 2009
- [8] C. Whitelaw, N. Garg, & S. Argamon, "Using Appraisal Groups for Sentiment Analysis." In Proceedings of the 14th ACM Conference on Information and Knowledge Management, 2005, pp. 625-631.
- [9] A. Abraham, S. Das, & A. Konar, "Document Clustering Using Differential Evolution." 2006 IEEE International Conference on Evolutionary Computation, 2006, pp. 1784-1791.
- [10] R. Storn, & K. Price, "Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces." Journal of global Optimization, 1997, pp 341-359.
- [11] G. C. Onwubolu, & D. Davendra, Differential Evolution: A Handbook for Global Permutation-Based Combinatorial Optimization. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009
- [12] B. Pang, & L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts." Proceedings of the Association for Computational Linguistics, 2004, pp 271-278.
- [13] Yang, Y., & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. In International Conference on Machine Learning. Retrieved from <http://faculty.cs.byu.edu/~ringger/Winter2007-CS601R-2/papers/yang97comparative.pdf>

Sistema Multi-agente para la Detección de Fraudes en el Correo Electrónico

Christian J. Lucero^a, Perla J. Castro^a, María de Guadalupe Cota^a, Pedro Flores Pérez^a, Juan P. Soto Barrera^a

^aUniversidad de Sonora

Bld. Luis Encinas y Rosales S/N, Colonia Centro, CP 83000 Hermosillo, Sonora, México.

[[crizlucero](mailto:crizlucero@gmail.com), [janethkasztro](mailto:janethkasztro@gmail.com)]@gmail.com, [[lcota](mailto:lcota@gauss.mat.uson.mx), [pflores](mailto:pflores@gauss.mat.uson.mx), [jpsoto](mailto:jpsoto@gauss.mat.uson.mx)]

2013 Published by *DI²U₁₀₀ci* @ <http://www2.uaz.edu.mx/web/www/publicaciones>

Selection and peer-review under responsibility of the Organizing Committee of the CICOMP-2013, www.cicomp.org

Resumen

En este trabajo se presenta un sistema multi-agente para la detección de amenazas y fraudes a través del correo electrónico. Esta propuesta utiliza la tecnología de agentes software para identificar aquellos correos considerados como amenazas de fraude para los usuarios. Para esto, los agentes hacen uso de reglas tipo Prolog con el fin de detectar patrones que permitan identificar correos electrónicos maliciosos, y cuya intención sea obtener información confidencial de los usuarios mediante el uso de phishing y bots.

Palabras clave: Análisis de Contenido, Bots, Correo electrónico, Multi-agente, Phishing..

1. Introducción

En la actualidad el uso del correo electrónico es uno de los medios de intercambio de información más utilizados por los usuarios en internet. Por lo tanto, también es uno de los medios más utilizados por los crackers o maleantes cibernéticos para engañar a las personas y vulnerar sistemas. Existen varios métodos de robo de información, siendo uno de ellos el phishing, cuyo objetivo es obtener información confidencial de las personas a través de engaños [1]. Otro método comúnmente utilizado son los bots (abreviatura de robots), el cual es un programa informático que simula a un ser humano para interactuar con las personas y estafarlas [2]. Dicho programa envía archivos infectados

a través del correo electrónico. Con el fin de buscar una alternativa que contribuya a disminuir dicha problemática, en este trabajo se propone un sistema multi-agente para la detección de amenazas en el correo electrónico. En la Sección 2, se describen dos de los métodos más utilizados en el correo electrónico para engañar y estafar a las personas: phishing y bots. En la Sección 3, se modela el sistema multi-agente, para esto, se utiliza la metodología INGENIAS. Por otro lado, en la Sección 4 se describe el desarrollo e implementación del sistema propuesto. Por último, en la Sección 5 las conclusiones son presentadas.

2. Amenazas de estudio

En la actualidad podemos encontrar diversos tipos de programas utilizados para estafar a las personas a través de sistemas de gestión de correos electrónicos, por ejemplo: Adware, Phishing, Spyware, Riskware, Bots etc., así como métodos basados en ingeniería social, tales como el Phishing. Este trabajo se centra en identificar amenazas basadas en el phishing con ayuda de los bots. Es por esto que en los siguientes apartados se aborda más a detalle cada uno de estos temas.

