

Interacción humano-computadora, visión y control inteligente

Research in Computing Science

Series Editorial Board

Editors-in-Chief:

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

Associate Editors:

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

Editorial Coordination:

María Fernanda Ríos Zacarías

Research in Computing Science es una publicación trimestral, de circulación internacional, editada por el Centro de Investigación en Computación del IPN, para dar a conocer los avances de investigación científica y desarrollo tecnológico de la comunidad científica internacional. **Volumen 91**, mayo 2015. Tiraje: 500 ejemplares. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No. : 04-2005-121611550100-102, expedido por el Instituto Nacional de Derecho de Autor. *Certificado de Licitud de Título* No. 12897, *Certificado de licitud de Contenido* No. 10470, expedidos por la Comisión Calificadora de Publicaciones y Revistas Ilustradas. El contenido de los artículos es responsabilidad exclusiva de sus respectivos autores. Queda prohibida la reproducción total o parcial, por cualquier medio, sin el permiso expreso del editor, excepto para uso personal o de estudio haciendo cita explícita en la primera página de cada documento. Impreso en la Ciudad de México, en los Talleres Gráficos del IPN – Dirección de Publicaciones, Tres Guerras 27, Centro Histórico, México, D.F. Distribuida por el Centro de Investigación en Computación, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, México, D.F. Tel. 57 29 60 00, ext. 56571.

Editor responsable: *Grigori Sidorov, RFC SIGR651028L69*

Research in Computing Science is published by the Center for Computing Research of IPN. **Volume 91**, May 2015. Printing 500. The authors are responsible for the contents of their articles. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of Centre for Computing Research. Printed in Mexico City, in the IPN Graphic Workshop – Publication Office.

Interacción humano-computadora, visión y control inteligente

Grigori Sidorov (ed.)



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



Instituto Politécnico Nacional, Centro de Investigación en Computación
México 2015

ISSN: 1870-4069

Copyright © Instituto Politécnico Nacional 2015

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

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

<http://www.ipn.mx>

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

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

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

Indexed in LATINDEX and Periodica / Indexada en LATINDEX y Periódica

Printing: 500 / Tiraje: 500

Printed in Mexico / Impreso en México

Editorial

El propósito de este volumen de la revista RCS es reflejar las nuevas direcciones de investigación y aplicaciones de los métodos de la Inteligencia Artificial en algunas áreas específicas.

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

Este volumen contiene 14 artículos relacionados con varios aspectos del desarrollo de los métodos de Inteligencia Artificial y ejemplos de sus aplicaciones a varias tareas tales como:

- Interacción entre un humano y un robot móvil,
- Reconocimiento de emociones en el rostro humano,
- Reconocimiento óptico de caracteres,
- Clasificación automática de imágenes,
- Mejoramiento automático de imágenes,
- Transmisión de datos con esteganografía,
- Control difuso y adaptivo, entre otras.

Noten que en la gran mayoría de esas aplicaciones se usan técnicas de inteligencia artificial, tales como lógica difusa, algoritmos genéticos, etc.

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

En este número especial de la revista RCS, a nombre de la comunidad académica del Centro de Investigación e Innovación en Tecnologías de la Información y Comunicación (INFOTEC) y de la SMIA expresamos nuestro agradecimiento al Dr. Sergio Carrera Riva Palacio, Director Ejecutivo, y Dr. Juan Carlos Téllez Mosqueda, Director Adjunto de Innovación y Conocimiento, por apoyar de manera ingente la investigación y el desarrollo de la ciencia y la tecnología, sustentado todo ello en la responsabilidad y el compromiso social.

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

Grigori Sidorov
Mayo 2015

Table of Contents

	Page
Towards an Indoor Location System using Indoor-light as Information Source9 <i>Carlos E. Galván-Tejada, Jorge I. Galván-Tejada, Selene Ramírez-Rosales, F. J. Martínez-Ruiz, S. Paulizeth Maldonado-Luján</i>	9
Implementación de reconocimiento de objetos por color y forma en un robot móvil21 <i>Efraín Ernesto Arévalo-Vázquez, Arturo Zúñiga-López, Juan Villegas-Cortez, Carlos Avilés-Cruz</i>	21
Reconstrucción de caracteres para mejora de sistemas OCR con RNA's33 <i>Daniela Moctezuma, Oscar S. Siordia</i>	33
Factores en el reconocimiento facial de emociones y la integración de optimización evolutiva45 <i>Luis-Alberto Pérez-Gaspar, Santiago-Omar Caballero-Morales, Felipe Trujillo-Romero</i>	45
Implementación y control de un conmutador de imágenes multifocales57 <i>Leonel Morales-Botello, J. Humberto Sossa-Azuela, Elsa Rubio-Espino</i>	57
Clasificador de pigmentación de la ballena azul aplicando el método SIFT en medios no controlados69 <i>Rosa I. Ramos-Arredondo, Blanca E. Carvajal-Gómez, Francisco J. Gallegos-Funes, Diane Gendron-Laniel</i>	69
Módulo de generación de aplicaciones multi-dispositivo a partir del procesamiento de imágenes81 <i>Laura Sánchez Morales, Viviana Yarel Rosales Morales, Giner Alor Hernández, Rubén Posada Gómez, Hilarión Muñoz Contreras, Ulises Juárez Martínez</i>	81
Implementación del modelo Retinex aplicado al procesamiento de imágenes subacuáticas para mejorar su contenido cromático95 <i>Mario Dehesa, Alberto J. Rosales, Francisco J. Gallegos, Samuel Souverville, Isabel V. Hernández</i>	95
Transmisión de video simultaneo en ancho de banda limitado aplicando esteganografía105 <i>Julio C. Suarez-Tapia, Blanca E. Carvajal-Gómez, Chadwick Carreto-Arellano</i>	105
Diseño óptimo para transmisión de fuerza en un efector final117 <i>Eric Santiago-Valentín, Adrián Solano-Palma, Pedro Bautista-Camino, José Marco Antonio Rueda-Meléndez, Edgar Alfredo Portilla-Flores</i>	117
Control par calculado difuso basado en pasividad para seguimiento de trayectorias de robots manipuladores131 <i>S.M. Orozco-Soto, J.C. Ramos-Fernández</i>	131

Control adaptativo para navegación de una silla de ruedas inteligente empleando lógica difusa tipo 1 y tipo 2	143
<i>Mario Rojas, Pedro Ponce, Arturo Molina</i>	
Enriquecimiento de señal para activación de frenos ABS en automóviles	157
<i>José Vázquez Quezada, Juan Villegas-Cortez, Iván Vázquez-Álvarez, Carlos Avilés-Cruz, Ismael Osuna-Galán, Yolanda Pérez-Pimentel</i>	
Implementación de un control multiresolución empleando un dispositivo háptico.....	167
<i>E. Mendoza-Monjaraz, J. A. Cruz-Tolentino, A. Jarillo-Silva, J. Pacheco-Mendoza</i>	

Towards an Indoor Location System using Indoor-light as Information Source

Carlos E. Galván-Tejada¹, Jorge I. Galván-Tejada¹, Selene Ramírez-Rosales¹,
F. J. Martínez-Ruiz¹, S. Paulizeth Maldonado-Luján²

¹ Universidad Autónoma de Zacatecas, Unidad Académica de Ingeniería Eléctrica,
Zacatecas, Zac., México

² Universidad Autónoma de Durango, Campus Zacatecas, Maestría en Diseño
Arquitectónico, Zacatecas, Zac., México

`ericgalvan@uaz.edu.mx`

Abstract. The location of individuals is a key element for commercial and assistive location-based applications. Since the Global Positioning System, the most effective technology for positioning a mobile object in the outdoors does not work in indoor environments. Several technological approaches have been proposed to tackle this problem. In this paper, we present an alternative approach based on the use of indoor lighting variations, feature extraction of the signal and a deterministic selection to get a model in order to estimate localization of an individual in indoor environments.

Keywords: indoor location, mobile systems, indoor luminosity, location estimation.

1 Introduction

Indoor location systems using natural signals (environmental signals) are becoming relevant in new approaches. Because it can be assumed that a natural signal must be present everywhere, avoiding the constraint of the coverage of non natural signals (i.e. WiFi, RFID) and the need for extra infrastructure deployment. However, natural signals generate a lot of data that must be processed in order to get information of them. This treatment implies more computing time and storage problems. To tackle these problems, Galván-Tejada et al. proposes a work developing ILS [10], exploring feature extraction and feature selection, in order to reduce the amount of data required to develop an ILS using natural signals, in our case, magnetic-field signal.

Nevertheless, from our experience, we know that using Genetic Algorithms during the selection phase, despite performance, can carry a high computational cost, given the stochastic nature of the approach. In this paper, we use the methodology proposed in [9] to extract features from the signal, and a feature selection thought a *Chi-Squared Filter* [15] to avoid stochasticity and finally be used to classify using well-known Random Forest Classifier [3]. Finally, we

propose using luminosity. Since this signal in controlled environments is more robust than other natural signals in terms of changes.

This paper is organized as follows. In section 2, we present a briefly review of Indoor Location Systems development approaches (ILS). The methodology to develop an ILS is explained in section 3. In section 4 we described our proposal. The results of experiments are presented in section 5, and finally, our conclusions and future work are presented in section 6.

2 Related work

Given the importance of the location context variable to develop ILS, several technological approaches have been proposed to develop ILS using different approaches. For instance, the use of technology that must be deployed as RFID, Wi-Fi, Bluetooth, Ultra Wide-band (UWB), Ultrasonic Sensors, among others [16,4]. Depending on the technique and the technology, these approaches allows us to estimate user location. Nevertheless, the use of signals that must be generated, as Wi-Fi or Bluetooth, present one main disadvantage. They require a dedicated infrastructure (e.g. access points, array of sensors or cameras) that must be deployed to develop an ILS. On the other hand, if infrastructure is already deployed. We have to consider other issues (limited Coverage, scalability, etc).

Therefore, new technological approaches have been proposed to tackle these issues. Approaches using the signals that are already available in the indoor environment [20,2,14,19,6]. For instance, the use of barometers to detect location is one of the options that emerges to know a "high-precision indoors location system". Barometers can be used it on smart phones in different kinds of buildings like hospitals, airports, etc. However, has several disadvantages. These ILS are very sensitive to "floor selection". they are just focused on plane location. The challenge is knowing, exactly in what floor of the building users are located. Besides that, we cannot know accurately location in the horizontal axe . To tackle this disadvantage, Xia et al [22] proposes the use of multi-reference barometer floor positioning (MBFP) as the answers under some conditions, also using a mobile device. This method proposes "comparing the altitude with the height of the floors in a building" to estimate in which floor is the user. Also, this method "is robust and less sensitive to factors such as temperature and humidity". The method MBFP is instable, the results that it have in the experimental face are different, it uses different sensors in the same altitude and environment. MBFP provides facilities in the new lifestyle using a sensor in the mobile device, but it needs a combination with another method to get better results.

Other available environmental signal is the light, which can be generated artificially with bulbs or naturally by the sun. It can be sensed in lumens using a sensor, which nowadays is included in almost all the smartphones in the market. For instance, Rahman et al [17] propose a Visible Light Communication (VCL) system using lighting LEDs and image sensors, several LEDs transmit the 3D coordinate information and it is demodulated by the image sensors. The 3D

array of LED helps to send location by two image sensors through two optical sensors. Also the positioning on the VCL has a limited area of response. This limitation introduces a positioning error with a few centimeters of difference.

These works use irregularities of available natural signals to estimate the location. However, huge amount of data must be analyzed to estimate the location, and it must be combined with other techniques to improve the accuracy. Galvín-Tejada et al [10] we propose a methodology to reduce the amount of data and use a single signal to develop an ILS using a natural signal (i.e. geomagnetic-field signal), but from our experience we know that feature selection, carried on through a Genetic Algorithm (GA) is computationally expensive.

The present work enhances Galvín-Tejada et al methodology through a deterministic method in the feature selection step, which is used to reduce the amount of information required. Finally, GAs are avoided since they are the most time consuming process of the methodology.

3 Methodology

In this section, we present the new version of the methodology presented in *Galvín-Tejada et al.* [9] and [10]. This methodology is modified in order to use light as information source. Beside that GAs are not used. This methodology consists of three steps described below.

3.1 First step: data collection

In data collection step, lumen data of an indoor environment is collected. To record this data, user must walk around the indoor environment carrying on a device running our mobile application, which is able to record lumen data through a lumen sensor. This walk must be done in an average walk speed during 10 seconds. The recording device must be positioned with the sensor up side, regardless any other position in the spatial axes as is shown in Figure 1. The data lectures are done at a 100 lectures per second frequency. We call *signatures* this basic set of data that represent a room.

The lumen/light sensor returns a float value indicating lumens in the environment at the moment when the lecture is done.

3.2 Second step: signature collection

The minimal number of lumen signatures is calculated using the equation 1 presented by Eberhardt [8], where x is the minimal number of needed signatures and N is the number of independent features that will compose the final lumen fingerprints.

$$x = \log_2(N) + 1 \quad (1)$$

To avoid the spatial scaling an Z norm normalization is applied to each light lecture signature using the equation 2, $z_{i,d}$ where $z_{i,d}$ is the normalized reading,

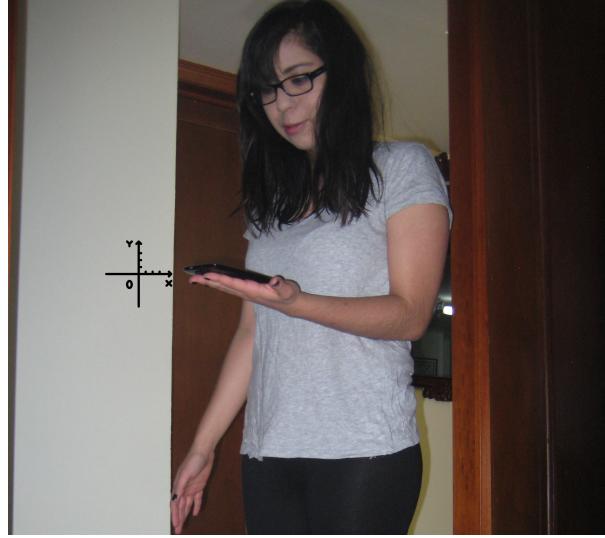


Fig. 1. Disposition of sensor in order to record light lectures.

$r_{i,d}$ refers to the i^{th} observation of the signature in dimension d μ_d is the mean value of the signature for dimension d and σ_d is the standard deviation of the signature for dimension d .

$$\forall i \in m : z_{i,d} = \frac{r_{i,d} - \mu_d}{\sigma_d} \quad (2)$$

Equation 2 is applied for all dimensions in R^d

Once light signatures are collected in the rooms and stored in a data set created during the Signature Collection. Feature extraction and selection are carried on to get the classification model that allows us to estimate the location of users.

3.3 Third step: feature extraction

The feature extraction is a process that consists in carrying out an efficient data reduction while preserving the appropriate amount of signal information. In this sense, the feature extraction step allows us to reduce the amount of data to develop the indoor location estimation model. To apply this step we must select the features that will be extracted from the signal, these activities are described below.

Signal features Following Galvn-Tejada et al methodology and given the fact that behavior of lumen signal (under the assumption of a stochastic behavior). We select 46 features collected from literature review in digital signal processing

(DSP). Besides first and second order statistic parameters [13,1,5,11] shown in Table 1, where 16 features are from temporal domain and 30 ones are from spectral domain. *Temporal features* are computed from the waveform of lumen signal generated from the signature previously recorded, and *Spectral features* extracted from the signal after performing a P-point Fast Fourier Transform to each signature [21]; to achieve it, we apply Equation (3), where ES_i is the i^{th} energy signature of the normalized lumen's signal, and NS_i is the i^{th} normalized signature.

We keep the first 10 components from the spectral evolution, because from observation (we presume) that the light phenomena tends to be chaotic too.

$$\forall i \in n : ES_i = FFT(NS_i) \quad (3)$$

Table 1. Extracted features

Features	Temporal Domain	Frequency Domain
Kurtosis	*	*
Mean	*	*
Median	*	*
Standard Deviation	*	*
Variance	*	*
Coefficient of Variation (CV)	*	*
Inverse CV	*	*
1,2,3 Quartile	*	*
1,5,95,99 Percentile	*	*
Trimmed Mean	*	*
Shannon Entropy		*
Slope		*
Spectral Flatness		*
Spectral Centroid		*
Skewness		*
1-10 Spectrum Components		*

Once all the features are computed, all of them are merged into a data set of features that summarizes the behavior of the signal, reducing the amount of data from 1000 data points to 46 per signature.

To keep all the features in a range between 0 to 1 and make them have the same impact in the model development, a percentile rank is applied using Equation (4), where x is the x^{th} feature.

$$PR = \frac{trunc(rank(x))}{length(x)} \quad (4)$$

Model development: feature selection For getting a model with less features that allows us to estimate user location. A feature selection process is applied. Feature selection use multiply techniques. For instance, filter, wrapper and embedded methods [18]. From literature [7] we propose apply a *filter method*, because it returns features that can be used with different classifiers and is less computationally expensive.

Filter methods that allow us to get a feature selection at low computational cost includes: *Chi-Squared Filter* [15], because is capable of evaluate the data in function of weights; features with a low weight are removed based on the chi squared statistics [12]. Once the features are selected, these can be used with several classifiers to get a model to estimate the location, which is the core of an ILS.

4 Experimentation

The experimentation was carried on in a residential home shown in Figure2, using available data from AAAMI research group website ³. We select four rooms because are contiguous. This characteristic involved a extra complication. The spaces share light data in some parts of rooms. The light is provided by 7 bulbs composed by 107 LED's warm light.

In order to determine the minimal number of signatures required, the equation 1 was applied. We consider 46 features per each room, which means 184 variables, getting a total of 8.5 lumen signatures, finally rounded to 9.

Once we got the minimal number of signatures. We choose a Samsung Galaxy S3 i900 to record lumen data, using an Android application, previously developed in order to read lumen sensor.

To perform feature extraction step, we develop a script in R Project for Statistical Computing software ⁴, a free multi-platform software (GNU project) environment for statistical computing, Getting a matrix composed of 36 rows (one for each signature) and 46 columns (one for each feature). Percentile rank is applied to all the values to keep them between 0 to 1.

After feature matrix normalized. The filter method (Chi-Square Filter) proposed is used to select features. This step select 6 features from the original 46 extracted from the signatures.

The Chi-Square Filter is applied using the CARET package [12] in R, which does the work in five phases:

1. Data equalization and normalization.
2. Use the filter to obtain significative features.
3. Obtain classification rate using the filtered model with Random Forest.
4. Obtain the classification rate using 46 selected features with Random Forest.
5. Calculate the Confusion Matrix for comparison.

³ <http://aaami.mty.itesm.mx/>

⁴ <http://www-r-project.org/>

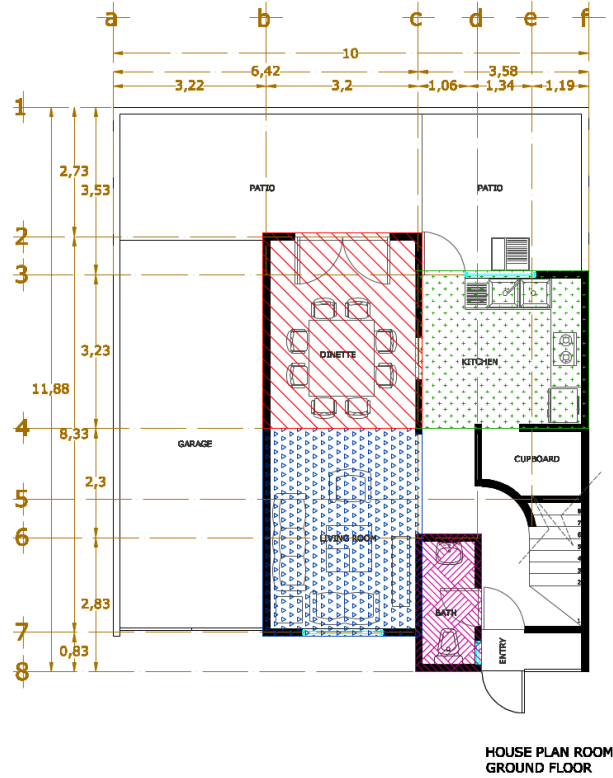


Fig. 2. Plane of residence used in our experiment. bathroom in pink, living room in blue, dining room in red and kitchen in green.

With selected features, a Random Forest classifier composed of 5,000 trees is used to evaluate indoor location model conformed by features selected by filter method. This classifier was proposed by Breiman et al. [3]; and selected because it provides tree ensembles that depends on the values of a random feature vector and provides the same distribution to all the trees included in the forest. The decision of this classifier depends on the decision of several trees.

All classification process was performed using Random Forest R package.

5 Results

After experimental test, the final filtered model was a 17 feature model of classification. Table 3 shows the confusion matrix acquired from the classification model of the random forest using 17 features model from data set. The obtained model can classify 95.83 percent of the samples, correctly. In table 2 filtered features model is presented.

Table 2. Filtered Features

Features	Temporal Domain	Frequency Domain
Mean		*
Median	*	
Standard Deviation	*	
Variance		*
Coefficient of Variation (CV)	*	*
Inverse CV	*	*
1,2,3 Quartile	*	
1,5,95,99 Percentile	*	
Slope		*
6 Spectrum Component		*

Table 3. Signatures Classification For Different Locations Using filtered features

	Bathroom	Dining Room	Kitchen	Living Room
Bathroom	6	0	0	0
Dining Room	0	6	0	0
Kitchen	1	0	5	0
Living Room	0	0	0	6

The forty six features were tested in order to produce a classification model. Table 4 shows the confusion matrix acquired from classification model of the random forest using 46 features model. The obtained model can classify correctly 91.66 percent of the samples.

Table 4. Signatures Classification For Different Locations

	Bathroom	Dining Room	Kitchen	Living Room
Bathroom	4	0	2	0
Dining Room	0	6	0	0
Kitchen	0	0	6	0
Living Room	0	0	0	6

The obtained features are used to generate a random forest classification model which is compared against the 46 features random forest model. As shown in Table 5.

Table 5. Comparison of Different Approaches

Approach	Classification Rate (in percentage)
Filtered Model	95.83
Unfiltered features	91.66

6 Conclusions and future work

This paper presents a novel approach where light-based models for geo spatial positioning are used for the first time. Instead of using stochastic techniques. Like in the first version of the methodology. A classification model based on spatial characteristics of light emission and the behavior of the same in different environments within a residence space was introduced.

The following important aspects regarding the classifiers were identified :

- *Sensitivity and Specificity*: Since the model found by using Chi Square filter is relatively large, their sensitivity and specificity are high, as can be seen in its confusion matrix , resulting in a very high percentage of correct classification of spaces that were used for this study. It is clear that reducing variables by filtering not only improves speed of classifier , it helps increase the percentage of prediction. Because some statistically rivalry between variables. Therefore, causing a lower model performance.
- *Computational cost* compared to the use of genetic algorithms, the calculation time is much smaller. In this version, it was reached an acceptable performance through fewer variables and more concise models. In our tests the GA version has required 144 minutes (with 2 features) while this version required 2 minutes and 5 seconds (with 17 features) .The improvement is considerable.
- *Time is the most important issue to take into account*. Along the evaluation performed in this work, we also show that time is the variable with higher variability. Time became the most important characteristic to take into account when developing an ILS that requiers recalibration.

For future work the possibility of sampling a larger space is proposed. Taking into account different types of lighting in each scene and purpose of their interiors. Also, using other tools to improve the performance of mobile phones and tablets .

References

1. Agostini, G., Longari, M., Pollastri, E.: Musical instrument timbres classification with spectral features. EURASIP Journal on Applied Signal Processing 2003, 5–14 (2003)

2. Bilke, A., Sieck, J.: Using the magnetic field for indoor localisation on a mobile phone. In: Krisp, J.M. (ed.) *Progress in Location-Based Services*, pp. 195–208. *Lecture Notes in Geoinformation and Cartography*, Springer Berlin Heidelberg (2013)
3. Breiman, L.: Random forests. *Machine learning* 45(1), 5–32 (2001)
4. Chen, L., Pei, L., Kuusniemi, H., Chen, Y., Krger, T., Chen, R.: Bayesian fusion for indoor positioning using bluetooth fingerprints. *Wireless Personal Communications* 70(4), 1735–1745 (2013), <http://dx.doi.org/10.1007/s11277-012-0777-1>
5. Chen, P.C., Pavlidis, T.: Segmentation by texture using a co-occurrence matrix and a split-and-merge algorithm. *Computer graphics and image processing* 10(2), 172–182 (1979)
6. Chung, J., Donahoe, M., Schmandt, C., Kim, I.J., Razavai, P., Wiseman, M.: Indoor location sensing using geo-magnetism. In: *Proceedings of the 9th international conference on Mobile systems, applications, and services*. pp. 141–154. *MobiSys '11*, ACM, New York, NY, USA (2011)
7. Delgado-Contreras, J.R., García-Vázquez, J.P., Brena, R.F., Galván-Tejada, C.E., Galván-Tejada, J.I.: Feature selection for place classification through environmental sounds. *Procedia Computer Science* 37, 40–47 (2014)
8. Eberhardt, F.: A sufficient condition for pooling data. *Synthese* 163(3), 433–442 (2008), <http://dx.doi.org/10.1007/s11229-007-9293-3>
9. Galván-Tejada, C.E., García-Vázquez, J.P., Brena, R.: Magnetic-Field Feature Extraction for Indoor Location Estimation. In: *Ubiquitous Computing and Ambient Intelligence. Context-Awareness and Context-Driven Interaction*, pp. 9–16. Springer (2013)
10. Galván-Tejada, C.E., García-Vázquez, J.P., Brena, R.F.: Magnetic-field feature reduction for indoor location estimation applying multivariate models. In: *MICAI (Special Sessions)*. pp. 128–132. IEEE (2013)
11. Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on* (6), 610–621 (1973)
12. from Jed Wing, M.K.C., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., the R Core Team: caret: Classification and Regression Training (2014), <http://CRAN.R-project.org/package=caret>, r package version 6.0-24
13. Lambrou, T., Kudumakis, P., Speller, R., Sandler, M., Linney, A.: Classification of audio signals using statistical features on time and wavelet transform domains. In: *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. vol. 6, pp. 3621–3624 (1998)
14. Li, B., Gallagher, T., Rizos, C., Dempster, A.G.: Using geomagnetic field for indoor positioning. *Journal of Applied Geodesy* 7(4), 299–308 (2013)
15. Liu, H., Setiono, R.: Chi2: Feature selection and discretization of numeric attributes. In: *2012 IEEE 24th International Conference on Tools with Artificial Intelligence*. pp. 388–388. IEEE Computer Society (1995)
16. Ni, L., Liu, Y., Lau, Y., Patil, A.: Landmarc: Indoor location sensing using active rfid. *Wireless Networks* 10(6), 701–710 (2004)
17. Rahman, M.S., Haque, M.M., Kim, K.D.: Indoor positioning by led visible light communication and image sensors. *International Journal of Electrical and Computer Engineering (IJECE)* 1(2), 161–170 (2011)
18. Saeys, Y., Inza, I., Larrañaga, P.: A review of feature selection techniques in bioinformatics. *bioinformatics* 23(19), 2507–2517 (2007)

19. Storms, W., Shockley, J., Raquet, J.: Magnetic field navigation in an indoor environment. In: Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS), 2010. pp. 1–10 (oct 2010)
20. Subbu, K.P., Gozick, B., Dantu, R.: Locateme: Magnetic-fields-based indoor localization using smartphones. *ACM Trans. Intell. Syst. Technol.* 4(4), 73:1–73:27 (Oct 2013), <http://doi.acm.org/10.1145/2508037.2508054>
21. Tsai, W.H., Tu, Y.M., Ma, C.H.: An FFT-based fast melody comparison method for query-by-singing/humming systems. *Pattern Recognition Letters* 33(16), 2285–2291 (2012), <http://www.sciencedirect.com/science/article/pii/S016786551200284X>
22. Xia, H., Wang, X., Qiao, Y., Jian, J., Chang, Y.: Using multiple barometers to detect the floor location of smart phones with built-in barometric sensors for indoor positioning. *Sensors* 15(4), 7857–7877 (2015)

Implementación de reconocimiento de objetos por color y forma en un robot móvil

Efraín Ernesto Arévalo-Vázquez, Arturo Zúñiga-López, Juan Villegas-Cortez,
Carlos Avilés-Cruz

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

al206200579@alumnos.azc.uam.mx, {azl, juanvc, caviles}@correo.azc.uam.mx

Resumen. La visión artificial la componen un conjunto de procesos destinados a realizar el análisis de imágenes. Estos procesos son: captación de imágenes, memorización de la información, procesado e interpretación de los resultados. En este artículo se presenta un avance en el aprovechamiento del poder de procesamiento de los nuevos sistemas embebidos para controlar un robot móvil por medio de un algoritmo de visión más que por sensores electrónicos. En este caso particular se buscó desde un robot móvil analizar un objeto con base a su color y forma, tal que el dispositivo pueda moverse de manera autónoma hacia el objeto para analizarlo, y que realice movimientos para buscar y encontrar un objeto idéntico a él, obtenido del resultado del procesamiento de imágenes.

Palabras clave: visión por computadora, sistemas embebidos, reconocimiento de objetos, robots móviles.

1. Introducción

Los sistemas de visión artificial han evolucionado poco a poco a medida que la tecnología ha evolucionado y permitido la creación de cámaras, computadoras y algoritmos de procesamiento más potentes. De forma básica un sistema de visión artificial está conformado por varios subsistemas capaces de realizar dos funciones fundamentales: captar la información de la escena real mediante la proyección en una imagen y analizar las imágenes para extraer la información que contienen [7] [5].

El mayor peso del análisis de las imágenes captadas se encuentra a nivel de software, lo que proporciona una mayor flexibilidad en cuanto a lo que se puede implementar con base a los resultados obtenidos [8]. Esto lo podemos aprender de los resultados usados en la aplicación de algoritmos de visión para el reconocimiento de instrumental médico [6], el reconocimiento de marcas de localización para rutas en robots móviles [3], y dándole libertad de movimiento a los robots móviles, usando toma de decisión con base en sistemas difusos [4], y también en la aplicación para desplazamiento en 3D con visión [7]. En los trabajos mencionados se muestran técnicas diversas para el análisis del entorno

de desplazamiento, y en nuestro caso nos enfocamos en el análisis espacial del espacio de color de las imágenes a usar del entorno del robot móvil [9], como una etapa importante, pero también considerando la implementación de rasgos característicos del objeto de estudio, como lo son los momentos de Hu [11]. Existen aplicaciones robustas de visión para posicionamiento con base únicamente a la visión [10], pero para nuestro propósito no se considera el posicionamiento, sino la detección y reconocimiento de un objeto específico en el campo de visión de una cámara montada en un robot móvil diferencial, todo desde el paradigma de la imagen digital y el procesamiento digital de imágenes.

En éste artículo se muestra la realización del diseño e implementación de un sistema embebido que permite controlar un robot móvil que analiza, por color y forma un objeto muestra, mediante una cámara digital, se dé la vuelta y avance mientras identifica diferentes objetos con colores y formas distintas, hasta detenerse enfrente del objeto buscado. El algoritmo de reconocimiento usa los momentos invariantes de Hu [11] [5], para poder tener invarianza suficiente para lograr nuestro objetivo. En todo momento se puede monitorear parte del procesamiento de las imágenes a través de una computadora por medio de una conexión inalámbrica a una red LAN, gracias a un servidor HTTP para este propósito.

En la sección 2 presentamos el desarrollo de esta propuesta, en la sección 3 mostramos los resultados de los experimentos realizados, y finalmente las conclusiones con las perspectivas del trabajo futuro se muestran en la sección 4.

2. Desarrollo

El desarrollo del sistema se realiza en tres etapas: *(i)* diseño e implementación de una metodología de detección e identificación del objeto de interés usando Procesamiento Digital de Imágenes, *(ii)* construcción del robot móvil diferencial con la cámara, y *(iii)* implementación del sistema de monitoreo en tiempo real. A continuación se describe cada una de estas etapas.

2.1. Metodología de detección e identificación del objeto de interés usando Procesamiento Digital de Imágenes.

Aquí se muestra cómo se analizan las imágenes, capturadas con una cámara web ordinaria, con el fin de buscar y detectar marca, color y forma de un objeto, para posteriormente estar enviando las imágenes para su visualización en el sistema de monitoreo. Las imágenes se tomaron con resolución de 320 x 240 píxeles a 32 bits de color. Las etapas del procesamiento digital se muestran en la figura 1, ahí se puede seguir con detalle cada uno de los bloques en su propósito y funcionamiento.

Inicio

Al encender el dispositivo: se carga el sistema operativo, se conecta automáticamente al modem por la conexión inalámbrica, inicio automático al servicio de monitoreo a través del servidor web y el servicio de conexión remota. El usuario

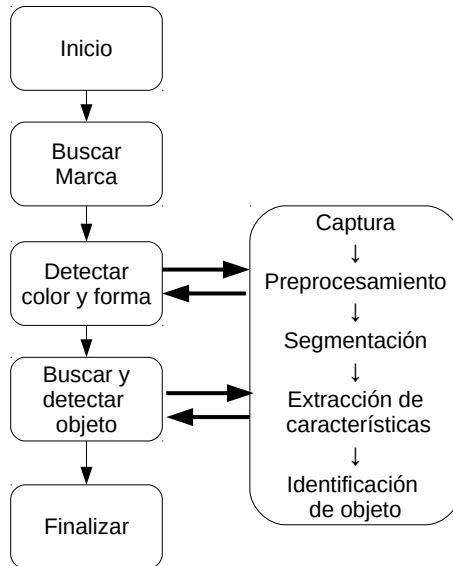


Fig. 1: Metodología propuesta para la detección y reconocimiento de objetos en un robot móvil diferencial.

se conecta vía SSH Secure Shell Client y se ejecuta el comando para iniciar el programa. Al iniciar el programa se verifica la conexión y el funcionamiento de la cámara web, en caso de error el programa se interrumpe y manda un aviso.

Buscar marca

Se utiliza una marca distintiva para la colocación del objeto y para que el sistema identifique su ubicación, la cual se coloca frente al objeto y este es analizado. Consta de una hoja carta de color blanco colocada de manera horizontal con cuatro cuadrados huecos de contorno negro dibujados cerca de las esquinas de la hoja. El tamaño de los cuadrados es de 3cm por lado y el grosor de la línea es de 3mm . El tamaño puede variar, pero no deben quedar cubiertos al colocar el objeto a analizar. El objeto se coloca cerca de la hoja y en la parte media de los dos cuadrados inferiores. Si no se coloca en esta posición, el sistema no podrá analizarlo correctamente.

Al iniciar el programa, el robot busca la marca de los cuatro cuadrados, una vez encontrada toma las coordenadas de dos cuadrados inferiores para centrar la imagen y acercarse lo suficiente para analizar el color y la forma del objeto. La muestra se toma de un segmento interno del objeto para poder crear la muestra correcta; si no se logra establecer una muestra correcta, el robot retrocede y vuelve a acercarse.

En ésta etapa el procesamiento de imágenes se utiliza para reconocer la marca de los cuatro cuadrados, y obtener las coordenadas de los mismos para enviar las instrucciones de movimiento y centrar la marca en cuadro de visión de la

cámara. Las coordenadas nos servirán para ubicar un rectángulo en el centro de los cuadros inferiores, que es donde se encuentra el objeto.

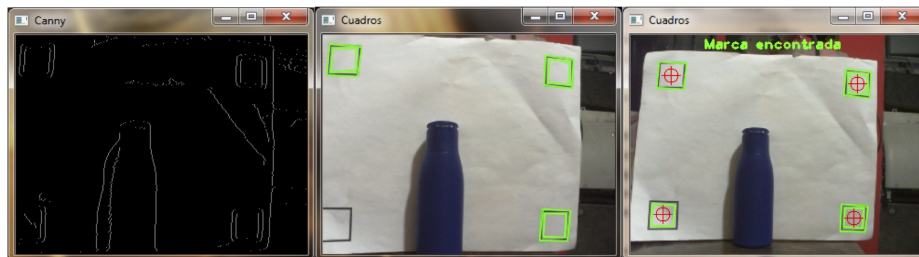


Fig. 2: Ejemplo del funcionamiento del uso de la plantilla usada y las marcas sobre la imagen capturada, para ubicar la posición del objeto de interés.

Captura

Se obtiene una imagen de la cámara en la configuración establecida la cual se utiliza para buscar la marca de los cuatro cuadrados.

Pre-procesamiento

Se aplican filtros, el primero es convertir la imagen a escala de grises, el segundo reduce la resolución de la imagen a la mitad y se vuelve a ampliar a la resolución original para eliminar ruido, y el tercero aplica el filtro Gaussiano [5] para suavizar la imagen y eliminar el mayor ruido posible.

Segmentación

La imagen filtrada es segmentada mediante el algoritmo de Canny [5], buscando reducir los datos de la imagen de interés con la información que nos interesa, para nuestro caso lograr una imagen binaria que muestra todos los bordes encontrados en color blanco y el fondo negro, preservando la información estructural importante en la imagen. Este algoritmo halla bordes con base a buscar máximos locales del gradiente de la imagen, calculándolo usando la derivada de un filtro Gaussiano, usa dos valores de umbral para localizar bordes fuerte o débilmente marcados en la imagen, y en esto está la ventaja de usar Canny en lugar de otros métodos ya que tiene buena detección ante ruido.

Extracción de características

En la imagen binaria se buscan contornos, ya que los cuadrados de la marca son figuras cerradas. De éste proceso se obtiene un conjunto de coordenadas que representan cada uno de los contornos encontrados en la imagen. Éste arreglo se procesa para encontrar figuras geométricas, que en este caso son cuadrados.

Identificación de objetos

Los contornos encontrados se analizan para detectar si son cuadrados o no, y almacena las coordenadas de cada uno de los cuadrados encontrados. Se dibuja un recuadro verde sobre la imagen original, resaltando cada cuadro encontrado. Cuando se detecta los cuatro cuadrados, se informa que se ha encontrado la

marca, y se guardan las coordenadas de los centros de cada uno de ellos. En la imagen original se dibuja la frase “Marca encontrada” y se resalta en rojo el interior de cada cuadro.

Una vez obtenida la marca, se identifican los dos cuadrados inferiores y se calculan dos datos: la distancia entre los centros y la coordenada x del punto medio de dicha distancia. El punto x nos sirve para centrar la imagen con respecto a la horizontal del campo de visión de la cámara. La distancia obtenida es utilizada para calcular la relación con respecto al ancho de la imagen y poder acercar el robot a la marca.

Los movimientos laterales que puede realizar el robot son determinados por la posición del punto x con respecto al centro de la imagen, moviéndose primero a la izquierda o a la derecha hasta centrarlo en un margen de error de 16 píxeles. Una vez centrado, avanza hacia adelante hasta que la relación calculada no sea menor a 0.73. Una vez centrado el robot y cerca de la marca se detendrá el sistema de movimiento y el proceso de ‘Búsqueda de Marca’, y se iniciará el análisis del objeto en la etapa ‘Detectar color y forma’.

Detectar color y forma

En esta etapa se dibuja un rectángulo en el centro de los cuadrados inferiores de 25 píxeles de lado el cual debe cubrir una sección interna de la imagen del objeto a analizar. De esta sección se toma la muestra de los colores de cada píxel y se usa para crear el histograma de colores y la máscara que ayuda a separar la imagen del objeto y el fondo de la imagen. Una vez separados se analiza la forma del objeto y se guardan sus valores característicos que sirven para identificar el objeto.

Captura. Se obtiene una imagen de la cámara y se utiliza las coordenadas de la sección para crear una segunda imagen que solo contenga la sección recortada.

Pre-procesamiento. Se reduce la resolución de la imagen a la mitad y se vuelve a ampliar a la resolución original para eliminar ruido. Creando una imagen llamada ‘corte’, de la imagen original a partir de la sección del rectángulo central. De la imagen ‘corte’ se obtienen los mínimos y máximos de los valores de tono, saturación y brillo (HSV) de cada píxel. Estos valores se usan para crear el histograma de colores y el filtro para la imagen original. La imagen original se transforma de color RGB al modelo HSV usando los valores mínimos y máximos obtenidos del corte, y se obtiene una imagen en color HSV. De la imagen HSV se crea una imagen binaria que es la primera máscara en la que los píxeles que se encuentran dentro del rango de mínimos y máximos que se representan en blanco, y los demás en negro. Se crea una imagen llamada hue (tinte) mezclando los canales tomados del corte de la imagen original, para realzar la tonalidad de colores del objeto y ayude a los cambios leves de iluminación. Se aplica un filtro a la imagen hue en base a los valores del histograma para eliminar el fondo, obteniendo una imagen en escala de grises que representa la máscara del color predominante del objeto. La imagen se convierte a binaria. Esta imagen se combina con la primera máscara y se obtiene otra con un mejor detalle del objeto. Ésta última imagen contiene un mayor detalle de la forma del objeto y los filtros aplicados ayudan a eliminar del fondo todo lo que no coincida con

la tonalidad del color. En la figura 3 se muestra un ejemplo del procesamiento digital aplicado al cuadro de imagen.

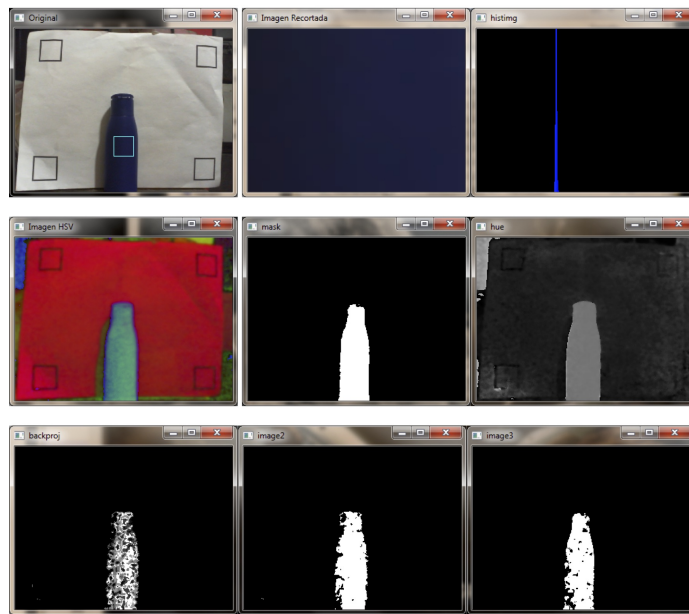


Fig. 3: Metodología propuesta para la detección y reconocimiento de objetos en un robot móvil diferencial.

Segmentación. En la imagen del paso anterior se buscan todos los contornos, obteniendo como resultado solo uno, el del objeto. Si no se encuentra ningún contorno quiere decir que hubo una mala muestra y se debe volver a tomar.

Extracción de características. Una vez detectados los contornos, se analizan en busca del contorno con mayor área, y se guarda en un vector que contenga todos los puntos del contorno del objeto. Este vector se guarda para ser comparado con el objeto a buscar en la siguiente etapa. También se guarda el histograma y los valores máximos y mínimos HSV para usarlos como filtro de color.

Identificación de objetos. El contorno guardado se utiliza para identificar el objeto en la siguiente etapa.

Buscar y detectar el objeto

Una vez completada la etapa 'Detectar color y forma', el robot gira a la derecha hasta perder de vista el objeto analizado y comienza la búsqueda de otro objeto con las mismas características. El robot se mueve lentamente a la derecha hasta que aparezca un objeto similar en el campo de visión de la cámara. Se utilizan los filtros obtenidos de la etapa anterior; por lo cual solo aparecerá en la imagen binaria cualquier objeto con la misma tonalidad que el objeto a buscar. Una

vez encontrado un objeto, se compara con el contorno guardado y si su forma es similar, se moverá el robot para centrarlo en la cámara y marcarlo como encontrado.

Captura. Se obtiene una imagen de la cámara.

Pre-procesamiento. Se utilizan los mismos filtros de la etapa anterior, tanto el histograma de colores como los mínimos y máximos valores de la muestra de color. De esta manera cualquier objeto que no pertenezca al rango de colores de la muestra, no se verá en la imagen filtrada, y el robot seguirá girando a la derecha. Un objeto de color similar aparecerá en ventana dependiendo su grado de coincidencia.

El robot seguirá girando a la derecha hasta que aparezca en la imagen un objeto del mismo color, cuando esto suceda, se guarda el contorno para que en la siguiente fase del procesamiento sea analizado.

Segmentación. De la misma manera que en la etapa anterior, se busca el contorno de la imagen para extraer sus características.

Extracción de características. Del contorno obtenido se extraen los puntos que lo conforman, incluyendo las coordenadas de su centro de masa.

Identificación de objetos. Se compara con el contorno del objeto a buscar por medio de sus momentos invariantes Hu [9], obteniendo un valor numérico de dicha comparación. Se coloca una señal en la imagen original, utilizando las coordenadas de su ubicación y el número obtenido de la comparación de contornos. El número representa la diferencia entre la similitud de ambas figuras; sí es mayor a 40 entonces son diferentes y entre más cercano a cero mayor el parecido. El carro seguirá girando a la derecha, pero deja la señal para indicar la mínima coincidencia. En la figura 4 se muestran cómo funciona la identificación de un objeto, aún ante presencia de oclusiones.

Finalizar Si encuentra un objeto similar, entonces inicia movimientos laterales para centrar la imagen y colocarse frente al objeto y así dejar de moverse. En este momento se da por encontrado el objeto.

2.2. Construcción del robot móvil diferencial con la cámara

Existen varios diseños de ruedas para elegir cuando se quiere construir una plataforma móvil sobre ruedas, para el desarrollo se utiliza el diseño diferencial. Se compone de 2 ruedas en un eje común, cada rueda se controla independientemente, puede realizar movimientos en línea recta, en arco y sobre su propio eje de contacto de rodamiento, requiere de una o dos ruedas adicionales para balance o estabilidad. Sencillo mecánicamente, puede presentar problemas de estabilidad y su cinemática es sencilla, para lograr el movimiento en línea recta requiere que las dos ruedas de tracción giren a la misma velocidad.

La implementación del sistema embebido se realizó integrando los componentes del proyecto, para que trabajen como un solo sistema. La parte del hardware está compuesto por la tarjeta de desarrollo Raspberry PI [2], La cámara web, la tarjeta de red inalámbrica y el sistema de movimiento. En la figura 5 se muestran imágenes de la construcción del robot móvil experimental aquí explicado.

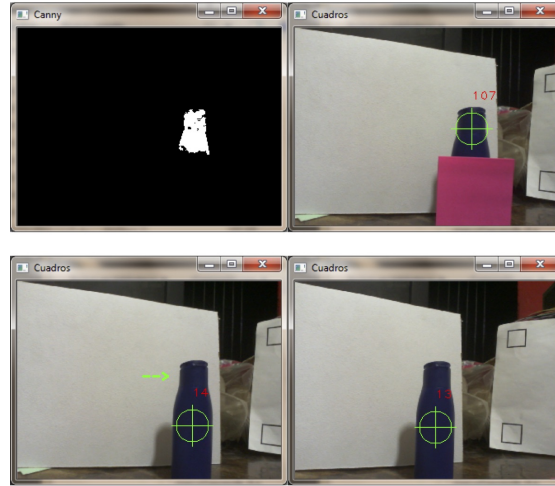


Fig. 4: Imágenes de monitoreo sobre el funcionamiento para la detección e identificación de un objeto.

La parte del software la integran el sistema operativo instalado en la tarjeta SD del Raspberry Pi, la librería de visión por computadora OpenCV [1], Servidor Web Lighttpd, el cual es parte del sistema de monitoreo, y el código fuente del programa que representa el sistema de procesamiento de imágenes.

Los pasos a seguir para la implementación son los siguientes:

- Se instala el sistema operativo Raspbian (Linux Debian Wheezy), la librería OpenCV 2.4.9, el servidor Lighttpd.
- Conectamos la cámara web y la tarjeta inalámbrica a los puertos USB de la Raspberry PI. La cámara Web y la tarjeta inalámbrica son compatibles con el sistema; por tanto no es necesario realizar una configuración adicional.
- Conectamos los puertos Gpio's al sistema de movimiento.
- Implementación del Sistema de monitoreo, se creó una página web con cuatro recuadros en los cuales se muestran las imágenes del sistema de procesamiento de imágenes. Estableciendo que cada una se actualice cada 200 ms.
- El sistema de procesamiento de imágenes se implementó dentro una carpeta, ubicada dentro de la tarjeta de desarrollo Raspberry PI, el cual se compila y ejecuta en el mismo sistema.

2.3. Sistema de monitoreo en tiempo real de la implementación

El sistema de monitoreo fue instalado en el dispositivo Raspberry Pi. El servidor Web Lighttpd se configuró en el puerto 80 para acceder por medio de cualquier navegador web escribiendo la dirección IP del Raspberry Pi, el cual

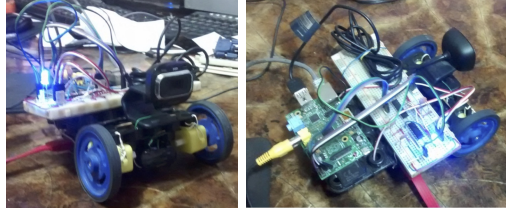


Fig. 5: Detalle de construcción del robot móvil diferencial.

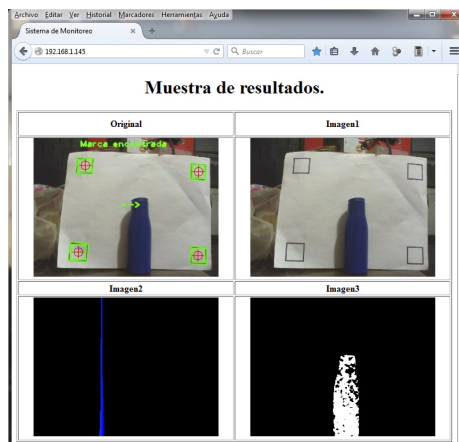


Fig. 6: Ventana del sistema de monitoreo del sistema, accesando con cualquier navegador Web en una red LAN.

representa la página web con cuatro imágenes que muestran los resultados del procesamiento que se lleva a cabo, como se muestra en la figura 6.

3. Resultados

En la figura 6 se muestran las pantallas de los resultados experimentales que se obtuvieron, con el reconocimiento esperado para condiciones de iluminación suficiente, a fin de lograr el nivel de detección deseado.

En la primera imagen (esquina superior izquierda) se muestra su color original con señales que destacan el resultado del procesamiento de imágenes, la segunda imagen (esquina superior derecha) muestra los cambios del proceso de filtrado, la tercera imagen (esquina inferior izquierda) muestra el histograma de colores del objeto analizado; y la cuarta (esquina inferior derecha) muestra el contorno guardado del objeto que será la base para buscar coincidencias.

Las imágenes se van actualizando cada 200 ms y son recibidas del programa principal en aproximadamente la misma cantidad de tiempo. La página solo proporciona imágenes, no interfiere en el funcionamiento del sistema.

4. Conclusiones

En este artículo presentamos una propuesta de fácil y rápida implementación, de bajo costo de la implementación de un algoritmo de visión por computadora en un sistema embebido. El sistema tiene operación autónoma, tomando las decisiones de movimiento del robot móvil con base al análisis y procesamiento de las imágenes capturadas desde su cámara web, teniendo un procesamiento fluido de 5 cuadros de imagen por segundo, logrando la detección de objetos en tiempo real. El modelo aquí propuesto se basa primero en obtener una muestra de color, y después su contorno. Los objetos que son mejores para éste sistema son aquellos que ofrecen texturas opacas.

A futuro planteamos continuar la mejora del sistema, refinando el algoritmo para diferentes propósitos. Adicionalmente el sistema puede mejorarse en su desempeño usando un sistema embebido con mayor capacidad de procesamiento.

Referencias

1. Adrian Kaehler, G.B.: Learning OpenCV. O'Reilly Media (2013)
2. Foundation, R.P.: Raspberry pi, <https://www.raspberrypi.org>
3. Gonzalez, E.: Study of behaviors subsumed with the generation of motor schemas in lego nxt 2.0. In: Engineering Applications (WEA), 2012 Workshop on. pp. 1–6 (May 2012)
4. Gonzalez, F., Guarnizo, J., Benavides, G.: Emulation system for a distribution center using mobile robot, controlled by artificial vision and fuzzy logic. Latin America Transactions, IEEE (Revista IEEE America Latina) 12(4), 557–563 (June 2014)
5. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1992)
6. Humberto Sossa, Roberto A. Vazquez, R.B.: Reconocimiento y localización de instrumental médico usando análisis automatizado de imágenes. Revista mexicana de ingeniería biomédica 26, 75–85 (2005)
7. Kim, J.H., Lyou, J.: The development of an artificial vision based navigation system for helicopter using modified ndgps/ins integration. In: Control, Automation and Systems, 2007. ICCAS '07. International Conference on. pp. 2114–2118 (Oct 2007)
8. Malpartida, E.D.S.: Sistema de visión artificial para el reconocimiento y manipulación de objetos utilizando un brazo robot. Master's thesis, Escuela de Graduados, Pontificia Universidad Católica del Perú (2003)
9. Rao, A., Srihari, R., Zhang, Z.: Spatial color histograms for content-based image retrieval. In: Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on. pp. 183–186 (1999)
10. Villanueva-Escudero, C., Villegas-Cortez, J., Zúñiga-López, A., Avilés-Cruz, C.: Monocular visual odometry based navigation for a differential mobile robot with android os. In: Gelbukh, A., Espinoza, F., Galicia-Haro, S. (eds.) Human-Inspired

- Computing and Its Applications, Lecture Notes in Computer Science, vol. 8856, pp. 281–292. Springer International Publishing (2014)
11. Zhang, R., Wang, L.: An image matching evolutionary algorithm based on hu invariant moments. In: Image Analysis and Signal Processing (IASP), 2011 International Conference on. pp. 113–117 (Oct 2011)

Reconstrucción de caracteres para mejora de sistemas OCR con RNAs

Daniela Moctezuma, Oscar S. Siordia

Centro de investigación en Geografía y Geomática Ing. Jorge L. Tamayo AC,
México D.F., México

{dmoctezuma, osanchez}@centrogeo.edu.mx

<http://www.centrogeo.org.mx>

Resumen. Numerosos avances dentro de la Inteligencia Artificial han estado inspirados en las características de los procesos biológicos de las redes neuronales. Estas características, por las ventajas que presentan, han intentado ser emuladas dentro de una computadora para imitar las tareas que son capaces de realizar los seres humanos. Siendo el OCR (Optical Character Recognition) una habilidad humana, la intención de utilizar RNAs (Redes Neuronales Artificiales) para un sistema de reconocimiento de caracteres, va en relación de cómo las RNAs tratan de simular los procesos que ocurren en el cerebro humano. El presente trabajo implementa un sistema OCR mediante RNAs, así como una modificación al mismo, implementando un paradigma para la reconstrucción del patrón de entrada (con una red Hopfield) antepuesto al paradigma utilizado para el reconocimiento (red Perceptrón). Las pruebas realizadas comparan los resultados entre el sistema OCR original y el modificado para este trabajo, utilizando tanto números como letras. Las ventajas de la modificación planteada para este trabajo pueden ser ampliadas a cualquier sistema OCR ya implementado sin importar el método de reconocimiento utilizado. Los resultados muestran una importante mejora en el porcentaje de reconocimiento con el método propuesto respecto al método comparado.

Palabras clave: RNA, OCR, red Hopfield.

1. Introducción

El reconocimiento de caracteres es una habilidad humana cuya emulación en sistemas de cómputo ha recibido mucha atención en los campos de PDI (Procesamiento Digital de Imágenes) y RP (Reconocimiento de Patrones). Como se muestra en la Figura 1, el PDI se hace necesario para el pre-procesamiento de los datos sea cual sea la técnica de reconocimiento a utilizar; sin embargo, esta última puede variar de acuerdo a los requerimientos del problema. Un sistema OCR puede ser definido como el conjunto de procesos mecánicos o electrónicos de conversión de caracteres, escritos o impresos, en un formato que sea editable en una computadora [7]. En un principio eran consideradas como

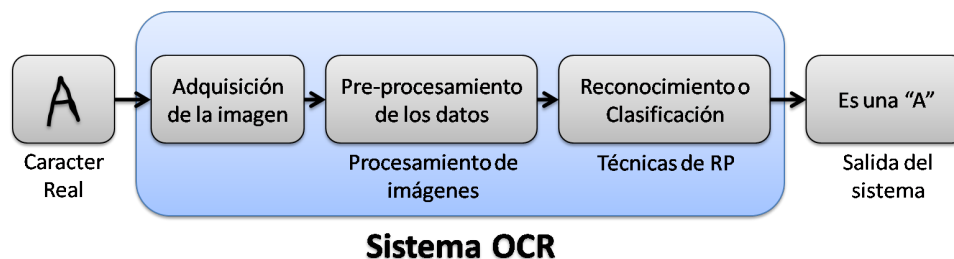


Fig. 1. Diagrama a bloques de un sistema OCR

OCRs únicamente las técnicas de reconocimiento que utilizaban lentes o espejos para realizar su tarea y, por otro lado, aquellas técnicas que utilizaban algoritmos computacionales eran conocidas como DCR (Digital Character Recognition); sin embargo, al existir muy pocas técnicas de OCR reales, esta palabra fue redefinida para incluir ambas técnicas. Los primeros sistemas OCR datan de 1929 y es un campo de investigación muy activo hasta la fecha. Los métodos clásicos de RP no han sido suficientes para el reconocimiento de caracteres debido a las siguientes razones [1]:

- Un mismo carácter difiere en tamaños, formas y estilos de una persona a otra e incluso en la misma persona en diferentes momentos.
- Al tratarse de imágenes, los caracteres pueden contener ruido que dificulte el reconocimiento.
- No hay un conjunto de reglas que definan la apariencia de los caracteres visuales y, por lo tanto, su reconocimiento.

Estos problemas pueden ser resueltos mediante la aplicación de RNAs que, gracias a sus características, permiten resolver problemas de reconocimiento con tolerancia a ruido e invariancia a escalas y rotaciones que pueden ser traducidos a formas y estilos [2]. Estas ventajas han hecho al OCR una de las aplicaciones más comunes en el área de RNAs y ha sido utilizado en los últimos años como el ejemplo práctico por excelencia en el campo académico (por ejemplo los trabajos presentados en [3],[4], [5] y [8]) .

Casi cualquier paradigma de RNAs puede ser aplicado para la solución de este problema. Sin embargo, la mayoría de ellos requieren de un proceso de aprendizaje exhaustivo para garantizar un reconocimiento acertado. En este trabajo se propone una mejora para reducir el tiempo de entrenamiento de un sistema OCR y mejorar la eficacia de los resultados. Lo anterior mediante la implementación de una red de reconstrucción antepuesta a la de clasificación de un sistema ya probado [1]. En la sección 2 de este trabajo, se exponen los detalles de la implementación del sistema OCR mencionado. En la sección 3, se presentan las modificaciones realizadas y la justificación de las mismas. Los resultados obtenidos con esta modificación se muestran en la sección 4 donde se hace también una comparación de ambos sistemas. Finalmente, en la sección 5 se presentan las conclusiones y el trabajo futuro.

2. Sistema de Reconocimiento Original

En [1], se presenta un sistema OCR de fácil implementación que se apoya en una RNA de tipo perceptrón con una sola capa y una neurona por cada caracter que se desea reconocer. En la Figura 2, se muestra un esquema del sistema para el reconocimiento de n caracteres. La entrada I es una imagen que contiene

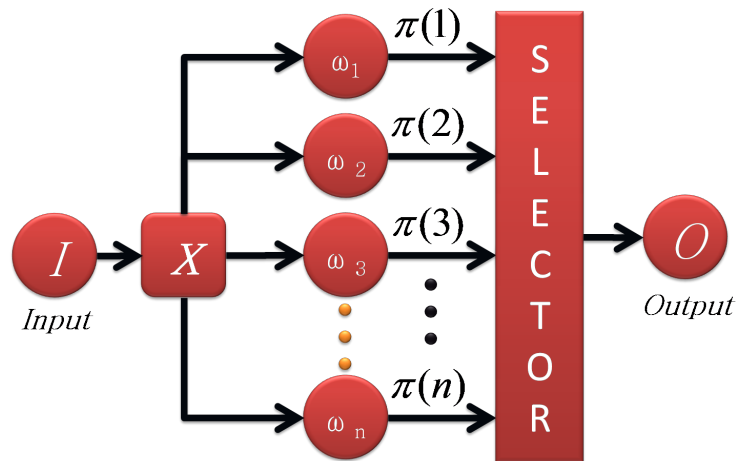


Fig. 2. Esquema del sistema OCR original

el caracter real a reconocer. Esta imagen debe ser convertida a una matriz M binaria de dimensiones $a \times b$ como se muestra en la Figura 3.

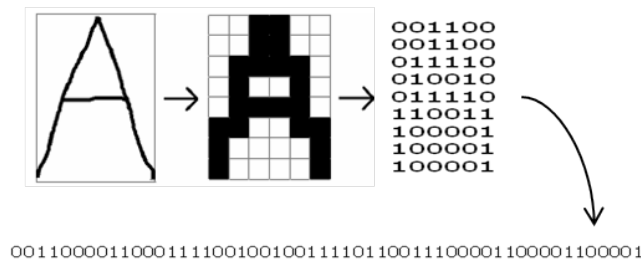


Fig. 3. Proceso de conversión de un carácter real a una matriz binaria

Por cada uno de los caracteres que el sistema sea capaz de reconocer, se requiere una neurona wn . Cada neurona contendrá una matriz de pesos Wn de las mismas dimensiones que M y su salida $\pi(n)$ será el factor de reconocimiento dada una nueva entrada. El conjunto de factores será la entrada de un selector,

que no es más que una red del tipo winner takes all, que indicará en su salida O cuál de las neuronas ha tenido un factor de reconocimiento mayor y, por lo tanto, qué carácter ha sido reconocido por el sistema OCR. Para el aprendizaje, se modificará la matriz de pesos Wn a razón de M con el siguiente método:

```
for all  $i = 1$  to  $x$  do  
  for all  $j = 1$  to  $y$  do  
    if  $M(i, j) == 0$  then then  
       $Wn(i, j) - -$   
    else  
       $Wn(i, j) ++$   
    end if  
  end for  
end for
```

Lo anterior, se realizará siempre que exista una nueva entrada I y se desee entrenar una neurona wn para su reconocimiento. El factor de reconocimiento $\pi(n)$ de cada una de las neuronas, es el cociente de la puntuación del candidato $\psi(n)$ y la puntuación del ideal $\mu(n)$ como se muestra en la ecuación 1.

$$\pi(n) = \frac{\psi(n)}{\mu(n)} \quad (1)$$

La puntuación del candidato, es una medida de qué tanto se parece la entrada M actual a la matriz de pesos Wn de la neurona n y se calcula de la siguiente manera:

```
for all  $i = 1$  to  $x$  do  
  for all  $j = 1$  to  $y$  do  
     $\psi(n) = \psi(n) + Wn(i, j) * M(i, j)$   
  end for  
end for
```

La puntuación del ideal, es una medida de la calificación que tendría el patrón de entrada M si fuera perfecto, es decir si solo tuviera 1's en las posiciones donde la matriz de pesos Wn tiene valores positivos. Este factor para la neurona n se calcula de la siguiente manera:

```
for all  $i = 1$  to  $x$  do  
  for all  $j = 1$  to  $y$  do  
    if  $Wn(i, j) > 0$  then  
       $\mu(n) = \mu(n) + Wn(i, j)$   
    end if  
  end for  
end for
```

El factor de reconocimiento tendrá siempre un rango de $(-\infty, 1]$. Para este caso, el sistema fue implementado para trabajar con matrices de 32×32 ya que permite un reconocimiento más preciso y abre la posibilidad de realizar una mejor comparación entre el sistema original y el que se propone.

3. Desarrollo

En la Figura 4, se muestra el esquema del sistema OCR con la modificación propuesta.

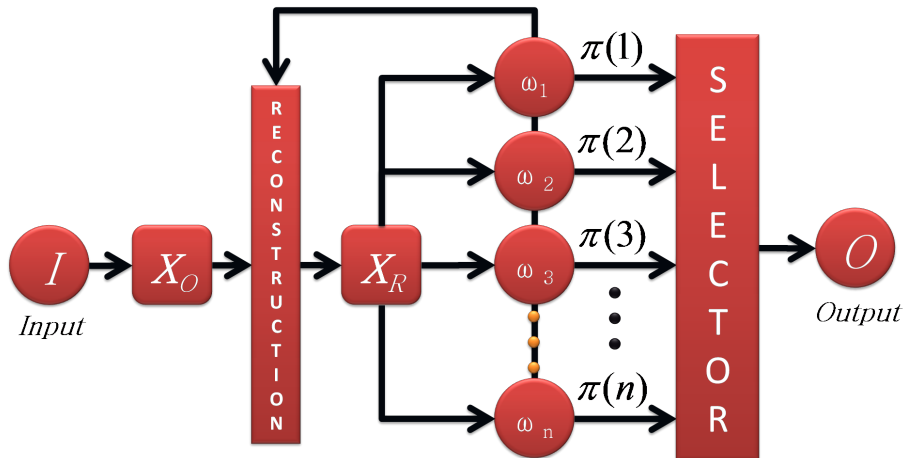


Fig. 4. Esquema del sistema OCR con la modificación propuesta

Esta modificación añade al OCR un módulo de reconstrucción que basa su operación en la información contenida en las neuronas del sistema original por lo que puede ser aplicado en cualquier sistema ya existente como una mejora al mismo. El módulo de reconstrucción se trata de una RNA de tipo *Hopfield* [6] la cual tiene como entrada la matriz binaria M_O y como salida la matriz reconstruida M_R del mismo tipo. El resto del proceso de reconocimiento será idéntico al descrito en la sección 2.

3.1. Aprendizaje del módulo de reconstrucción

Continuando con la terminología utilizada para describir el sistema OCR original, a continuación se describirá el proceso de aprendizaje del módulo de la red *Hopfield*. Para esta etapa, es necesario calcular la matriz de pesos de la nueva red basándose en la información aprendida por la red perceptrón. Para este caso es necesaria una matriz de 1024×1024 que será calculada por medio de la Ecuación 2.

$$W(i, j) = \sum_{s=1}^n W_s(\text{mod}(32, i), \text{div}(32, i) + 1 * W_s(\text{mod}(32, j), \text{div}(32, j) + 1) \quad (2)$$

La matriz de pesos de la red Hopfield se trata de una matriz simétrica cuadrada. Esta matriz deberá ser actualizada cada vez que el sistema OCR original

agregue una nueva neurona a su red perceptrón para reconocer un caracter adicional. Una vez que la matriz de pesos ha sido calculada, es necesaria la etapa de reconocimiento para realizar la reconstrucción de la matriz de entrada, es decir, obtener M_R a partir de M_O . Las funciones *div* y *mod* fueron aplicadas debido a que es necesaria la conversión de una matriz bidimensional a un vector unidimensional.

3.2. Reconocimiento del módulo de reconstrucción

El reconocimiento de una red *Hopfield*, trata de un proceso iterativo que permite la reconstrucción de una matriz de entrada en base a los patrones aprendidos durante el proceso de aprendizaje. Para lograr esta reconstrucción se utiliza la Ecuación 3.

$$M_0(t+1) = fh \left\{ \sum_{j=1}^n w(i,j) M_o(t) \right\} \quad (3)$$

Este proceso se repetirá indefinidamente hasta que $M_O(t) = M_O(t+1)$, en ese momento se tendrá una reconstrucción aceptable de acuerdo a los datos aprendidos por la red. La función *fh* es una función de umbralización definida en la Ecuación 4.

$$fh(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases} \quad (4)$$

La función de umbralización se utiliza para optimizar la calidad de la imagen para su posterior procesamiento, esta función puede ser modificada de acuerdo al problema que se esté resolviendo (calidad de adquisición de la imagen), sin embargo en base a trabajos previos se ha demostrado que estos valores son ideales para trabajar con sistemas OCR.

4. Experimentos y resultados

Las pruebas realizadas para este trabajo incluyeron caracteres tanto de dígitos como de letras. En la Figura 5 se muestra la colección de imágenes utilizadas para el entrenamiento del sistema OCR. En cada prueba se entrenó al sistema OCR una sola vez por cada uno de los patrones de números o letras. El sistema permite realizar un reconocimiento con o sin haber realizado una reconstrucción para fines comparativos. En las Figuras 6, 7, 8, 9, 10, 11, 12 y 13 se muestran diversas pruebas realizadas en el sistema indicando, en cada caso, si se trata de un reconocimiento por medio del sistema OCR original y por el propuesto. Para la primera prueba, fue dibujado a mano el dígito 1. En este caso, el sistema realizó un reconocimiento parcial del número con un 36% de certeza como se muestra en la Figura 6. En la Figura 7, se presenta el reconocimiento utilizando el módulo de reconstrucción. Para este caso, la



Fig. 5. Base de datos de imágenes utilizadas en el entrenamiento

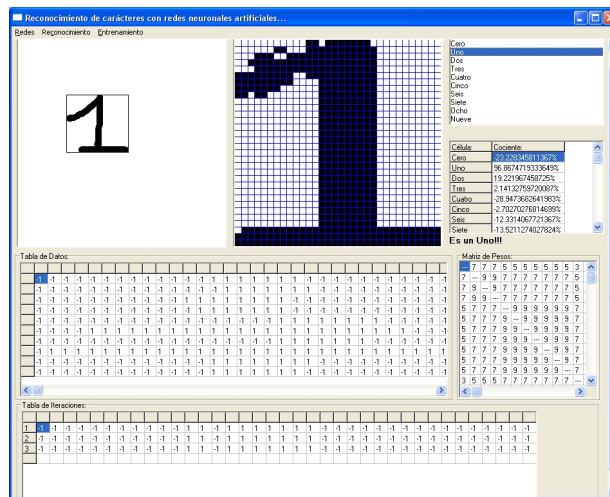


Fig. 6. Prueba con el dígito 1 en el sistema OCR original.

certeza del reconocimiento del número aumentó hasta un 97%. Para la segunda prueba con dígitos numéricos, se intentó realizar un dibujo del número 4 que se diferenciara en mayor proporción al dígito representado en la base de datos de imágenes. En la Figura 8 se muestra el reconocimiento realizado para esta prueba en el sistema OCR original, para este caso se obtuvo un porcentaje de reconocimiento del 19%. En la Figura 9, se presenta la reconstrucción realizada por medio de la red *Hopfield* y el reconocimiento del dígito con una certeza de 100%. Fueron realizadas otras pruebas con dígitos numéricos utilizando bases de datos de imágenes o de números escritos a mano, sin embargo, aunque los dígitos de entrada del sistema son siempre escritos a mano, los resultados de reconocimiento fueron muy superiores cuando la base de datos de entrenamiento se trata de caracteres con una tipografía estándar. También se utilizaron letras para los experimentos y, al igual que los resultados anteriores, los resultados del reconocimiento fueron mejores al utilizar el módulo de reconstrucción propuesto.

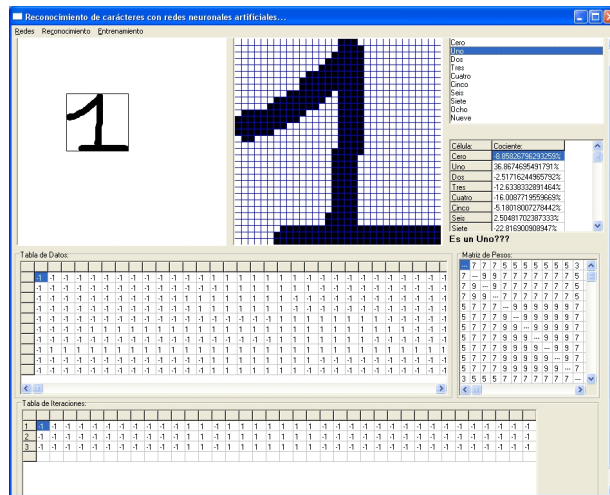


Fig. 7. Prueba con el dígito 1 en el sistema OCR modificado

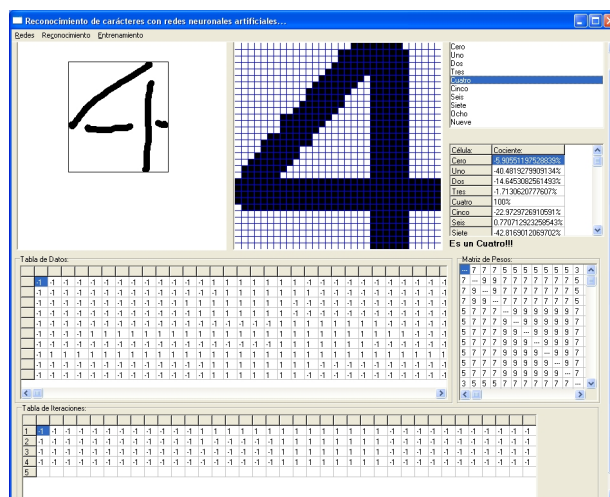


Fig. 8. Prueba con el dígito 4 en el sistema OCR original

En la figura 10 se presenta el reconocimiento de una letra A dibujada de forma asimétrica para dificultar el reconocimiento y reconstrucción del carácter. En el sistema OCR original se logra un reconocimiento con una certeza de 27.8%. En la figura 11, se muestra la reconstrucción realizada para la letra asimétrica y su reconocimiento con el sistema OCR modificado con una certeza del 100%. La última prueba realizada fue con la letra E. Para esta prueba se intentó dibujar una letra E cuyas características principales variaran respecto a la letra utilizada en el entrenamiento. Por ello, se dibujó una E redondeada en todas sus líneas y se intentó obtener un reconocimiento por medio del sistema OCR original. En

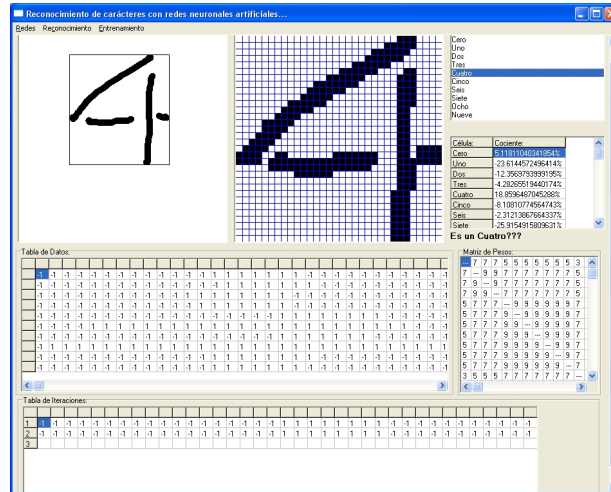


Fig. 9. Prueba con el dígito 4 en el sistema OCR modificado

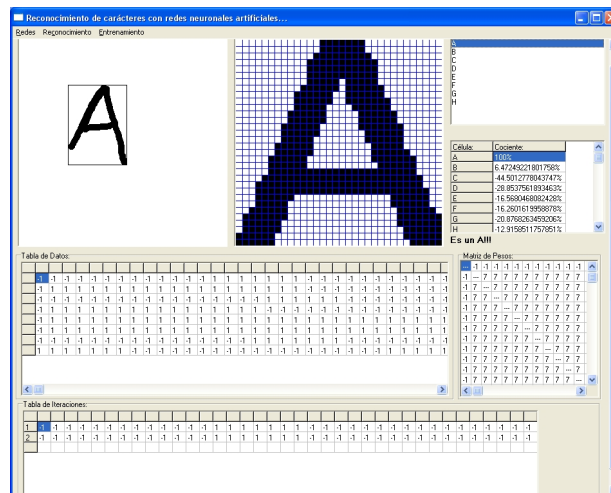


Fig. 10. Prueba con la letra A en el sistema OCR original

la Figura 12, se muestra el reconocimiento parcial de la letra E con una certeza del 31.16%. Finalmente, en la Figura 13 se muestra el reconocimiento logrado después de la reconstrucción (con el método propuesto) donde la certeza fue mejorada hasta un 70.9%

Considerando todos los resultados se puede resumir que el porcentaje de acierto del método propuesto fue de un 97.25% de reconocimiento contra un 30% de reconocimiento del sistema comparado.

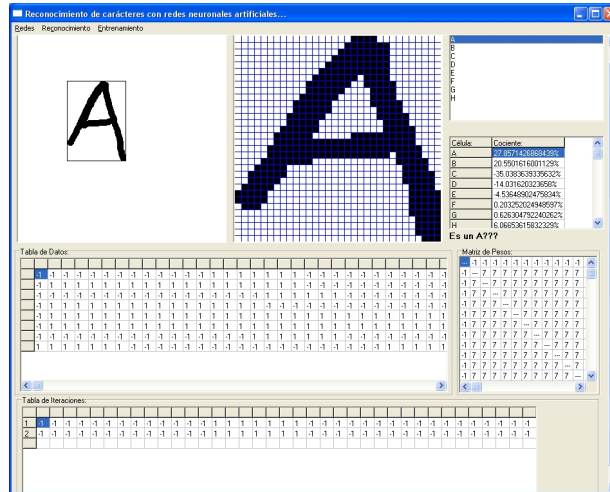


Fig. 11. Prueba con la letra A en el sistema OCR modificado

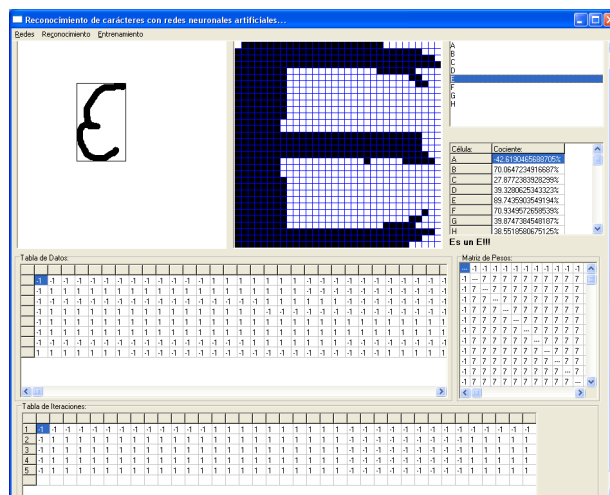


Fig. 12. Prueba con la letra E en el sistema OCR original

5. Conclusiones y trabajos futuros

La modificación propuesta para el sistema OCR es la unión de dos paradigmas que normalmente trabajan separados. Al agregar una red *Hopfield* para la reconstrucción de los datos de entrada, en este caso el caracter a reconocer, aumentó el porcentaje de la eficacia del sistema en la etapa de reconocimiento. El sistema probado fue entrenado únicamente con un patrón de cada una de las letras o números mientras la mayoría de los sistemas OCR requieren de un entrenamiento exhaustivo para lograr resultados aceptables. Los resultados

Daniela Moctezuma, Oscar S. Siordia

6. Hopfield, J.: Neural networks and physical systems with emergent collective computational abilities. *PNAS*, 79, 2554–2558 (1982)
7. Carolyn Strange, Daniel McNamara, Josh Wodak, Ian Wood: Mining for the Meanings of a Murder: The Impact of OCR Quality on the Use of Digitized Historical Newspapers. *Digital Humanities Quarterly*, 8 (2014)
8. Shoichi Taniguchi: Duplicate bibliographic record detection with an OCR-converted source of information. *J. Information Science*, 39, 153–168 (2013)

Factores en el reconocimiento facial de emociones y la integración de optimización evolutiva

Luis-Alberto Pérez-Gaspar, Santiago-Omar Caballero-Morales,
Felipe Trujillo-Romero

Universidad Tecnológica de la Mixteca, Huajuapán de León, Oaxaca,
México

`luis_335450@hotmail.com, scaballero@mixteco.utm.mx, ftrujillo@mixteco.utm.mx`

Resumen. En este artículo se presenta la integración de los Algoritmos Genéticos (Genetic Algorithms, GA) con Análisis de Componente Principal (Principal Component Analysis, PCA) y Redes Neuronales Artificiales (Artificial Neural Networks, ANNs) como un método para mejorar el reconocimiento de emociones en expresiones faciales. Para la evaluación del método dos bases de datos fueron consideradas: (1) una base de datos de emociones faciales creada a partir de usuarios Mexicanos, y (2) una base de datos estándar (Japanese Female Facial Expression, JAFFE). La eficiencia de tres sistemas de reconocimiento fue estudiada: (a) un sistema basado en PCA (Sistema PCA), (b) un sistema basado en PCA para reducción de dimensionalidad y ANNs como técnica de reconocimiento (Sistema PCA-ANN), y (c) un sistema PCA-ANN en donde la ANN es optimizada con GA (Sistema PCA-GA-ANN). El desempeño de cada sistema fue evaluado con ambas bases y resultados con mejoras estadísticamente significativas fueron obtenidos al integrar el GA.

Keywords: reconocimiento facial de emociones, algoritmos genéticos, redes neuronales, análisis de componente principal.

1. Introducción

En el campo de la interacción Humano-Robot y Humano-Computadora la tarea del reconocimiento de emociones se ha convertido en un reto importante. Esto dado que en los humanos la habilidad para interpretar las emociones es vital para lograr una comunicación efectiva. Esta capacidad es buscada para alcanzar una comunicación más natural y eficiente con sistemas artificiales [10].

Investigaciones en la detección de emociones humanas basadas en expresiones faciales ha conducido al desarrollo de varias técnicas. En [2] un sistema de reconocimiento de emociones integrando PCA y ANNs fue presentado. Las regiones de los ojos y de la boca fueron consideradas importantes para el reconocimiento. El algoritmo Canny para detección de bordes fue implementado para identificar

estas regiones y PCA fue usado para la extracción de características. Después una ANN fue construida para clasificar cada característica dentro de siete emociones consideradas (Neutro, Miedo, Felicidad, Tristeza, Enojo, Disgusto y Sorpresa). Las tasas de reconocimiento obtenidas fueron de 46.00% a 80.00% con la base de datos FEEDTUM (Facial Expressions and Emotions Database from the Technical University of Munich).

En [4] estas emociones fueron reconocidas mediante PCA y la base de datos JAFFE. Previo al reconocimiento las imágenes tuvieron un ajuste de tamaño y fueron pre-procesadas con el algoritmo Sobel para detección de bordes. La tasa de reconocimiento total de este enfoque fue alrededor del 67.14%. En cambio en [6] estas emociones fueron reconocidas a partir de video obtenido por una webcam. Para ello se emplearon las técnicas de PCA, Patrones Locales Binarios (Local Binary Pattern, LBP) y Máquinas de Soporte Vectorial (Support Vector Machine, SVM). Para SVM la tasa de reconocimiento fue alrededor del 71.50%, mientras que para el sistema híbrido fue del 91.25%. El reconocimiento para el sistema integrado PCA+LBP+SVM fue del 93.75%. Otro sistema que utilizó PCA y JAFFE fue presentado en [13] el cual reportó una tasa de clasificación total de 91.16% para las mismas emociones.

Una comparación entre 2D-LDA (2D Linear Discriminant Analysis), SVM, PCA y RBFN (Radial Basis Function Network) fue presentada en [11] para el reconocimiento de las siete emociones de la base de datos JAFFE con extracción de fondo. La combinación de 2D-LDA con SVM presentó tasas de reconocimiento de 95.10% (con la estrategia de deja-uno-fuera) y 94.13% (con la estrategia de validación cruzada). En [9] fue presentado el reconocimiento de cinco emociones (Enojo, Felicidad, Tristeza, Miedo y Neutro) en video mediante la técnica de ANNs. Tres regiones de interés fueron consideradas para la tarea de reconocimiento: ojo izquierdo, ojo derecho y boca. Una ANN fue construida para cada región y cada emoción. Las tasas de reconocimiento fueron de 73.00% a 87.00%. Finalmente en [8] el reconocimiento de cuatro emociones (Enojo, Felicidad, Tristeza y Neutro) fue llevado a cabo con una ANN optimizada por un Algoritmo Genético sobre la base de datos JAFFE. Para dicho trabajo las regiones de los ojos y boca fueron extraídas y una ANN para el modelado de error de reconocimiento fue integrada para mejorar el desempeño de la ANN principal. La tasa de reconocimiento total reportada fue del 85.00%.