2.1. Phishing

El termino phishing se empezó a definir alrededor de 1995, y consiste en enviar correos electrónicos con el fin de obtener información personal del usuario, para posteriormente utilizarla en algún tipo de fraude [3]. Por ejemplo, el contenido de un correo suelen incluir un enlace a un sitio conocido pero re-direccionando a una página web falsa. De esta manera, el usuario se confía y proporciona información personal a un sitio web falso en donde se realiza la estafa [4]. Los phishing más conocidos son las “estafas nigerianas”. Este tipo de estafas tienen al menos 5 variantes [5]:

- Estafa basada en el depósito de una fuerte cantidad de dinero.
- Estafa por lotería.
- Estafa por herencia de familiar desconocido.
- Estafa de un prisionero de guerra.
- Estafa de venta de celulares o recargas de tiempo aire.

2.2. Bots

Un bot (abreviatura de robot) es un programa o script que realiza tareas automáticamente. Una de estas tareas consiste en imitar el comportamiento del ser humano al intercambiar información con otras personas. Por lo general, los bots se utilizan para enviar correos electrónicos a diversos usuarios, creando continuidad en el intercambio de mensajes, y haciendo creer al usuario que está interactuando con una persona real [2]. Este tipo de fraude es conocido como “Fraud bots”, lo cuales envían correos de manera insistente hasta lograr engañar a los usuarios y así obtener información personal y bancaria, ofreciéndoles falsos premios, herencias, etc., [6]. Debido a que este tipo de fraudes representa

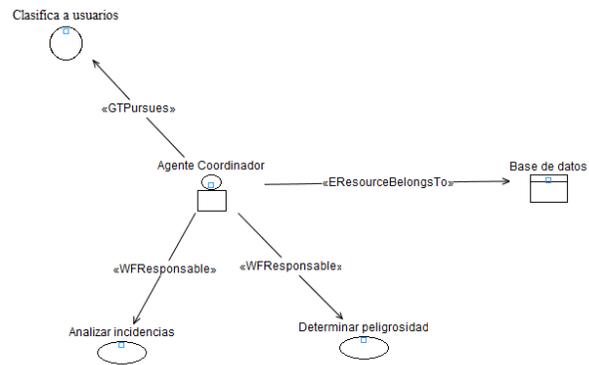


Figura 1. Agente Coordinador.

uno de los métodos menos costosos para las asociaciones delictivas, han logrado convertirse en los ataques más utilizados para engañar a las personas a través del correo electrónico.

3. Modelo del Sistema Multi-agente

Como una alternativa para proteger a los usuarios de las amenazas descritas en la sección anterior, en este apartado se presenta el modelo del sistema multi-agente propuesto siguiendo la metodología de INGENIAS [7]. El sistema se compone de un conjunto de agentes los cuales interactúan y realizan tareas enfocadas en proteger al usuario del correo electrónico de posibles ataques. Para esto, se cuenta con un agente encargado de analizar correos electrónicos entrantes. Otro agente que de acuerdo a las incidencias del remitente decide mover aquellos correos considerados como posibles amenazas a la carpeta de “No deseados”. Y por último, un agente que proporcione seguridad en el envío de mensajes entre los agentes. A continuación se describe a detalle cada uno de estos agentes:

- Agente coordinador. Es el agente encargado de coordinar a los agentes que intervienen en el sistema. Además determina si el remitente del correo electrónico entrante es una amenaza para el usuario. Para esto, el agente coordinador analiza las incidencias presentadas y determina la peligrosidad de las mismas (ver Figura 1).
- Agente local. Es el agente encargado de detectar aquellos correos identificados como phishing (Figura 2).

Para esto, el agente obtiene los correos entrantes y analiza el contenido del mensaje recibido. Dicho análisis

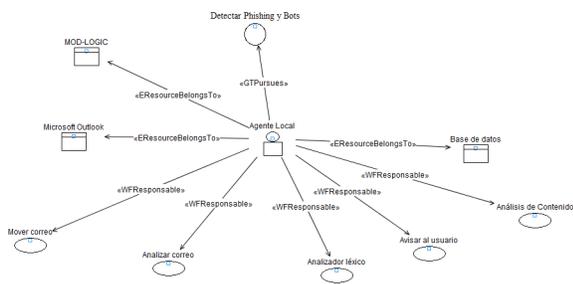


Figura 2. Agente Local.