Mientras la mayoría de los trabajos revisados utilizan bases de datos establecidas como JAFFE y FEEDTUM pocos trabajos hacen uso de información propia para sus experimentos. También no es conocida la arquitectura más adecuada para una ANN cuando más de una base de datos es usada para entrenamiento y prueba. El presente artículo explora sobre estas cuestiones desarrollando para ello una base de datos propia con usuarios Mexicanos (MX) y cuatro emociones principales: Enojo, Felicidad, Neutro y Tristeza [16,17].

La base de datos MX fue creada para ser añadida a la base de datos estándar JAFFE. Sistemas de reconocimiento basados en PCA y ANN fueron construidos con estas bases de datos. Para mejorar el reconocimiento de emociones un GA fue implementado como se presenta en [8] para identificar una arquitectura

adecuada para la ANN con PCA como técnica de extracción. De esta forma, considerando diferentes esquemas de entrenamiento/prueba con ambas bases de datos se observó que se pueden obtener mejoras estadísticamente significativas con más arquitecturas para la ANN. En contraste con [8] en este trabajo la optimización es llevada a cabo con diferentes esquemas de entrenamiento/prueba y la extracción de características es realizada. Esto conduce a una ANN más robusta para diferentes bases de datos.

El presente artículo se encuentra estructurado de la siguiente manera: en la Sección 2 se hace una descripción de la base de datos MX y las técnicas de pre-procesamiento aplicadas. Después en la Sección 3 se describe la aplicación de PCA para la reducción de dimensionalidad (extracción de características) y reconocimiento. En la Sección 4 se describe el GA diseñado para mejorar la estructura de la ANN para reconocimiento. El desempeño de los sistemas de reconocimiento desarrollados con las bases de datos emocionales es analizado en la Sección 5. Finalmente las conclusiones son presentadas en la Sección 6.

2. Bases de datos de emociones faciales

2.1. Base de datos mexicana (MX)

El contar con una base de datos emocional de usuarios Mexicanos fue importante para el desarrollo del presente trabajo. Esto porque las expresiones para la misma emoción pueden ser diferentes entre culturas y personas. También porque las características faciales entre personas de diferentes nacionalidades podrían afectar el desempeño del entrenamiento/prueba del sistema de reconocimiento. Para tener muestras representativas de rostros de usuarios Mexicanos, personas de las regiones Este y Suroeste de México fueron reclutadas.

Nueve usuarios Mexicanos (tres hombres y seis mujeres) fueron reclutados para la base de datos MX a los cuales se les pidió que expresaran las emociones de Enojo (EN), Felicidad (FE), Neutro (NE) y Tristeza (TR) [16,17]. Para tener consistencia con una base de datos como JAFFE, tres muestras fueron capturadas para cada emoción. Las muestras fueron tomadas con un fondo blanco y en condiciones estándar de iluminación [5]. Finalmente esto condujo a una base de datos de 108 imágenes (9 usuarios \times 3 muestras \times 4 emociones). Algunos ejemplos de la base de datos MX son presentados en la Figura 1.

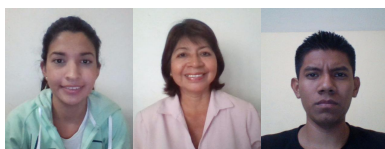


Fig. 1. Muestra de Imágenes de la Base de Datos Emocional Mexicana (MX).

2.2. Base de datos estándar (JAFFE)

La base de datos JAFFE (Japanese Female Facial Expression) ha sido ampliamente usada para el reconocimiento de emociones [4,7,11,13]. Esta base de datos consiste de 213 imágenes de 10 mujeres Japonesas que presentan aproximadamente tres diferentes expresiones para cada una de las siguientes emociones: Enojo, Felicidad, Neutro, Tristeza, Disgusto, Miedo y Sorpresa. En la Figura 2 algunos ejemplos de esta base de datos son presentados.



Fig. 2. Muestra de Imágenes de la Base de Datos Emocional JAFFE.

2.3. Pre-procesamiento

Una expresión emocional involucra la deformación de algunas partes de la cara como cejas, ojos y boca. En tal caso la información acerca del cabello, cabeza y hombros no es relevante para el proceso de reconocimiento. En su forma original las imágenes de la base de datos MX son en color y contienen información de estas características y del fondo. Un sistema detector de rostros basado en el algoritmo de Viola y Jones [15] fue implementado para adquirir y delimitar la región del rostro de los usuarios Mexicanos. El mismo sistema de detección de rostros fue aplicado a la base de datos JAFFE. La región facial adquirida después fue cambiada de tamaño a 256×256 píxeles. Para la base de datos MX una conversión de RGB a escala de grises con una corrección de contraste por un factor de 0.7 fue realizado. Algunos ejemplos de las regiones faciales extraídas son mostrados en la Figura 3.

3. PCA para extracción de características y reconocimiento

Un método común para reducir la dimensionalidad y obtener las características distintivas de una imagen es el Análisis de Componente Principal (PCA) [14]. Para la extracción de características cada imagen fue transformada en un vector columna P_{ij} en donde i =índice de la emoción y j =índice del usuario. Debido a que cuatro emociones fueron consideradas (EN=Enojo, FE=Felicidad, NE=Neutro, TR=Tristeza) $i=1, \dots, I$ en donde $I=4$. $j=1, \dots, J$ en donde $J=9$ para la base de datos MX y $J=10$ para la base de datos JAFFE. Los vectores P_{ij} que representan las imágenes de una base de datos fueron arreglados como se



Fig. 3. Muestra de Imágenes Pre-procesadas de las Bases de Datos MX y JAFFE.

presenta en la Figura 4 para crear la matriz S , la cual es la base para la reducción de dimensionalidad. El número total de columnas en S es obtenido como $J \times I$ mientras que el número de filas es $H = 256 \times 256$.

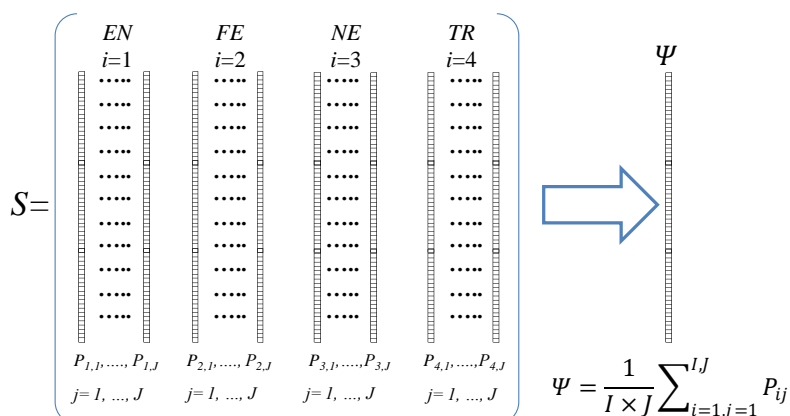


Fig. 4. Arreglo de Vectores (Matriz S) antes de la Reducción de Dimensión.

Después de que S es creada, el vector columna que corresponde a la media, Ψ , es obtenido y después es restado de cada vector columna P_{ij} en S . Esto produce la matriz A de vectores columna $\Phi_{ij} = P_{ij} - \Psi$ en donde:

$$A = [\Phi_{11} \dots \Phi_{1J} \quad \Phi_{21} \dots \Phi_{2J} \quad \Phi_{31} \dots \Phi_{3J} \quad \Phi_{41} \dots \Phi_{4J}] \quad (1)$$

De la matriz A la matriz de covarianza L es obtenida como $L = A^T A$ [12]. Esta matriz es usada para obtener los eigenvectores $v = eig(L)$ los cuales son la base para los eigenrostros definidos por $u = Av$. Finalmente al considerar R como el número de eigenrostros (aquellos con los eigenvalores más altos) un nuevo rostro Γ puede ser transformado en sus componentes de eigenrostro de la

siguiente manera:

$$\Omega = u^T(\Gamma - \Psi) = [\omega_1, \omega_2, \dots, \omega_R]^T \quad (2)$$

La reducción de dimensionalidad es lograda al ser la dimensión de Ω igual al número de muestras de entrenamiento ($I \times J$) [3]. También Ω representa las características de las muestras de entrenamiento las cuales pueden ser usadas para reconocimiento: los pesos ω_r describen las contribuciones de cada eigenrostro en la representación de la imagen del rostro de entrada. Este vector puede ser usado para reconocimiento de rostros/emociones encontrando la distancia euclidiana e más pequeña entre los vectores de pesos del rostro de entrada y los rostros de entrenamiento de la siguiente manera:

$$e = \|\Omega_{input} - \Omega\| \quad (3)$$

Si la etiqueta de la imagen consiste en la identidad del usuario entonces esto puede ser usado para el reconocimiento de rostros. Sin embargo, si la etiqueta de la imagen consiste de solamente el estado emocional expresado por el rostro (sin importar la identidad del usuario) entonces esto puede ser usado para el reconocimiento de emociones.

4. Sistema de reconocimiento integrado con PCA y ANN

Con PCA una reducción de dimensionalidad es obtenida la cual es representada por Ω . Esta matriz es normalizada y dada como entrada a una ANN para el entrenamiento. El algoritmo RP de propagación hacia atrás (backpropagation) fue usado para el proceso de entrenamiento. Para la ANN la salida fue considerada como un arreglo de vectores binarios representando la emoción asociada a cada vector P_{ij} (o ω_r). En la Figura 5 la estructura para la entrada y los datos de entrenamiento para la ANN es mostrada.

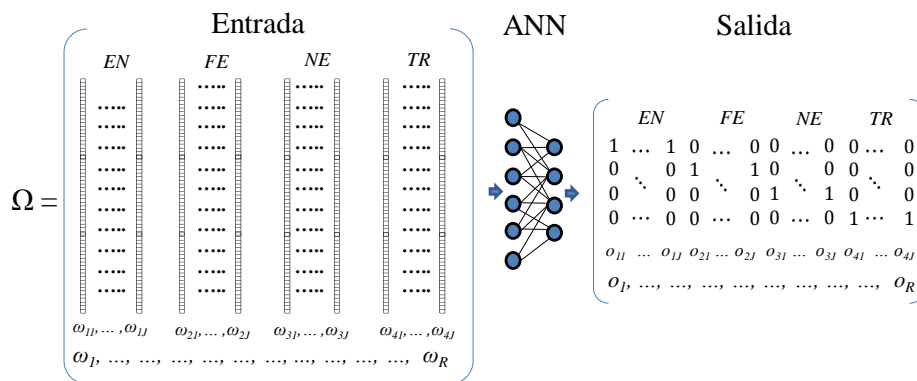


Fig. 5. Estructura de los Vectores de Entrada y Salida para la ANN.

4.1. Optimización de la ANN con algoritmos genéticos

Cuando una ANN es construida es necesario indicar su estructura, número de neuronas y función de activación para las capas ocultas. En la mayoría de los casos estos parámetros son propuestos de acuerdo a lo reportado por trabajos similares o son basados en experiencia previa. Con el objetivo de lograr un reconocimiento más alto con el clasificador en este trabajo la selección de estos parámetros se realizó mediante un Algoritmo Genético (GA).

Los GA se basan en el proceso natural de sobrevivencia de los individuos más aptos en una población. Los individuos con más aptitud para sobrevivir/adaptarse a un entorno se ganan un derecho (o son más probables) a reproducirse con otros individuos de igual o mejor aptitud. Estos individuos “padres” generan “hijos” que heredan las características de sus padres que les dieron una mejor aptitud para sobrevivir/adaptarse al entorno. Estas características se van mejorando en cada ciclo de reproducción de manera generacional. Este proceso es adaptado mediante los GA para resolver problemas de búsqueda de soluciones óptimas, en donde los individuos representan posibles soluciones.

El diagrama general de operación y módulos principales de un GA se presentan en la Figura 6. Para este trabajo la configuración de cada módulo se basó en el GA presentado en [8]. Dado el tamaño de la Población Inicial del GA este algoritmo se puede clasificar como un micro-GA [1]. Este tipo de GA puede converger de manera rápida después de algunas generaciones, generando soluciones de igual calidad que un GA convencional que puede tener poblaciones de hasta 1000 individuos [1].

5. Experimentos

Para los experimentos de reconocimiento de emociones una validación cruzada fue llevada a cabo. Las imágenes disponibles por emoción fueron enumeradas de la 1 a la 3 y los esquemas presentados en la Tabla 1 fueron considerados para el entrenamiento, optimización del GA y evaluación del sistema (prueba). Esto fue realizado para determinar la influencia de la muestra considerada para entrenamiento sobre el desempeño del reconocimiento con una muestra diferente (muestra de prueba).

Tabla 1. Esquemas de Imágenes Considerados para Entrenamiento, Optimización y Evaluación de los Sistemas de Reconocimiento.

Esquema	S ₁	S ₂	S ₃	S ₄	S ₅	S ₆
Entrenamiento	1	1	2	2	3	3
Optimización GA	3	2	3	1	2	1
Prueba	2	3	1	3	1	2

Para el reconocimiento de emociones se consideraron los siguientes factores:

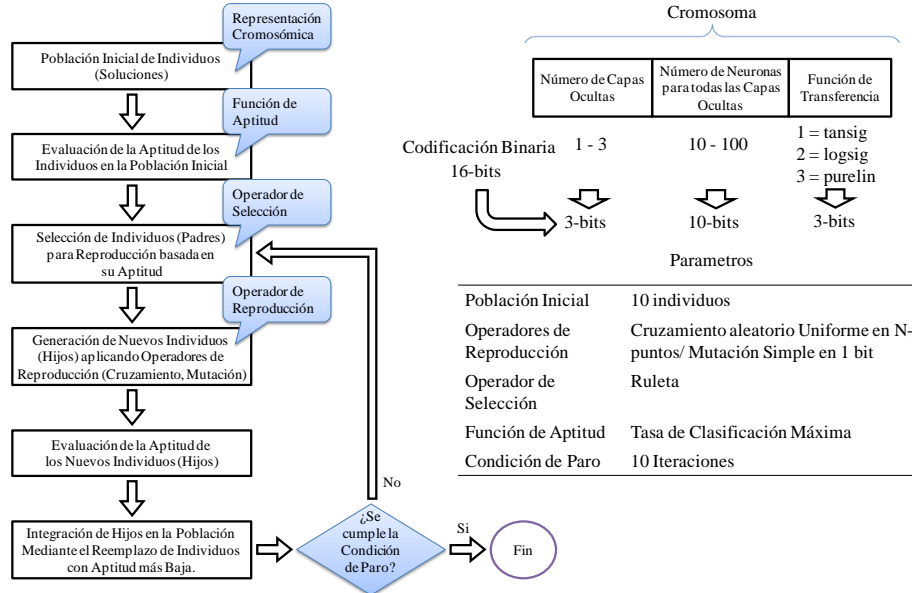


Fig. 6. Diagrama, Cromosoma y Parámetros de Configuración del Algoritmo Genético.

- Sistema de Reconocimiento: (a) sistema basado en PCA (Sistema PCA) como se presenta en la Sección 3, (b) sistema basado en PCA para reducción de dimensionalidad y ANN como técnica de reconocimiento (Sistema PCA-ANN) como se presenta en la Sección 4, y (c) Sistema PCA-ANN en donde la ANN es optimizada con el GA (Sistema PCA-GA-ANN) como se presenta en la Sección 4.1.
- Emociones: (a) Enojo, (b) Felicidad, (c) Neutro, (d) Tristeza.
- Base de Datos de Emociones Faciales: (a) base de datos propia de usuarios Mexicanos (MX), (b) base de datos estándar (JAFFE), y (c) la unión de las bases de datos MX+JAFFE.
- Esquema de Entrenamiento-Optimización-Prueba: S_1 , S_2 , S_3 , S_4 , S_5 y S_6 como se presenta en la Tabla 1.

Un análisis factorial con Minitab (v17.2.1) fue realizado para determinar el impacto de cada factor en la tarea de reconocimiento de emociones. En la Figura 7 las Gráficas de Interacción para la tasa media de reconocimiento de emociones a través de todos los factores son presentadas.

Considerando el “Sistema de Reconocimiento” (Método de Clasificación) como el factor principal las Figuras 7(a),7(b) y 7(c) presentan la siguiente información:

- Las tasas de reconocimiento más altas para Enojo, Felicidad y Neutro son obtenidas con el sistema integrado PCA-GA-ANN en comparación con los

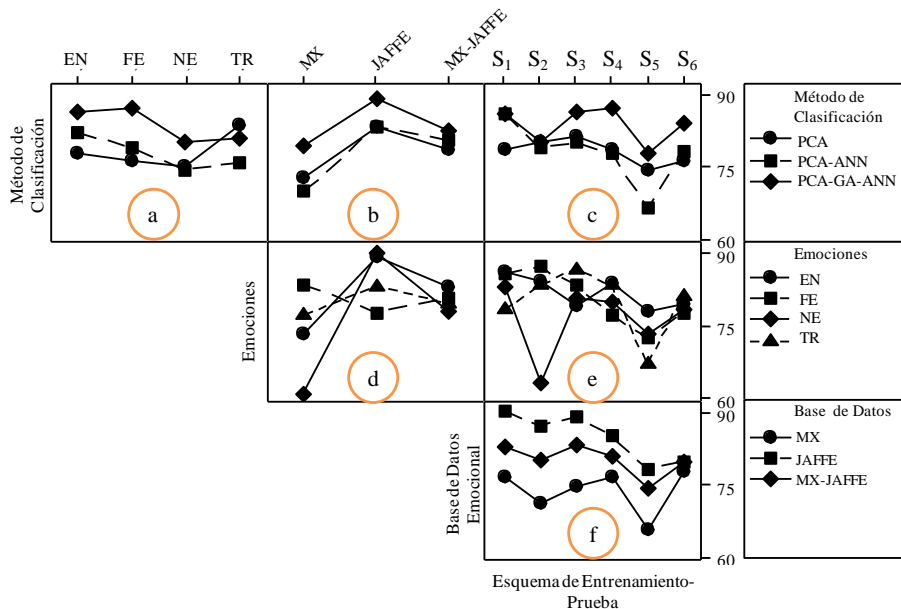


Fig. 7. Gráficas de Interacción de la Tasa Promedio de Reconocimiento en Muestras de Prueba.

sistemas PCA y PCA-ANN. Para Tristeza la tasa de reconocimiento más alta es obtenida con el Sistema PCA.

- Las tasas de reconocimiento más altas son obtenidas para la base de datos JAFFE en comparación con la base de datos MX. Para ambas bases de datos una tasa mayor de reconocimiento es conseguida con el Sistema PCA-GA-ANN.
- Independientemente del esquema de Entrenamiento-Optimización-Prueba el Sistema PCA-GA-ANN presenta una mayor tasa de reconocimiento que los otros sistemas. Sin embargo hay una variabilidad significativa entre las tasas de reconocimiento de estos esquemas.

Cuando se consideran las “Emociones” como el factor principal en la Figura 7(d) se observa que Felicidad es la emoción que mejor se reconoce en la base de datos MX. En tanto esta emoción tiene la tasa de reconocimiento más baja en la base de datos JAFFE. La emoción Neutro tiene la tasa de reconocimiento más baja en la base MX y se reconoce mejor en la base de datos JAFFE. La Figura 7(e) da evidencia del efecto de los esquemas de Entrenamiento-Prueba sobre el desempeño del reconocimiento. La tasa de reconocimiento para Neutro es severamente afectada bajo el esquema S_2 (entrenando con la imagen 1, optimizando con la imagen 2, y probando con la imagen 3). Una disminución general en la tasa de reconocimiento es observada bajo el esquema S_5 (entrenando con la imagen 3, optimizando con la imagen 2, y probando con la imagen 1).

Finalmente cuando se considera la “Base de Datos de Emociones Faciales” como el factor principal la Figura 7(f) presenta la base de datos JAFFE con la mayor tasa de reconocimiento a través de todos los esquemas de Entrenamiento-Optimización-Prueba. La base de datos MX tiene la tasa más baja a través de todos los esquemas. También se observa una disminución general en la tasa de reconocimiento bajo el esquema S_5 . En general el Sistema PCA-GA-ANN presenta la mayor tasa promedio de reconocimiento (83.7462%) en comparación con el Sistema PCA-ANN (77.9083%) y el Sistema PCA (78.2501%). Minitab v17.2.1 fue utilizado para realizar una prueba ANOVA para evaluar la significancia estadística de estos resultados. Esta prueba concluyó que el desempeño del Sistema PCA-GA-ANN es estadísticamente diferente del desempeño de los Sistemas PCA y PCA-ANN con $p < 0.05$ ($p=0.003$).

Un tema importante acerca de estos experimentos es la estructura de la ANN obtenida para cada esquema de Entrenamiento-Prueba en el Sistema PCA-GA-ANN. La Tabla 2 presenta un resumen general de estas estructuras para cada esquema y base de datos.

Tabla 2. Desempeño del Sistema PCA-GA-ANN con Estructura Específica y Promedio para la ANN.

Base de Datos MX										
Estructura Específica para la ANN para cada Esquema de Entrenamiento-Optimización-Prueba					Estructura Promedio para la ANN					
Ent	Opt	Pru	Estructura	EN	FE	NE	TR	Promedio		
1	2	3	[1 71 3]	80.37	88.88	45.55	61.85	69.16	88.89	
1	3	2	[2 78 3]	84.07	88.51	72.96	77.40	80.74	88.89	
2	1	3	[1 56 3]	88.88	100.00	77.77	79.25	86.48	88.89	
2	3	1	[2 64 3]	72.96	97.03	77.77	72.96	80.18	77.78	
3	1	2	[1 30 3]	82.59	88.88	77.77	100.00	87.31	88.89	
3	2	1	[1 42 3]	59.25	90.00	60.37	79.62	72.31	55.56	
Estructura Promedio = [1.33 56.83 3] ≈ [1 57 3]				Tasa de Reconocimiento Global = 79.36					Tasa de Reconocimiento Global = 82.41	

Base de Datos JAFFE										
Estructura Específica para la ANN para cada Esquema de Entrenamiento-Optimización-Prueba					Estructura Promedio para la ANN					
Ent	Opt	Pru	Estructura	EN	FE	NE	TR	Promedio		
1	2	3	[3 62 3]	94.00	95.66	89.66	92.66	93.00	100.00	
1	3	2	[3 73 3]	93.33	90.66	100.00	82.66	91.66	100.00	
2	1	3	[3 50 2]	92.66	79.33	94.00	91.66	89.41	100.00	
2	3	1	[3 28 3]	100.00	85.66	100.00	90.00	93.92	100.00	
3	1	2	[2 59 2]	84.33	80.66	87.00	79.00	82.75	90.00	
3	2	1	[1 18 2]	98.00	71.33	96.33	75.00	85.17	100.00	
Estructura Promedio = [2.5 48.33 2.5] ≈ [3 48 3]				Tasa de Reconocimiento Global = 89.32					Tasa de Reconocimiento Global = 94.58	

Base de Datos MX + JAFFE										
Estructura Específica para la ANN para cada Esquema de Entrenamiento-Optimización-Prueba					Estructura Promedio para la ANN					
Ent	Opt	Pru	Estructura	EN	FE	NE	TR	Promedio		
1	2	3	[3 52 3]	89.12	86.14	60.52	78.77	78.64	89.47	
1	3	2	[2 79 3]	91.40	91.40	84.03	74.38	85.30	94.74	
2	1	3	[1 22 2]	88.94	85.61	82.28	88.42	86.31	94.74	
2	3	1	[2 56 3]	90.87	83.85	80.70	87.71	85.78	94.74	
3	1	2	[2 35 3]	83.85	83.15	83.50	80.52	82.76	89.47	
3	2	1	[1 11 2]	81.57	83.33	75.43	65.96	76.57	94.74	
Estructura Promedio = [1.8 42.5 2.66] ≈ [2 43 3]				Tasa de Reconocimiento Global = 82.56					Tasa de Reconocimiento Global = 86.40	
Tasa de Reconocimiento Total = 83.75					Tasa de Reconocimiento Total = 87.80					

En general el GA determinó más capas ocultas (3) para la base de datos JAFFE que para la base MX (1). Cuando ambas bases se unen el GA determinó un valor intermedio. El número promedio de neuronas estuvo dentro del rango de 43 a 57 neuronas y para todas las bases de datos y esquemas de prueba la tercera

función de transferencia (purelin) fue definida. Tomando los valores promedio de las ANNs una estructura general de ANN fue estimada para cada base de datos y todos los esquemas de Entrenamiento-Prueba. Como se presenta en la Tabla 2 con la ANN general para ambas bases de datos la tasa de reconocimiento total incrementó de 83.7462 % a 87.7979 %.

6. Conclusiones

Sin importar el sistema de reconocimiento una tasa mayor fue obtenida con la base de datos JAFFE en comparación con la base de datos propia MX. Como se presenta en la Tabla 2 con el sistema propuesto PCA-GA-ANN una tasa de reconocimiento de 94.58 % fue alcanzada con la base de datos JAFFE y 82.41 % con la base de datos MX a través de diferentes esquemas de Entrenamiento-Prueba. La integración del GA para optimización de la estructura de la ANN condujo a mejoras estadísticamente significativas sobre las tasas de reconocimiento para ambas bases.

Una situación interesante respecto a la base de datos JAFFE es que fue creada con condiciones adecuadas de postura e iluminación. Estas condiciones no fueron muy estrictas para la creación de la base MX ya que fue considerado el representar a los usuarios en ambientes menos restringidos. Esto puede explicar la diferencia entre las tasas de reconocimiento presentadas en la Figura 7 y la Tabla 2. Otra explicación puede ser que los usuarios Mexicanos tienen sus formas propias y diferentes de expresar sus emociones. Particularmente cuando ambas bases de datos son consideradas juntas la extracción de características con PCA puede conducir a una combinación de emociones.

También es importante probar cualquier técnica para el reconocimiento de emociones con diferentes esquemas de imágenes para entrenamiento y prueba. Esto es porque como se presenta en la Figura 7 estos esquemas influyen en el desempeño total del sistema de reconocimiento.

La información presentada en la Figura 7 también mostró que el reconocimiento de una emoción puede ser dependiente de las características de la base de datos y del sistema de reconocimiento mismo. De esta forma, el desarrollo de un sistema con múltiples reconocedores (por ejemplo, PCA para Tristeza y ANN para Enojo, Felicidad y Neutro) pueden ser una aproximación adecuada para mejorar el reconocimiento de emociones. También la integración de un sistema de reconocimiento de emociones basado en voz puede llevar a mejores resultados.

Referencias

1. Bakare, G.A., Venayagamoorthy, G.K., Aliyu, U.O.: Reactive Power and Voltage Control of the Nigerian Grid System Using Micro-Genetic Algorithm. In: Proc. of the Power Engineering Society General Meeting. pp. 1916–1922 (2005)
2. Filko, D., Martinović, G.: Emotion recognition system by a neural network based facial expression analysis. *AUTOMATIKA* 54(2), 263–272 (2013)

3. Gil, A., Benavides, M., Guilarte, Y., Marquez, M.: Sistema para el reconocimiento e identificación de rostros a través de fotografías. *Revista Ciencia e Ingeniería* 29(2), 131–136 (2008)
4. Gosavi, A.P., Khot, S.R.: Facial expression recognition using principal component analysis. *Int. Journal of Soft Computing and Engineering* 3(4), 258–262 (2013)
5. López, J.M., Cearreta, I., Garay, N., López de Ipiña, K., Beristain, A.: Creación de una base de datos emocional bilingüe y multimodal. In: *Proc. of the 7th Spanish Human Computer Interaction Conference, Interaccion 2006*. vol. 6, pp. 55–66 (2006)
6. Luo, Y., Wu, C., Zhang, Y.: Facial expression recognition based on fusion feature of PCA and LBP with SVM. *Optik* 124, 2767–2770 (2013)
7. Lyons, M., Budynek, J., Akamatsu, S.: Automatic classification of single facial images. *IEEE Trans. Patt. Anal. Mach. Intell.* 21, 1357–1362 (1999)
8. Pérez-Gaspar, L., Caballero-Morales, S.O., Trujillo-Romero, F.: Error modelling approach based on artificial neural network for face emotion recognition. *Research in Computing Science* 78, 21–30 (2014)
9. Rao, K.S., Saroj, V.K., Maity, S., Koolagudi, S.G.: Recognition of emotions from video using neural networks models. *Expert Systems with Applications* 38, 13181–13185 (2011)
10. Ratliff, M., Patterson, E.: Emotion recognition using facial expressions with active appearance models. In: *Proc. of the IASTED International Conference on Human-Computer Interaction*. pp. 92–138 (2008)
11. Shih, F.Y., Chuang, C.-F., Wang, P.S.P.: Performance comparisons of facial expression recognition in JAFFE database. *International Journal of Pattern Recognition and Artificial Intelligence* 22(3), 445–459 (2008)
12. Slakovic, M., Jevtic, D.: Face recognition using eigenface approach. *Serbian Journal of Electrical Engineering* 9(1), 121–130 (2012)
13. Thuseethan, S., kuhanesan, S.: Eigenface based recognition of emotion variant faces. *Computer Engineering and Intelligent Systems* 5(7), 31–37 (2014)
14. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3(1), 71–86 (1991)
15. Viola, P., Jones, M. J.: Robust real-time face detection. *International Journal of Computer Vision* 57(2), 137–154 (2004)
16. Yildirim, S., Bulut, M., Lee, C.M., Kazemzadeh, A., Busso, C., Deng, Z., Lee, S., Narayanan, S.: An acoustic study of emotions expressed in speech. In: *Proc. Int. Conf. Spoken Language Processing (ICSLP '04)*. vol. 1, pp. 2193–2196 (2004)
17. Yu, F., Chang, E., Xu, Y. Q., Shum, H.Y.: Emotion detection from speech to enrich multimedia content. In: *Proc. IEEE Pacific-Rim Conf. Multimedia 2001*. vol. 1, pp. 550–557 (2001)

Implementación y control de un conmutador de imágenes multifocales

Leonel Morales-Botello, J. Humberto Sossa-Azuela, Elsa Rubio-Espino

Instituto Politécnico Nacional, Centro de Investigación en Computación,
México

lmoralesbotello@gmail.com, hsossa@cic.ipn.mx, erubio@cic.ipn.mx

Resumen. En este trabajo se presenta un sistema basado en control difuso que selecciona, entre un conjunto de imágenes con enfoques en diferentes planos, aquella imagen en la cual el objeto de interés presenta mayor nitidez. Este tipo de imágenes generalmente asociadas a las cámaras plenópticas, proveen información útil en el procesamiento digital de imágenes. El sistema difuso tiene como entradas el nivel de saturación correspondiente a la transformación HSV de la imagen, y la distancia existente entre el valor mínimo y máximo en su matriz de color RGB. Este sistema trabaja con matrices representantes de una zona que el usuario selecciona y obtiene el plano de la imagen con mayor nitidez.

Palabras clave: plenóptica, lógica difusa, control difuso, procesamiento digital de imágenes, HSV.

1. Introducción

Una cámara plenóptica es aquel dispositivo que puede tomar fotografías y ser enfocadas después de haber sido tomadas, siendo Lytro la primera marca que ha lanzado una cámara plenóptica de consumo personal [6].

La principal ventaja de una cámara plenóptica es que puede adquirir la información de profundidad (o diferentes distancias focales) en un sólo disparo, algo que una cámara convencional no lo puede hacer. Sin embargo, las cámaras plenópticas demandan un poder de cómputo considerablemente alto, siendo un reto el mantener la relación entre calidad y resolución asequible existente en este tipo de cámaras. Por lo cual, se ha intentado forzar a los fabricantes de sensores ópticos, a repensar los estándares de imágenes digitales actuales. La información de una cámara plenóptica (denominado en inglés como *light-field camera*), matemáticamente hablando es una función que los especialistas en óptica llaman plenóptica.

Siendo una función de cinco dimensiones parametrizada como la siguiente igualdad $P_{\tau} = P(x, y, z, \theta, \varphi)$, donde, los valores de x y y especifican la posición de cada punto en la imagen, z la profundidad de cada elemento de la imagen y tanto θ como φ representan cada rayo entrante al sensor óptico, describen el total de luz que incide en una región dada en cualquier momento. Al medir la

luz una región libre de cualquier obstrucción, se puede obtener la información necesaria para averiguar de dónde proviene un rayo de luz o en su caso hacia donde se dirige con sólo tomar cuatro de los cinco parámetros de P_τ . Por esta razón, se pueden generar numerosas imágenes correspondientes a las diversas separaciones existentes entre la lente y el sensor óptico, donde, el objeto de interés puede encontrarse en los llamados primeros planos o en el caso de las vistas distantes llamadas de imágenes de horizonte lejano [3].

Recientemente, las cámaras plenópticas han sido el foco de atención en la industria de la telefonía celular, por la capacidad de manipular diferentes enfoques de las imágenes posterior al disparo, obteniendo las imágenes más nítidas según la zona de interés del usuario. Sin embargo, el uso de la información proveída por las cámaras plenópticas en el ambiente científico [2], se ha extendido en aplicaciones de procesamiento digital de imágenes, tales como el reconocimiento de rostros en multitudes, reconstrucción 3D de objetos o la creación de imágenes con súper resolución [7,4,1].

2. Bases teóricas

El presente trabajo se basa en las premisas básicas del procesamiento digital de imágenes, así como en el uso de espacios de color, imágenes con profundidad de campo y también en el diseño de sistemas expertos con control difuso. Por ello, en esta sección se exponen las bases necesarias para comprender el manejo y procesamiento de imágenes multifocales.

2.1. Procesamiento digital de imágenes

Una imagen digital denotada como una función $f(x, y)$ es un arreglo bidimensional $\in \mathbb{Z} \times \mathbb{Z}$ si (x, y) son enteros de $\mathbb{Z} \times \mathbb{Z}$ y es una f es una función que asigna a cada par (x, y) un número de \mathbb{Z} .

Por otra parte, un píxel es denominado como la menor unidad homogénea en color que forma parte de una imagen digital, ya sea esta imagen obtenida de una fotografía, un fotograma de un video o un gráfico.

La vecindad de un píxel p , es la composición de píxeles que lo rodean directamente en sus coordenadas (x, y) y son conocidos como los vecinos de p .

En el caso de las vecindades existen dos variantes, las cuales, se denomina como vecindad 4 conectada a los píxeles que tienen 2 vecinos horizontales y 2 vecinos verticales ejemplificados en (ec. 1), y se define como vecindad 8 conectada a los píxeles que tienen como vecinos a (ec. 1) y además a los píxeles contenidos en sus diagonales (ec. 2).

$$\begin{aligned} &(x + 1, y) \\ &(x - 1, y) \\ &(x, y + 1) \\ &(x, y - 1) \end{aligned} \tag{1}$$

$$\begin{aligned}
 &(x + 1, y + 1) \\
 &(x + 1, y - 1) \\
 &(x - 1, y + 1) \\
 &(x - 1, y - 1)
 \end{aligned}
 \tag{2}$$

En la Fig. 1 se muestra una representación general de un píxel con su vecindad 4 y 8 conectada.

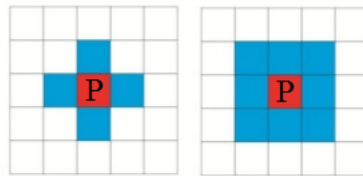


Fig. 1. Píxel con vecindad 4 y 8 conectada.

En el caso del estudio de la plenóptica, no es suficiente el concepto de vecindad para discernir de los elementos que pertenecen a diferentes planos en el elemento z de P_τ . Por lo cual, se propone el uso del concepto región y sección de un píxel. Una región conectada R , es un conjunto de puntos $p(x, y)$ tales que existe un camino que comienza en $P1$ y termina en $P2$ para todo $P1$ y $P2 \in \mathbb{R}$.

El contorno externo de una región R es el conjunto de píxeles de \mathbb{R} con, al menos, un vecino de fondo y que al momento de ser recorrido, en el sentido horario, la región se encuentra a la derecha. La sección de un píxel p es una región conectada donde p es el centro de la sección, siendo n la distancia que existe entre el centro de la sección y el contorno externo. La Fig. 2 muestra la representación gráfica de las secciones con distancia $n = 0, 1, 2, 3$.

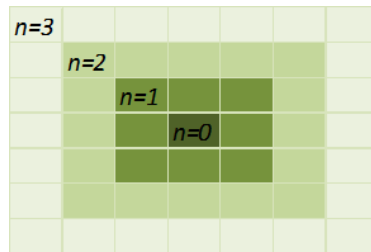


Fig. 2. Ejemplificación de sección de un píxel con $n = 3$.

2.2. Control difuso

El control difuso es una estrategia de control basada en conceptos de lógica difusa, la cual procesa información de forma cualitativa en lugar de hacerlo cuantitativamente.

Lo anterior permite emular y reproducir la experiencia obtenida por el conocimiento de expertos en un sistema. Tales características han permitido que dicho esquema de control, en ciertas aplicaciones, brinden mejores resultados que las estrategias convencionales.

En la lógica formal se evalúan proposiciones en un lenguaje simbólico, de manera que es capaz de abstraerse del mismo sus contenidos. De tal modo, a cualquier proposición se le es posible asignar un valor de verdad (sea el caso de falso o verdadero).

Dado lo anterior, mientras un conjunto en lógica formal se representa con la simple enumeración o una caracterización de sus elementos, en la teoría difusa se representa como el conjunto de pares ordenados de elementos con su correspondiente valor de pertenencia al conjunto, es decir, siendo F un conjunto difuso, éste se define como $F = \{(u, \mu_F(u)) | u \in U\}$.

El valor de posibilidad de pertenencia a un conjunto se conoce como grado de membresía $\mu_i(j)$, la cual es una función que asigna a un elemento j en el intervalo $[0, 1]$ que define la posibilidad de pertenencia al conjunto i . En otras palabras, $\mu_F : U \rightarrow [0, 1]$.

De acuerdo con [5], se dicen tres los pasos para poder llevar a nivel implementación un controlador difuso (Fig. 3): la fuzificación, el mecanismo de inferencia y la defuzificación. El primero de ellos consiste en la traducción de variables medibles en elementos difusos (cualitativos). El mecanismo de inferencia desprende conclusiones a partir de las variables fuzificadas. Dichas conclusiones no son cuantificables aún, y deben pasar por un proceso de defuzificación para poder ser empleadas como acciones de control de forma automática.

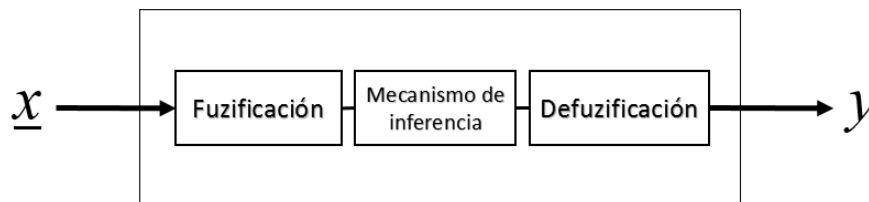


Fig. 3. Estructura general de los sistemas difusos.

2.3. Modelo de color hsv

En la teoría de color existen diversos tipos de espacios utilizados comúnmente por dispositivos que muestran o utilizan imágenes. El modelo HSV fue creado en 1978 por Alvy Ray Smith [8]. Generalmente, el espacio de color HSV se considera de una transformación no lineal del espacio RGB. Dado que es más confiable que el modelo RGB en progresiones de color y el cálculo a éste espacio de color es muy simple se ha adoptado ampliamente en el procesamiento digital de imágenes para el desarrollo de aplicaciones que requieran un filtro por color. En la Fig. 4 se muestra el espacio de color HSV donde el modelo de color se basa en:

1. Matiz H: Es un ángulo entre 0° y 360° , que corresponde a un color RGB. Para el color blanco se puede poner cualquier matiz y saturación, siempre que el valor de luminosidad sea 1, para el color negro se puede poner cualquier color y saturación, siempre que se ponga un valor de 0.
2. Saturación S: Representa la distancia al eje de brillo negro-blanco, los valores posibles van de 0 a 1.
3. Valor V: Representa la altura en el eje blanco-negro. Los valores posibles van de 0 a 1, el valor 0 siempre es negro y dependiendo de la saturación, 1 podría ser blanco o un color más o menos saturado.

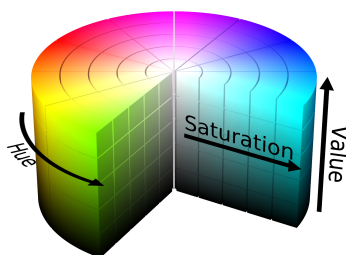


Fig. 4. Modelo de color hsv.

3. Desarrollo

En este trabajo un plano dentro de un fotograma, se define como una imagen cuya área interés (segmento local) presenta una mayor nitidez y una escena es el conjunto de planos con número de elementos n dentro de la función $P(x, y, z, \theta, \varphi)$ donde $z \in \mathbb{Z}^+$.

La Fig. 5 ejemplifica una escena compuesta por tres planos, donde el primer plano está enfocando un cuádrimotor, en el segundo plano el desenfoque del cuádrimotor empieza a ser notorio y el enfoque se traslada a los robots Bioloid, para el tercer plano el robot NAO está totalmente enfocado junto con el fondo, sin embargo el cuádrimotor ha llegado a su máximo desenfoque.



Fig. 5. Escena compuesta de tres planos.

3.1. Declaración de conjuntos difusos

El control difuso transforma las políticas de control indicadas por un lenguaje natural a funciones numéricas a través de conjuntos difusos e inferencias difusas, utilizando la computadora para realizar el control determinado [9].

La variable de entrada para el controlador difuso es la diferencia entre valores de RGB de la región seleccionada que se denominó DV perceptible en la Fig. 6(a) a y como segundo parámetro se escogió la saturación de la imagen resultante de la transformación HSV de la región, designada como S (véase Fig. 6 (b)), la variable de salida es la cantidad de nitidez N mostrada en la Fig. 7. Por lo tanto, se obtiene un sistema difuso de doble entrada y una salida para completar la tarea.

En la designación de valores para los conjuntos difusos se seleccionó el valor de $[0, 255]$ como el rango básico en enteros para la argumentación de DV y cinco subconjuntos difusos: MP (muy pequeña), P (pequeña), M (mediana), G (grande) y MG (muy grande). Por otro lado, se seleccionó $[0, 1]$ como el rango de argumentación para S y cuatro subconjuntos: B (bajo), M (medio), A (alto) y MA (muy alto).

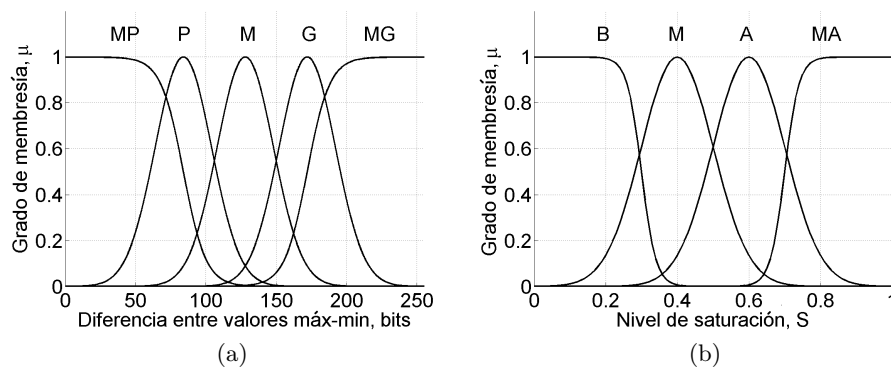


Fig. 6. (a) Funciones de membresía de los subconjuntos de DV. (b) Funciones de membresía de los subconjuntos de S.

Para el rango de salidas se seleccionó $[0, 5]$ con cuatro subconjuntos: NN (no nítido), PN (poco nítido), N (nítido) y MN (muy nítido), utilizando campanas de Gauss y curvas sigmoides como funciones de membresía de los subconjuntos DV y S, mientras que la salida N tiene singletons como funciones de membresía.

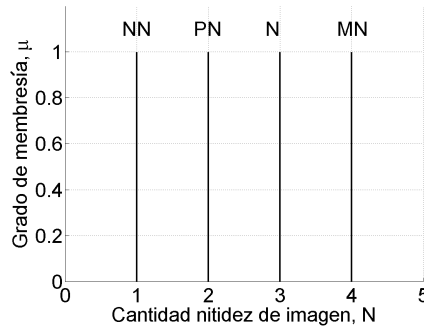


Fig. 7. Singletons de salida del conjunto N.

3.2. Establecimiento de reglas difusas

En el desarrollo de este trabajo se establecieron las siguientes reglas de control de acuerdo a la experiencia que ha sido acumulada por parte del uso del sistema.

- si DV=MP y S=B entonces N=NN
- si DV=MP y S=M entonces N=NN
- si DV=MP y S=A entonces N=PN
- si DV=MP y S=MA entonces N=N
- si DV=P y S=B entonces N=NN
- si DV=P y S=M entonces N=PN
- si DV=P y S=A entonces N=PN
- si DV=P y S=MA entonces N=N
- si DV=M y S=B entonces N=PN
- si DV=M y S=M entonces N=PN
- si DV=M y S=A entonces N=N
- si DV=M y S=MA entonces N=N
- si DV=G y S=B entonces N=PN
- si DV=G y S=M entonces N=N
- si DV=G y S=A entonces N=N
- si DV=G y S=MA entonces N=MN

La Tabla 1 muestra las reglas difusas que se proponen para la selección de la cantidad de nitidez según el área de interés del usuario.

Tabla 1. Reglas para controlador difuso

Nitidez		Diferencia entre valores				
		MP	P	M	G	MG
Saturación	B	NN	NN	PN	PN	N
	M	NN	PN	PN	N	N
	A	PN	PN	N	N	MN
	MA	N	N	N	MN	MN

Existen varios métodos para defuzificar las variables, en este trabajo se utilizó el método del centroide.

3.3. Algoritmo

En este apartado se describe el algoritmo que fue utilizado para encontrar el plano dentro de la imagen con mayor nitidez.

1. Obtener una escena $P(x, y, z)$ de N planos con distintos enfoques. Donde x y y forman la posición de cada píxel y n es el n -ésimo plano de la escena.
2. Seleccionar un píxel de la imagen en la escena.
 - a) El usuario selecciona un píxel de la imagen $P(x, y, z)$ que representa la zona de interés.
 - b) Crear una matriz M_n de 11×11 píxeles para cada plano de la escena. Esta matriz se forma por los valores de color en RGB. Calcular la diferencia entre el valor mínimo y el máximo de cada matriz para cada plano (DV).
 - c) Transformar la imagen a sus valores en HSV tomando el elemento de la matriz con mayor saturación como referencia.
3. Encontrar el plano más detallado para la vecindad seleccionada.
 - a) Los valores DV y S son las entradas del controlador difuso, el cual calcula el grado de membresía y entrega la salida.
4. Mostrar la imagen con el plano ganador.
 - a) Al obtener el vector de los N valores representativos de la escena con N planos, se elige el valor más grande. Un valor más grande en el vector de planos es sinónimo de un mayor detalle para ese plano. Luego entonces, se ha seleccionado el plano de la escena con mayor nitidez con respecto al área de interés seleccionada por del usuario.

En esta sección se muestra una escena que consta de tres planos diferentes emulando el efecto de una cámara plenóptica. Las Fig. 8 muestra el enfoque correspondiente a los planos 1, 2 y 3 respectivamente. En la Tabla 2 se muestra el grado de nitidez que tiene cada plano según la imagen. Este tipo de imágenes se pueden encontrar en [6], o bien, se pueden obtener al manipular los enfoques de las cámaras en dispositivos móviles manualmente.

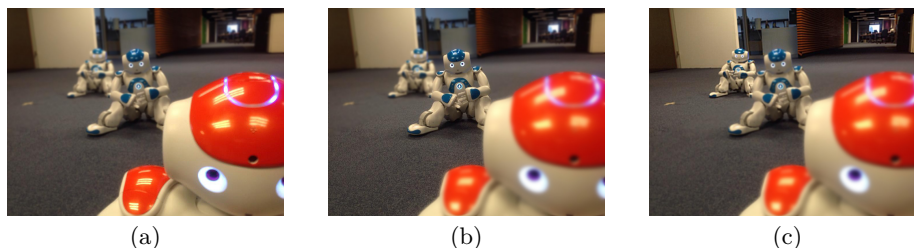


Fig. 8. (a) Imagen tomada de escena de prueba con robot más cercano enfocado. (b) Imagen de la escena de prueba con el robot central enfocado. (c) Imagen de la escena con el robot de fondo enfocado.

Tabla 2. Grados de enfoque (nitidez) en planos de Fig.9.

Figuras	Planos		
	P1	P2	P3
Fig. 9(a)	MN	N	NN
Fig. 9(b)	PN	MN	NN
Fig. 9(c)	NN	PN	N

El esquema de la Fig. 9 es una representación abstracta del algoritmo descrito anteriormente, siendo el uso del control difuso la parte principal del trabajo. El programa desarrollado debe ser alimentado con escenas del tipo a la conformada por la Fig. 8. Al inicio el programa muestra una imagen para que el usuario de un clic sobre el objeto de su interés, de acuerdo con el algoritmo descrito en la sección 2.3 el programa encuentra el plano donde el objeto se encuentra enfocado y lo muestra.

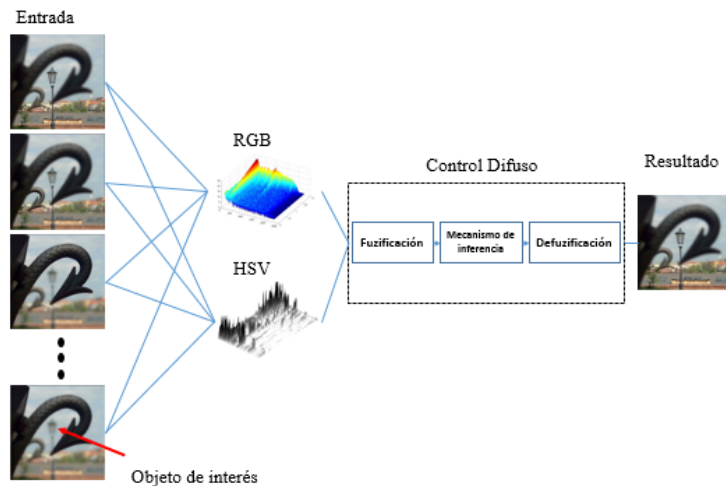


Fig. 9. Sistema difuso propuesto para elegir un objeto de interés en una imagen con múltiples focos.

4. Resultados

La experimentación se llevó a cabo con una escena compuesta de 4 imágenes de tamaño 770×770 píxeles, en una PC de las siguientes características: procesador Intel i7 2600 @ 3.40 GHz, 8GB memoria RAM.

Las pruebas realizadas se hicieron con 4 zonas de interés correspondientes a la nitidez de cada plano, obteniendo los resultados mostrados en Tabla 3.

Tabla 3. Resultados experimentales y tiempo promedio de procesamiento.

Zona	No. Experimentos	Plano buscado	Plano devuelto	Tiempo promedio
1	10	1	[1, 1, 4, 1, 1, 1, 1, 1, 1, 1]	0.93s
2	10	2	[2, 2, 2, 2, 2, 2, 4, 1, 2, 2]	0.95s
3	10	3	[3, 3, 3, 3, 3, 3, 3, 3, 3, 3]	0.98s
4	10	4	[4, 4, 4, 4, 4, 4, 4, 4, 4, 4]	0.96s

En la escena encontrada en Fig. 10, de la cual se hicieron las pruebas del algoritmo propuesto, se encuentra dividida en 4 planos. En estos planos existen diferentes objetos de interés, es decir, en el plano 1 está contenido de la lente de la cámara hasta la cabeza del tornillo. En el caso del plano 2, se encuentra la flecha curvada de mayor área dentro de la imagen. Para el caso del plano número 3 está una flecha curvada de menor área y como plano 4 se encuentra el área del farol y los objetos detrás de él, como es el caso de las azoteas de los edificios.

La zonas de experimentación muestran un conflicto reportado en Tabla 3, en el cual los planos devueltos tienden a irse al fondo de la imagen, y esto se presenta en zonas donde la imagen no cuenta con los suficientes datos para la matriz de RGB.

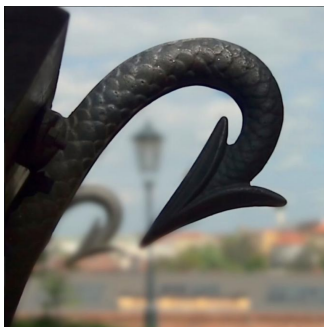


Fig. 10. Escena de experimentación

En la Fig. 11 se visualiza las posibles salidas que las escenas pueden generar siempre y cuando cumplan con los criterios descritos. Además, se muestra una superficie que es el resultado de una simulación hecha en MATLAB. En esta figura (Fig. 11) se puede observar que el sistema siempre entrega una escena que contiene la mayor información respecto a la nitidez de las imágenes ya que la superficie no presenta grandes picos ni oscilaciones.

La adición de niveles de clasificación creará superficies de control más suaves. No obstante, para la experimentación realizada en este trabajo, fueron suficientes las 5 entradas para DV, 4 entradas para el nivel de saturación y 4 salidas, que es la escena con enfoque en 4 planos.

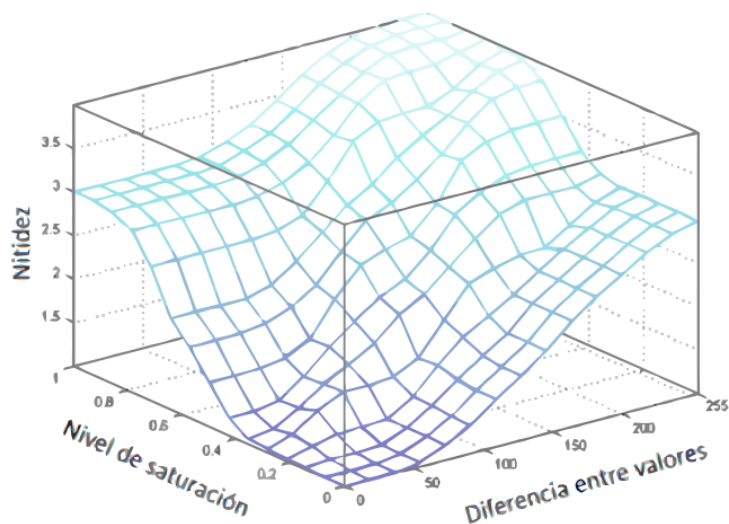


Fig. 11. Superficie de interacción entre los conjuntos de entrada y salida.

5. Conclusiones y trabajo a futuro

La selección de las correctas funciones de membresía y la distribución de los conjuntos difusos, determina el buen desempeño del sistema de control difuso, sin embargo, para la creación de las reglas de inferencia se requiere que el diseñador tenga conocimientos sólidos acerca del problema a resolver, en este caso técnicas de procesamiento digital de imágenes y manejo de variables en los sensores ópticos.

La técnica propuesta basada en la relación que hay entre la diferencia de valor de RGB entre sus píxeles y el nivel de saturación obtenido por la transformación no lineal al espacio HSV de la imagen, permite mejorar el proceso de selección de una imagen nítida dentro de las imágenes de una escena.

Los conjuntos difusos permitieron disuadir la mayor parte de los elementos considerados falsos positivos encontrados en las regiones de interés donde la distribución de color es completamente uniforme. No obstante, en este trabajo el objetivo del control difuso, que generalmente es aplicado en sistemas simples de control, no es emular el funcionamiento del control convencional sino utilizado para encontrar el mejor resultado en el índice de las imágenes que componen una escena.

Por otra parte, como trabajo a futuro, existen diversos métodos de defuzificación los cuales pueden ser manipulados para encontrar mejores resultados o incluso mejorar el tiempo de respuesta al tiempo presentado en este trabajo. De igual forma, trabajar con imágenes de mayor tamaño, ya que conforme avanza la tecnología, la cantidad de píxeles obtenidos por las cámaras comerciales así como los teléfonos inteligentes aumentan rápidamente.

Agradecimientos. Los autores agradecen al IPN-CIC bajo los proyectos SIP 20151187, SIP 20151769 y a CONACYT bajo el proyecto 155014 por el apoyo económico recibido para llevar a cabo esta investigación.

Referencias

1. Bishop, T., Favaro, P.: The light field camera: Extended depth of field, aliasing, and superresolution. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 34(5), 972–986 (May 2012)
2. Chen, H.: Research on light field camera and music emotion recognition. In: *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*. pp. 1558–1559 (June 2009)
3. Harris, M.: Focusing on everything. *Spectrum, IEEE* 49(5), 44–50 (May 2012)
4. Iwane, T.: Light field camera and integral 3d display: 3d image reconstruction based on lightfield data. In: *Information Optics (WIO), 2014 13th Workshop on*. pp. 1–4 (July 2014)
5. Lilly, J.H.: *Fuzzy control and identification*. John Wiley & Sons (2011)
6. Lytro: www.lytro.com (2012)
7. Raghavendra, R., Yang, B., Raja, K., Busch, C.: pp. 1–8 (June 2013)
8. Smith, A.R.: Color gamut transform pairs. In: *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques*. pp. 12–19. SIGGRAPH '78, ACM, New York, NY, USA (1978), <http://doi.acm.org/10.1145/800248.807361>
9. Wang, G., Song, H.: Study of the mine drainage fuzzy control system based on matlab. In: *Intelligent Human-Machine Systems and Cybernetics, 2009. IHMSC '09. International Conference on*. vol. 1, pp. 99–102 (Aug 2009)

Clasificador de pigmentación de la ballena azul aplicando el método SIFT en medios no controlados

Rosa I. Ramos-Arredondo¹, Blanca E. Carvajal-Gómez²,
Francisco J. Gallegos-Funes³, Diane Gendron-Laniel⁴

¹ Escuela Superior de Cómputo, Instituto Politécnico Nacional, SEPI-ESCOM,
México D.F.

² Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas,
Instituto Politécnico Nacional, México D.F.

³ Escuela Superior de Ingeniería Mecánica y Eléctrica, Instituto Politécnico Nacional,
ESIME SEPI-Electrónica, México D.F.

⁴ Centro Interdisciplinario de Ciencias Marinas, Instituto Politécnico Nacional,
La Paz, B.C.S., México

alesija@gmail.com, {becarvajal, fgallegosf, dgendron}@ipn.mx

Resumen. Clasificar las imágenes recolectadas por los especialistas marinos en medios no controlados de la tonalidad de pigmentación de la piel de la ballena azul, haciendo uso del método de foto-identificación; es necesario aplicar un método para la extracción de características que identifique a cada imagen con el método SIFT (Scale-invariant feature transform). Se analiza el rendimiento del clasificador y el desarrollo del método en conjunto; valorando la precisión, la sensibilidad, la especificidad y su rango de falso positivos. Al clasificar 500 muestras de cinco tonalidades de pigmentación (clara, clara-moteada, franja, balanceada y oscura), se obtuvo un 81.80 % de precisión para la tonalidad balanceada y 80.40 % con la tonalidad clara-moteada, tomando 100 muestras de entrenamiento. Al aumentar el número de muestras de entrenamiento a 150, la tonalidad balanceada se mantiene con un 81.20% de precisión, la tonalidad clara-moteada aumentando su precisión a 81%. Al disminuir el número de muestras de entrenamiento a 50, la precisión de la tonalidad oscura fue del 79.80 % y 78.60% de la tonalidad balanceada.

Palabras claves: clasificadores, procesamiento de imágenes, extracción de características, ballena azul.

1. Introducción

La identificación de la ballena azul por medio del método de foto-identificación se basa en clasificar a la especie por medio de características representativas que las hacen únicas e irrepetibles entre otras especies. [1] Para el caso de la ballena azul, la

identificación de la tonalidad de la piel de la ballena azul como la forma de la aleta dorsal son características esenciales para la clasificación de las fotografías recolectadas en campo por los especialistas marinos. En el Laboratorio de Ecología de Cetáceos y Quelonios del Centro Interdisciplinario de Ciencias Marinas (CICIMAR) cuenta con un acervo de imágenes de la ballena azul, clasificadas por tonalidades de pigmentación y por la forma de la aleta dorsal; que han sido recolectados por más de 25 años. La recolección de las imágenes de la ballena azul en campo es complicado controlar las condiciones y el medio de exposición del individuo de estudio y las condiciones de su hábitat. Las condiciones generales son; estar a más de 100 m de distancias del individuo de estudio y recolectar imágenes en condiciones climáticas favorables de iluminación.

El proceso de clasificar las imágenes por la tonalidad de pigmentación y la forma de la aleta dorsal por parte de los especialistas marinos es una tarea, en donde se involucran de al menos tres especialistas con experiencia en la especie del individuo de estudio para determinar de forma veraz y sin dar a equivocaciones. Es necesario el desarrollo de métodos y herramientas en la clasificación de las imágenes, además que se recolecta un gran volumen de imágenes en cada avistamiento y no todas son candidatas para ser utilizadas en la identificación.

En el estudio de las imágenes a clasificar se comprobó que el factor de ser en medios no controlados la exposición de las imágenes; provocara que el elemento a identificar se encontrará en diferentes proporciones, posición y escalas. El empleo del método SIFT para la extracción de características de las imágenes es por ser robusto a transformaciones locales y globales, como también a perturbaciones de cambios en escalamiento, rotación y posicionamiento. Elementos a considerar en el momento de clasificar una gran cantidad de imágenes.

El presente trabajo de investigación se encuentra dividido en secciones; en la primera sección métodos y materiales; se enfocará en exponer los métodos que se necesitarán para la clasificación de los tonos de pigmentación de la ballena azul y la aplicación del método SIFT, como los materiales involucrados para la realización de los experimentos. En la segunda sección resultados, se muestran los datos obtenidos a partir de los experimentos, mostrados en las tablas. En la tercera sección, se discuten los resultados obtenidos y se enfoca en efectuar las comparativas entre los procedimientos experimentados.



Fig. 1. Diagrama de bloques del proceso general para clasificar a la ballena azul por su tono de pigmentación.

2. Métodos y materiales

En esta sección nos enfocaremos en el desarrollo de los métodos y materiales para el clasificador de las imágenes por tonos de pigmentación de la piel de la ballena azul. En la Figura 1, se muestra el diagrama de bloques de las etapas y el flujo que nos permitirá clasificar a las imágenes de la ballena azul por sus tonos de pigmentación.



Fig. 2. Imágenes del acervo del CICIMAR de la ballena azul, a) Imagen mostrando el flanco izquierdo de la ballena azul, b) imagen de la ballena azul en donde se muestra el lomo sin su aleta dorsal, c) Imagen mostrando una porción reducida del lomo de la ballena azul

2.1 Imágenes a clasificar

En el CICIMAR cuenta con un acervo de imágenes previamente clasificadas por la forma de la aleta dorsal y sus tonos de pigmentación de la piel de la ballena azul. En este trabajo de investigación nos centramos en la clasificación de las imágenes de la ballena azul por su tono de pigmentación y para el caso de estudio tomaremos cinco clases (clara, clara-moteada, franja, balanceada y oscura) [2]. Se tomarán un total de 846 imágenes y su distribución entre clases es; clara (101), clara moteada (304), franja (116), balanceada (174), oscura (151). Información adicional, la clase clara-moteada y

la clase franja presenta un patrón además de un tono de pigmentación. Esta consideración no altera el estudio de la investigación, ni se añadirá alguna etapa extra.

2.2 Procesamiento de las muestras

Las imágenes del acervo a utilizar no tiene una dimensión concreta y al ser recolectadas la porción de la ballena azul en la imagen es variado como su posición, su ángulo y en algunos casos la distancia en la que fue tomada.

Selección de la región de interés. La selección de la región de interés (ROI, region of interest) de las imágenes en la ballena azul es para reducir el número píxeles al analizar, como el número de características a clasificar del tono de pigmentación en conjunto. Por tal motivo se reduce las tareas computacionales al analizar la imagen completa. Se tomará una región cuadrada de 100 píxeles cuadrados en el lomo de la ballena azul y no próxima de la aleta dorsal para evitar un análisis adicional como; segmentación, realce de píxeles entre otros.



Fig. 3. Selección de la región de interés a analizar de la ballena azul

Las regiones de interés de las imágenes del acervo se ocuparán para su clasificación y una porción de estas, se utilizará para la etapa de entrenamiento. Una muestra puede describirse como I en donde, $f(x,y)$ representa el nivel de intensidad en las coordenadas (x,y) . Véase en la ecuación 1. La región de interés estará acotada en una región cuadra de 100 píxeles cuadrados.

$$I = f(x,y). \quad (1)$$

Extracción de características con SIFT. Las muestras a clasificar como las muestras para un previo entrenamiento, se les extraerá características, que en la etapa de clasificación se ocupara para identificar. Es cuando el método SIFT es utilizado y nos proporcione un arreglo de descriptores (descriptor) y puntos claves (keypoints). El algoritmo consiste en cuatro etapas: 1) selección de escala-espacio de picos (scale-space peak selection), b) localización de punto clave (keypoint localization), c) asignación de orientación (orientation assignment), d) descriptores de punto clave (keypoint descriptor).[3] El aplicar este método a las muestras a clasificar es porque nos ubica puntos de interés, además de ser un algoritmo invariante a cambio de escala, rotación de la imagen y pequeños cambios de iluminación; condiciones que se da en la recolección de imágenes de la ballena azul en medio no controlados.

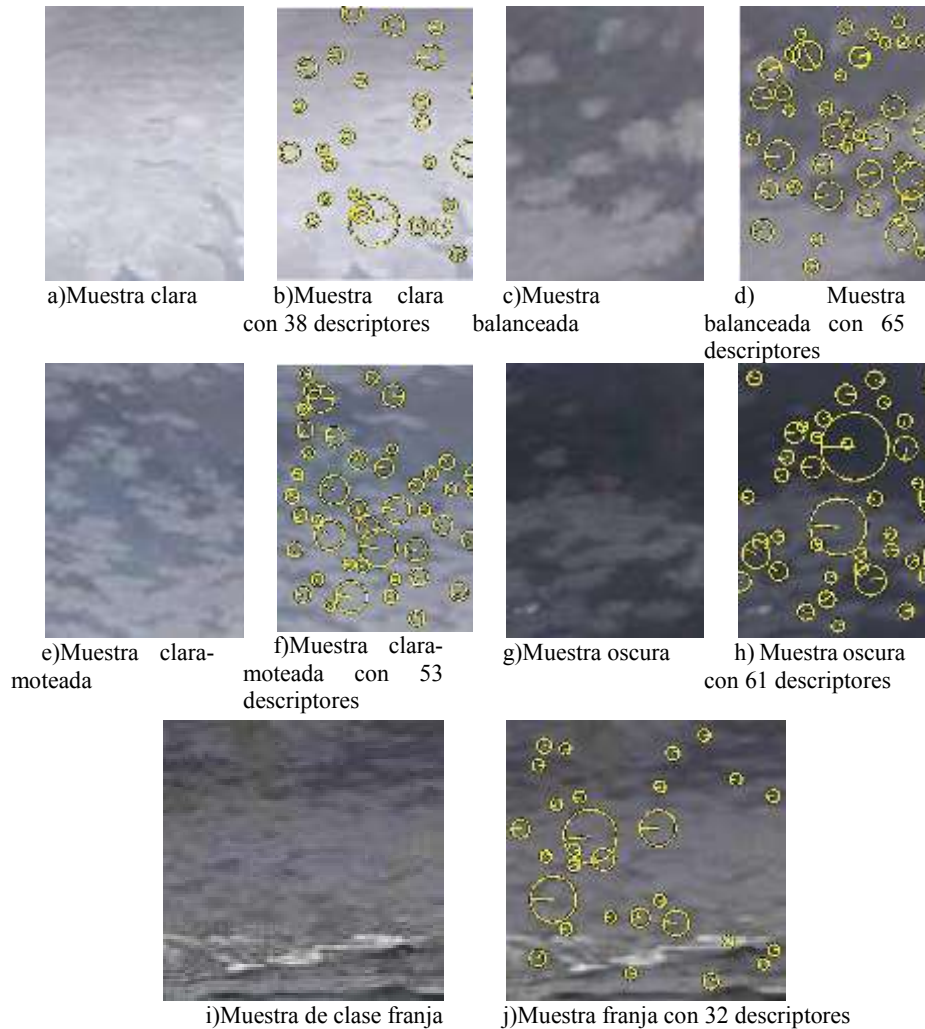


Fig. 4. Muestras de cada uno de los tonos de pigmentación a estudiar y su representación de los descriptores de cada uno, como su magnitud y ángulo. a),c),e), g),i) muestras de los tonos de pigmentación y b),d),f),h),j) su representación de los descriptores.

2.3 Clasificación

Todas las muestras (imágenes que clasificar e imágenes que se ocuparán en el entrenamiento) después de obtener la extracción de características se compone por una estructura (frame) y descriptores. El número de descriptores es variable para cada imagen. La correspondencia entre los descriptores de las imágenes en la clasificación es obteniendo la distancia euclidiana entre los descriptores de la imagen más cercana y la siguiente, obteniendo solo el número de puntos de encuentro. La clasificación de las

imágenes de la ballena azul en la literatura se encuentra el estudio por medio de la forma de la aleta dorsal, en donde se emplea un clasificador supervisado, con una previa selección de muestras para el entrenamiento [4].

Entrenamiento. Las muestras que se seleccionan para el entrenamiento son imágenes representativas de cada tono de pigmentación o clase (termino que se ocupará en el desarrollo del clasificador). La selección de muestras de entrenamiento se ocupa como punto de referencia para la clasificación de muestras a identificar. Las muestras a clasificar y las de entrenamiento después del proceso de extracción de características empleando el método SIFT, tiene un vector de características y tomaremos el vector descriptor $\#d_i = |d_i|$, donde i , es el número de elementos que lo componen. Donde \bar{C}_j , es el valor representativo por cada j clases; las clases utilizadas son clara, clara moteada y oscura. Véase en la ecuación 2.

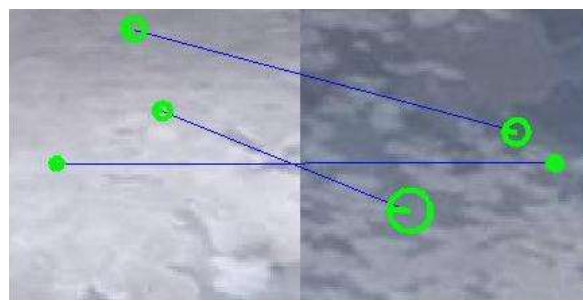
$$\bar{C}_j = \frac{\sum_{i=1}^n |d_i|}{n}. \quad (2)$$

El vector de entrenamiento E que se empleara en la clasificación de muestras estará compuesto por $k=5$ valores, véase en la ecuación 3.

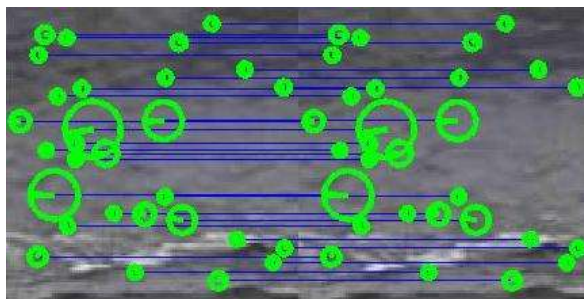
$$E = (\bar{C}_1, \bar{C}_2, \dots, \bar{C}_k). \quad (3)$$

Se tomara el mismo número de muestras de entrenamiento de cada clase para evitar una sobre clasificación. Además, de tomar un número pequeño con relación al número de muestras a clasificar. De las muestras de entrenamiento solo se estudiara sus descriptores.

Conteo de encuentros. Las muestras a clasificar se analizaran por medio de su vector descriptor $\#d_i$, el cual se comparará su distancia con los vectores descriptores de las muestras de entrenamiento de cada una de las clases. Posteriormente se clasificará tomando el valor máximo con relación al vector de entrenamiento E ; proporcionando la pertenencia a una de las clases (tonos de pigmentación de la piel de la ballena azul). En la figura 5 se ilustra los puntos de encuentro entre muestras de diferentes tonos de pigmentación y de iguales tonos.



a) Puntos de encuentro entre una muestra clara y clara moteada



b) Puntos de encuentros entre dos muestras franja

Fig. 5. Ejemplo de puntos de encuentro entre dos muestras.

3. Resultados

Se ocuparon cien muestras de cada tono de pigmentación de la piel de la ballena azul para el estudio (clara, moteada, franja, balanceada y oscura), analizando un total de 500 muestras a clasificar. Para las muestras de entrenamiento se analizaron con 50,100 y 150 muestras, tomando de forma equitativa entre los tonos de pigmentación de las muestras a clasificar.

En la Tabla 1, se muestra la clasificación de 500 muestras con relación de 50 muestras de entrenamiento (10 muestras de cada clase), dando la clase clara el mayor número de clasificación en comparación de las clases moteada y balanceada.

Tabla 1. Clasificación de cinco clases utilizando cincuenta muestras de entrenamiento.

Tono de pigmentación (Número de muestras)	Clasificación de Clases				
	Clara	Moteada	Franja	Balanceada	Oscura
Clara(100)	59	8	21	2	10
Moteada(100)	45	13	27	7	8
Franja(100)	43	7	39	4	7
Balanceada(100)	51	8	17	13	11
Oscura(100)	40	3	15	7	35

En la Tabla 2, se muestran las matrices de confusión de cada una de las clases, proporcionando la relación entre las 100 muestras a clasificar de forma supervisada y las 400 muestras de otras clases a analizar.

Tabla 2. Matriz de confusión del clasificador de cinco clases utilizando cincuenta muestras de entrenamiento

	Clara		Moteada		Franja		Balanceada		Oscura	
	Condición		Condición		Condición		Condición		Condición	
	+	-	+	-	+	-	+	-	+	-
Prueba Positiva	59	179	13	26	39	80	13	20	35	36
Prueba Negativa	41	221	87	374	61	320	87	380	65	364
Total	100	400	100	400	100	400	100	400	100	400

En la Tabla 3, se muestra la clasificación de 500 muestras con relación de 100 muestras de entrenamiento (20 muestras de cada clase), proporcionando la clase balanceada caso nulos en la clasificación con otras clases.

Tabla 3. Clasificación de cinco clases utilizando cien muestras de entrenamiento.

Tono de pigmentación (Número de muestras)	Clasificación de Clases				
	Clara	Moteada	Franja	Balanceada	Oscura
Clara(100)	62	2	18	0	18
Moteada(100)	36	10	32	1	21
Franja(100)	31	2	51	0	16
Balanceada(100)	37	4	21	10	28
Oscura(100)	30	0	20	0	50

En la Tabla 4 se muestran las matrices de confusión de cada una de las clases. Corroborando el caso de la clase balanceada y moteada como las clases con menor número de errores en la clasificación supervisada.

Tabla 4. Matriz de confusión del clasificador de cinco clases utilizando cien muestras de entrenamiento.

	Clara		Moteada		Franja		Balanceada		Oscura	
	Condición		Condición		Condición		Condición		Condición	
	+	-	+	-	+	-	+	-	+	-
Prueba Positiva	62	134	10	8	51	91	10	1	50	83
Prueba Negativa	38	266	90	392	49	309	90	399	50	317
Total	100	400	100	400	100	400	100	400	100	400

En la Tabla 5, se muestra la clasificación de 500 muestras con relación de 150 muestras de entrenamiento (30 muestras de cada clase), proporcionando la clase balanceada caso nulos en la clasificación con otras clases.

Tabla 5. Clasificación de cinco clases utilizando ciento cincuenta muestras de entrenamiento.

Tono de pigmentación (Número de muestras)	Clasificación de Clases				
	Clara	Moteada	Franja	Balanceada	Oscura
Clara(100)	69	1	17	0	13
Moteada(100)	44	13	29	0	14
Franja(100)	44	1	41	0	14
Balanceada(100)	48	4	17	7	24
Oscura(100)	32	2	16	1	49

En la Tabla 6, se muestran las matrices de confusión de cada una de las clases. Corroborando el caso de la clase clara con el mayor número de errores en la clasificación supervisada, en cambio la clase balanceada y moteada siguen presentando el menor número de errores.

Tabla 6. Matriz de confusión del clasificador de cinco clases utilizando ciento cincuenta muestras de entrenamiento.

	Clara		Moteada		Franja		Balanceada		Oscura	
	Condición		Condición		Condición		Condición		Condición	
	+	-	+	-	+	-	+	-	+	-
Prueba Positiva	69	168	13	8	41	79	7	1	49	65
Prueba Negativa	31	232	87	392	59	321	93	399	51	335
Total	100	400	100	400	100	400	100	400	100	400

4. Discusión

El estudio del rendimiento de la clasificación de la tonalidad de la pigmentación de la piel de la ballena azul hace uso de cuatro parámetros; verdadero positivo (VP) se prueba de forma positiva y la condición es positiva, falso positivo (FP) se prueba de forma positiva pero la condición es negativa, verdadero negativo (VN) se prueba de forma negativa dando negativa la condición y falso negativo (FN) se prueba de forma negativa dando positiva la condición. Estos parámetros se encuentran concentrados en la matriz de confusión de cada clase del clasificador. Son utilizados en las siguientes cuatro evaluaciones; especificidad (SP), sensibilidad (SN), la razón de falso positivos (FPR) y la precisión (ACC) [5-7].

En la Tabla 7, se concentra las evaluaciones de cada una de las clases tomando 50 muestras de entrenamiento. La clase con menor precisión es la clase clara como también en cuanto a sensibilidad. En cambio la clase moteada, balanceada y oscura superan el 77% de precisión.

Tabla 7. Rendimiento de la clasificación utilizando cincuenta muestras de entrenamiento.

	Clara	Moteada	Franja	Balanceada	Oscura
Especificidad (%)	84.35	81.12	83.98	81.37	84.84
Sensibilidad (%)	24.78	33.33	32.77	39.39	49.29
FRP (%)	15.64	18.87	16.01	18.62	15.15
Precisión (%)	56	77.40	71.80	78.60	79.80

En la Tabla 8, se concentra las evaluaciones de cada una de las clases tomando 100 muestras de entrenamiento. La clase moteada y balanceada obtiene una precisión arriba del 80%, la clase franja y oscura su precisión es del 72% y 73.40%.

Tabla 8. Rendimiento de la clasificación utilizando cien muestras de entrenamiento.

	Clara	Moteada	Franja	Balanceada	Oscura
Especificidad (%)	87.50	81.32	86.31	81.59	86.37
Sensibilidad (%)	31.63	55.55	35.91	90.90	37.59
FRP (%)	12.50	18.67	13.68	18.40	13.62
Precisión (%)	65.60	80.40	72	81.80	73.40

En la Tabla 9, se concentra las evaluaciones de cada una de las clases tomando 150 muestras de entrenamiento. La clase moteada y balanceada obtiene una precisión arriba del 80%, la clase franja y oscura su precisión es del 72% y 76.80%.

Tabla 9. Rendimiento de la clasificación utilizando ciento cincuenta muestras de entrenamiento.

	Clara	Moteada	Franja	Balanceda	Oscura
Especificidad (%)	88.21	81.83	84.47	81.09	86.78
Sensibilidad (%)	29.11	61.90	34.16	87.50	42.98
FRP (%)	11.78	18.16	15.52	18.90	13.21
Precisión (%)	60.20	81	72.40	81.20	76.80

5. Conclusión

En este trabajo de investigación, se llegó al análisis de cinco clases de tonos de pigmentación de la piel de la ballena azul (clara, clara-moteada, franja, balanceada y oscura), dando la clase balanceada la más alta en precisión independientemente del número de muestras en el entrenamiento del clasificador empleando el método SIFT en la extracción de características. Se observó que entre las clases moteada, franja y oscura se mantenía la precisión y se mejoró al aumentar el número de muestras de entrenamiento, con 100 y 150 muestras.

En el caso específico de la clase clara, el análisis nunca alcanza un porcentaje en precisión aceptable y fue la clase en donde se obtuvo un mayor número de errores de clasificación. La clase clara presenta características muy similares entre otras clase, por tal motivo se propone en futuros estudios; el reducir el número de clases a clasificar.

Agradecimientos. Los autores agradecen al Instituto Politécnico Nacional de México (IPN) y al Consejo Nacional de Ciencia y Tecnología de México (CONACYT) por la ayuda y el apoyo con número de proyecto 221284. Como la colaboración y apoyo otorgado por parte de la Dra. María Chantal Diane Gendron Laniel y a su equipo de profesionales colaboradores.

Referencias

1. Randall S. Wells.: Identification Methods. In: Perrin, William F., Bernd Wursig, and J. G. M. Thewissen. (eds.) Encyclopedia of marine mammals, pp.593-596. Academic Press, Diego (2009)
2. Gendron, D., Ugalde De, A.: A new classification method to simplify blue whale photo-identification technique. Journal of Cetacean Research and Management, pp. 79-84. (2012)
3. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. In: International journal of computer vision, vol. 60.2, pp. 91-110. Kluwer Academic Publishers (2004)
4. Ramos-Arredondo, R. I., Carvajal-Gómez, B. E., Gallegos-Funes, F. J., Gendron-Laniel, D.: Multi-Spatial Classifier for Blue Whale Images using Photo-Identification Method. In: Castro Espinoza, F. (eds.) MICAI 2014: Advances in

- Artificial Intelligence, vol. 82, pp. 31-40. Center for Computing Research of IPN (2014)
5. Fathi, A., Naghsh-Nilchi, A.R.: Automatic wavelet-based retinal blood vessels segmentation and vessel diameter estimation. *Biomedical Signal Processing and Control*. vol. 8, pp.71–80. (2013)
 6. Trejo-Salazar, D., Carvajal-Gómez, B. E., Gallegos-Funes, F. J.: Algoritmo de segmentación para imágenes reales para la clasificación de cetáceos. En: M. González-Mendoza y F. Castro-Espinoza (eds.). En: Congreso Mexicano de Inteligencia Artificial (COMIA 2013): Research in Computing Science. vol. 62, pp. 209-218, (2013).
 7. Fraz, M.M., Barman, S.A., Remagnino, P., Hoppe, A., Basit, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G.: An approach to localize the retinal blood vessels using bit planes and centerline detection. *Computer Methods and Programs in Biomedicine*, vol. 108, pp: 600–616, 2012

Módulo de generación de aplicaciones multi-dispositivo a partir del procesamiento de imágenes

Laura Sánchez Morales , Viviana Yarel Rosales Morales, Giner Alor Hernández,
Rubén Posada Gómez, Hilarión Muñoz Contreras, Ulises Juárez Martínez

División de Estudios de Posgrado e Investigación, Instituto Tecnológico de Orizaba,
Orizaba, Veracruz, México

lauransanchezmorales@gmail.com, viviana_rosales@outlook.com, galor@itorizaba.edu.mx,
rposada@itorizaba.edu.mx, hmunoz@itorizaba.edu.mx, ujuarez@itorizaba.edu.mx

Resumen. En la actualidad existen diferentes esquemas de generación automática de software como son MDA (*Model Driven Architecture*), FDD (*Feature-driven developmen*), RAD (*Rapid Application Development*), por mencionar algunos. Sin embargo hasta el momento no se ha abordado la generación de software mediante el uso de técnicas de inteligencia artificial como son el procesamiento digital de imágenes y reconocimiento de patrones. En este trabajo se propone el desarrollo de un componente de software que utilice técnicas de procesamiento de imágenes y reconocimiento de patrones para la generación automática de aplicaciones multi-dispositivo. Para lograr este objetivo, se identifican los elementos de una imagen y se genera el código de una aplicación multi-dispositivo con los elementos detectados previamente seleccionados en dicha imagen. Finalmente se presenta un caso de estudio que permite describir la funcionalidad del módulo propuesto.

Palabras clave: aplicaciones multi-dispositivo, generación automática de software, procesamiento de imágenes.