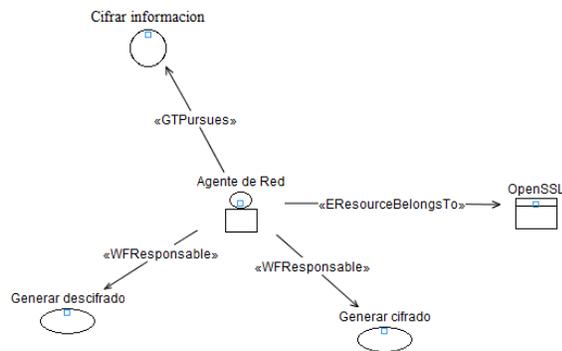


Figura 3. Agente de Red.

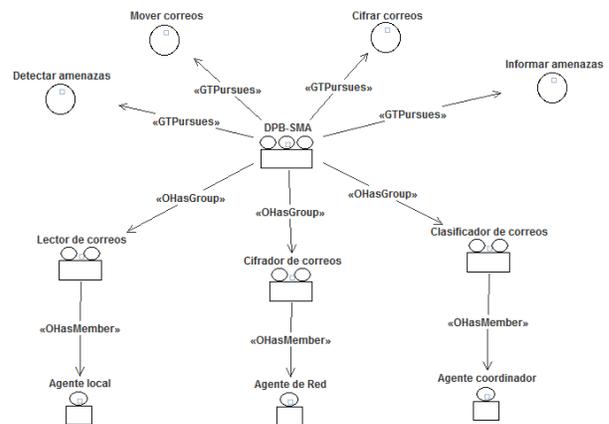


Figura 4. Modelo de la Organización.

permite al agente enviar alertas al usuario en caso de que detecte una amenaza. Para realizar el análisis de las palabras, el agente local utiliza la aplicación MOD-LOGIC.

- Agente de red. Es el agente encargado de proveer la seguridad durante el intercambio de información entre el agente local y el agente coordinador (ver Figura 3).

Una vez descritos los agentes que componen el sistema, a continuación se ilustra el modelo de la organización los agentes (ver Figura 4).

Por cuestiones de espacio el modelo del entorno y el de objetivos y tareas del sistema, se explican en [8].

4. Prototipo

El prototipo propuesto fue implementado en .NET, con el uso de dll's para acelerar el proceso de análisis.

Detalles de implementación:

Como parte de la programación del conocimiento de los agentes, el sistema utiliza reglas tipo Prolog y técnicas de análisis de contenido (apartado 4.1). Por otra parte, y con el fin de proteger la información que se

intercambia a través de la red se utilizarán las librerías OpenSSH [9] y OpenSSL [10]. Además se diseñó un analizador léxico (tokenizador), el cual descompone la frase en tokens para un manejo más sencillo en el análisis de contenido (apartado 4.2), una base de datos para el control de sesiones de agentes, un glosario con una clasificación ontológica de las palabras que son consideradas con cierto grado de peligrosidad para el usuario, y una pizarra para el control de ocurrencias de aquellos eventos considerados como amenazas para la seguridad del sistema. La base de datos almacena las palabras claves consideradas “peligrosas”, así como el registro de las amenazas en un historial.

4.1. Análisis de Contenido

El análisis de contenido es un conjunto de técnicas utilizadas por los agentes para distinguir el significado simbólico de los mensajes. Estos mensajes, por lo regular, no tienen un único significado, ya que en ocasiones cambia su semántica según el contexto en que se presenta cierta información. Por lo tanto, el objetivo del análisis de contenido consiste en realizar una inspección en las frases para identificar aquellas que representen un peligro para el sistema o para el usuario. Para esto, el sistema considera aquellos casos y frases que por lo general son utilizadas para cometer fraudes. A continuación se describen algunos de los casos considerados, así como las frases asociadas a cada uno de ellos (ver Tabla 1).