1. Introducción

El desarrollo de software está en constante evolución, y a través del tiempo se propusieron diversos enfoques de desarrollo como FDD (*Feature-driven development*), MDA (*Model Driven Architecture*), RAD (*Rapid Application Development*), por mencionar algunos; con la finalidad de agilizar el tiempo de desarrollo sin comprometer la calidad del producto final. Sin embargo, la mayoría de las herramientas empleadas por desarrolladores, y todos aquellos interesados en el desarrollo de software, ofrecen características de desarrollo muy similares entre sí pero son pocas las que, con la finalidad de satisfacer estas necesidades implementan técnicas de inteligencia artificial como el reconocimiento de patrones en imágenes. Desde esta perspectiva surge uno de los principales motivos para realizar el presente trabajo.

Por otra parte las aplicaciones multi-dispositivo como su nombre lo indica, permiten el desarrollo en un lenguaje o conjunto de lenguajes y su posterior ejecución en diferentes dispositivos de hardware. Sin embargo esto no implica que se hayan cubierto

todas las necesidades para la generación automática de código, por el contrario, cada vez surgen más alternativas para la generación automática de software. El procesamiento digital de imágenes es el conjunto de técnicas que se aplican a las imágenes digitales con el objetivo de mejorar su calidad o facilitar la búsqueda de información inmersa en ellas. En este sentido el Tratamiento Digital de Imágenes contempla el procesamiento y el análisis de imágenes. Este procesamiento se refiere a la realización de transformaciones y a la restauración y mejoramiento de las imágenes. El análisis consiste en la extracción de propiedades y características de las imágenes, así como la clasificación e identificación y el reconocimiento de patrones [1]. De ahí que la importancia del procesamiento y análisis de imágenes digitales esté presente en diversas áreas, tales como la medicina [2], biología [3], astronomía [4], fotografía [5], historia [6] y geología [7], entre otras. Esto se debe a que la obtención de imágenes no está limitada por el dominio de aplicación. Sin embargo en la ingeniería de software hasta el momento no se reportan métodos o técnicas ampliamente definidos sobre el tratamiento de imágenes para el desarrollo automático de software. Con base en lo anterior este trabajo propone el desarrollo de un componente de software que identifique elementos en una imagen mediante técnicas de procesamiento de imágenes y genere el código de una aplicación multi-dispositivo con los elementos identificados en la imagen.

Este documento está estructurado de la siguiente manera, en la sección 2 se presenta el estado del arte referente a los diversos trabajos relacionados para la generación automática de software. En la sección 3 se describe el proceso planteado para la generación de software. En la sección 4 se presentan parte de las especificaciones definidas para la identificación de elementos en interfaces. En la sección 5 se presenta un caso de estudio empleando una imagen que representa la portada principal de una página Web. Finalmente en la sección 6 se presentan las conclusiones y trabajo a futuro.

2. Estado del arte

A continuación se presenta la revisión del estado del arte sobre los trabajos relevantes en el área de la generación automática de código para el desarrollo de software, las cuales se encuentran clasificadas por herramientas para desarrollo o generación de aplicaciones y técnicas de reconocimiento y procesamiento de imágenes.

2.1. Herramientas para desarrollo o generación de aplicaciones

En [8] se presentó una herramienta para la generación automática de código de automoción de sistemas embebidos. El código se genera sobre la base de una máquina de estados finitos.

En [9] se presentó SolidFX, un marco de trabajo de ingeniería inversa para realizar análisis de código en C++, calcular métricas específicas en los sistemas, extraer datos de grandes bases de datos y analizar de manera visual e interactiva los resultados de la misma manera que se hace en la mayor parte de los IDE de desarrollo.

En [10] se presentó un marco de trabajo y una herramienta de edición como una extensión del lenguaje UIML. El desarrollador diseña su interfaz gráfica de usuario de

manera abstracta y posteriormente a través de la aplicación de técnicas de transformación basadas en gramáticas de grafos, dichas interfaces se transforman en interfaces de usuario concretas.

En [11] se propuso un nuevo enfoque el cual se basa en el desarrollo de una capa adicional Orientada a Aspectos que codifica un DSML (*Domain Specific Language*) para Frameworks basados en aplicaciones, de tal manera que se elimine la necesidad de implementar un generador de código.

En [12] se presentó un proceso para la generación de código fuente de RIAs (*Rich Internet Applications*) multi-dispositivo que complementa el proceso de Fases para el Desarrollo de RIAs denominado PPRD por sus siglas en inglés (*Phases Process for RIAs Development*).

En [13] se presentaron las herramientas Lonworks, KNX y BCU SDK Tools que ayudan a los desarrolladores a crear modelos de sistemas de automatización del hogar por medio de un lenguaje específico de dominio que se transforma en el código para plataformas de automatización del hogar específicas.

En [14] se describió una herramienta de generación de código para aplicaciones de lenguaje de procedimiento basado en el procesamiento distribuido. Los programas de aplicación, junto con las primitivas de partición se convierten en implementaciones concretas independientemente ejecutables.

En [15] se describió un método para automatizar el proceso de desarrollo de nuevas estructuras de datos, definiciones de meta-datos, y software de traducción a través de herramientas automatizadas. Con esto se generan automáticamente sistemas de colaboración.

La mayoría de las herramientas para el desarrollo o generación de aplicaciones encontradas en la literatura reportan diversas técnicas sin embargo no se observa que ofrezcan generar código a partir de técnicas de procesamiento de imágenes.

2.2. Técnicas de reconocimiento y procesamiento de imágenes

En [16] se presentó un estudio comparativo del procesamiento de imágenes de 3 diferentes técnicas: *Multiplicative Homomorphic Image Processing* (MHIP), *Log-Ratio Image Processing* (LRIP) y *Logarithmic Image Processing* (LIP). En [17] se presentó un trabajo que contiene una revisión de los más recientes, así como de los clásicos métodos de registro de imágenes con el objetivo de proporcionar una fuente de referencia completa para los investigadores involucrados en el registro de imágenes, con independencia de las áreas de aplicación específicas. En [18] se presentó el sistema BioPro que emplea el método de Programación Orientada a imágenes también propuesto, el cual utiliza gráficos como una herramienta de diseño de software para aplicaciones Web.

En [19] se presentó el Framework AgentSketch para la interpretación de símbolos esbozados que explota en gran medida la información contextual para la resolución de ambigüedades en las imágenes. AgentSketch contempla bocetos de varios dominios uno de ellos los diagramas UML, específicamente de casos de uso.

En [20] se propuso un concepto y la arquitectura de un producto genérico basado en un reconocedor geométrico para la agrupación y segmentación de trazos. Reconoce

componentes individuales, diagramas esbozados como un conjunto y permite resolver las ambigüedades mediante análisis sintáctico y semántico.

En [21] se presentó un enfoque de reconocimiento visual y desarrollo de aplicaciones a partir de modelos para diagramas de ingeniería. El enfoque es una red neural convolucional aprovechada como un reconocedor de símbolos de ingeniería entrenable capaz de aprender las características visuales de las categorías de los símbolos definidos en algunos diagramas prototipo proporcionados por el usuario.

En [22] se discutió la noción de un diagrama de historia de estado. Una combinación de un metamodelado en UML y un marco de trabajo se utilizan para dar semántica precisa a los diagramas de la historia del estado y de los artefactos de forma orientada al análisis. Un diagrama de historia de estado o SHD es en realidad un diagrama de estados de transición.

En [23] se enfatizó el uso de *skeletons* en el procesamiento de imágenes digitales, para la realización de operaciones de procesamiento de imágenes, el esqueleto es una herramienta mucho más esencial y altamente adaptable. Los esqueletos son descriptores importantes en la representación de objetos y el reconocimiento.

En [24] se presentó un análisis de procesamiento de imagen paralela y distribuida con amplios detalles, se presentan resultados de un estudio de procesamiento de imágenes paralelo y distribuido con énfasis en los mecanismos, herramientas, tecnología, APIs utilizadas, dominios de aplicación y trabajos de investigación en curso.

Las diversas técnicas de reconocimiento y procesamiento de imágenes se reportan en diversos campos de aplicación, no obstante el momento no se han utilizado para cubrir las necesidades en la generación de código para aplicaciones multi-dispositivo. Por lo anterior, es necesario enfatizar la importancia del desarrollo de nuevos métodos que permitan crear aplicaciones de software de manera sencilla y rápida, y que posteriormente estos métodos se utilicen en herramientas para el desarrollo de aplicaciones.

3. Proceso de generación de software

En esta sección se describe el procedimiento de generación de software a partir del procesamiento de imágenes y reconocimiento de patrones.

Para el proceso de generación de software se utilizaron los siguientes algoritmos de procesamiento de imágenes:

Operaciones morfológicas: Un operador morfológico utiliza un elemento de estructuración para procesar una imagen. Las operaciones morfológicas se pueden utilizar en imágenes en binario y en escala de grises [25].

Operaciones de convolución: La convolución es una operación matemática que es fundamental para muchos operadores comunes de procesamiento de imágenes. Su campo de aplicación es amplio uno de ellos es el procesado lineal de imágenes [26].

Detección de bordes: Los bordes pueden ser detectados mediante la aplicación de un filtro de frecuencia de paso alto. Se utiliza ampliamente en la segmentación de la imagen cuando se quiere dividir la imagen en las zonas correspondientes a los diferentes objetos, llamadas zonas de interés [27].

Filtro de mediana: Un filtro de mediana es un filtro digital no lineal, que es capaz de preservar los cambios de señal agudos y es muy eficaz en la eliminación de ruido de impulso. Este algoritmo sustituye el valor de un pixel por el valor de la media de los pixeles vecinos. Es capaz de mejorar ciertas características de una imagen que posibiliten efectuar operaciones del procesado sobre ella [28].

El proceso de generación de software a partir del procesamiento de imágenes, consta de 8 pasos que a continuación se describen:

1) Entrada de la imagen a procesar, las imágenes deben estar delimitadas previamente, las imágenes soportadas son modelos de diseños ADVs (*Abstract Data Views*) que permiten especificar clara y formalmente interfaces de usuario separadas de los componentes de la aplicación de un sistema de software [29]. Los ADVs pueden ser generados por el usuario en cualquier herramienta de dibujo. Un ejemplo de estos ADVs se presentan posteriormente. Los formatos de imagen permitidos son: GIF, JPG o JPEG y PNG [30].

2) Validación de la imagen de entrada para asegurar que la imagen es una interfaz. La validación de la imagen consiste en aplicar los algoritmos de filtro de mediana, operaciones morfológicas, operaciones de convolución y detección de bordes para obtener los distintos elementos dentro de la imagen. Los elementos que se obtienen para las aplicaciones Web son: form, textarea, radiobutton, checkbox, list, por mencionar algunos. Los elementos que se obtienen para las aplicaciones para dispositivos móviles son: dataspinner, radiobuttongroup, spinnerlist, toggleswitch, button, checkbox, list, label, img, radiobutton, textarea, e input text. Si dentro de la imagen se localizan algunos de estos elementos entonces la imagen es válida para continuar con el proceso de generación de software.

3) Selección por parte del usuario de los elementos identificados que formarán parte del código a generar.

4) Selección del tipo de aplicación a generar: una aplicación Web o una aplicación móvil. Esta decisión se toma con la finalidad de que la generación de código se realice según las características de cada una de estas aplicaciones. Otro factor relevante que debe considerarse en esta fase son los elementos encontrados en la imagen de entrada, ya que no todas las imágenes tendrán elementos válidos para generar cualquier tipo de aplicación.

5) Configuración de la aplicación: de acuerdo con el tipo de aplicación seleccionada en el paso anterior, se selecciona la configuración general de la aplicación (título principal, lenguaje, color, descripción, entre otras).

6) Generación de un documento XML con etiquetas para: nombre de la aplicación, autor (s), etiquetas con la configuración de acuerdo a cada tipo de aplicación y lenguaje seleccionado esto es; resolución, orientación y elementos de plantilla (header, body, footer, entre otros); finalmente el conjunto de etiquetas con la representación de cada uno de los elementos seleccionados por el usuario.

7) Generación del código a partir de su representación XML: Se procesa el documento XML por medio de un documento XML Schema generando su equivalente en código HTML 5 de las etiquetas asociadas a cada elemento en la imagen, dicho código incrustado en frameworks tales como Sencha Touch un framework de HTML 5 que permite desarrollar aplicaciones web para diferentes dispositivos móviles [31]; JQuery Mobile un framework que permite agregar complejidad y enriquecer páginas [32]; y PhoneGap, un framework de desarrollo de aplicaciones web móviles que

permite a los desarrolladores construir aplicaciones web basadas en HTML5 y JavaScript con envoltorios para más de seis plataformas móviles, incluyendo iOS, Android y BlackBerry [33]. Por mencionar algunos; siguiendo con la descripción de este paso, la estructura generada se agrupa en clases de acuerdo con el lenguaje de programación seleccionado por el usuario.

8) Entrega al usuario de un archivo ZIP con el código fuente generado.

El proceso de generación de software se muestra a continuación en la Fig.1 donde se puede observar la secuencia de cada paso antes descrito.

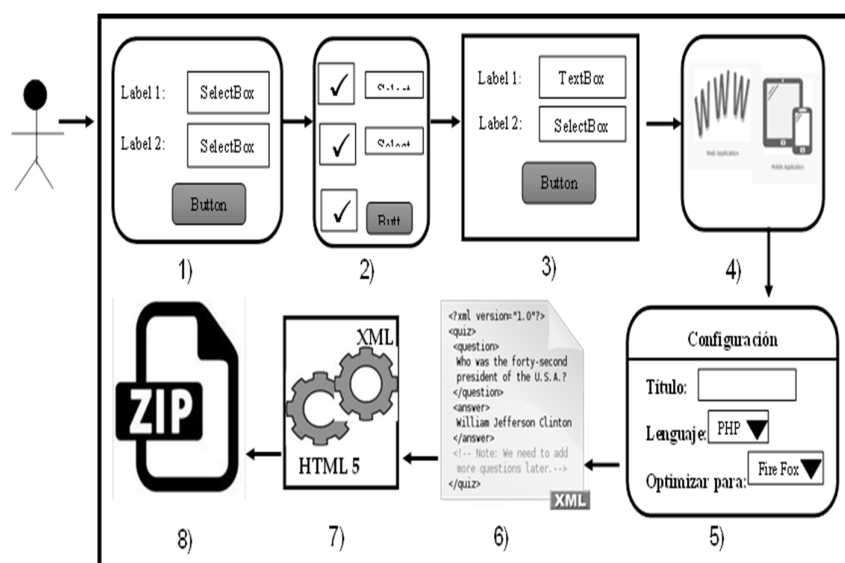


Fig. 1. Proceso de generación de software a partir del procesamiento de imágenes.

4. Proceso de identificación de elementos de interfaces

En principio se definen un conjunto de reglas para la identificación de cada elemento de la interfaz. La finalidad de definir el conjunto de reglas parte del hecho de tener un grupo de especificaciones para cada elemento que sirvan como identificadores únicos al momento de realizar el procesamiento de la imagen, dichas reglas están representadas en un árbol el cual emplea el lenguaje RuleML en su versión 1.0 [34].

En total se generaron 25 reglas considerando para ello 13 elementos como parte de una interfaz Web los cuales son: input text, button, checkbox, select, input date, input email, img, label, a (ancla), list, radiobutton, textarea y form. Para las interfaces para dispositivos móviles las reglas generadas son 12 y los elementos son: dataspinner, radiobuttongroup, spinnerlist, toggleswitch, button, checkbox, list, label, img, radiobutton, textarea, e input text. De acuerdo con la notación de RuleML 1.0 parte de los archivos XML generados para la representación de reglas, se muestran en la Fig. 2

donde, (a) es la representación de la regla para el elemento radiobutton de una interfaz Web, y (b) la regla para el elemento dataspinner de una interfaz móvil. RuleML es una iniciativa abierta en la que se busca establecer un sistema de reglas de inferencia lógica a partir de ontologías y documentos RDF con su propio lenguaje de especificación y ejecución.

```

<Implies>
  <If>
    <And>
      <Atom>
        <Var>Forma externa</Var>
        <Rel>Cuadrado</Rel>
      </Atom>
      <Atom>
        <Var>Figura interna</Var>
        <Rel>Círculo</Rel>
        <Ind>Posiblemente rellena</Ind>
      </Atom>
    </And>
  </If>
  <Then>
    <Atom>
      <Var>Forma externa</Var>
      <Var>Figura interna</Var>
      <Rel>Radiobutton</Rel>
    </Atom>
  </Then>
</Implies>
  
```

a)

```

<Implies>
  <If>
    <And>
      <Atom>
        <Var>Forma externa</Var>
        <Rel>Rectángulo</Rel>
        <Ind>Esquinas redondeadas</Ind>
      </Atom>
      <Atom>
        <Var>Figuras interna</Var>
        <Rel>Rectángulos</Rel>
        <Ind>6 en dos filas</Ind>
      </Atom>
    </And>
  </If>
  <Then>
    <Atom>
      <Var>Forma externa</Var>
      <Var>Figuras interna</Var>
      <Rel>Dataspinner</Rel>
    </Atom>
  </Then>
</Implies>
  
```

b)

Fig. 2. a) Regla para el elemento *radiobutton* de una interfaz Web, b) Regla para el elemento *dataspinner* de una interfaz móvil.

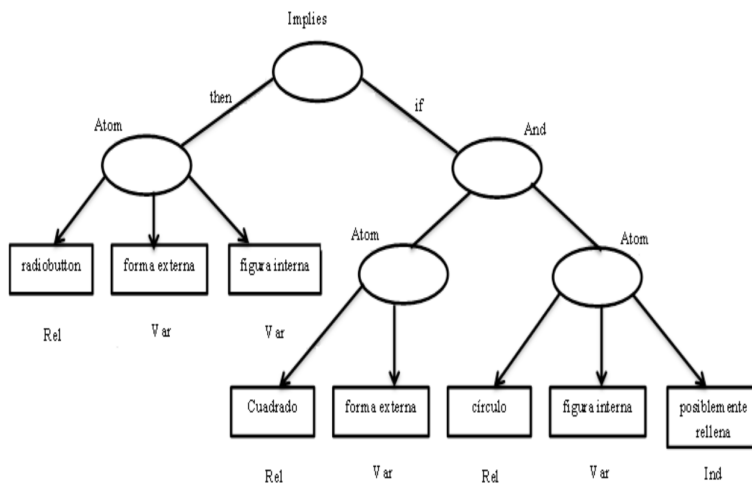


Fig. 3. Árbol parcial de reglas para el elemento *radiobutton* de una interfaz Web.

En la Fig. 3 se presenta un ejemplo del árbol de la regla definida para el elemento radiobutton de una interfaz Web. En la Fig. 4 se presenta el árbol de la regla para el elemento *dataspinner* de una interfaz móvil:

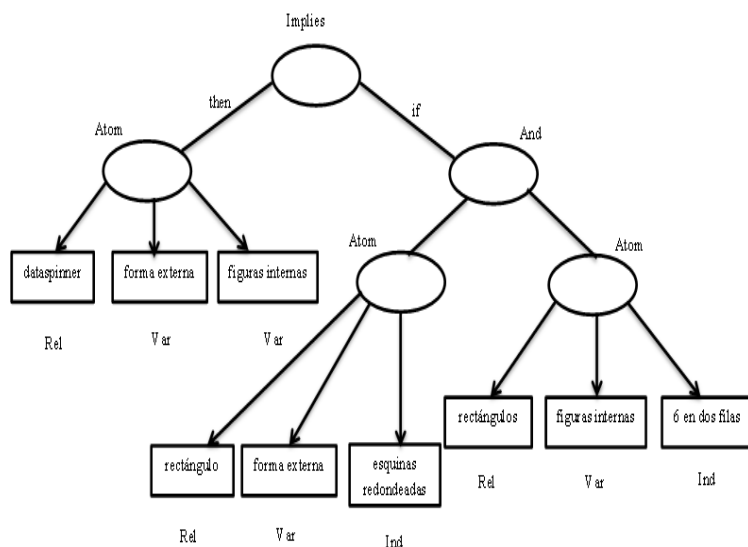


Fig. 4. Árbol parcial de reglas para el elemento *dataspinner* de una interfaz móvil.

Los arboles de reglas definidos en RuleML sirvieron como guía para su implementación en Matlab (*Matrix Laboratory*) [35], un software matemático que proporciona una biblioteca de funciones: *Image Processing Toolbox* para el procesamiento de imágenes. La *Image Processing Toolbox* de Matlab provee un conjunto de algoritmos bastante amplio, así como herramientas gráficas para procesamiento, análisis y visualización de imágenes digitales. Los principales algoritmos y funciones empleados para la identificación de elementos en el procesamiento de imágenes con Matlab son:

- 1) **bwlabel**: función que permite etiquetar los píxeles de cada objeto de manera única para su análisis, así como algunas de sus propiedades tales como la obtención de la imagen del objeto.
- 2) **bwmorph**: para realizar algunas operaciones morfológicas sobre la imagen del objeto.
- 3) **edge**: empleada para obtener los bordes de la imagen, a través del algoritmo de *Canny*.
- 4) **houghlines**: aplicada para la detección de líneas horizontales y verticales.
- 5) **corner**: permite obtener las coordenadas de las esquinas en una imagen.

5. Caso de estudio: generación de código a partir de una imagen que representa la portada principal de una página Web

En esta sección se presenta un caso de estudio como prueba de concepto del módulo propuesto. El caso de estudio representa la generación de código de la portada principal de una página Web a partir de la representación de su interfaz en una imagen. De acuerdo con las especificaciones establecidas para cada elemento se muestra a continuación un ejemplo que representan las especificaciones empleándose ya un diseño de interfaz. En la Fig. 5 se observa el ADV (*Abstract Data View*) de la portada principal de una página Web que consta de los siguientes elementos empezando de arriba hacia abajo: un elemento imagen que representa el banner, cuatro elementos ancla que representan un menú, un elemento label para dar la bienvenida, cuatro elementos imagen que representan un carrusel de imágenes y un elemento label con el pie de página.

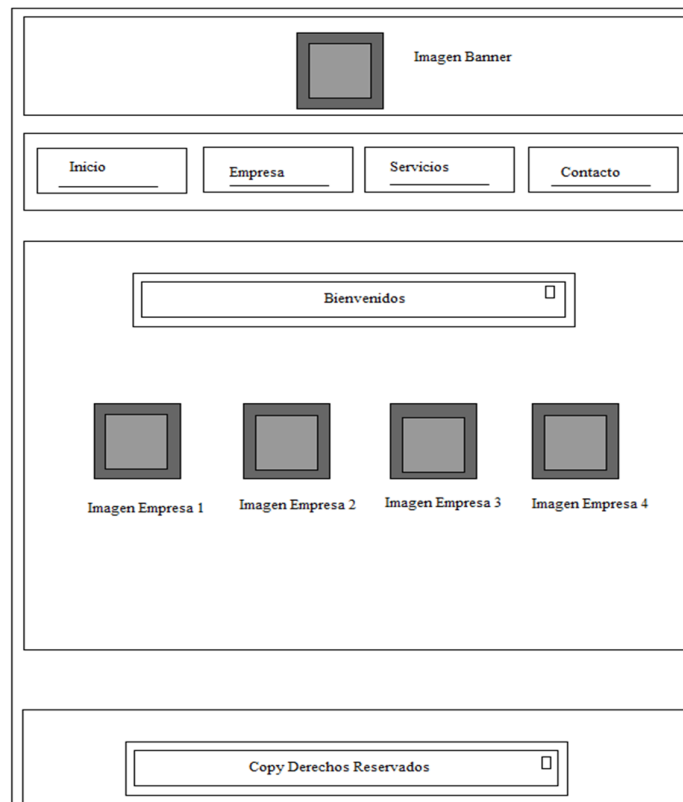


Fig. 5. Portada principal de una página Web.

A continuación se presenta la funcionalidad de un prototipo Web desarrollado en PHP que soporta el proceso de generación de software a partir del procesamiento de imágenes presentado en la Figura 1. Este prototipo es la herramienta generadora de código donde se ve como una caja negra, el parámetro de entrada es una imagen y el parámetro de salida es una aplicación multi-dispositivo empaquetada. La herramienta emplea un módulo ejecutable desarrollado en Matlab para el procesamiento de la interfaz, este módulo se invoca desde una clase en Java, la cual esta encapsulada en un servicio web, y es invocado desde este cliente PHP para generar el código. Los requerimientos mínimos para ejecutar el módulo de procesamiento son un equipo con 6 GB en memoria RAM, un procesador Core (TM) i7 CPU 2.80 GHz, y un sistema operativo de 64 bits.

Detallando el funcionamiento de la herramienta tenemos en la Fig. 6 (a) el formulario para que el usuario seleccione la imagen que se someterá al procesamiento, en este caso la Fig. 5 es la imagen de entrada. Después de cargar la imagen se analiza con la finalidad de reconocer elementos dentro de la imagen, como resultado se muestra al usuario el conjunto de elementos detectados en ella como se ilustra en la Fig. 6 (b), de estos elementos el usuario selecciona todos aquellos que desea sean parte de su aplicación. El paso 3 de la Fig. 6 (c) indica al usuario algunas características de la imagen y los elementos que selecciono en el paso anterior y de los cuales se generará el código.

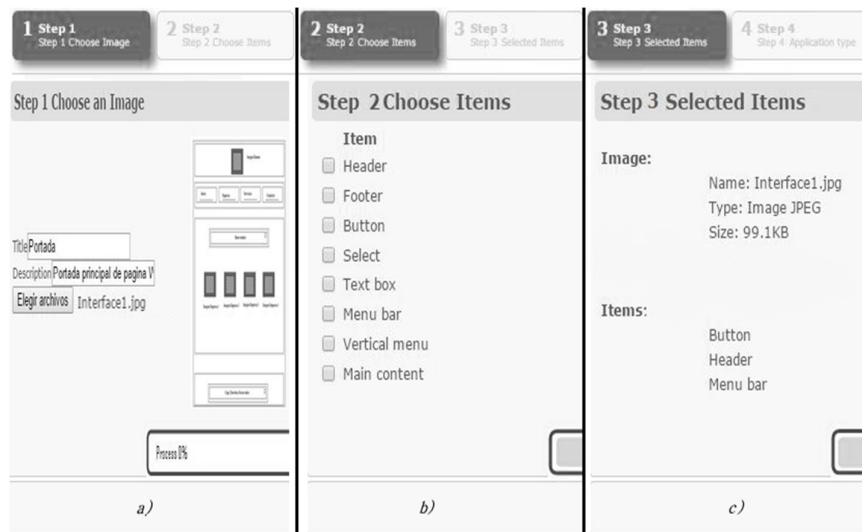


Fig. 6. a) Paso uno: selección de la imagen, b) Paso dos: selección de los elementos detectados, c) Paso tres: elementos seleccionados.

En la Fig. 7 (a) se muestra el paso 4 donde el usuario selecciona el tipo de aplicación que se generará: aplicación Web, aplicación de escritorio o aplicación móvil.

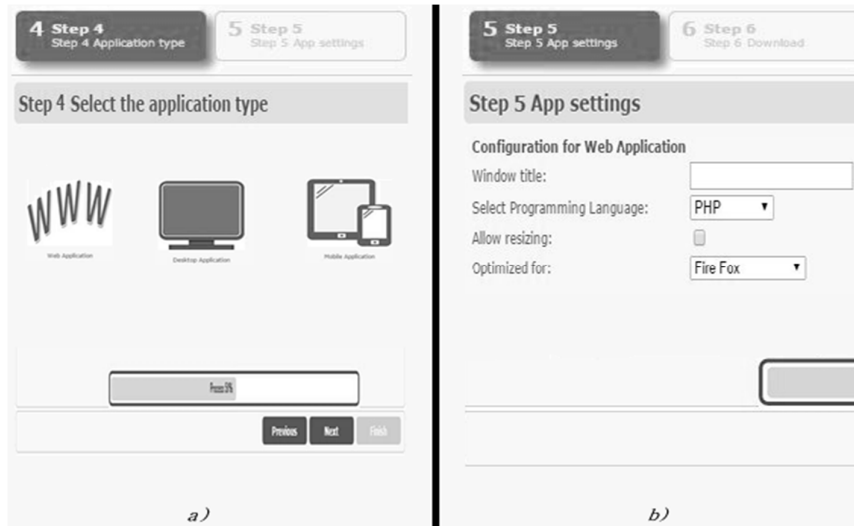


Fig. 7. a) Paso cuatro: selección del tipo de aplicación, b) Paso cinco: primera parte de la configuración de características generales de la aplicación.



Fig. 8. a) Paso cinco: segunda parte de la configuración de características generales de la aplicación y generación de código, b) Paso seis: descarga del código generado.

Después de seleccionar el tipo de aplicación en la Fig. 7 (b) el usuario configura algunos parámetros de acuerdo con el tipo de aplicación que seleccionó en el paso anterior. Para este ejemplo la configuración se muestra para una aplicación Web por lo que los parámetros solicitados son un título, el lenguaje de programación en que se debe generar el código, y el explorador para el que se desea optimizar la vista de nuestra interfaz.

En la Fig. 8 (a) el usuario configura algunas características más, tales como el color, ancho y alto, además de dar nombre y una pequeña descripción de la aplicación. Una vez enviados los datos de configuración se transforma cada elemento en su correspondiente código XML y se genera el código en el lenguaje seleccionado previamente. Finalmente en la Fig. 8 (b) el usuario puede descargar un archivo ZIP con el código en el lenguaje que seleccionó en los pasos anteriores.

6. Conclusiones y trabajo a futuro

Con base en la combinación de distintos algoritmos de procesamientos de imágenes se obtiene el reconocimiento de los distintos elementos que forman parte de una imagen representativa de una interfaz de usuario con lo que se tiene un primer aporte para una nueva y novedosa forma de desarrollar software.

Los aportes de este trabajo son prometedores en el área de la ingeniería de software en conjunción con técnicas de inteligencia artificial donde con la aplicación de un componente de software se apoyara a los desarrolladores facilitándoles la tarea de crear por ejemplo, prototipos de aplicaciones para presentarlas al usuario antes de iniciar con el desarrollo de una aplicación final, permitiendo así ahorrar tiempo y recursos económicos.

Como trabajo a futuro se pretende desarrollar una herramienta que permita a los usuarios generar las interfaces y por supuesto la generación automática del código de dicha interfaz. Además se espera aceptar que las imágenes provengan de bocetos hechos a mano o bien una cámara, por mencionar algunos ejemplos.

Por otra parte los resultados presentados aún son parciales con lo que se espera continuar con la puesta en marcha de distintos casos de estudio para evaluar correctamente el comportamiento del producto final, esto es; la correcta generación del código y porcentajes de error o eficiencia en el reconocimiento de elementos lo que permitirá continuar con la adaptación del procesamiento para otro tipo de imágenes.

Agradecimientos. Los autores agradecen el apoyado por el Consejo Nacional de Ciencia y Tecnología (CONACYT), Tecnológico Nacional de México (TecNM) y la Secretaría de Educación Pública (SEP) a través de PRODEP para la realización de este proyecto.

Referencias

1. De la Rosa, R.: Procesamiento de Imágenes Digitales. In: X Congreso Nacional en Informática y Computación del Instituto Tecnológico de Puebla (2007)
2. Eklund, A., Dufort, P., Forsberg, D., LaConte, S.M.: Medical image processing on the GPU—Past, present and future. *Medical image analysis* 17 (8), pp. 1073–1094 (2013)
3. Benoit, A., Caplier, A., Durette, B., Héroult, J.: Using human visual system modeling for bio-inspired low level image processing. *Computer vision and Image understanding* 114 (7), pp. 758–773 (2010)

4. Rué, F., Bijaoui, A.: A multiscale vision model applied to astronomical images. *Vistas in Astronomy*, 40(4), pp. 495–502 (1996)
5. Piszczek, M.: Laser Photography Examples of Processing of Image Information. *Acta Physica Polonica, A*, 124 (3), pp. 546–549 (2013)
6. Galván, J.V., Bertolino, S.R., Riveros, J.A.; Castellano, G.: Methodology for processing backscattered electron images. Application to Aguada archaeological paints. *Micron* 40 (8), pp. 793–799 (2009)
7. Obara, B.: An image processing algorithm for the reversed transformation of rotated microscope images. *Computers & geosciences*, 33(7), pp. 853–859 (2007)
8. Lindlar, F., Zimmermann, A.: A Code Generation Tool for Embedded Automotive Systems Based on Finite State Machines. In: *Industrial Informatics, INDIN 2008, 6th IEEE International Conference*, pp. 1539–1544 (2008)
9. Telea, A., Byelas, H., Voinea, L.: A Framework for Reverse Engineering Large C++ Code Bases. *Electronic Notes in Theoretical Computer Science*, Elsevier, pp. 143–159 (2009)
10. Iñesta, L., Aquino, N., Sánchez, J.: Framework and authoring tool for an extension of the UIML language. *Advances in Engineering Software*, Elsevier, pp. 1287–1296 (2009)
11. Santos, A., Koskimies, K., Lopes, A.: Automating the construction of domain-specific modeling languages for Object-Oriented Frameworks. *The Journal of Systems and Software*, Elsevier, pp. 1078–1093 (2010)
12. Colombo, M.L.O., Alor, H.G., Rodríguez, G.A.: A Novel Approach for Generating Multi-device Rich Internet Applications. In: *CONIELECOMP*, pp. 361–367 (2012)
13. Sánchez, P., Jiménez, M., Rosique, F., Álvarez, B., Iborra, A.: A Framework for developing home automation systems: From requirements to code. *The Journal of Systems and Software*, 84, Elsevier, pp. 1008–1021 (2011)
14. Sairaman, V., Ranganathan, N., Singh, N.: An Automatic Code Generation Tool for Partitioned software in Distributed Systems. In: *Proceedings of the 19th International Conference on VLSI Design held jointly with 5th International Conference on Embedded Systems Design (VLSID '06)*, pp. 477–480 (2006)
15. Hartrum, T.: Automated Code Generation Tools for Collaboration Systems. *Collaborative Technologies and Systems, International Symposium*, pp. 183–190 (2007)
16. Pinoli, J.: A general comparative study of the multiplicative homomorphic, log-ratio and logarithmic image processing approaches. *Signal Processing* 58, Elsevier, 11–45 (1997)
17. Zitová, B., Flusser, J.: Image registration methods: a survey. *Image and Vision Computing*, Elsevier, pp. 977–1000 (2003)
18. Shimomura, T.: Visual design and programming for Web applications. *Journal of Visual Languages and Computing*, Elsevier, pp. 213–230 (2005)
19. Casella, G., Deufemia, V., Mascardi, V., Costagliola, G., Martelli, M.: An agent-based framework for sketched symbol interpretation. *Journal of Visual Languages and Computing*, 19, Elsevier, pp. 225–257 (2008)
20. Brieler, F., Minas, M.: A model-based recognition engine for sketched diagrams. *Journal of Visual Languages and Computing*, 21, Elsevier, pp. 81–97 (2010)
21. Luoting, F., Levent, K.: From engineering diagrams to engineering models: Visual recognition and applications. *Computer-Aided Design*, 43, Elsevier, pp. 278–292 (2011)
22. Draheim, D., Weber, G., Lutteroth, C.: Finite state history modeling and its precise UML-based semantics. In: *Proceedings of the 2006 international conference on Advances in Conceptual Modeling: theory and practice (CoMoGIS'06)*, Springer-Verlag, pp. 43–52 (2006)
23. Komala, J., Punithavalli, M.: A survey on skeletons in digital image processing. In: *International Conference on Digital Image Processing (ICDIP '09)*, IEEE Computer Society, Washington, DC, pp. 260–269 (2009)
24. Prajapati, H., Vij, S.: Analytical Study of Parallel and Distributed Image Processing. In: *Image Information Processing (ICIIP) International Conference*, pp. 1–6 (2011)

25. Vincent, L.: Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *Image Processing, IEEE Transactions on*, 2(2), pp. 176–201 (1993)
26. Socarrás, L. M. G., Sarmiento, A. J. C., Solano, S. S., Jiménez, P. B.: Diseño de bloques de convolución para procesamiento de imágenes con FPGA. *Ingeniería Electrónica, Automática y Comunicaciones*, 32(3), pp. 56–69 (2011)
27. Betancur, A., J. A., Prieto O., F. A., Osorio L., G. A.: Segmentación de frutos de café mediante métodos de crecimiento de regiones. *Revista Facultad Nacional de Agronomía, Medellín*, 59(1), pp. 3311–3333 (2006)
28. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing (3rd Edition)*. Prentice Hall (2007).
29. Cowan, D. D., Lucena, C. J. P. D.: Abstract data views: An interface specification concept to enhance design for reuse. *Software Engineering, IEEE Transactions on*, 21(3), pp. 229–243 (1995)
30. Ordoñez, S.C.A.: Formatos de imagen digital. *Revista Digital Universitaria UNAM*, 10 (10) (2005)
31. Clark, J. E., Johnson, B. P.: *Sencha touch 2 mobile javascript framework*. Packt Publishing Ltd (2013)
32. David, M.: *Developing websites with jQuery mobile*. Taylor & Francis (2011)
33. Ghatol, R., Patel, Y.: *Beginning PhoneGap: Mobile Web Framework for JavaScript and HTML5*. Apress (2012)
34. Boley, H., Paschke, A., Shafiq, O.: RuleML 1.0: the overarching specification of web rules. *Lecture Notes in Computer Science*, 6403(4), pp. 162–178 (2010)
35. Guide, M. U. S.: *The mathworks*. Natick, MA, 5, 333 (1998)

Implementación del modelo Retinex aplicado al procesamiento de imágenes subacuáticas para mejorar su contenido cromático

Mario Dehesa, Alberto J. Rosales, Francisco J. Gallegos, Samuel Souverville,
Isabel V. Hernández

Instituto Politécnico Nacional, ESIME Zacatenco,
México DF, México

m_dehesa@hotmail.com, arosales@ipn.mx

Resumen. Una de las principales dificultades para una correcta captura de imágenes subacuáticas utilizando medios electrónicos (cámara fotográfica o de video), se presenta en el mismo ambiente subacuático en donde la iluminación y el tono de la fuente de luz cambian dependiendo de la profundidad del escenario para la captura de la imagen, esto debido a las diferentes longitudes de onda que se logran filtrar a diferentes niveles de profundidad, lo que trae como consecuencia la captura de diferentes tonos con condiciones de iluminación no uniformes, debido a la dispersión de la luz incidente y a las diferentes longitudes de onda capturadas por el sensor del dispositivo de adquisición, lo que provoca que los colores sean modificados. Como solución, se propone el uso del modelo Retinex, este modelo se basa en el entendimiento del Sistema Visual Humano (SVH-HVS) y hace uso del fenómeno de constancia del color, esto para disminuir el efecto producido mediante los problemas identificados anteriormente. Éste modelo se utilizará como una etapa de pre-procesamiento para que de esta manera se manipule la escena capturada, los resultados se verán mediante la modificación del contenido cromático de la imagen, principalmente en las áreas con poca iluminación. Finalmente, en este artículo se evaluarán los resultados experimentales de dos modelos aplicados a imágenes subacuáticas, Retinex simple y Retinex multiescala.

Palabras clave: cromaticidad, brillantez, fuente de iluminación, sistema de Visión, modelos Retinex.

1. Introducción

Un problema que se observa en los medios electrónicos para la captura de imágenes es que son muy sensibles a las diferentes fuentes de luz y los fenómenos naturales como pueden ser lluvia, niebla, contaminación, etc. Las imágenes capturadas con cámaras fotográficas y de video en presencia de estos fenómenos muestran dificultades en el reconocimiento de regiones y áreas de interés para determinadas aplicaciones, este tipo

de fenómenos produce una captura deficiente de la información, además de la presencia inherente del ruido propio añadido por el sistema de captura.

El modelo del Sistema de Visión Humano (HVS) tiene como característica propia la posibilidad de tener tanto las componentes de cromaticidad de los objetos, y de manera separada, la longitud de onda de la fuente de iluminación (componentes de luminosidad y de cromaticidad). Estas componentes podrán ser procesadas de manera independiente sin afectar las características en las cuales no se requiere manipular la información de luminosidad y de cromaticidad al mismo tiempo; por ejemplo, si en la noche se ilumina un espacio con un foco de filamento de tungsteno, éste nos dará una luz amarilla, pero alrededor del espacio de iluminación, los colores mantendrán su respectivo balance de color. Esta investigación tiene importantes aplicaciones prácticas, una de las principales es la que encontramos para la reproducción de imágenes fotográficas y de video. Además el modelo de constancia del color tiene aplicaciones en el campo de la radiografía médica, la fotografía submarina y la fotografía forense, entre otras [1], este fenómeno fue demostrado por primera vez por el Dr. Edwin H. Land. El Dr. Land demostró como el HVS es capaz de detectar la cantidad de energía que impacta a un objeto desde una fuente luminosa desconocida; esto produjo el desarrollo de la teoría de la “Constancia del Color” [2].

2. Imágenes y modelos

En este artículo se presenta el modelo Retinex aplicado al tratamiento de imágenes subacuáticas que se obtuvieron de la página de la NASA - Langley Research Center, y se usaron en el proyecto “Retinex Image Processing” [3].

El modelo Retinex propuesto por el Dr. Edwin Land, tiene la característica de utilizar operadores matemáticos, que asemejan el fenómeno de constancia del color presente en el HVS. La constancia del color (Sección 2.1) es una característica deseable en el pre-procesado de imágenes digitales. Para cumplir este propósito existen en la actualidad diferentes investigaciones que se han dedicado al mejoramiento del modelo Retinex propuesto por el Dr. Land [4].

Estos modelos reciben como entrada en cada pixel los valores del color rojo, verde, y azul, en donde el algoritmo Retinex es el encargado de realizar la estimación de la reflectancia en cada pixel [5], con el objetivo de poder estimar el valor de la fuente de iluminación, como se muestra en la ecuación (1).

$$f_i(x, y) = G(x, y)R_i(x, y)I_i, \quad (1)$$

donde I_i es la fuente de iluminación estimada en la escena como se muestra en la sección 2.2, $f_i(x, y)$ es la intensidad de cada pixel de una imagen en la posición (x, y) , $G(x, y)$ es un factor que depende de las características de la escena, $R_i(x, y)$ es la reflectancia de un punto del objeto en la misma posición. El sub-índice i corresponde al canal de color en la imagen ($i = \{\text{Rojo } (R), \text{ Verde } (G), \text{ Azul } (B)\}$).

2.1. Constancia del color

El modelo Retinex se basa principalmente en como el HVS comprende el color, el cual percibe de una escena el producto de la reflectancia y la distribución espectral de la iluminación, como se mostró en la ecuación (1) [2]. Desde el punto de vista fotométrico, un objeto de color rojo que es iluminado con una fuente luminosa en color verde, deberá producir la misma distribución espectral que un objeto de color verde iluminado con una fuente luminosa de color rojo [6]. Pero aún bajo estas condiciones, el color del objeto permanece estable independientemente de los cambios en la longitud de onda de la fuente luminosa, y esto se debe gracias a una teoría conocida como “Constancia del Color” [6].

La constancia del color se apoya en el campo de la radiometría en donde se describe como los receptores presentes en la retina miden la energía total captada en cada uno de los sensores presentes en la retina, esto se puede calcular integrando todas las longitudes de onda (λ), considerando tres tipos de conos sensibles en la banda del espectro magnético visible en el color rojo, verde y azul. En donde $S_i(\lambda)$ responde a una curva de tres receptores, el subíndice i representa cada uno de los canales de color ($i \in \{R, G, B\}$). Siendo $E(\lambda)$ la irradiancia, es decir, la potencia incidente sobre los receptores por unidad de superficie por determinada longitud de onda (λ). Entonces la energía medida (Q_i) por un solo receptor se calcular como en la ecuación (2)

$$Q_i = \int S_i(\lambda)E(\lambda) d\lambda. \quad (2)$$

El fenómeno de Constancia del Color es el encargado de separar la reflectancia de la iluminación de la escena [7], con lo cual se podrá definir la característica espectral de la fuente de luz, con esto se podrá calcular un valor aproximado de la reflectancia, como se muestra en la ecuación (3). En donde la irradiancia de la luz que es capturada por los sensores en la retina $E(\lambda)$, es el producto del porcentaje de la luz reflejada por un objeto $R(\lambda)$ y la irradiancia incidente en el objeto $L(\lambda)$, la irradiancia dependerá de las coordenadas (x, y). La teoría de la Constancia del Color desde el punto de vista de la percepción del HVS utiliza la distribución espacial y cromática de cada pixel para definir los colores de la escena [8].

$$E(\lambda) = R(\lambda)L(\lambda). \quad (3)$$

La función de un sensor electrónico de video es realizar la conversión de la escena desde el punto de vista fotométrico, el cual captura el producto de la reflectancia por la iluminación. Lo que se realiza con el modelo Retinex, es imitar el mecanismo del fenómeno Constancia del Color que posee el HVS. Un punto muy importante en el mecanismo de la teoría de Constancia del Color es que no puede determinarse el color de un objeto de manera aislada, es decir, se necesita de la comparación de otras longitudes de onda reflejadas por otros objetos presentes en la escena. Por lo tanto la energía Q_i puede ser modelada por la ecuación (4).

$$Q_i = \int S_i(\lambda)R(\lambda)L(\lambda) d\lambda. \quad (4)$$

De esta manera los receptores transmiten la información a las células bipolares situadas en la segunda capa. Las células bipolares a su vez forman conexiones sinápticas con las células ganglionares de la retina. En la retina existen alrededor de 126 millones de foto-receptores, en donde 120 millones son bastones y 6 millones son conos. La mayoría de los conos se encuentran dentro de una pequeña área llamada fovea, que se encuentra en el centro de la retina.

2.2. Modelo Retinex

El Dr. Land desarrollo la teoría para el modelo Retinex de percepción del color, de acuerdo con este modelo, tres receptores diferentes miden la energía en diferentes puntos del espectro visible, donde cada conjunto de receptores actúa como una unidad para procesar la energía medida. En un principio el Dr. Land supuso que el proceso de interpretación del color se iniciaba en la retina con la interpretación de la corteza visual, a este sistema retino-cerebral que procesa una parte del espectro visible lo llamó Retinex, en donde los operadores que forman parte del modelo Retinex tienen la cualidad de imitar a los receptores de luz del HVS [2].

El modelo iterativo fue descrito por el Dr. Land, en el cual el modelo Retinex se describe mediante operadores biológicos encargados de la recepción de la luz y se aplican de forma iterativa a la imagen a lo largo de una trayectoria o un conjunto de trayectorias y se compara el valor de cada pixel con los pixeles vecinos [2]. Las variaciones que podemos encontrar en la literatura para este modelo se dan en la forma en cómo se van a elegir los puntos de comparación y el orden en que serán comparados [9]. Para poder encontrar la mejor aproximación del color blanco, se realiza la búsqueda de la máxima intensidad en cada canal, como se muestra en la ecuación (5).

$$I_i = \max\{f_i(x, y)\}. \quad (5)$$

La ecuación (5) se puede cambiar calculando el histograma para cada componente de color, considerando a la fuente de luz como el punto en donde existe una alta acumulación de energía. En la ecuación (1), el algoritmo Retinex simple toma en cuenta el valor máximo en cada componente de color, $\max\{f_i(x, y)\}$, para toda la imagen, y lo utiliza como el valor representativo del color blanco a ser utilizado en la respectiva componente de la imagen [10].

El otro modelo Retinex de filtrado implementado es el modelo Retinex multiescala (MSR) [11]. Este modelo se basa en el modelo Retinex de Escala Simple para obtener un Retinex multiescala (MSR) que nos da como resultado el mejoramiento del rango dinámico y el tono, el modelo matemático lo podemos observar en la ecuación (6).

$$R_i(x_1, x_2) = \sum_{k=1}^K W_k \{\log I_i(x_1, x_2) - \log [F_k(x_1, x_2) * I_i(x_1, x_2)]\}. \quad (6)$$

La ecuación (6) nos representa un algoritmo que es capaz de reducir los cambios bruscos de brillo, color y saturación. En donde i nos representa el canal actual, $*$ expresa la convolución, \log es la función logaritmo natural, (x_1, x_2) es la coordenada espacial del pixel, I_i es la imagen actual, R_i es la salida del proceso, W_k es el peso asociado con F_k , K es el número de función envolvente, en donde la función envolvente F_k está definida por la ecuación (7) [12].

$$F_k(x_1, x_2) = K \exp[-(x_1^2 + x_2^2)/\sigma_k^2], \quad (7)$$

donde, σ_k es la desviación estándar típica de las envolventes Gaussianas y su magnitud controla la extensión de la envolvente, y toda la función es normalizada según k .

Para medir la eficiencia de mejoramiento del color para cada método Retinex propuesto, se utilizara un criterio de evaluación de cromaticidad que nos permita medir la magnitud del vector de cromaticidad en el espacio de color CIELab.

3. Método de evaluación del color

Los métodos de evaluación para medir la calidad de una imagen están muy relacionados con la comparación de una versión original o bajo algún criterio de imagen ideal que se propone en la literatura, la cuantificación de la distorsión de las imágenes es necesaria en diversos campos del procesamiento de imágenes. Pero la cuantificación del color incorpora elementos psicofísicos relacionados con la percepción humana y computacionalmente la evaluación se vuelve compleja. Sin embargo, los criterios basados en modelos de sistemas de visión humana (como pueden ser CIELab, HSB, HSL y otros) tienen mejores resultados en comparación con las evaluaciones basadas en la distorsión de la imagen [13].

Para realizar la evaluación de los modelos Retinex se utiliza el espacio de color CIELab, principalmente por su uniformidad de percepción y su utilización en los modelos RGB y sRGB. El espacio de color CIELab establece que para observar los colores, estos deben observarse bajo un fondo que va del color blanco al gris con una fuente de iluminación estándar tipo D_{65} que es una fuente de luz con características similares a la luz en el medio día, y la cual tiene una temperatura de 6,504 °K [14]. Cualquier fuente que cumpla con estos datos se le denomina fuente D_{65} .

Las tres componentes del sistema CIELab se representan con los vectores L^* , a^* , y b^* , en donde estos vectores representan Luminosidad, tonalidad de rojo a verde, y tonalidad de amarillo a azul respectivamente [15].

Para poder relacionar el espacio de color RGB al espacio CIELab, especialmente las bandas amarillo-azul y rojo-verde. Estas bandas están directamente relacionadas con las diferencias de las señales en los conos presentes en la retina, como se muestran en la ecuación (8).

$$C_1 = R_a - G_a; C_2 = G_a - B_a; C_3 = B_a - R_a. \quad (8)$$

El cálculo de los vectores a, y b se muestran en la ecuación (9) [14],

$$\begin{aligned} a &= [C_1 - C_2/11] = [R_a - 12G_a/11 + B_a/11], \\ b &= (1/2)[C_2 - C_1 + C_1 - C_3]/(4.5) = (1/9)[R_a + C_2 + 2B_a]. \end{aligned} \quad (9)$$

La cromaticidad es una componente representada en las coordenadas cilíndricas del espacio de color CIELab, esta componente está representada por un vector de magnitud entre dos componentes cromáticas del espacio de color CIELab como se observa en la Fig. 1, mientras más grande sea su valor en magnitud, mayor será la distancia que existe desde el origen del espacio CIELab al pixel en observación. Esto es, entre más grande

es el valor en magnitud del vector de cromaticidad, estará indicando la saturación del color o que color es más intenso. La ecuación para la obtención del vector de cromaticidad del CIELab se muestra en la ecuación (10), que muestra el modelo característico del espacio CIELab:

$$C_{ab} = \sqrt{a^2 + b^2}, \quad (10)$$

donde C_{ab} representa el vector de cromaticidad, a representa la percepción cromática rojo-verde y b representa la percepción cromática amarillo-azul, como se muestra en la Fig. 1.

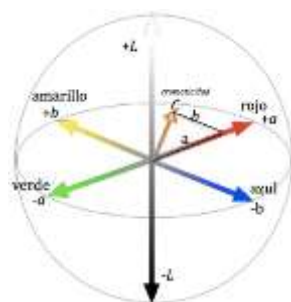


Fig. 1. Coordenadas del espacio de color CIELab.

La cromaticidad es un atributo relacionado con la intensidad del color, por tanto, entre mayor magnitud del vector de cromaticidad, mayor intensidad en el color. En esta evaluación se considera el promedio de cromaticidad en la imagen como el indicador de calidad percibida por el HVS, esto considerando que el ser humano tiende de manera natural en apreciar mejor las imágenes cuyos colores se encuentran más saturados considerándolas de mejor calidad. Por lo tanto para los resultados obtenidos, entre mayor sea el promedio de cromaticidad presente en la imagen, el color resultante de la imagen procesada será mejor [16].

4. Resultados


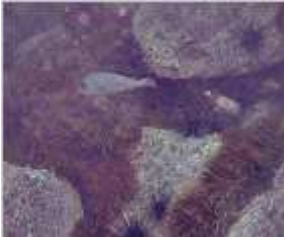


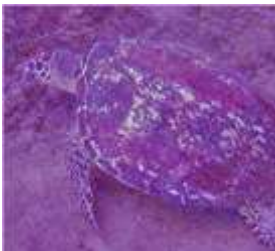

Las imágenes utilizadas como objetos de prueba, fueron procesadas utilizando dos modelos Retinex diferentes, y como resultado podemos observar subjetivamente y mediante el criterio de cuantificación sugerido, como las zonas oscuras así como la corrección del color fueron mejoradas. Posteriormente para su evaluación se realizó una transformación de la imagen del espacio RGB al espacio CIELab en sus coordenadas cilíndricas, y se calculó el promedio de la cromaticidad de la imagen. El mejor resultado se indica por el mayor valor obtenido que corresponde al algoritmo con un mejor desempeño, ya que mientras más grande es el valor del vector de cromaticidad tenemos un color más saturado, como se muestra en los resultados de cromaticidad al evaluar las imágenes procesadas con el algoritmo Retinex simple y multiescala. Las características del equipo de cómputo utilizado para la construcción de estos modelos




son las siguientes: Sistema operativo Mac OS X ver. 10.6.8, Procesador Intel Core 2 Duo 2.33 GHz con memoria RAM de 3GB. Lo anterior nos sirve como tercer criterio para la medida de la calidad de los algoritmos propuestos, al poder obtener los tiempos de procesamiento que toma cada variación del modelo Retinex, esto con el objetivo de cuantificar el tiempo de uso de los recursos computacionales por cada metodología, esto se denotará por la variable “Tp” y estará en segundos (s).

En la Tabla 1 podemos ver los resultados objetivos y subjetivos para cada imagen subacuática propuesta, al ser procesadas con los modelos Retinex simple y Retinex multiescala (MSR). El cálculo para la evaluación de cromaticidad en cada imagen se realizó utilizando la ecuación (10), primero se realizó la transformación del espacio de color RGB de la imagen a procesar al espacio de color CIELab, y se utilizaron las componentes a y b, promediados por las dimensiones de la imagen (M, N), donde M representa el número de filas y N el número de columnas, como se muestra en la ecuación (11):

$$C_{MN} = (1/MN) \sum \sqrt{a_{(x,y)}^2 + b_{(x,y)}^2}. \quad (11)$$

Tabla 1. Resultados de evaluación de los modelos Retinex.

Imagen original	Retinex Simple	Retinex MSR
 <p>Cromaticidad:170.64 Dimensiones: 500x416 Espacio de color: RGB Formato: JPG</p>	 <p>Cromaticidad:182.27 Tp: 1.3 s</p>	 <p>Cromaticidad:182.51 Tp: 6.4 s</p>
 <p>Cromaticidad: 201.60 Dimensiones: 500x454 Espacio de color: RGB</p>	 <p>Cromaticidad:188.10 Tp: 1.3 s</p>	 <p>Cromaticidad: 198.26 Tp: 6.2 s</p>

Formato: JPG		
		
Cromaticidad: 179.96 Dimensiones: 500x418 Espacio de color: RGB Formato: JPG	Cromaticidad: 187.65 Tp: 1.4 s	Cromaticidad: 183.08 Tp: 6.5 s

Se observa en los valores obtenidos de la evaluación cromática, como los modelos Retinex presentan diferente rendimiento comparando sus resultados de cromaticidad, en donde si el resultado de cromaticidad es mayor que el de la imagen procesada quiere decir que el contenido de colores es mayor, estos resultados dependieron principalmente de la iluminación homogénea o heterogénea (entendiéndose como iluminación homogénea en aquellas imágenes en donde se presenta una iluminación uniforme sobre toda la imagen y en la iluminación heterogénea se puede observar una iluminación variada sobre la misma) y del contenido de longitudes de onda de la fuente luminosa en la imagen.

5. Conclusiones

Para el procesamiento de imágenes subacuáticas, el método Retinex es una herramienta que sirvió para realizar correcciones de color en imágenes que han sido expuestas a diferentes fuentes de iluminación con diferente longitud de onda. En los resultados obtenidos podemos darnos cuenta que la eficiencia de los modelos Retinex está relacionada directamente con el contenido de longitudes de onda del que está compuesta la iluminación que afecta nuestra imagen, en donde cada modelo depende directamente de la iluminación homogénea o heterogénea y debido a estas variantes los resultados pueden tener una interpretación errónea al ser procesados con un método Retinex en particular.

En los resultados cromáticos, se observa como el método Retinex simple realiza las mejoras del contenido de color comparado el resultado con la cromaticidad de la imagen antes de procesarla, en donde la imagen esta iluminada de forma homogénea, mientras que el algoritmo multiescala realiza una mejor corrección del color de la imagen teniendo una iluminación heterogénea en la escena en comparación con el algoritmo Retinex simple.

Retinex puede ser aplicado a imágenes subacuáticas como una herramienta que realiza la corrección del color, debido a que el contenido de longitudes de onda que iluminan la escena depende de la profundidad a la que se captura la imagen ocasionando

que los colores capturados no correspondan a los colores que se obtienen bajo una iluminación tipo D_{65} .

Agradecimientos. Los autores agradecen al Instituto Politécnico Nacional de México y al CONACyT por su apoyo para la realización del presente trabajo de investigación.

Referencias

1. G. Hines, U. Rahman, G. Woodell: Single-Scale Retinex Using Digital Signal Processors. In: Global Signal Processing Expo (GSPx) (2004)
2. E. H. Land, J. J. McCann: Lightness and Retinex Theory. Journal of the Optical Society of America (1971)
3. D. Jobson, A. Moore, G. Woodell: NASA/Langley Research Center/Electromagnetics and Sensors Research Branch. [Online] <http://dragon.larc.nasa.gov> (2015)
4. J. M. Morel, A. B. Petro, C. Sbert: Fast implementation of color constancy algorithms. In: SPIE (2009)
5. M. Bertalmio, V. Caselles, E. Provenzi: Issues About Retinex Theory and Contrast Enhancement. Science+Business Media, Springer (2009)
6. M. Ebner: Color Constancy. Würzburg, Germany: Wiley-IS&T Series in Imaging Science and Technology (2007)
7. S. Seki: Vision of the Brain. Oxford: Wiley-Blackwell (1993)
8. A. Rizzi, C. Gatta: From Retinex to Automatic Color Equalization: issues in developing a new algorithm for unsupervised color equalization. Journal of Electronic Imaging (2004)
9. B. Funt, F. Ciurea: Retinex in matLaB™. Journal of electronic imaging (2004)
10. E. H. Land: The Retinex Theory of Color Vision. Scientific American (1977)
11. D. J. Jobson, Z. Rahman, G. A. Woodell: A Multiscale Retinex for Briding the Gap between Color Images and the Human Observation of Scenes. IEEE Transactions on Image Processing (1997)
12. Z. Rahman, D. J. Jobson, G. A. Woodell: Retinex Processing for Automatic Image Enhancement. Journal of Electronic Imaging (2004)
13. X. Zhang, B. A. Wandell: Colour image fidelity metrics evaluated using image distortion maps. Imaging Science and Techology Program, Department of Psychology (1998)
14. J. Schanda: Colorimetry: Understanding the CIE System. Wiley-Interscience (2007)
15. M. D. Fairchild: Color Appearance Models. Second Edition, England, John Wiley & Sons Ltd. (2005)
16. V. Tsagaris, G. Ghirstoulas, V. Anastassopoulos: A Measure for Evaluation of the Information Content in color images. In: IEEE Inter. Conf. Imag. Process. (2005)

Transmisión de video simultaneo en ancho de banda limitado aplicando esteganografía

Julio C. Suarez-Tapia¹, Blanca E. Carvajal-Gómez^{1,2},
Chadwick Carreto-Arellano¹

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

² Instituto Politécnico Nacional, Escuela Superior de Cómputo,
SEPI-ESCOM, México DF

{jcsuarez, becarvajal}@ipn.mx

Resumen. El problema de la transferencia de datos en medios limitados en ancho de banda, y la seguridad en la transferencia de estos, que en algunos casos como los medios inalámbricos deben de considerarse y administrarse sus recursos, han generado diversas propuestas de solución para este problema, sobre todo en infraestructura y en sistemas de administración de recursos. Sin embargo, en algunos casos como en infraestructura se invierte mucho tiempo y dinero. En este trabajo se presenta un algoritmo esteganográfico en video sin audio en formato MPEG-4, para la transferencia de información de otro video de igual formato. El objetivo de este trabajo es presentar una doble propuesta de solución a estas limitaciones. Logrando el envío de video estaremos ahorrando ancho de banda del medio, así como logrando la aplicación de una técnica de seguridad sobre la información enviada. La propuesta consiste en aplicar un kernel de 3x3 para la estimación del campo de varianza en cada imagen, aunado a la descomposición wavelet Haar para poder detectar las zonas ruidosas de la secuencia de video e insertar los datos correspondientes de las imágenes del video a ocultar. Los resultados presentados están en términos cuantitativos y cualitativos de la imagen. Verificamos la calidad del algoritmo con las métricas del PSNR, NCD, MAE, Q, RMS, así como los histogramas que muestran que no hay alteración de la distribución de color en las imágenes.

Palabras clave: esteganografía, ancho de banda, recursos limitados, video, wavelet, Haar, MPEG-4.

1. Introducción

Actualmente aumenta la demanda de utilización de nuevos servicios inalámbricos, como los de sistemas de comunicaciones móviles, redes de difusión de televisión digital terrestre o los diversos sistemas de acceso inalámbrico de banda ancha, como el enlace local inalámbrico, conocida como la conexión de última milla [1, 2]. Con la rápida

evolución tecnológica y el fenómeno de la convergencia, de los contenidos multimedia y de los dispositivos electrónicos, se está creando un entorno dinámico, es indispensable para el desarrollo de una adecuada y oportuna planificación y eficaz gestión, a los fines de optimizar su uso del ancho de banda. Los dispositivos móviles tienen muchas restricciones y limitaciones en relación al consumo de la energía en comparación con los dispositivos de una red cableada [1, 2]. La conservación de la energía de ambientes inalámbricos es un asunto importante que debe tomarse en cuenta. Las redes de este tipo están expuestas a muchos factores que contrarrestan el uso óptimo de la energía, tales como la continua comunicación entre los dispositivos, alojamiento de recursos y memoria, el uso eficiente de la batería de alimentación, el tráfico, etc. [3]

El uso eficiente de recursos energéticos tiene como objetivos la viabilidad económica, es decir que la tecnología sea económica; responsabilidad social mediante la construcción de tecnología que contribuya a minimizar los problemas de consumo irracional de energía y minimizar el impacto en el ambiente. En el contexto del cómputo móvil, redes inalámbricas y telefonía inalámbrica, se han realizado múltiples esfuerzos por optimizar el uso de recursos, tales como el caudal de comunicación y la energía.

Ejemplo de esto ha sido la telefonía celular, las redes ManNets y las redes Ad Hoc, cuya utilidad en sus inicios ha sido restringida por la capacidad limitada de almacenamiento de energía y mínimas optimizaciones para su administración. Por tal motivo, en el contexto de las redes inalámbricas, las optimizaciones han sido enfocadas a minimizar el número de dispositivos activos y de mensajes transmitidos por difusión, tomando como criterios de calidad el mantenimiento de conectividad y calidad en el servicio (QoS) en los flujos de comunicación [1].

Minimizar el número de dispositivos activos es logrado a través de la elección de un conjunto dominante de comunicación, es decir, un subconjunto mínimo de dispositivos que actúen con suficiente capacidad de almacenamiento de mensajes y energía [1]. El coste para desplegar una red inalámbrica es normalmente dominado por gastos de bienes inmuebles, planificación, mantenimiento, red de distribución, energía, etc. Por consiguiente crear la infraestructura de una red inalámbrica con pocos recursos es complicado, en lugares como las escuelas que hay muchas solicitudes a servicios multimedia el coste de desplegado en dispositivos móviles aumenta [1, 2].

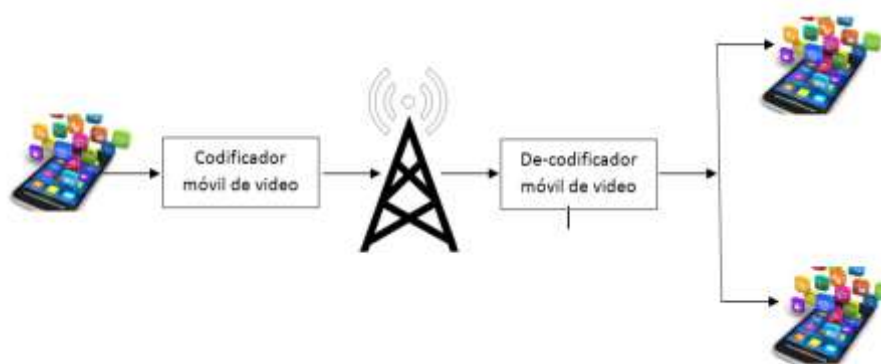


Fig. 1. Arquitectura del sistema propuesto.

Por consiguiente proponemos el uso de los algoritmos esteganográficos como una herramienta alterna a la ya utilizada comúnmente como protección intelectual, a una

herramienta eficaz del uso de la banda ancha en lugares donde las solicitudes de información son interrumpidas por el número de usuarios conectados. Proponiendo la siguiente arquitectura de solución colaboramos en una solución alterna para la transmisión y descarga de video de manera simultánea en dispositivos móviles, Figura 1.

Este trabajo está dividido en las siguientes secciones: sección 2 materiales y métodos, sección 3 teoría y cálculos, sección 4 discusión y sección 5 conclusiones.

2. Materiales y métodos

El ocultamiento de la información consiste en el proceso de integrar información o datos dentro de archivos digitales como: música, video e imágenes; estos archivos son conocidos como archivo anfitrión (host) [4]. Esta técnica es útil para el envío de información altamente confidencial como información de negocios y/o información personal durante la comunicación multimedia. En un sistema de imagen oculta, la imagen usada para llevar la información es llamada imagen anfitriona (cover image o host image). La imagen resultante, la cual ya se le inserto la información secreta deseada es conocida como estego-imagen [4-6]. Una de los resultados esperados es que la estego-imagen sea perceptualmente idéntica a la imagen portadora, con el fin de que los datos insertados no sean expuestos de manera tan fácil por el medio de transmisión. Posteriormente la estego-imagen se envía al receptor de la información [4-6]. Entonces finalmente el receptor podrá extraer la información oculta mediante la generación de una llave. Los datos ocultos son imperceptibles e imposibles de extraer por receptores no deseados o que no cuenten con la llave. La esteganografía en nuestros días es una técnica que se ha ido convirtiendo poco a poco en una herramienta de vital importancia para la protección de los derechos de autor, la cual permite un proceso de autenticación, el cual permite la distribución y uso legal de diferentes tipos de materiales. Una técnica esteganográfica es usualmente evaluada en términos de la calidad visual, de la capacidad de inserción, la calidad y porcentaje de la recuperación de los datos insertados; en otras palabras, un algoritmo esteganográfico ideal debe de tener una gran capacidad de ocultamiento y una excelente calidad visual del estego-objeto generado. La manera más razonable para hacer frente a esta disyuntiva es proponer técnicas que puedan proporcionar un equilibrio entre ambas [3, 7, 8], sin apartar de la vista el hecho de que se debe de recuperar totalmente los datos insertados en la estego-imagen. A juzgar por el hecho de la sensibilidad de la visión humana la cual debe de ser considerada en el diseño del algoritmo esteganográfico, los podemos categorizar los esquemas en tres tipos: (1) Los sistemas de alta capacidad de la inserción con calidad de imagen aceptable [4], (2) los regímenes de calidad de imagen alta con una capacidad moderada de inserción [4], y esquemas de alta eficiencia de inserción [4] con una ligera distorsión [4]. Por lo general, hay algunas técnicas para realizar

estego-imagen como: en el dominio del espacio, dominio de la frecuencia y métodos adaptativos (toman las mejores cualidades de los otros dominios y trabajan en ellos simultáneamente). El método más común para el dominio espacial es la modificación del bit menos significativo (LSB), el cual consiste en modificar el bit menos significativo en cada pixel de la imagen [4]. En el dominio de la frecuencia, se encuentra la transformada discreta de Fourier (DFT), la transformada del coseno discreto (DCT), y la transformada wavelet discreta (DWT), se utilizan para transformar los valores de los píxeles espaciales en coeficientes de frecuencia [4]. El ojo humano es menos sensible al ruido en la sub-banda de altas frecuencias [5, 6]. En aplicaciones donde el dominio de la frecuencia está implicado, dependiendo de la naturaleza de la imagen utilizada como portador, esta imagen puede ser alterada significativamente cuando se aplica un filtro digital como detectores de ruido. En esta investigación proponemos un algoritmo esteganográfico adaptativo montado sobre un dispositivo móvil, el cual va a ser utilizado como codificador para realizar la transmisión de 2 videos simultáneos sin audio. El estego-video deberá de llegar al receptor, donde también es un dispositivo móvil que servirá como receptor-decodificador del estego-video. El algoritmo esteganográfico propuesto y la manipulación del video es descrito en esta sección. El esquema propuesto consta de las siguientes etapas, Figura 2.

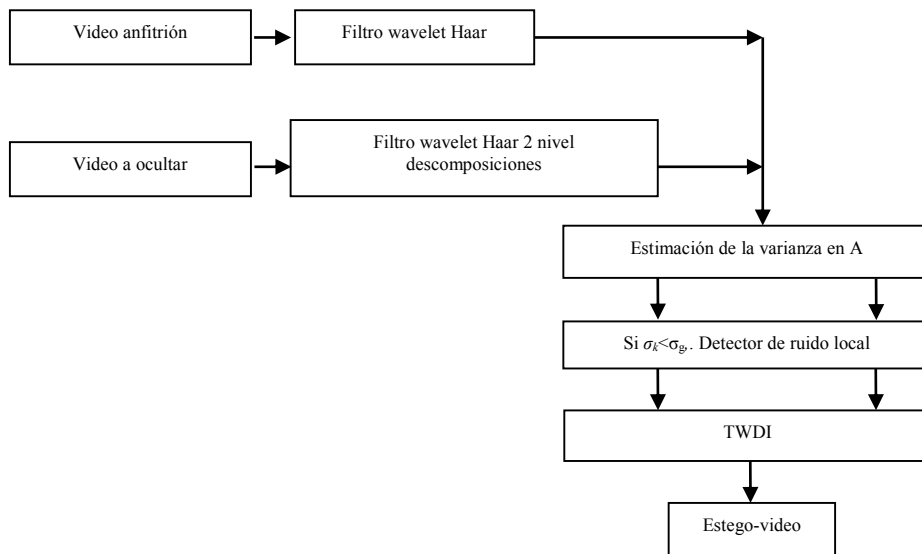


Fig. 2. Diagrama a bloques del algoritmo esteganográfico propuesto.

Este algoritmo consta de 5 etapas que serán explicadas a continuación:

Etapas de adquisición de video: El video adquirido consta de hasta 30 frames/segundo, es decir; 30 imágenes por segundo son transmitidos en un video de 5

minutos en un formato *.mp4. MPEG-4 es una de las técnicas más recientes de compresión de video y permite relaciones de compresión mucho menor que el pasado MPEG-2. MPEG-4 es ideal para aplicaciones de bajo ancho de banda, exactamente como se requiere para la transmisión de video en una red inalámbrica, separando el video en imágenes se puede analizar cada imagen por separado, para poder hacer la inserción de la información. La descomposición se realiza tanto en el video anfitrión como en el video a ocultar [7, 8].

Etapa de descomposición wavelet: El algoritmo esteganográfico propuesto se representa en la Figura 1. Este algoritmo se aplica en cada canal de la imagen RGB del video portador. De la Figura 1, el bloque de la redundancia de los enfoques utiliza un nivel de descomposición con la wavelet Haar. Esta transformada, una de sus características es que suaviza la imagen portadora a través de una operación de convolución doble (primero de descomposición, y después de la reconstrucción) de los coeficientes de aproximación (A), detalles horizontales (DH), detalles diagonales (DD) y detalles verticales (DV). Así en las muestras de la imagen anfitriona [5,6], proporciona mayor capacidad de ocultamiento y preservación de los detalles de la imagen [9-12].

Etapa de estimación simple del campo de varianza (ESCV): Para hacer el cálculo de la estimación simple del campo de varianza en la matriz de horizontales obtenida en la descomposición wavelet de la imagen portadora, se realiza el cálculo de la varianza general (σ_g) de cada sub-imagen [6],

$$\sigma_g = \sqrt{\frac{\sum_{m=1}^N (y_m - \bar{y})^2}{N}} \quad , \quad (1)$$

donde y_m es el m-esimo elemento de la sub-imagen portadora, $\bar{y} = \frac{\sum_{m=1}^n y_m}{N}$ es el valor medio de la actual imagen portadora, y N es el número total de elementos en la sub-imagen portadora. Para el cálculo de la estimación local actual, se aplica una ventana de 3x3 sobre la doble descomposición wavelet Haar de la sub-imagen del video a ocultar [6],

$$\sigma_v = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad , \quad (2)$$

donde x_i es el i-esimo elemento del actual kernel de la sub-imagen a ocultar, $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ es el valor medio del actual kernel, y $n=9$ es el número de elementos en la muestra del kernel. Obtenidos los valores de la estimación local y global del campo de varianza para cada uno de las imágenes del video portador como del video a ocultar entonces, se propone el siguiente umbral $\sigma_k < \sigma_g$.

Formalizando: “Si la estimación simple local del campo de varianza de la imagen a ocultar es menor a la estimación simple global del campo de varianza de la imagen

portadora se oculta la información correspondiente”. Con este criterio garantizamos la localización de las zonas ruidosas correspondientes.

Etapa de obtención de la estego-imagen: Para la obtención del estego-video, se van uniendo las estego-imágenes en un archivo alterno, donde se aplica la transformada wavelet discreta inversa en cada una de las imágenes portadoras para poder integrar finalmente el video MPEG-4, y así hacer el envío del estego-video.

Etapa de de-codificación del estego-video: Para poder recuperar los datos insertados en el estego-video, se sigue el mismo procedimiento realizado para la inserción, aplicando la condición siguiente: $\sigma_k < \sigma_g$. Finalmente se aplica dos niveles de recuperación de la transformada inversa utilizando la wavelet Haar.

3. Resultados

Los criterios usados para comparar esta relación de calidad con la imagen sin modificar son: la relación señal a ruido pico (PSNR), el error medio absoluto (MAE), la correlación (COI), el índice de calidad (Q), la desviación de color normalizada (NCD), y la capacidad de inserción (HC). Las pruebas se realizan en cada imagen que compone el video para finalmente obtener un promedio final de desempeño del algoritmo esteganográfico propuesto. La más común medida de distorsión es el PSNR [13],

$$\text{PSNR} = 10 \cdot \log \left[\frac{(255)^2}{\text{MSE}} \right], \text{ dB}, \quad (3)$$

donde, $\text{MSE} = \frac{1}{M_1 M_2} \sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \|y(i, j) - x(i, j)\|_{L_2}^2$ es el error medio cuadrático, donde M_1, M_2 son las dimensiones de la imagen, $y(i, j)$ es el vector 3D de los valores del pixel (i, j) de la estego-imagen, $x(i, j)$ es el correspondiente pixel en la imagen anfitriona, y $\| \cdot \|_{L_1}$, $\| \cdot \|_{L_2}$ son los vectores normales, respectivamente. La desviación del color normalizado (NCD) es usado para la cuantificación del error del color perceptual [13],

$$\text{NCD} = \frac{\sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \|\Delta E_{Luv}(i, j)\|_{L_2}}{\sum_{i=1}^{M_1} \sum_{j=1}^{M_2} \|E_{Luv}^*(i, j)\|_{L_2}} \quad (4)$$

Aquí, $\|\Delta E_{Luv}(i, j)\|_{L_2} = \left[(\Delta L^*)^2 + (\Delta u^*)^2 + (\Delta v^*)^2 \right]^{1/2}$ es la norma del error del color; ΔL^* , Δu^* , y Δv^* son las diferencias en los componentes L^* , u^* , y v^* ,

respectivamente, entre los dos vectores de color que representa a la estego-imagen y la imagen anfitriona para cada pixel (i,j) de una imagen, y $\|E_{Luv}^*(i,j)\|_{L_2} = \left[(L^*)^2 + (u^*)^2 + (v^*)^2 \right]^{1/2}$ es la norma o magnitud del pixel de la imagen anfitriona en el espacio $L^* u^* v^*$; El índice de calidad (Q) provee la calidad de la estego-imagen [13],

$$Q = \frac{4\sigma_{xy}xy}{(\sigma_x^2 + \sigma_y^2)(x^2 + y^2)}, \quad (5)$$

donde x y y son los valores medios de la imagen anfitriona y de la estego-imagen, respectivamente, σ_x^2 y σ_y^2 son las varianzas de la imagen anfitriona y de la estego-imagen, respectivamente, y $\sigma_{xy} = \frac{1}{N-1} \sum_i (x_i - x)(y_i - y)$. La capacidad de inserción (HC), determina el número de bits que pueden insertarse en la imagen portadora [5, 6, 13],

$$HC = MSE \times \frac{\text{number of samples in embedding band}}{\text{number of bits of secure data}}. \quad (6)$$

Las pruebas se realizaron sobre dos videos calidad MPEG-4, con una velocidad de 30 frames/segundo, y una tasa de 4Mbps/segundo, estándar establecido por la norma MPEG-4 para el video anfitrión, y para el video a ocultar presenta una velocidad de 30 imágenes/segundo con una tasa también de 4Mbps/segundo. Las tablas 1 y 2 muestran los resultados comparativos del estego-video, comparándolo con técnicas clásicas de esteganografía, como la LSB, la optimización del LSB (LSBO) [14,15]. A partir de estos resultados podremos validar si la técnica propuesta cumple con los tres compromisos básicos de las técnicas esteganográficas: *i*) capacidad de inserción, *ii*) calidad en la estego-imagen y *iii*) recuperación total de la imagen oculta. En la tabla 1 y 2 se muestra los resultados de desempeño del algoritmo propuesto en términos de PSNR, MAE, COI, Q, NCD y en HC para la tabla 1.

Tabla 1. Resultados comparativos de desempeño del algoritmo propuesto para el estego-video con el video a ocultar.

Criterios	Algoritmos esteganográficos				Estimación simple del campo de varianza
	4LSB	3LSB	4LSBO	3LSBO	
PSNR (dB)	36.168	36.169	30.111	30.112	54.7468
MAE	2.9911	2.9902	9.7307	9.7315	0.0721
CORR	99.78	99.78	99.76	99.76	99.9990
Q	0.9976	0.9976	0.9948	0.9948	0.9997
NCD	0.00073	0.00073	0.0022	0.0022	6.1616 e-6
HC (KB)	51,200	38,400	51,200	38,400	183,911

En la tabla 1 se puede observar que el algoritmo aquí propuesto cumple con los tres compromisos marcados para los algoritmos esteganográficos, si lo comparamos con los otros métodos mostrados podemos observar que la propuesta presenta: mayor capacidad de ocultamiento, mejor calidad de la estego-imagen que conforma el estego-video, y para el último compromiso podemos observar que la calidad de la imagen recuperada en términos de Q y COI supera a las métodos tomados de comparativos, sin embargo para el MAE se encuentra con un valor intermedio entre el LSB y el LSBO, para el PSNR se encuentra en segundo lugar después del 3LSB.

Tabla 2. Resultados comparativos del video recuperado.

Criterios	Algoritmos esteganográficos				
	4LSB	3LSB	4LSBO	3LSBO	Estimación simple del campo de varianza
PSNR (dB)	28.2910	29.1210	28.5090	28.144	28.7083
MAE	2.7469	2.7479	8.1255	8.1255	4.5247
CORR	99.26	99.25	99.24	99.17	99.34
Q	0.9926	0.9922	0.9924	0.9920	0.9950
NCD	0.0278	0.0203	0.0335	0.0332	0.0190

Finalmente en la Figura 3 se muestra la imagen de error de la estego-imagen con la imagen anfitriona.

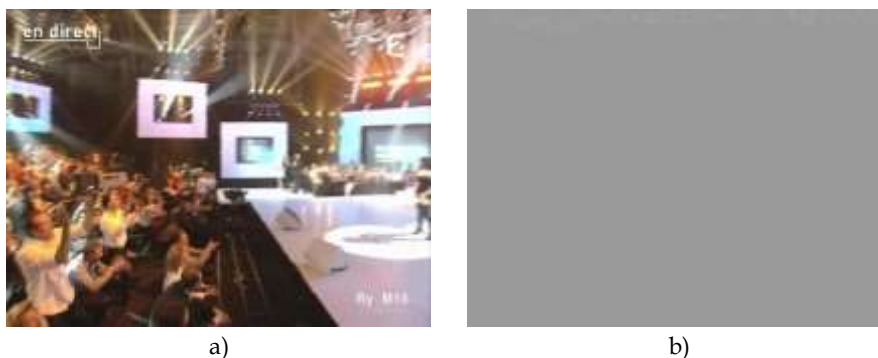


Fig. 3. a) estego-imagen, b) imagen error obtenido a partir de la imagen original con la estego-imagen.

Para poder validar el envío simultáneo de video, para este caso de estudio se cálculo el ancho de banda (AB) consumido por la transmisión del video anfitrión sobre un canal de transmisión, se toma como escenario real la situación de una cableada LAN, del Instituto Politécnico Nacional (IPN), el cual en horas picos de tráfico demora el envío y recepción de archivos multimedia, para el cálculo del AB se obtiene a partir de [16],

$$AB = Velocidad \times tamaño\ de\ la\ imagen\ del\ video\ promedio \times porcentaje\ de\ actividad \times 8 \quad (7)$$

Donde *velocidad* esta dado por el número de frames/segundo, *tamaño de la imagen* es el estándar marcado por la norma MPEG-4 que es de 720x560 para este caso, *porcentaje de actividad* es en horas pico de hasta el 80%. Así que obtenemos el siguiente resultado, para un video que dure 120 minutos con las características marcadas por la norma,

Tabla 3. Resultados comparativos del video recuperado.

Características de video	Aprovechamiento de Ancho de Banda				
	Video 1	Video 2	Video 3		Total
Velocidad (Kb/seg)	240	360	240	----	
Frames/Segundo	30	30	30	----	
Porcentaje de actividad(%)	80	78	90	----	
AB total por transmisión de video	1.57Mbps	2.30Mbps	1.769Mbps	=	5.56Mbps

Este resultado se multiplica por el número de usuarios conectados a la red por hora. Así que si tenemos una población de 500 usuarios máximos efectivos conectados, el AB necesario para satisfacer será la máxima demanda presentada en la tabla 3 por 500 usuarios, dando un total de 1,150 Mbps. Actualmente el IPN no cuenta con esta infraestructura para satisfacer esta demanda y su población va en aumento. Para la propuesta de solución en esta investigación si se toma el *Video 2* como video portador, con esta técnica al menos podría enviarse simultáneamente otro video u otro tipo de datos multimedia ocupando solamente el AB calculado de 2.30 Mbps, sin llegar a ocupar los 1,150 Mbps obtenidos por 500 usuarios en hora pico.