Existe una infinidad de casos que se pueden abordar, sin embargo, en este prototipo sólo se consideraron los casos descritos anteriormente, ya que es importante reconocer que nuestro idioma es muy extenso y por lo tanto su análisis muy complejo. Como resultado del estudio realizado de las frases consideradas como ame-

Tabla 1. Casos y frases fraudulentas utilizadas en los contenidos de los mensajes de correo electrónicos.

Caso	Frases fraudulentas
Solicitud de información personal	<p>¿Cómo se llama la calle dónde vives?</p> <p>Te mandaré un pequeño presente, a dónde te lo envío?</p> <p>Creo que me quitaron mi cuenta del Facebook, préstame la tuya.</p> <p>Te has ganado un premio ingresa tu número de cuenta.</p>
Solicitud de información bancaria	<p>Has sido uno de los 5 seleccionados del Sorteo Tec 2012, mándanos tu número de cuenta para que puedas disfrutar tu regalo.</p> <p>Necesitamos tu NIP para saber si ganaste la rifa.</p>
Avisos falsos sobre premios y ofertas	<p>Usted ha sido especialmente seleccionado para escuchar esta oferta.</p> <p>Si compra nuestro producto, obtendrá un maravilloso premio o bono gratis.</p> <p>Usted ha ganado uno de cinco valiosos premios.</p> <p>Usted ha ganado una importante suma de dinero en un sorteo de lotería extranjera.</p>

Tabla 2. Palabras claves.

Adorable	Crédito	Fabuloso	Mandar
Agradable	Creer	Facebook	Metroflog
Alta	Crei	Fantástico	Mirar
Ardiente	Cuenta	Feo	Mostrar
Asombroso	Débito	Foto	Muestro
Banamex	Decir	Fraude	MySpace
Banco	Depositar	Ganar	Necesitar
Bancaran	Desagradable	Gracioso	NIP
Banorte	Dice	Gratis	Nómina
Bello	Dinero	Grotesco	Número
Bonito	Elevado	Hermoso	Oferta
Buena	Eliminar	Horrible	Pagar
Cargamos	Encantador	Increible	Pensar
Caro	Enlace	Ingresar	Perfil
Checar	Enviar	Lindo	Piensa
Comprar	Error	Lotería	Premio
Correcto	Especial	Lujoso	Santander
Costoso	Etiquetar	Mal	Scotiabank
Twitter	Suscripción	Suba	Seleccionar
Valioso	Taggear	Sube	Sorprendente
Ver	Tarjeta	Subir	Sorteo
Vivir	Tuenti	Súper	

nazas de fraude, se hizo una selección de palabras que proporcionan información significativa en la interpretación del contenido del mensaje, con el fin de detectar patrones que permitan clasificar los mensajes considerados como amenazas o riesgos para los usuarios (Tabla 2).

Debido a que una oración se compone de sustantivos, verbos y adjetivos, las palabras fueron clasificadas siguiendo dicho esquema, para después obtener su núcleo y cambiar cada letra por un símbolo, de esta manera, se evitan problemas de repetición de palabras. En la Tabla 3 se enlistan los verbos, sustantivos y adjetivos seleccionados. La lista puede variar dependiendo del tiempo de conjugación presentado en las frases.

Además se agregaron las palabras con mala ortografía, ya que es común que este problema se presente en la mensajería instantánea.

4.2. Tokenizador

Este módulo es el encargado de separar las palabras, para esto, el tokenizador selecciona cada palabra para codificarla utilizando una simbología previamente establecida (Tabla 4) y la compara con las contenidas en la base de datos. Por medio de consultas determina si la palabra pertenece al conjunto de verbos, sustantivos o

adjetivos. Después de esto, las palabras son ordenadas para después ser evaluadas con reglas tipo Prolog, y así analizar los mensajes para determinar la peligrosidad del mismo. Se debe de tener en cuenta el orden en el cambio de los caracteres, por ejemplo: en la palabra **quiere**, su codificación debe ser la siguiente **6e-e**, pero al tomar el orden de la tabla sería de la forma **qu4e-e**, lo que daría una codificación incorrecta y ocuparía más espacio del necesario.