4. Discusión

El algoritmo esteganográfico aquí propuesto para la transmisión simultánea de video en un ancho de banda limitado, dio resultados satisfactorios. Estos servirán para poder ocultar datos multimedia con una capacidad para la portadora de hasta 183, 911 KB. Este algoritmo se irá perfeccionando para mejorar la recuperación de la información oculta, para así mejorar el PSNR de esta. De los resultados experimentales expuestos en este artículo, se puede ver que el método propuesto provee los tres compromisos básicos de los algoritmos esteganográficos que son: capacidad de inserción, calidad de la estego-imagen y la recuperación de la información oculta. Los resultados cuantitativos podemos observar que se tiene de hasta un 99.9% de correlación de la imagen original vs estego-imagen, de igual manera se observa que para la calidad de la imagen se conserva hasta un 99.9% de la calidad visual que se puede ver en la Figura 3 b), que se presenta la imagen de error de la estego-imagen y de la imagen a ocultar. Se tendrán que ir haciendo más pruebas para validar este método propuesto en diferentes escenarios, hasta este momento se ha validado la descarga simultanea del

video anfitrión que contiene la información oculta, disminuyendo así en horas picos el consumo de ancho de banda limitado para este caso de estudio.

5. Conclusión

El método propuesto para el ocultamiento de archivos multimedia en anchos de banda limitado, usa la desviación estándar general y particular de un kernel de 3x3 como detectores de ruidos en imágenes, en conjunto aplicando la descomposición de los filtros wavelet Haar proveen una mayor eficiencia de los recursos en medios limitados para así tener un mayor aprovechamiento de estos como lo es para este caso de estudio del ancho de banda limitado y que exista alta demanda de solicitudes en horas pico. Se presenta un compromiso integral entre el procesamiento del video portador descomponiéndolo en imágenes que sirven de anfitrionas a información multimedia. Los resultados obtenidos avalan estos compromisos anteriormente mencionados que son: calidad de la imagen, capacidad de ocultamiento y recuperación de la información insertada en el video anfitrión. Al obtener de manera satisfactoria este compromiso se puede aplicar en redes limitadas para el máximo aprovechamiento del ancho de banda total.

Agradecimientos. El equipo de trabajo agradece al IPN y a CONACYT en general por el soporte de esta investigación.

Referencias

1. Sgardoni V., Nix Andrew R.: Raptor Code-Aware Link Adaptation for Spectrally Efficient Unicast Video Streaming over Mobile Broadband Networks, *IEEE Transactions on Mobile Computing*, Vol. 14, No. 2, pp. 401–415 (2015)
2. Li Bo, Wang Peng, Zhang Yongfēi: Classification-based Multi-client Video. *Journal of Multimedia*, Vol. 8, No. 4, pp. 315–323.
3. TRICALCAR. www.wilac.net/tricalcar (2015)
4. Yuan-Hui, Yu, Chin-Chen Chang, Yu-Chen Hu: Hiding secret data in images via predictive coding. *Pattern Recognition*, 38, pp. 691–705 (2005)
5. Carvajal-Gómez, B.E., Gallegos-Funes, F.J., López-Bonilla, J.L.: Método simple de ocultamiento de datos para imágenes RGB basado en Estimación de la Varianza. In: *COMIA* (2012)
6. Carvajal-Gómez, B.E., Gallegos-Funes, F.J., Rosales-Silva, A.J.: Color local complexity estimation based steganographic (CLCES) method. *Expert Systems with Applications*, 40, pp. 1132–1142 (2013)
7. Yesilyurt, M., Yalman, Y., Turan Ozcerit, A.: A Robust Watermarking Method for Mpeg-4 Based on Kurtosis. *The Computer Journal*, doi: 10.1093/comjnl/bxu112 (2014)
8. Moccagattaole, T. R.: MPEG-4 video verification model: Status and directions. *International Journal of Imaging Systems and Technology*, Vol. 8, No. 5, pp. 468–479 (1997)
9. Daubechies, I.: Orthornormal bases of compactly supported wavelets. *Communications on Pure Applied Mathematics*, 41, pp. 909–996 (1988)

10. Debnath, L.: Wavelets and Signal Processing. Birkhauser, Berlin, 106 p. (2002)
11. Ginesu, G, Massidda, F., Giusto, D.D.: A multi-factors approach for image quality assessment based on a human visual system model. Signal Processing: Image communication, Vol. 21, No. 4, pp. 316–333 (2006)
12. Huang, K.Q., Wu, Z.Y., Fung, G.S.K., Chan, H.Y.: Color image denoising with wavelet thresholding based on human visual system model. Signal Processing: Image communication, Vol. 20, No. 2, pp. 115–127 (2005)
13. Carvajal-Gómez, B. E., Gallegos-Funes, F. J., Rosales-Silva, A. J., López-Bonilla J. L.: Adjust of energy with compactly supported orthogonal wavelet for steganographic algorithms using the scaling function $1/\sqrt{2^J}$. International Journal of Physical Sciences, Vol. 8, No. 4, pp. 157–166, doi: 10.5897/IJPS12.516 (2013)
14. Chan, C. K., Cheng, L. M.: Hiding data in images by simple LSB substitution. Pattern Recognition, vol. 37, pp. 469–474 (2004)
15. Yu, Y. H., Chang, C. C., Lin, I. C.: A new steganographic method for color and grayscale image hiding. Computer Vision and Image Understanding, Vol. 107, No. 3, pp. 183–194 (2007)
16. Cortés Hernández, G. A.: Cómo calcular el ancho de banda de un enlace para video. Primera parte, ww.rnd.com.ar (2015)

Diseño óptimo para transmisión de fuerza en un efector final

Eric Santiago-Valentín*, Adrián Solano-Palma, Pedro Bautista-Camino,
José Marco Antonio Rueda-Meléndez, Edgar Alfredo Portilla-Flores

Instituto Politécnico Nacional,
Centro de Innovación y Desarrollo Tecnológico en Cómputo,
México D.F., México

e.santiago.valentin@gmail.com, asolanop@ipn.mx, peterbc1@gmail.com,
jrueda@ipn.mx, aportilla@ipn.mx

Resumen. En este trabajo se presenta una solución novedosa para el diseño óptimo de un efector final tipo pinza, para la transmisión de fuerza constante en su espacio de trabajo o agarre. Para tal fin se propone un problema de optimización numérica con restricciones asociado al mecanismo, el cual se resuelve utilizando el algoritmo de evolución diferencial. La función objetivo propuesta es la transmisión de fuerza constante por parte del eslabón de agarre. Los resultados obtenidos muestran una transferencia de fuerza constante con diferentes configuraciones del efector final, comprobándose la naturaleza multimodal del problema.

Palabras clave: efector final, evolución diferencial, optimización, restricciones, síntesis, transmisión de fuerza.

1. Introducción

Actualmente, el proceso de diseño industrial obliga al ingeniero a utilizar metodologías alternas para obtener sistemas mejores y más eficientes. Una de éstas técnicas consiste en traducir el problema original de diseño en un problema de optimización numérica, con el propósito de obtener combinaciones válidas de valores para los parámetros que describen al sistema, buscando un desempeño óptimo del mismo. Sin embargo, en la mayoría de los casos se tienen problemas duros de optimización, cuya solución presenta una alta complejidad. En los últimos años los algoritmos evolutivos han tenido un gran auge, aplicándose exitosamente a la solución de problemas de ingeniería del mundo real. Un algoritmo líder dentro de esta clase de metodologías es el propuesto por Rainer Storn en 1995 [2], denominado Evolución Diferencial (ED); desde entonces su eficacia ha sido puesta a prueba una y otra vez, demostrando su utilidad en problemas de optimización que en muchas ocasiones no pueden resolverse por métodos clásicos y/o presentan una alta dificultad de implementación.

* Autor para correspondencia.

La optimización en aplicaciones industriales y tecnológicas tiene una alta importancia, debido a la disminución potencial de costos que se puede obtener, que va desde el ahorro de material hasta la reducción de los tiempos de ejecución. Los efectores finales tipo pinza (*grippers*) de dos dedos, cuya función principal es la manipulación de objetos, se utilizan ampliamente en automatización [1]. En el proceso de diseño de estos elementos es importante considerar aspectos tales como el tipo y las dimensiones del objeto a manipular, o el peso final del efector y su tipo de accionamiento (eléctrico o neumático), por mencionar algunos; todo ello con el fin de obtener una interfaz adecuada entre el sistema automatizado, el espacio de trabajo del efector final y el objeto de interés.

El diseño óptimo de grippers ha sido abordado desde diferentes perspectivas. En [3] se propone un algoritmo genético para diseñar un efector final robótico, mientras que Cuadrado et al. [4] desarrollaron una solución para grippers de dos dedos, aplicando una función nativa de Matlab®. Saravanan et al. [5] presentaron un estudio para el diseño óptimo de este elemento por medio de tres diferentes algoritmos evolutivos, planteándolo como un problema multi-objetivo. En [9] se hizo la optimización de dos efectores finales comerciales mediante un problema multi-objetivo utilizando el *toolbox* de Matlab®. En [6] se plantea el problema como multi-objetivo, resolviéndolo mediante el algoritmo de colonia artificial de hormigas, obteniendo una alta precisión en los resultados. Portilla et al. [7] resolvieron el caso de un efector final de tres dedos utilizando el algoritmo de forrajeo de bacterias, planteando el problema como mono-objetivo; de igual forma, en [8] se propuso un problema de optimización de un gripper para una planta embotelladora, utilizando ED pero aplicándolo únicamente a una de las piezas mecánicas del sistema y no a su totalidad.

En este trabajo se presenta una solución novedosa al problema de síntesis dimensional de un efector final tipo pinza, para la transmisión de fuerza constante en su espacio de trabajo o agarre, utilizando el algoritmo de ED. La organización del artículo es la siguiente: en la Sección 2 se plantea el problema y se analiza el mecanismo para la obtención de las ecuaciones que describen al sistema; en la Sección 3 se detalla la estrategia de optimización aplicada, basada en el algoritmo de ED, mientras que en la Sección 4 se definen los parámetros del sistema para su optimización. En la Sección 5 se describe el algoritmo de ED así como su implementación computacional; en la Sección 6 se revisan los resultados obtenidos y, finalmente, en la Sección 7 se presentan las conclusiones.

2. Problema de síntesis del efector final

Sea el efector final tipo pinza que se muestra en la Figura 1, integrado por los siguientes elementos: pieza base (1), tornillo de potencia y tuerca (2), eslabón de agarre (3) y eslabón de acoplamiento (4). El principio operacional del efector es el siguiente: fijando un sistema coordenado de referencia en la pieza base, la posición mínima de agarre ocurre cuando el desplazamiento de la tuerca es cercano a cero, es decir, cuando la posición relativa del elemento 2 es mínima respecto del origen de dicho sistema coordenado; así mismo, la máxima posición

se obtiene cuando la distancia entre la tuerca y el sistema de referencia es máxima. Una vez satisfecha la condición del rango de apertura del efector, se debe asegurar en dicho rango una fuerza de agarre constante sobre el objeto de interés.

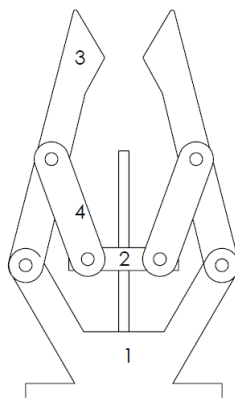


Fig. 1. Efector final

2.1. Cinemática del mecanismo

En la Figura 2 se observa un diagrama esquemático del mecanismo del efector final, donde cada vector r_i está relacionado con el i -ésimo eslabón del mecanismo. Del mecanismo propuesto se establece la ecuación de cierre de circuito como:

$$\mathbf{r}_1 + \mathbf{r}_3 = \mathbf{r}_2 + \mathbf{r}_0 + \mathbf{r}_4 \quad (1)$$

Aplicando notación polar a cada término de (1), se obtiene:

$$r_1 e^{j\theta_1} + r_3 e^{j\theta_3} = r_2 e^{j\theta_2} + r_0 e^{j\theta_0} + r_4 e^{j\theta_4} \quad (2)$$

Usando la ecuación de Euler en (2) y separando las partes real e imaginaria:

$$\begin{aligned} r_1 \cos\theta_1 + r_3 \cos\theta_3 &= -r_0 + r_4 \cos\theta_4 \\ r_1 \sin\theta_1 + r_3 \sin\theta_3 &= r_2 + r_4 \sin\theta_4 \end{aligned} \quad (3)$$

Para obtener la posición angular θ_3 , el lado izquierdo del sistema de ecuaciones (3) se expresa en términos de θ_4 :

$$\begin{aligned} r_4 \cos\theta_4 &= r_0 + r_1 \cos\theta_1 + r_3 \cos\theta_3 \\ r_4 \sin\theta_4 &= r_1 \sin\theta_1 + r_3 \sin\theta_3 - r_2 \end{aligned} \quad (4)$$

Elevando al cuadrado (4) y sumando sus términos se obtiene la ecuación de Freudenstein en su forma compacta [10], la cual se establece como:

$$A \cos \theta_3 + B \sin \theta_3 + C = 0 \tag{5}$$

donde:

$$A = 2r_3(r_0 + r_1 \cos \theta_1) \tag{6}$$

$$B = 2r_3(r_1 \sin \theta_1 - r_2) \tag{7}$$

$$C = r_0^2 + r_1^2 + r_2^2 + r_3^2 - r_4^2 + 2r_0r_1 \cos \theta_1 - 2r_1r_2 \sin \theta_1 \tag{8}$$

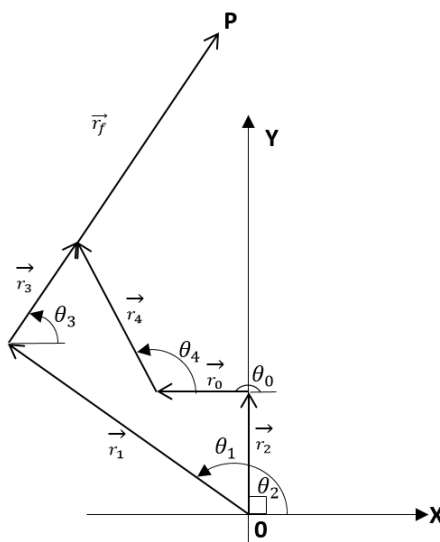


Fig. 2. Diagrama esquemático del efector final

El ángulo θ_3 puede ser calculado como una función de los parámetros A , B y C . Dicha solución puede ser obtenida al expresar $\sin \theta_3$ y $\cos \theta_3$ en términos de $\tan(\theta_3/2)$ como sigue:

$$\sin \theta_3 = \frac{2 \tan(\theta_3/2)}{1 + \tan^2(\theta_3/2)}, \quad \cos \theta_3 = \frac{1 - \tan^2(\theta_3/2)}{1 + \tan^2(\theta_3/2)} \tag{9}$$

sustituyendo éstas en (5), se obtiene una ecuación lineal de segundo orden:

$$[C - A] \tan^2(\theta_3/2) + [2B] \tan(\theta_3/2) + A + C = 0 \tag{10}$$

Resolviendo (10), la posición angular θ_3 esta dada por (11).

$$\theta_3 = 2 \arctan \left[\frac{-B \pm \sqrt{B^2 + A^2 - C^2}}{C - A} \right] \tag{11}$$

Un procedimiento similar al anterior se lleva a cabo para obtener θ_4 ; a partir de (3) se llega a la ecuación de Freudenstein en su forma compacta:

$$D\cos\theta_4 + E\sin\theta_4 + F = 0 \quad (12)$$

donde:

$$D = -2r_4(r_0 + r_1\cos\theta_1) \quad (13)$$

$$E = 2r_4(r_2 - r_1\sin\theta_1) \quad (14)$$

$$F = r_0^2 + r_1^2 + r_2^2 + r_4^2 - r_3^2 + 2r_0r_1\cos\theta_1 - 2r_1r_2\sin\theta_1 \quad (15)$$

Por lo tanto, la posición angular θ_4 puede calcularse como:

$$\theta_4 = 2\arctan \left[\frac{-E \pm \sqrt{D^2 + E^2 - F^2}}{F - D} \right] \quad (16)$$

Finalmente, la posición de los extremos del efector final P esta dada por:

$$P_x = r_1\cos\theta_1 + r_f\cos\theta_3 \quad (17)$$

$$P_y = r_1\sin\theta_1 + r_f\sin\theta_3 \quad (18)$$

2.2. Análisis de fuerzas en el efector final

Como se ha mencionado previamente, uno de los aspectos más importantes al diseñar un efector final tiene que ver con la fuerza de agarre o con la transmisión de la misma. La Figura 3 muestra la distribución de fuerzas en los elementos mecánicos del efector, donde P representa la fuerza de entrada o impulsión del sistema, F_T es la fuerza de agarre ejercida por el efector sobre el objeto de interés y F_{kj} representa la fuerza que ejerce el k -ésimo elemento sobre el j -ésimo. Como se puede observar:

$$F_{42i} = F_{42} \quad (19)$$

$$F_{24} = F_{34} \quad (20)$$

$$F_{34} = \frac{P}{2\sin\theta} \quad (21)$$

$$\theta = \pi - \theta_4 \quad (22)$$

Así mismo, tomando el momento de fuerza respecto del punto A y considerando $\sum M = 0$:

$$F_T = \frac{r_3\sin(\theta_4 - \theta_3)}{2r_f\sin\theta_3\sin\theta_4} P \quad (23)$$

3. Estrategias de optimización

Una vez que se han establecido apropiadamente la cinemática y la dependencia de la fuerza con las relaciones geométricas del mecanismo, el problema de diseño se puede definir como un problema de optimización numérica, por lo que se requiere especificar tanto las relaciones matemáticas que permitan evaluar el desempeño del sistema como las restricciones a los que estará sujeto.

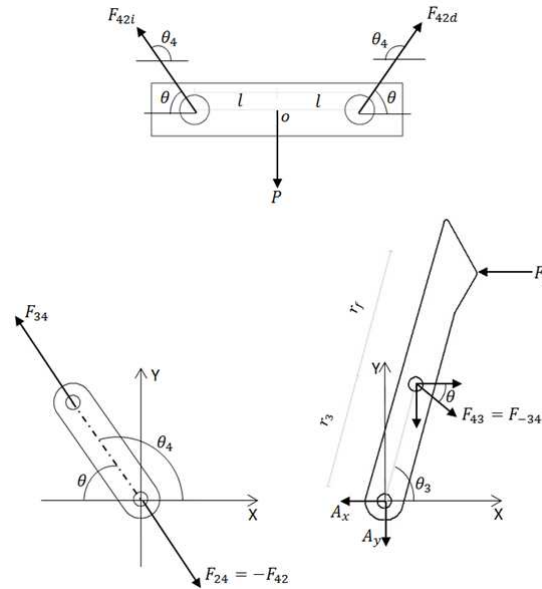


Fig. 3. Diagrama de la distribución de fuerzas en el efector final

3.1. Función objetivo

Un criterio para cuantificar el desempeño del efector final considera que la fuerza de sujeción debe ser constante en todo su espacio de trabajo. Como se vió en la sección anterior, la posición del extremo del efector depende de la tuerca en el tornillo de potencia, requiriéndose que la fuerza en los extremos de máximo y mínimo desplazamiento sea constante. Proponiendo la función objetivo de (24), ésta debe tener un valor mínimo para asegurar una fuerza casi constante en todo el espacio de trabajo; así, el valor ideal de la función es cero.

$$f(r_2) = (F_T(r_{2min}) - F_T(r_{2max}))^2 \quad (24)$$

Finalmente, es necesario definir el espacio de trabajo del efector final, para lo cual se tiene que:

- X_{min} : Dimensión mínima del objeto de interés.
- X_{max} : Dimensión máxima del objeto de interés.
- X_G : Rango máximo de desplazamiento de los extremos del efector.
- r_{2min} : Valor mínimo de la posición de la tuerca.
- r_{2max} : Valor máximo de la posición de la tuerca.
- R_{2max} : Rango de variación de la posición de la tuerca.

3.2. Restricciones de diseño

Debido a que el efector final es la interfaz entre un sistema automatizado y el mundo real, es necesario que la interacción entre ambos dominios sea correcta. En

este sentido, se debe asegurar la sujeción apropiada para las posiciones mínima y máxima del espacio de trabajo, es decir:

$$P_x(r_{2min}) \leq X_{min} \quad (25)$$

$$P_x(r_{2min}) \geq 0 \quad (26)$$

$$P_x(r_{2max}) \geq X_{max} \quad (27)$$

$$P_x(r_{2min}) \leq X_G \quad (28)$$

Considerando una geometría esférica para el objeto a asir, se debe cumplir que la coordenada P_y del extremo del efector sea mayor que la dimensión máxima del objeto más el rango de variación del tornillo de potencia y la tuerca, esto es:

$$P_y(r_{2max}) \geq X_{max} + R_{2max} \quad (29)$$

4. Diseño óptimo del mecanismo

Como se mencionó, en este trabajo se lleva a cabo el diseño óptimo para transmisión de fuerza de un efector final. Para ello se parametrizó el sistema, ya que una descripción apropiada de las variables permite al diseñador una amplia posibilidad de reconfiguración del sistema.

4.1. Variables de diseño

Sea el vector de variables de diseño para el efector final, establecido como:

$$\mathbf{p} = [p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8]^T \quad (30)$$

$$= [r_1, r_{2min}, r_{2max}, r_3, r_4, r_0, r_f, \theta_1]^T \quad (31)$$

donde las variables r_1, r_3, r_4, r_0, r_f corresponden a las longitudes de las barras del mecanismo, r_{2min} y r_{2max} a las posiciones de la tuerca en la mínima y máxima apertura y finalmente θ_1 al ángulo del extremo de la pieza base respecto del origen del sistema coordenado.

4.2. Problema de optimización

Sea el problema de optimización numérica mono-objetivo descrito por (32) hasta (46), para obtener la solución al problema de diseño de síntesis para la transmisión óptima de fuerza del efector final:

$$\begin{aligned} \text{Min } f(\mathbf{p}) &= [F_T(r_{2min}) - F_T(r_{2max})]^2 \\ \mathbf{p} &\in \mathbb{R}^8 \end{aligned} \quad (32)$$

con las cotas:

$$0 \leq p_i \leq 150, \quad i = 1, 4, 5, 6, 7 \quad (33)$$

$$0 \leq p_i \leq R_{2max}, \quad i = 2, 3 \quad (34)$$

$$\frac{\pi}{2} \leq p_i \leq \pi, \quad i = 8 \quad (35)$$

sujeto a:

$$g_1(\mathbf{p}) = |P_x(r_{2min})| - X_{min} \leq 0 \quad (36)$$

$$g_2(\mathbf{p}) = -|P_x(r_{2min})| \leq 0 \quad (37)$$

$$g_3(\mathbf{p}) = |P_x(r_{2max})| - X_G \leq 0 \quad (38)$$

$$g_4(\mathbf{p}) = X_{max} - |P_x(r_{2max})| \leq 0 \quad (39)$$

$$g_5(\mathbf{p}) = X_{max} + R_{2max} - P_y(r_{2max}) \leq 0 \quad (40)$$

$$g_6(\mathbf{p}) = p_2 - p_3 \leq 0 \quad (41)$$

$$g_7(\mathbf{p}) = p_4 - p_7 \leq 0 \quad (42)$$

y con el espacio de trabajo:

$$X_{min} = 20 \quad (43)$$

$$X_{max} = 150 \quad (44)$$

$$X_G = 200 \quad (45)$$

$$R_{2max} = 50 \quad (46)$$

5. Algoritmo de optimización

El algoritmo de Evolución Diferencial (ED) es una de las técnicas metaheurísticas más populares, y se ha aplicado para resolver diversos problemas no lineales, no diferenciables y multimodales [11]. Dicho algoritmo toma una población inicial aleatoria de soluciones, y en cada generación se producen nuevos individuos candidatos aplicando operadores de reproducción (cruza y mutación). La aptitud de cada descendiente se evalúa para que compita con el individuo padre, y así determinar cuál de ellos se conservará para la generación siguiente. Una de las principales ventajas de la ED es su número de parámetros de control, ya que solamente se requieren tres parámetros de entrada para controlar el proceso de búsqueda; esto es, el tamaño de la población NP , la constante de diferenciación F que controla la amplificación de la variación diferencial y el parámetro de control de cruce CR [12]. Las características generales de esta técnica son:

- Representación de soluciones como individuos
- Selección de padres
- Recombinación o cruce
- Mutación
- Selección de sobrevivientes y variantes

El pseudocódigo correspondiente a la ED se muestra en el Algoritmo 5.1; ahora bien, en la etapa de competencia para la sustitución generacional se utilizan las reglas de factibilidad de Deb [13], las cuales se listan a continuación:

1. Entre dos individuos factibles, se escoge al de mejor función objetivo.
2. Entre dos individuos no factibles, se escoge al que tenga un valor menor en la suma de violaciones a las restricciones.

3. Entre un individuo factible y otro no factible, se escoge al factible.

Algoritmo 5.1: Evolución Diferencial

```

1 Generar una población inicial aleatoria de tamaño NP;
2 Evaluar la aptitud y factibilidad de la población inicial;
3 repeat
4     Seleccionar un padre y dos individuos adicionales;
5     Realizar la cruza;
6     Generar hijo con mutación uniforme;
7     Evaluar la aptitud y factibilidad del hijo generado;
8     if el hijo es mejor que el padre con base en las reglas de Deb then
9         el hijo reemplaza al padre en la siguiente generación
10    else
11        se conserva individuo origen
12 until satisfacer condición de paro o terminar total de generaciones;

```

5.1. Implementación computacional

La implementación del algoritmo se realizó en Matlab R2013a®, sobre una plataforma computacional con las siguientes características: procesador Intel Core i7 @2.2GHz, con 8GB de memoria RAM y sistema operativo Windows 8.

En el programa se implementó un módulo para el cálculo de la función objetivo y las restricciones, con el fin de evaluar la aptitud de los individuos. La factibilidad se deduce con base en el cálculo de la suma de violación de restricciones (SVR); dicho valor indica si el individuo se encuentra en la zona factible [13]. Si las variables del vector de diseño cumplen con las restricciones se calcula la función objetivo; en caso contrario a ésta se le asigna un valor grande como penalización para fines de ahorro computacional (FO=1000). Así mismo, las características del problema se muestran en el Cuadro 1, donde li es el número de restricciones de desigualdad lineales, ni son las restricciones de desigualdad no lineales, le las restricciones de igualdad lineales, y ne las restricciones de igualdad no lineales.

Tabla 1. Características del problema

Max-Eval	N	Tipo Función	li	ni	le	ne
30,000	8	No Separable	7	0	0	2

En las simulaciones se comprobó que el tamaño de la población (NP) y número de generaciones (GMAX) son los parámetros más importantes para sintonizar el rendimiento del algoritmo. En este trabajo se utiliza como condición

de paro el número de evaluaciones (EVAL) en lugar del número de generaciones, para efectos de futuras comparaciones contra otros métodos de optimización.

6. Resultados

Después de algunas pruebas preliminares para la calibración del algoritmo se realizó un conjunto de treinta simulaciones, cuyos mejores resultados se muestran en el Cuadro 3. Los parámetros empleados fueron: tamaño de población $NP = 16$, número de evaluaciones $EVAL = 30,000$, factor de cruza $CR = [0.8, 1.0]$ por ejecución, y factor de escala $F = [0.3, 0.9]$ por generación; estos últimos dos parámetros se calcularon de forma aleatoria entre los rangos mencionados. Se puede observar que la mayoría de las simulaciones alcanzaron el valor ideal de la función objetivo ($FO = 0$) con respecto a lo planteado en la Sección 3.1, mientras que en los últimos dos casos se obtuvieron valores muy cercanos al valor óptimo. Desde el punto de vista algorítmico estos también se consideran casos de éxito, ya que ambos valores se encuentran dentro de la tolerancia establecida entre el óptimo conocido y el resultado, que en este caso fue de 1×10^{-4} .

Como un dato relevante se puede apreciar que el problema es multimodal, ya que se obtuvieron diferentes vectores de diseño con el mismo valor óptimo de la función objetivo. Por lo tanto, para esta clase de problemas el enfoque de la sintonización del algoritmo no debe centrarse en la exploración de la zona factible sino en la explotación de la misma; esto es, no es necesario utilizar un conjunto grande de individuos en la configuración inicial ya que con pocas soluciones se realiza una buena explotación de la zona factible con menos generaciones.

Tabla 2. Estadísticas de las simulaciones numéricas

Mejor	0
Mediana	0
Peor	1.316873E-06
Promedio	4.389577E-08
Desviación estándar	2.363859E-07

Los resultados del análisis estadístico correspondiente a las treinta simulaciones se incluyen en el Cuadro 2, donde se observa la solidez de las las soluciones obtenidas. Como se mencionó anteriormente se trata de un problema multimodal ya que la mayoría de las soluciones cumplen con el valor óptimo ideal de la función objetivo ($FO = 0$), con la ventaja de que dichas soluciones producen diferentes configuraciones del mecanismo. Esto se muestra en la Figura 4, en donde se aprecia el modelado de tres diferentes configuraciones de diseño tomadas del Cuadro 3, dichos modelos se realizaron con SolidWorks 2013®.

En la Figura 5 se muestra el comportamiento de la función objetivo en tres casos distintos; únicamente se incluye la parte esencial de la gráfica por razones

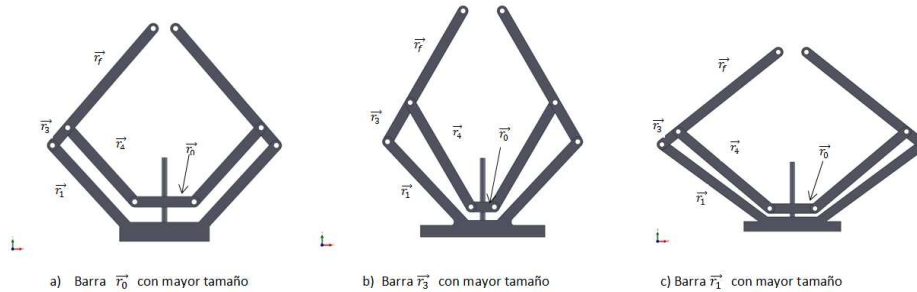


Fig. 4. Modelado de tres configuraciones distintas del efector final

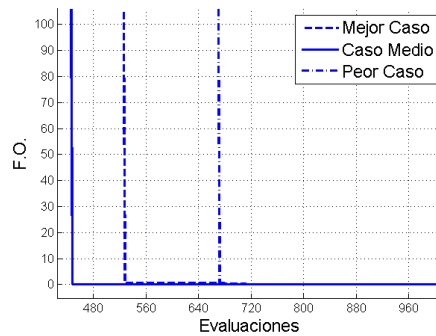


Fig. 5. Convergencia de la función objetivo hacia el valor óptimo

de visualización. Para el mejor caso se tomó la primera simulación del Cuadro 3, mientras que para los casos medio y peor se consideraron las dos soluciones con función objetivo diferente a cero. Considerando que las simulaciones se llevaron sobre 30,000 evaluaciones, se observa que en todos los casos el algoritmo presenta una convergencia muy rápida, entre las 400 y 650 evaluaciones. En la Figura 6 se muestra el comportamiento de los individuos que entran a la zona factible para los tres casos ya descritos; como se observa, nuevamente el algoritmo presenta un comportamiento muy estable ya que en todas las soluciones la totalidad de individuos entró a las zona factible después de un número relativamente corto de evaluaciones de la función objetivo (< 1500).

7. Conclusiones

En este trabajo se presenta una solución novedosa para la síntesis dimensional de un efector final, con transferencia constante de fuerza en todo el espacio de trabajo del mecanismo; dicha síntesis se llevó a cabo proponiendo un problema de optimización numérica el cual se resolvió utilizando el algoritmo de ED. Para este

Tabla 3. Mejores vectores de solución por cada simulación

N	r_1	r_{2min}	r_{2max}	r_3	r_4	r_0	r_f	θ_1	F.O.
1	125.6648	14.8219	49.8588	31.8815	78.1777	52.1401	148.8865	2.6537	0
2	132.1698	4.01253	48.0439	33.3737	110.0118	25.7951	149.6869	2.6122	0
3	110.2324	0.5953	48.3966	66.5903	132.3761	9.2966	149.7525	2.5073	0
4	123.4283	10.8431	48.5606	26.1936	107.3209	14.7612	146.0794	2.5258	0
5	111.9395	8.5666	49.8989	27.6270	102.0167	8.1545	146.0073	2.4823	0
6	118.4443	0.2539	49.9138	54.0889	122.0082	18.2222	149.8627	2.5485	0
7	108.3848	14.6305	47.4309	36.8179	102.9599	13.5255	149.8074	2.4559	0
8	109.7232	0.1354	48.9268	69.2182	142.7452	0.2063	149.6518	2.4521	0
9	117.4040	12.6399	47.0143	27.3896	111.9544	3.7957	149.5433	2.4668	0
10	111.1112	0.8470	49.8789	63.8318	136.8610	3.5517	149.6605	2.4586	0
11	104.6851	16.0943	49.2388	41.6483	109.7281	4.2353	149.9761	2.3857	0
12	119.6796	2.7717	42.1311	31.0122	120.6785	3.1324	148.6101	2.5256	0
13	143.9719	19.2837	47.6935	16.9678	61.93212	79.0936	149.6438	2.7446	0
14	137.6210	5.3280	48.3020	23.0749	118.8039	14.7306	144.3349	2.5958	0
15	130.8984	13.9086	49.6894	20.3562	100.1112	26.0418	148.8370	2.5755	0
16	146.6675	19.3563	49.8179	14.4672	130.6670	2.8650	149.8476	2.5628	0
17	125.8573	2.7284	49.7670	47.4768	110.2369	30.0987	149.7088	2.6155	0
18	118.2962	0.9036	49.5427	50.5869	130.9261	2.5862	149.4477	2.4894	0
19	145.1526	20.3894	49.1449	12.0728	116.7860	17.3613	149.7476	2.5925	0
20	148.4874	0.0173	45.2371	25.3724	141.7813	2.0523	149.8704	2.6161	0
21	126.0129	1.1487	48.9555	41.8592	100.4658	35.3734	149.8708	2.6562	0
22	123.3468	0.1855	49.3994	43.7044	131.2199	4.2993	148.1329	2.4976	0
23	107.0277	0.6777	49.9972	81.8748	147.6265	3.7487	149.8855	2.4635	0
24	114.3837	4.0352	49.6333	45.3223	123.3099	4.8782	149.9987	2.4612	0
25	124.6722	3.4510	49.6703	52.9403	101.1188	42.5436	149.7304	2.6740	0
26	126.9049	10.5596	49.9930	35.4906	81.6281	52.5143	149.3339	2.6739	0
27	118.0669	13.4724	48.6274	34.4630	100.2700	23.3688	149.2731	2.5284	0
28	113.7514	6.6159	49.8117	56.4621	115.5941	21.4816	149.3870	2.5377	0
29	130.8298	3.8528	5.93549	0.8488	104.9226	33.7778	149.9932	2.3785	1.61E-19
30	119.3823	35.8580	40.5029	1.9082	38.2731	75.0968	150	2.5201	1.31E-06

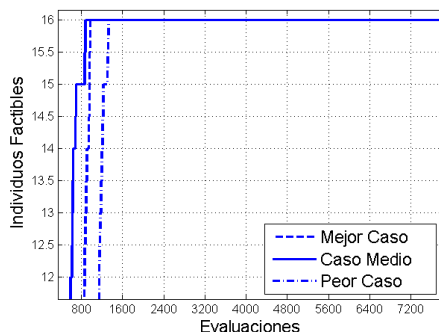


Fig. 6. Individuos factibles por evaluaciones

caso en particular se observó que, de acuerdo a los resultados obtenidos, existen varias soluciones que satisfacen el valor óptimo ideal de la función objetivo ($FO=0$), por lo que se puede concluir que es un problema de carácter multimodal. En este sentido, el algoritmo realiza una buena exploración y posterior explotación de la zona factible, para encontrar diversas soluciones óptimas; dichas soluciones representan diferentes configuraciones del mecanismo, que deben ser posteriormente analizadas desde el punto de vista de ingeniería para valorar si cumplen tanto con las restricciones de fabricación como de carácter estético. Lo anterior debido a que no todas las soluciones son físicamente construibles dadas las limitaciones de tamaño de algunos elementos, la cantidad de material que se emplearía para su construcción o la resistencia derivada de su tamaño y posición en el mecanismo, por mencionar algunos ejemplos.

Finalmente, se concluye que el uso de algoritmos evolutivos aplicados a problemas reales de ingeniería presenta un alto grado de confiabilidad y desempeño, siempre y cuando se haya realizado una sintonización adecuada para cada uno de los problemas a resolver. Así, esta clase de algoritmos presentan una opción viable de solución a problemas de optimización duros del mundo real, diferentes en su concepción a los problemas de *benchmark* con que usualmente se prueban estas técnicas de optimización; esto es, representan una herramienta de gran utilidad para problemas cuya solución no se conoce previamente.

Agradecimientos. Todos los autores agradecen el apoyo del Instituto Politécnico Nacional a través de la SIP vía el proyecto SIP-20151320. El primer autor agradece al CONACyT por la beca para estudios de posgrado en el CIDETEC-IPN.

Referencias

1. Sam, R. and Nefti, S.: Design and Development of Flexible Robotic Gripper for Handling Food Products. 10th Intl. Conf. on Control, Automation, Robotics and Vision, IEEE, 1689–1684, Hanoi, Vietnam, (2008)

2. Storn, R. and Price, K.: An Introduction to Differential Evolution, New Ideas in Optimization Differential Evolution - A simple and efficient adaptive scheme for global optimization over continuous spaces,)International Computer Science Institute, Muenchen, Germany (1995)
3. Osyczka, A., Krenich and Kara ,J.; Evolutionary multi criteria design optimization of robot grippers. Department of Mechanical Engineering, Cracow University of Technology, Polonia, (1999)
4. Cuadrado, J., Naya, M.A., Ceccarelli, M. and Carbone, G.: An Optimum Design Procedure for Two-Finger Grippers: A Case of Study. Universidad de La Coruña, España, (2001)
5. Saravanan, R., Ramabalan, S., Godwin Raja Ebenezer, N. and Dharmaraja, C.: Evolutionary multi criteria design optimization of robot grippers. Applied Soft Computing, Elsevier B.V., 159–172, Tamil Nadu, India, (2009)
6. Moses Mullar, S. and Satya Meher, R.: Optimizing of Robot Gripper Configurations Using Ant Colony Optimization. International Journal of Engineering Research and Technology, IJERT, 2655–2662, Andhra Pradesh, India, (2013)
7. Portilla-Flores E. A., Santiago-Valentin E., Solano Palma A., Vega Alvarado, E. y Calva Yáñez, M. B.: Cálculo de fuerza para agarre óptimo de un efector final de tres dedos utilizando el algoritmo de forrajeo de bacterias. Second International Conference on Advanced Mechatronics, Design, and Manufacturing Technology, 65–70, Colombia, (2014)
8. Chica Leal, A., Prada Jiménez, V., Rey Hernández, H. y Leon Hidalgo, C.: Diseño y construcción de un gripper implementando optimización. Second International Conference on Advanced Mechatronics, Design, and Manufacturing Technology, 219–222, Colombia, (2014)
9. Lanni, Ch., and Ceccarelli, M.: An Optimization Problem Algorithm for Kinematic Design of Mechanisms for Two-Finger Grippers. The Open Mechanical Engineering Journal, 49–62, Colombia, (2009)
10. Shigley J.E. y Uicker J.J. Jr.: Teoría de Máquinas y Mecanismos, McGraw Hill, México (1988)
11. Price K.V.: An Introduction to Differential Evolution, New Ideas in Optimization, Mc Graw Hill, UK (1999)
12. Boussaid, I., Lepagnot, J. and Siarry, P.: A Survey on Optimization Metaheuristics. Information Sciences. 237, 82–117 (2013)
13. Deb, K.: An Efficient Constraint Handling Method for Genetic Algorithms. Computer Methods in Applied Mechanics and Engineering. 186, 311–338 (2000)

Control par calculado difuso basado en pasividad para seguimiento de trayectorias de robots manipuladores

S.M. Orozco-Soto¹, J.C. Ramos-Fernández²

¹Universidad Tecnológica del Norte de Aguascalientes (UTNA),
Aguascalientes, México

²Universidad Politécnica de Pachuca (UPP),
Hidalgo, México

jramos@upp.edu.mx, santos.orozco@utna.edu.mx

Resumen. En este trabajo se presenta una alternativa para mejorar la técnica de control Par-Calculado para seguimiento de trayectorias de robots manipuladores. La estrategia de control propuesta utiliza lógica difusa para calcular la compensación requerida por el Control Par Calculado clásico ante variaciones en los parámetros del robot tales como masa o fricción. La innovación de esta propuesta es que la mejora del controlador Par Calculado se realiza en la parte de la compensación de las dinámicas en lugar de la parte tipo PD de la estructura del controlador. La estabilidad asintótica global se garantiza utilizando la teoría de la pasividad y el método directo de Lyapunov en el diseño del controlador. Los experimentos de simulación se realizaron utilizando el modelo dinámico de un robot manipulador de 2 GDL en configuración planar con presencia de pares gravitacionales. Los resultados satisfactorios, así como la sencillez de la estructura del controlador motivan a la implementación de esta propuesta en robots reales utilizando sistemas embebidos.

Palabras clave: par calculado, lógica difusa, seguimiento de trayectoria, pasividad, método directo de Lyapunov.