La búsqueda de las coincidencias con las tablas de la base de datos es considerando la inserción de las palabras codificadas pero en su raíz y no como palabra completa, esto es para evitar la existencia de los mismos datos en donde sólo cambian unos pocos caracteres. Por ejemplo, en la tabla 5, se muestran las palabras que contendrá la base de datos (como son originales, en su raíz léxica y codificada a la simbología propuesta) para algunos de los adjetivos.

En el siguiente algoritmo describe la acción a realizar por el tokenizador.

1. Función token(frase)
2. //Comentario: Función que separa las palabras y las busca en el diccionario.//
3. Obtiene la primera palabra;
4. Hacer
5. Convertir a minúsculas y omitir acentos de la
6. palabra;
7. Codificar la palabra a la simbología propuesta;
8. Buscar la palabra codificada en el diccionario;
9. Si encuentra palabra entonces
10. Aumenta contador;
11. Fin si
12. Mientras (obtiene la siguiente palabra)

Tabla 3. Palabras claves clasificadas.

Verbo	Sustantivo	Adjetivo
Bañearan	Banamex	Adorable
Cargamos	Banorte	Agradable
Checar	Crédito	Alta
Comprar	Cuenta	Ardiente
Creer	Débito	Asombroso
Creí	Dinero	Bello
Decir	Facebook	Bonito
Depositar	Foto	Buena
Dice	Lotería	Caro
Enviar	Metroflog	Correcto
Etiquetar	MySpace	Costoso
Eliminar	NIP	Desagradable
Ganar	Nómina	Elevado
Ingresar	Suscripción	Encantador
Mandar	Premio	Especial
Mirar	Santander	Fabuloso
Mostrar	Scotiabank	Fantástico
Muestro	Sorteo	Feo
Piensa	Tarjeta	Gracioso
Seleccionar	Tuenti	Gratis
Subir	Twitter	Grotesco
Sube	Banco	Hermoso
Suba	Fraude	Homble
Taggear	Número	Increible
Ver	Enlace	Lindo
Error	Oferta	Luioso

Tabla 4. Simbología establecida.

Sustituto	Letras o Sílabas
0	b, v
1	ca, ci, si, se, zi, ze
2	d, de
3	g, j, gu
4	i, y, ll
5	ca, co, cu, ka, ko, ku
6	qui, qi, ki, que, qe, ke, q, k
7	t, te
8	u, w
9	x, cs, xh, cc, sh, zh, ch
*	sa, so, su, za, zo, zu
^	s, es, z, ez
+	pe, p
?	N
-	f, ff
Se omite	H

Tabla 5. Tabla de adjetivos

Id adjetivo	palabra	original
0	a2o-a0le	Adorable
1	a3-a2a0le	Agradable
2	al7a	Alta
3	a-24en7	Ardiente
4	a*m0-o	Asombroso
5	0e4	Bello
6	0on47	Bonito
7	08en	Buena
8	5-	Caro
9	5-ec7	Correcto
10	5^7o	Costoso

4.3. Programación del Conocimiento

Para realizar la programación del conocimiento se utilizó *ProLog*, ya que por su forma de abstracción de objetos, permite evaluar diferentes premisas para tener una solución del problema rápido y exacto. Además se utiliza la librería *MOD-LOGIC*. Dicha librería es la que permite analizar los patrones para saber si en realidad el correo es amenazante. A continuación se muestra un ejemplo de las reglas *ProLog* en el sistema utilizando la librería *MOD-LOGIC*.

1. SEMÁNTICA DE PREDICADOS:
2. palabra(Word, número).
3. REGLAS:
- 4.
5. DATOS
6. palabra(ado-a0le,0).
7. palabra(al7a,2).
8. palabra(c-e247,100).

9. palabra(24ne-,101).
10. palabra(9e5,202).
11. palabra(2+o17,207).

Como se muestra en el código anterior, la búsqueda de las palabras se realiza a base de una plantilla, la cual cuenta con tres secciones:

- **Semántica de Predicados:** En esta sección se definen los significados de los argumentos de cada predicado incluido.
- **Reglas:** En este apartado se define un conjunto de reglas lógicas para determinar la peligrosidad del correo.
- **Datos:** Permite registrar datos relacionados con los hechos que forman parte de la base de conocimiento. Estos datos están conformados por dos argumentos: el primero es la palabra "neutra", y el segundo es el código de identificación dentro del sistema.

Al finalizar el análisis, se puede identificar si el correo electrónico en realidad es dañino. Esto es para identificar al remitente y tener alertado al usuario sobre lo ocurrido. Cabe mencionar que en esta versión del sistema, la sección de reglas no está descrita en la plantilla para MOD-LOGIC, debido a que la librería no proporciona todavía las estructuras condicionales, las cuales equivaldrían a usar "if-then". Por lo tanto, para el desarrollo de esta versión, la sección de reglas fueron programadas en C++. Por ejemplo, el análisis de la palabra **costosísimo**, primero se debe de tener codificada de la siguiente forma: **5⁷o11mo**. Una vez encontrado el tipo de la palabra, se busca en su estado neutral, que sería **5⁷o***, se obtendría el código de identificación, el cual será **10**. Una vez con el código de identificación, se analizarán en las reglas donde se harán las conjugaciones para determinar si en realidad el uso de esa palabra es amenazante.

4.4. Manejador de Pizarra

El manejador de pizarra es el módulo encargado de llevar un registro de los eventos generados en el sistema con el fin de:

- Tener un seguimiento de los mensajes que se consideren peligrosos
- Registrar a los usuarios que envían mensajes considerados peligrosos

5. Conclusiones

Una de las razones principales por las que consideramos muy importante nuestra propuesta fue porque aun existiendo muchos tipos de aplicaciones para la detección de malware, *bots* y *phishing*, éstas no son por completo seguras ya que sólo registra las actividades de las que tiene conocimiento y no de patrones que tienen algunas aplicaciones dañinas para los sistemas o los usuarios. El sistema propuesto se enfoca a las amenazas llamadas *bots* y *phishing*, ya que son las amenazas que más impactan a los usuarios, al robar su identidad y generar fraudes sin que los usuarios se percaten de lo que pasa. Por tal motivo en este artículo se describe el modelo y desarrollo de un prototipo que analiza los correos que reciben los usuarios, con el objetivo de encontrar la verdadera intención del mensaje recibido en un correo electrónico. Dicho prototipo constituye una propuesta de solución para la seguridad en los correos electrónicos, en el ambiente Windows-NT, con la herramienta de Microsoft Outlook 2010.

Referencias

- [1] Panda Security, 2012, Phishing, 20 de Enero 2012, <http://www.pandasecurity.com/spain/homeusers/security-info/cybercrime/phishing/>.
- [2] ALEGSA, 2011, Definición de Bot - ¿Qué es un Bot?, 20 de Enero 2012, <http://www.alegsa.com.ar/Dic/bot.php>
- [3] Phishing.org, 2012, URL: <http://www.phishing.org/history-of-phishing/>.
- [4] Asociación de Internautas, 2011, Guía rápida para reconocer páginas web falsas que simulan entidades bancarias, Disponible: <http://seguridad.internautas.org/html/863.html>. Último acceso: 06 Septiembre 2011.
- [5] Phishing activity trends report, 2005, Phishing activity trends report Anti-phishing working group, technical report. Disponible en: <http://www.Antiphishing.org/reports/apwgreportjan2006.pdf>
- [6] J. Rodríguez, 2012, Tipos y funcionamiento de bots maliciosos usados por los ciberdelincuentes, 21 Mayo 2012., <http://www.xatakaon.com/seguridad-en-redes/\tipos-y-funcionamiento-de-bots-maliciosos-\usados-por-los-ciberdelincuentes>.
- [7] J. Gómez-Sanz, 2002, Metodología de Desarrollo de Sistemas Multiagente (Ph.D. Tesis). Universidad Complutense de Madrid.
- [8] C. Lucero, 2013, Modelo de seguridad multi-agente para la detección de fraudes en correos electrónicos, Tesis de licenciatura, Universidad de Sonora.
- [9] OpenSSH. 2004, OpenSSH, 20 de Enero 2012, <http://www.openssh.com/es/index.html>
- [10] OpenSSL, 2009, OpenSSL: The Open Source toolkit for SSL/TLS, 20 de Enero 2012, <http://www.openssl.org>