1. Introducción

El controlador Par Calculado es preferido para seguimiento de trayectorias de robots manipuladores debido a que garantiza la estabilidad asintótica global y calcula la cantidad de energía exacta que requieren los actuadores de las articulaciones de los robots para llevar a cabo las trayectorias deseadas [1-3]. Sin embargo, este controlador es susceptible a variaciones en el modelo del robot, incertidumbre en los parámetros del mismo o dinámicas no modeladas, por lo que se han desarrollado diversas alternativas para mejorar el desempeño de esta técnica, tales como complementos robustos[4], algoritmos genéticos [5] o estrategias basadas en lógica difusa [6-9]. La mayoría de estas mejoras consisten en la adaptación de la parte proporcional-derivativa (PD) de la estructura del controlador en lugar de adaptar la parte de la compensación de las dinámicas del robot. En este trabajo, se presenta una alternativa para mejorar el

rendimiento del control Par Calculado ante variaciones en los parámetros del robot utilizando lógica difusa. La estrategia de control propuesta fue diseñada utilizando las propiedades generales del modelo dinámico de robots manipuladores, la Teoría de Pasividad y el Segundo Método de Lyapunov, para garantizar estabilidad asintótica global durante el seguimiento de trayectoria. Para probar el controlador diseñado, se llevaron a cabo algunos experimentos en simulación utilizando el modelo dinámico de un robot manipulador de 2 grados de libertad (GDL) en configuración planar afectado por los pares gravitacionales. Los resultados satisfactorios del controlador propuesto motivan a la implementación de esta alternativa utilizando un microcontrolador o un sistema embebido para controlar un robot real. Este artículo está organizado de la siguiente manera: en la Sección 2, se presentan algunas propiedades del modelo dinámico de robots manipuladores, que son útiles para el diseño del controlador. En la Sección 3, se plantea el problema de la susceptibilidad del Control Par Calculado a las variaciones en los parámetros; asimismo, se formulan los objetivos de control. La Sección 4 describe el diseño del controlador propuesto utilizando lógica difusa y la Teoría de Lyapunov. Los resultados satisfactorios a nivel simulación se presentan en la Sección 5. Finalmente, en la Sección 6 se presentan las conclusiones.

2. Propiedades del modelo dinámico de robots manipuladores

En esta sección, se presentan algunas propiedades del modelo dinámico de robots manipuladores que son útiles para el diseño del controlador. La dinámica de los robots manipuladores puede ser representada matemáticamente por [10]:

$$D(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) + F(\dot{q}) = \tau, \quad (1)$$

donde $D(q) \in \mathbb{R}^{n \times n}$ es la matriz de fuerzas inerciales, $C(q, \dot{q}) \in \mathbb{R}^{n \times n}$ es la matriz de fuerzas centrípetas y de Coriolis, $G(q) \in \mathbb{R}^n$ es el vector de pares gravitacionales, $F(\dot{q}) \in \mathbb{R}^n$ es el vector de fricción y $\tau \in \mathbb{R}^n$ es el vector de entradas de control. El modelo dinámico (1) cuenta con las siguientes propiedades:

Propiedad 1. La matriz de fuerzas inerciales es simétrica y definida positiva, i.e. [11]:

$$\begin{aligned} D(q) &\equiv D(q)^T, \\ D(q) &> 0. \end{aligned} \quad (2)$$

Propiedad 2. La matriz de fuerzas inerciales y la matriz de fuerzas centrípetas y de Coriolis tienen la siguiente relación (propiedad de antisimetría) [12]:

$$\dot{q}^T \left[\frac{1}{2} \dot{D}(q) - C(q, \dot{q}) \right] \dot{q} = 0, \quad (3)$$

Propiedad 3. El modelo dinámico (1) puede ser parametrizado como [12]:

$$D(q)\ddot{q} + C(q, \dot{q})\dot{q} + G(q) + F(\dot{q}) = Y(q, \dot{q}, \ddot{q})^T \Phi, \quad (4)$$

donde $Y(q, \dot{q}, \ddot{q})^T \in \mathbb{R}^{n \times p}$ es una matriz de regresión y $\Phi \in \mathbb{R}^n$ es un vector conformado por parámetros constantes.

Propiedad 4. La energía total del sistema robótico (1) está dada por su ecuación Hamiltoniana como se muestra a continuación [12]:

$$H(q, \dot{q}) = \frac{1}{2} \dot{q}^T D(q) \dot{q} + G(q), \quad (5)$$

Derivando (5) se obtiene:

$$\dot{H}(q, \dot{q}) = \dot{q}^T \tau, \quad (6)$$

Al integrar ambos lados de (6), se puede observar que el sistema (1) es pasivo, es decir:

$$H(q, \dot{q}) - H(0) = \int_0^t \dot{q}^T \tau dt. \quad (7)$$

3. Formulación del problema

El Control Par Calculado es preferido para seguimiento de trayectoria de robots manipuladores debido a que calcula la energía exacta para llevar a cabo la tarea deseada [3]; este controlador se representa matemáticamente como [11]:

$$\tau = K_P \tilde{q} + K_D \dot{\tilde{q}} + D(q) \ddot{q}_d + C(q, \dot{q}) \dot{q}_d + G(q) + F(\dot{q}), \quad (8)$$

donde $\dot{q}_d \in \mathbb{R}^n$ y $\ddot{q}_d \in \mathbb{R}^n$ son los vectores de velocidades y aceleraciones deseadas respectivamente. Aplicando la *Propiedad 3*, el controlador (8) se puede expresar como sigue [12]:

$$\tau = K_P \tilde{q} + K_D \dot{\tilde{q}} + Y(q, \dot{q}_d, \ddot{q}_d)^T \Phi. \quad (9)$$

Nótese que este controlador depende directamente del modelado adecuado y de la correcta identificación paramétrica del robot. Sin embargo, si alguno de los parámetros del vector Φ , tales como la masa o la fricción, varía, el controlador no alcanzará la posición deseada o no seguirá la trayectoria requerida debido a que fue diseñado para calcular el par necesario de acuerdo a diferentes valores de los parámetros del robot. De esta manera, se establecen los siguientes objetivos de control:

Objetivo de control 1: Asegurar que el error de posición $\tilde{q} \rightarrow 0$ mientras $t \rightarrow \infty$. El error de posición está dado por: $\tilde{q} = q_d - q$.

Objetivo de control 2: Si alguno de los parámetros del vector Φ cambia, adaptar el controlador (9) para garantizar que el Objetivo de control 1 se cumpla.

El Objetivo de control puede alcanzarse utilizando del Control Par Calculado clásico (9), ya que garantiza la estabilidad global asintótica durante el seguimiento de trayectoria. Para lograr el Objetivo de control 2, se propone utilizar una versión adaptable del controlador (9), la cual está basada en lógica difusa para compensar variaciones en los parámetros del modelo del robot manipulador.

4. Diseño del control par calculado difuso

En esta sección, se presenta el diseño de la técnica de control propuesta para alcanzar los objetivos de control mencionados en la sección anterior. Dicha propuesta es una versión adaptable del Control Par Calculado (9), misma que utiliza lógica difusa para especificar el grado de compensación que requiere el controlador ante una variación en los parámetros del modelo del robot. La variación en los parámetros es detectada utilizando un índice de variación, el cual es también la variable de entrada del sistema difuso; la salida del sistema difuso es una ganancia variable agregada al Control Par Calculado, la cual es útil para aumentar o disminuir energía al sistema dependiendo de las variaciones en los parámetros del robot y del par requerido para desarrollar la tarea deseada.

4.1. Variable fusificada

La variable utilizada como entrada del sistema difuso es el índice de variación dado por:

$$\Delta\Phi = \frac{\|\Phi_R\|}{\|\Phi\|}, \quad (10)$$

donde $\|\Phi_R\|$ es la norma Euclidiana del vector de parámetros con variaciones y $\|\Phi\|$ es la norma Euclidiana del vector de parámetros original. En la práctica, las normas de los vectores pueden ser sustituidas por mediciones de par. De esta forma, el índice de variación $\Delta\Phi$ se introduce al sistema difuso para calcular la energía requerida por los actuadores para realizar la tarea deseada en caso de una variación en los parámetros del robot.

4.2. Particiones difusas

El universo de discurso del sistema difuso fue seleccionado de acuerdo a los posibles valores del índice de variación, desde 0.6, que significa casi la mitad del par requerido por los actuadores, hasta 2, lo cual supone la posibilidad de duplicar dicho par. Se decidió utilizar 5 particiones difusas para fusificar el índice de variación, mismas que se ilustran en la Figura 1. La partición *High down* sugiere una pérdida considerable de masa o de fricción en el robot, mientras que *High up* supone un incremento significativo en la masa o la fricción. *Low up* y *Low down* abordan cambios razonables de los parámetros incluyendo rangos de hasta -20% y hasta $+30\%$ del par requerido. La partición *Minimum* maneja variaciones de hasta ± 5 . Nótese que todas las particiones

están configuradas de la forma *Hard Fuzzy*, lo cual significa que la suma de todos los valores de pertenencia da por resultado 1; dicha configuración facilita la implementación de este controlador en la mayoría de los dispositivos programables.

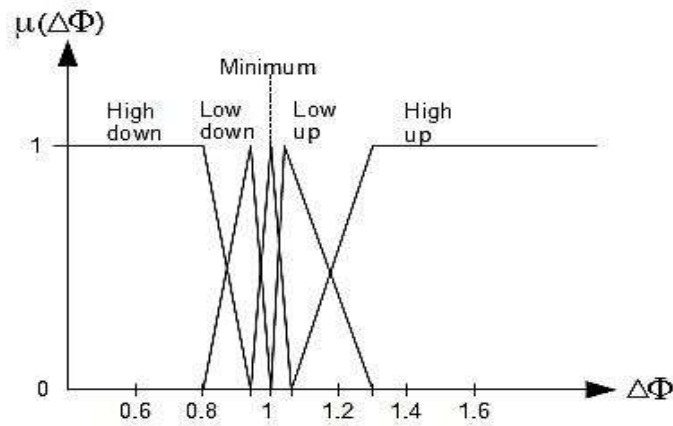


Fig. 1. Particiones difusas y universo de discurso del índice de variación.

4.3. Reglas y superficie difusas

Las reglas difusas fueron creadas realizando un mapeo del valor de pertenencia para cada partición difusa del índice de variación, hacia un valor duro, el cual es una ganancia llamada k_f . Las reglas difusas realizadas se pueden observar en la Tabla 1. Obsérvese que para pérdidas de masa o de fricción, el par debe ser reducido. Para incrementos de dichos parámetros, la compensación también debe incrementar. Para el caso de variaciones mínimas, el controlador no se ve afectado considerablemente, por lo que la parte tipo PD de la estructura del mismo debe ser capaz de manejar dichas perturbaciones; de esta forma, al no existir compensación, $k_f = 1$.

Tabla 1. Reglas difusas.

Si $\Delta\Phi$ es HIGH DOWN, entonces, k_f es 0.2
Si $\Delta\Phi$ es LOW DOWN, entonces, k_f es 0.8
Si $\Delta\Phi$ es MINIMUM, entonces, k_f es 1
Si $\Delta\Phi$ es LOW UP, entonces, k_f es 1.2
Si $\Delta\Phi$ es HIGH UP, entonces, k_f es 2

La superficie difusa que representa gráficamente el mapeo del índice de variación hacia la ganancia de compensación por medio de las reglas difusas se ilustra en la Figura 2, donde se puede apreciar que el máximo valor de k_f es capaz de compensar la energía necesaria incluso si el par incrementa al doble. En la práctica, el éxito del controlador

depende directamente de la capacidad de los actuadores y de la electrónica de potencia para suministrar la energía calculada.

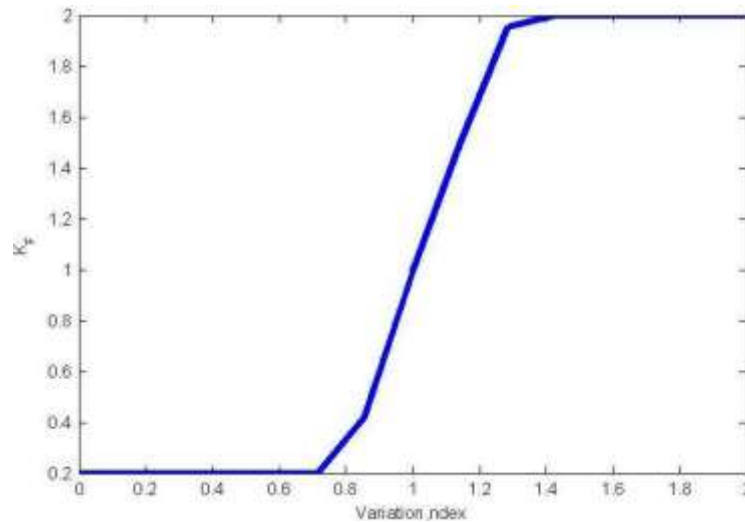


Fig. 2. Superficie difusa.

4.4. Defusificación

El método de defusificación seleccionado es la media ponderada, misma que se representa de la siguiente forma [13-14]:

$$K_F = \frac{\sum_{i=1}^5 \mu_i(\Delta\Phi) k_{f_i}}{\sum_{i=1}^5 \mu_i(\Delta\Phi)}, \quad (11)$$

donde K_F es una salida escalar que se utiliza como ganancia en el controlador, i es el número de particiones difusas, $\mu_i(\Delta\Phi)$ es el valor de pertenencia del índice de variación (10) para cada partición difusa y k_{f_i} es la salida correspondiente para cada partición. La ganancia K_F es el valor duro que representa la compensación requerida para mejorar el rendimiento del Control Par Calculado (9) ante variaciones en los parámetros del robot manipulador.

4.5. Diseño del controlador

El controlador fue diseñado utilizando el Método Directo de Lyapunov, mismo que se describe en el siguiente teorema [15]:

Teorema 4.1 (Estabilidad Global) *Una función escalar V del estado \mathbf{x} debe tener derivadas continuas de primer orden tales que:*

- $V(\mathbf{x})$ sea definida positiva

- $\dot{V}(\mathbf{x})$ sea definida negativa
- $V(\mathbf{x}) \rightarrow 0$ mientras $\|\mathbf{x}\| \rightarrow \infty$.

De esta forma, se propone la siguiente función candidata a ser Lyapunov:

$$V(q, \tilde{q}, \dot{\tilde{q}}) = \frac{1}{2} \dot{\tilde{q}}^T D(q) \dot{\tilde{q}} + \frac{1}{2} \tilde{q}^T K_P \tilde{q} . \quad (12)$$

Como K_P es una matriz diagonal constante, la función (12) satisface la primer condición del Teorema 4.1. Derivando (12) se obtiene:

$$\dot{V}(q, \dot{q}, \tilde{q}) = \dot{\tilde{q}}^T D(q) \ddot{\tilde{q}} + \frac{1}{2} \dot{\tilde{q}}^T \dot{D}(q) \dot{\tilde{q}} + \dot{\tilde{q}}^T K_P \tilde{q} . \quad (13)$$

Cabe mencionar que $D(q)\ddot{\tilde{q}} = D(q)\ddot{q}_d - D(q)\ddot{q}$, por lo tanto:

$$\begin{aligned} \dot{V}(q, \dot{q}, \tilde{q}) &= \dot{\tilde{q}}^T D(q) \ddot{q}_d - \dot{\tilde{q}}^T [\tau - C(q, \dot{q})\dot{q} - G(q) - F(\dot{q})] + \frac{1}{2} \dot{\tilde{q}}^T \dot{D}(q) \dot{\tilde{q}} \\ &+ \dot{\tilde{q}}^T K_P \tilde{q} . \end{aligned} \quad (14)$$

Como $C(q, \dot{q})\dot{q} = C(q, \dot{q})\dot{q}_d - C(q, \dot{q})\dot{\tilde{q}}$, se puede sustituir en (14) como sigue:

$$\begin{aligned} \dot{V}(q, \dot{q}, \tilde{q}) &= \dot{\tilde{q}}^T D(q) \ddot{q}_d - \dot{\tilde{q}}^T [\tau - C(q, \dot{q})\dot{q}_d + C(q, \dot{q})\dot{\tilde{q}} - G(q) - F(\dot{q})] \\ &+ \frac{1}{2} \dot{\tilde{q}}^T \dot{D}(q) \dot{\tilde{q}} + \dot{\tilde{q}}^T K_P \tilde{q} . \end{aligned} \quad (15)$$

Aplicando la *Propiedad de antisimetría*:

$$\dot{V}(q, \dot{q}, \tilde{q}) = \dot{\tilde{q}}^T [-\tau + D(q)\ddot{q}_d + C(q, \dot{q})\dot{q}_d + G(q) + F(\dot{q}) + K_P \tilde{q}] . \quad (16)$$

A pesar de que (19) no es una función definida negativa, puede ser forzada a serlo utilizando [12]:

$$-\dot{\tilde{q}}^T K_D \tilde{q} = \dot{\tilde{q}}^T [-\tau + D(q)\ddot{q}_d + C(q, \dot{q})\dot{q}_d + G(q) + F(\dot{q}) + K_P \tilde{q}] . \quad (17)$$

Entonces, la ley de control utilizada para satisfacer la segunda y tercera condición del Teorema 4.1 es:

$$\tau = K_P \tilde{q} + K_D \dot{\tilde{q}} + K_F Y(q, \dot{q}_d, \ddot{q}_d)^T \Phi \quad (18)$$

Nótese que cuando $K_F = 1$ (ausencia de compensación), el controlador (18) es el Control Par Calculado clásico (9).

5. Resultados

Los experimentos a nivel simulación se llevaron a cabo utilizando el modelo dinámico de un robot de 2 GDL afectado por pares gravitacionales con los siguientes parámetros:

Tabla 2. Parámetros del robot manipulador.

Parámetro	Símbolo	Valor
Masa del eslabón 1	m_1	2.8
Masa del eslabón 2	m_2	1.75
Longitud del eslabón 1	l_1	0.35
Longitud del eslabón 2	l_2	0.2
Longitud de la articulación 1 al centro de gravedad del eslabón 1	l_{c1}	0.21
Longitud de la articulación 2 al centro de gravedad del eslabón 2	l_{c2}	0.14
Momento de inercia del eslabón 1	I_1	0.45
Momento de inercia del eslabón 2	I_2	0.33
Aceleración de la gravedad	g	9.81
Coefficiente de fricción de la articulación 1	f_1	0.9
Coefficiente de fricción de la articulación 2	f_2	0.9

El vector Φ contiene todos los parámetros de la Tabla 2. La trayectoria que debe seguir el efector final del robot es un círculo con las siguientes ecuaciones paramétricas:

$$\begin{aligned} x_d &= 0.3 + 0.15 \cos(t) \\ y_d &= 0.2 + 0.15 \cos(t) \end{aligned} \quad (19)$$

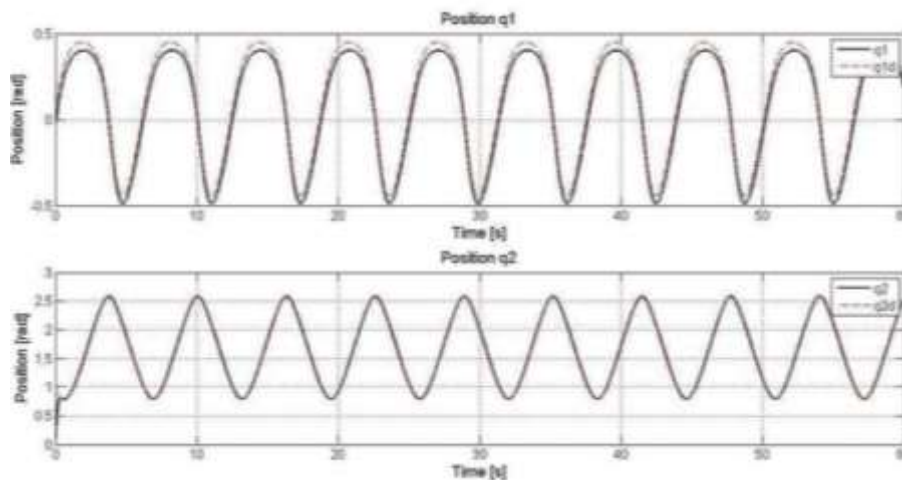


Fig. 3. Trayectorias articulares con Control Par Calculado clásico.

Lo cual sugiere que el robot debe realizar un círculo de 15cm de radio. El tiempo de simulación es de 60 segundos. El primer experimento fue realizado variando la fricción de las articulaciones de 0.9 a 1.5, utilizando el Control Par Calculado clásico; los resultados de este experimento se ilustran en la Figura 3 y en la Figura 4.

En la Figura 3 se aprecia que las trayectorias articulares no corresponden a las referencias; asimismo, en la Figura 4 se observa que el robot no realiza el círculo debido a la falta de energía por parte de los actuadores.

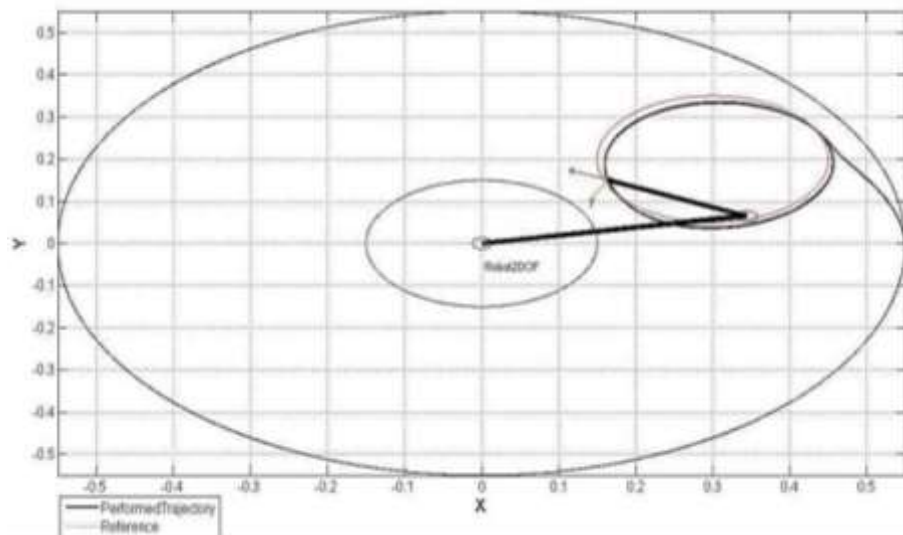


Fig. 4. Trayectoria del efector final con Control Par Calculado clásico.

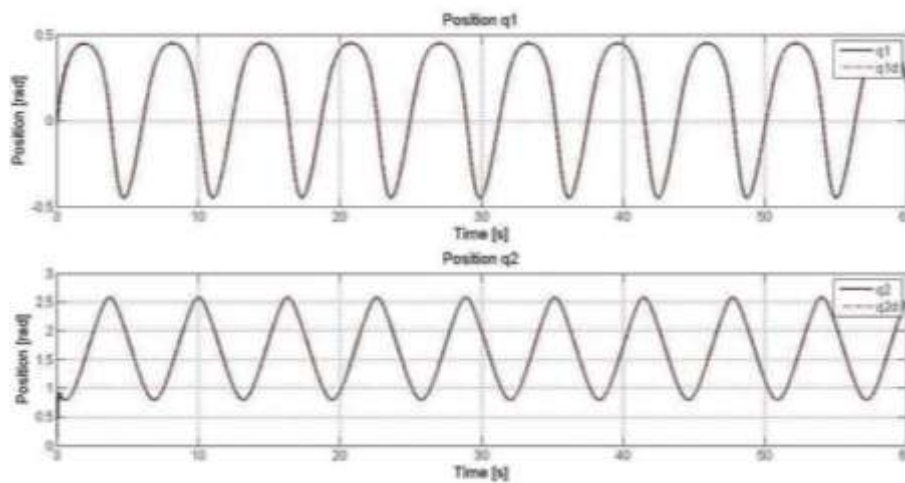


Fig. 5. Trayectorias articulares con Control Par Calculado difuso.

El segundo experimento se llevó a cabo utilizando el controlador propuesto en este trabajo, cuyos resultados se ilustran en la Figura 5 y en la Figura 6. En la Figura 5 se puede apreciar que las trayectorias articulares siguen correctamente las trayectorias

deseadas. En la Figura 6 se observa como la trayectoria del efector final del robot corresponde a la trayectoria deseada.

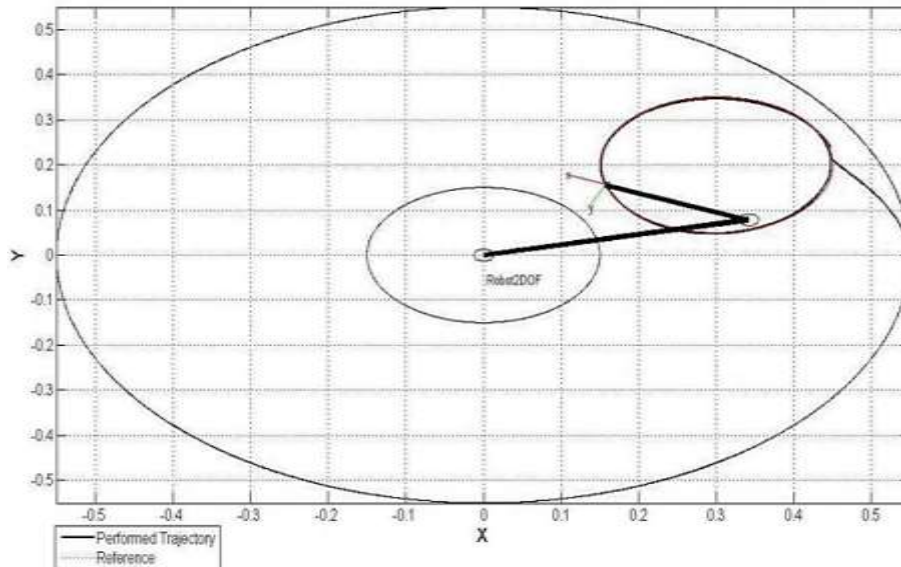


Fig. 6. Trayectoria del efector final con Control Par Calculado difuso.

6. Conclusiones

En este trabajo se presenta el diseño de una alternativa para mejorar el Control Par Calculado clásico utilizando lógica difusa. El controlador propuesto es capaz de compensar variaciones en los parámetros del robot tales como las masas de los eslabones o fricción en las articulaciones, utilizando la fusificación de un índice de variación, el cual se mapea hacia un dato duro que se utiliza como ganancia para la mejora del Control Par Calculado clásico. La estabilidad asintótica de este controlador se garantiza utilizando la Teoría de Pasividad y el Método Directo de Lyapunov para el diseño del mismo. La innovación que presenta esta estrategia de control es la mejora de la técnica clásica en la parte de la compensación de las dinámicas del robot en lugar realizarse en la parte tipo PD de la estructura.

Los resultados satisfactorios muestran que la alternativa propuesta es útil para compensar variaciones del robot cuando ya ha sido previamente parametrizado y puesto en marcha. En la práctica, el índice de variación puede ser calculado utilizando mediciones directas de par o de corriente en los actuadores. La sencillez de la estructura de este controlador y los resultados satisfactorios a nivel simulación motivan a su implementación utilizando microcontroladores comunes o hardware embebido; además, es recomendable asegurarse de que los actuadores y las interfaces de potencia utilizadas sean capaces de suministrar la energía calculada por el controlador propuesto para garantizar su correcto funcionamiento.

Referencias

1. M.A. Llama, R. Kelly and V. Santibañez: Stable Computed-Torque Control of Robot Manipulators via Fuzzy Self-Tuning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 30, no. 1, pp. 143–150 (2000)
2. B. R. Markiewicz: Analysis of the Computed Torque Drive Method and Comparison with Conventional Position Servo for a Computer-Controlled Manipulator. California Institute of Technology, Jet Propulsion Laboratory, Tech. Rep. 33-601 (1973)
3. P.K. Khosla and T. Kanade: Real-Time Implementation and Evaluation of the Computed Torque Scheme. *IEEE Robotics and Automation*, vol. 5, no. 2, pp. 254–253 (1989)
4. H. Berghuis, H. Nijmeijer and P. Löhner: An Addendum on Robust Control of Robots by the Computed Torque Methods. *Systems and Control Letters*, vol. 18, pp. 403–407 (1992)
5. B. Porter and N. N. Zadeh: Genetic Design of Computed-Torque/Fuzzy Logic Controllers for Robotic Manipulators. In: *Proc. IEEE International Symposium on Intelligent Control* (1995)
6. V. Santibañez, R. Kelly and M. Llama: Asymptotic Stable Tracking for Robot Manipulators via Sectorial Fuzzy Control. In: *Proc. 15th Triennial World Congress* (2002)
7. S. Soltani and F. Piltan: Design Artificial Nonlinear Controller Based on Computed Torque Like Controller with Tunable Gain. *World Applied Sciences Journal*, vol. 14, no. 9, pp. 1306–1312 (2011)
8. F. Piltan, N. Sulaiman, A. Jalali and F. D. Narouei: Design of Model Free Adaptive Fuzzy Computed Torque Controller: Applied to Nonlinear Second Order System. *International Journal of Robotics and Automation*, vol. 2, no. 4, pp. 232–244 (2011)
9. Y. Chen, G. Ma, S. Lin and J. Gao: Adaptive Fuzzy Computed-Torque Control for Robot Manipulator with Uncertain Dynamics. *International Journal of Advanced Robotic Systems*, vol. 9, no. 237, pp. 1–9 (2012)
10. M. W. Spong, S. Hutchinson and M. Vidyasagar: *Robot Modeling and Control*. John Wiley and Sons (2006)
11. R. Kelly y V. Santibañez: *Control de Movimiento de Robots Manipuladores*. Pearson Prentice Hall (2003)
12. J.P. Ordaz-Oliver, O.A. Domínguez-Ramírez, V. Parra-Vega y A. Jarillo-Silva: *Diseño de Estrategias de Control Basadas en la Pasividad para Sistemas Euler-Lagrange Aplicado en Robots Manipuladores*, Editorial Instituto Politécnico Nacional (2009)
13. S.M. Orozco-Soto: Intelligent Nonlinear PI Controller for a 3DOF Robot Manipulator Implemented on a DSP. In: *Proc. of 2nd International Congress on Instrumentation and Applied Sciences (ICIAS)* (2012)
14. L.X. Wang: *A Course on Fuzzy Systems and Control*, Prentice Hall (1997)
15. J.E. Slotine and W. Li: *Applied Nonlinear Control*, Prentice Hall (1991)

Control adaptativo para navegación de una silla de ruedas inteligente empleando lógica difusa tipo 1 y tipo 2

Mario Rojas, Pedro Ponce, Arturo Molina

Escuela de Ingeniería y Ciencias,
Tecnológico de Monterrey, Campus Ciudad de México,
México D.F., México

{mario.rojas,pedro.ponce,armolinagtz}@itesm.mx
<http://www.itesm.mx/>

Resumen. En este artículo se presentan dos alternativas para el control de navegación de una silla de ruedas eléctrica en condiciones de espacio reducido. Como primera opción, se propone una estructura empleando lógica difusa con universo de discurso ajustable. Dependiendo de la distancia promedio registrada por tres sensores ultrasónicos, se modifican los parámetros de los conjuntos difusos para expandirlos o contraerlos. Posteriormente, con los datos del ajuste máximo y mínimo de los conjuntos obtenidos, se definieron funciones difusas tipo 2 y se implementó un controlador de lógica difusa tipo 2. Las trayectorias obtenidas en las pruebas realizadas para ambas versiones muestran ventajas en la adaptabilidad al espacio disponible en el escenario de navegación.

Palabras clave: lógica difusa, tipo 2, sensores ultrasónicos, adaptativo, silla de ruedas.

1. Introducción

Los modelos comerciales de sillas de ruedas incluyen un joystick para manipularlas, pero existen pacientes que encuentran difícil utilizar esta interfaz debido a impedimentos físicos o cognitivos [1]. En muchas instituciones se han desarrollado prototipos de “sillas de ruedas inteligentes” que utilizan tecnología para auxiliar a las personas con discapacidad en tareas como evasión automática de obstáculos, navegación autónoma e interacción inteligente con el usuario [2,3]. En [4] se menciona que una “silla de ruedas inteligente” incorpora distintos sensores, software de control, interfaces de usuario, algoritmos de inteligencia artificial, computadoras o procesadores, así como algoritmos de navegación y planeación de trayectorias.

Dentro de los prototipos existen “sistemas semi-autónomos” que trabajan en colaboración con el usuario, el cual planea la trayectoria al destino y va guiando a la silla de ruedas con instrucciones sencillas por medio de alguna interfaz. El controlador se encarga de garantizar la seguridad del paciente durante el trayecto

a su objetivo. Se ha demostrado en distintos trabajos [5,6] que una técnica de control adecuada para los prototipos semi-autónomos es la lógica difusa o “borrosa”. Los conjuntos difusos fueron presentados por Zadeh en [7], dando pie al desarrollo de los sistemas difusos que son ampliamente usados como una técnica para tratar problemas de control donde no se tiene un modelo matemático y la toma de decisiones con información imprecisa.

Por otro lado, en muchos de los prototipos recientes de sillas de ruedas inteligentes es común encontrar la combinación de controladores difusos y sensores ultrasónicos [8,9,10,11,12,13,14,15]. El sensor ultrasónico detecta objetos en el camino y determina a que distancia se encuentran de la silla. Su principio de operación esta basado en el tiempo que tarda en regresar un eco emitido que fue reflejado por alguna superficie. No es raro que sean una alternativa ampliamente usada, ya que su costo es bajo y ofrecen una buena resolución.

Los usuarios de una sillas de ruedas (inteligente o no) desean poder desplazarse con toda confianza en cualquier lugar y que la silla se adapte a las condiciones de espacio. Este problema se puede atacar si se definen correctamente los conjuntos difusos del controlador, pero no es muy práctico estar cambiando los parámetros constantemente y en cada cambio de condiciones. Para resolver esta situación, en [16,17] se describe el principio de universo de discurso ajustable, el cual usa factores de escalamiento para ajustar las funciones de pertenencia y demuestra que este procedimiento tuvo una gran influencia en el desempeño del controlador difuso.

Utilizando la técnica del universo de discurso variable se implementó un controlador difuso adaptativo para la navegación de una silla de ruedas. Se instalaron tres sensores ultrasónicos en una silla de ruedas eléctrica de uso comercial, cuyos datos se usaron para programar reglas de evasión de obstáculos que modifican la velocidad y dirección del movimiento. Las funciones de pertenencia del controlador difuso cambian dependiendo del espacio disponible en el escenario. Para modificar esas funciones se utiliza otro control difuso de tipo SISO, el cual determina que tanto se tienen que expandir o contraer los conjuntos difusos del control de navegación.

También se exploró la alternativa de conjuntos difusos tipo 2 (T2) para modificar los parámetros de las funciones de pertenencia. Con los valores de los conjuntos ajustados por el controlador SISO, se definieron los límites superior e inferior de funciones difusas T2. Los conjuntos difusos T2 fueron presentados por Zadeh en [18], pero Karnik extendió la idea a sistemas difusos [19]. Usando estos sistemas, se busca definir la incertidumbre del espacio de navegación en las funciones T2.

Se hicieron pruebas de navegación con las implementaciones del controlador tipo 1 (T1) y T2 y se tomaron datos de las trayectorias descritas para establecer las características de cada versión. Los resultados muestran ventajas en la navegación debido a las metodologías de ajuste que se implementaron.

2. Metodología

2.1. Lógica difusa tipo 1 y tipo 2

La lógica difusa tipo 1 (T1) es una técnica desarrollada por el profesor Zadeh en la Universidad de Berkley, la cual amplía la teoría tradicional de inclusión o exclusión absoluta de un elemento en un conjunto. En la lógica difusa se considera que un elemento x pertenece a un conjunto A en cierto grado de inclusión, y este grado es descrito por una función de pertenencia designada como $\mu_A(x)$.

La estructura básica de un controlador difuso T1 se presenta en la figura 1a. Se puede ver en el diagrama de bloques que la información “nítida” en las entradas del sistema se “fuzzifica” en grados de pertenencia, usando variables lingüísticas y funciones de pertenencia que toman valores entre 0 y 1. Estas variables lingüísticas serán útiles para construir la base de reglas obtenidas de la experiencia de un ser humano. Las variables difusas de entrada serán los antecedentes y las de salida las consecuencias de las reglas. Los valores de las variables se relacionan con un método de inferencia que determina el valor de peso de la regla (*fire strenght*). Finalmente, este valor de peso se “defuzzifica” para obtener el valor “nítido” de la salida o salidas.

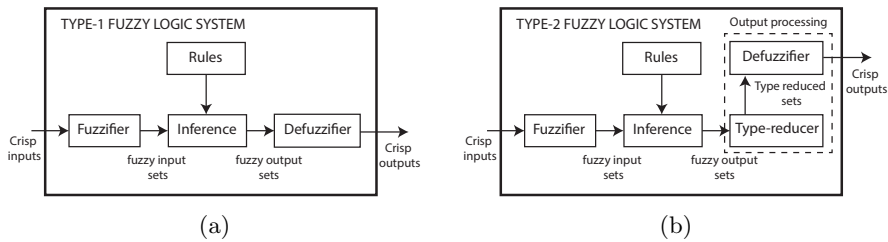


Fig. 1: Estructura de controladores difusos (a) tipo 1 (b) tipo 2.

Por otro lado, la lógica difusa T2 es una forma de lidiar con las incertidumbres que existen en el sistema [20]. Las incertidumbres se pueden expresar como “las palabras que tiene diferentes significados para cada persona”. De este modo, los conjuntos difusos T2 se pueden ver como una colección de distintas opiniones sobre una variable lingüística. Un conjunto difuso T2 se designa como \tilde{A} y su función de pertenencia es $\mu_{\tilde{A}}(x, u)$, donde $x \in X$ y $u \in J_x \subseteq [0, 1]$. J_x se conoce como la función de pertenencia primaria de x , y para representar la incertidumbre de J_x , se define una función de pertenencia secundaria $\mu_{\tilde{A}}(x, u) \in [0, 1]$. Se conoce como un conjunto Intervalo cuando la función de pertenencia secundaria $f_x(u) = 1, \forall u \in J_x \subseteq [0, 1]$. La función Intervalo T2 (I-T2) es una simplificación de la forma general que permite hacer una implementación computacional de sistemas difusos T2, ya que su costo es mucho menor que el de la forma general. Aún así, los sistemas basados en conjuntos I-T2 son más complejos que aquellos basados T1.

En [21] se mencionan algunas aplicaciones industriales de la lógica difusa T2 y en [22] se habla de una creciente tendencia de publicaciones usando esta técnica. En la figura 1b se presenta la estructura de un controlador difuso T2. Observe que es muy parecida a la estructura del T1 presentado en la figura 1a, pero las reglas están basadas en funciones de I-T2 y se incluye otro bloque llamado “reducción de tipo”, que reduce la salida a una función T1 para su defuzzificación.

La incertidumbre de una función de pertenencia \tilde{A} es una región llamada “huella de incertidumbre” (*footprint of uncertainty* o FOU). La FOU es la unión de todas las funciones de pertenencia primarias, y está delimitada por las funciones de pertenencia superior (*upper membership function* o UMF) e inferior (*lower membership function* o LMF), designadas como $\bar{\mu}_{\tilde{A}}(x)$ y $\underline{\mu}_{\tilde{A}}(x)$, respectivamente. Un esquema general de la FOU se muestra en la figura 2.

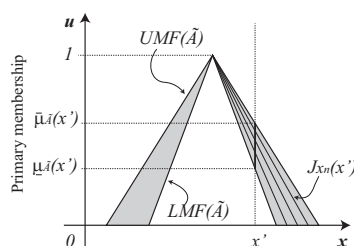


Fig. 2: La huella de incertidumbre.

Cambiar la estructura del controlador difuso de T1 a T2 no garantiza mejoras en su desempeño, pues es necesario entender cuáles son los efectos de la huella de incertidumbre. Con ayuda de la FOU se puede modificar algunas características de la respuesta del sistema, sin embargo, es necesario diseñarla cuidadosamente.

2.2. La silla de ruedas

Se adaptó una silla de ruedas eléctrica de la marca Quickie modelo P222-SE con el hardware mostrado en la figura 3. Se instalaron sensores ultrasónicos en los extremos derecho ($S1$) e izquierdo ($S2$), así como al frente en el centro ($S4$). Un microcontrolador reúne la información de distancia registrada por los sensores ultrasónicos (en centímetros) para mandarla a una computadora con LabVIEW, plataforma en la cual se programaron los controladores difusos. Así mismo, se usa un módulo NI-DAQ 6211 para proporcionar las señales de control a los motores de la silla de ruedas.

2.3. El controlador de navegación T1

El objetivo de la silla es avanzar siempre hacia adelante, pero cuando algún sensor detecta un objeto se deben modificar la velocidad y la dirección para eva-



Fig. 3: Los componentes del sistema.

dirlo. Las reglas del controlador se plantearon de acuerdo a los objetos detectados en el camino de la silla de ruedas y la distancia a la que se encuentran.

Para las tres variables de entrada de distancia $S1$, $S2$ y $S4$ se definieron los conjuntos trapezoidales C (“Close”) y F (“Far”), mostrados en las figuras 4a y 4b. Por otro lado, se definieron varios conjuntos para las salidas M (“Movement”) y D (“Direction”), que corresponden a los dos canales de voltaje analógico que permiten que la silla avance o gire. Para M se definieron los tres conjuntos difusos N (“Null”), SF (“Small-Forward”) y F (“Forward”); mientras que para D se usaron L (“Left”), SL (“Small-Left”), N (“Null”), SR (“Small-Right”) y R (“Right”). Las variables de salida se representaron con conjuntos triangulares y trapezoidales y se presentan en las figuras 4c y 4d.

Los conjuntos de salida se encuentran distribuidos dentro del rango de 4 a 8 volts, donde 5.8 representa el estado de reposo. En el caso de M , un aumento de voltaje representa movimiento de la silla hacia adelante, mientras que un descenso provoca movimiento hacia atrás. Similarmente con D , un incremento en el voltaje de 5.8 hace que la silla gire a la derecha, mientras que un decremento provoca que gire a la izquierda. La velocidad de movimiento es proporcional a cambio del voltaje de reposo.

Con las variables lingüísticas y los conjuntos definidos se plantearon las 8 reglas mostradas en la tabla 1 para el control de navegación. Por convención se definió que la silla gire a la derecha en caso de detectar un obstáculo que bloquea el sensor central (regla 7).

2.4. El controlador adaptativo T1

Con el fin de estimar el espacio disponible en el escenario de navegación se calcula el promedio de las distancias registradas por los sensores ultrasónicos. En trayectorias libres de obstáculos los conjuntos se expanden a su rango máximo.

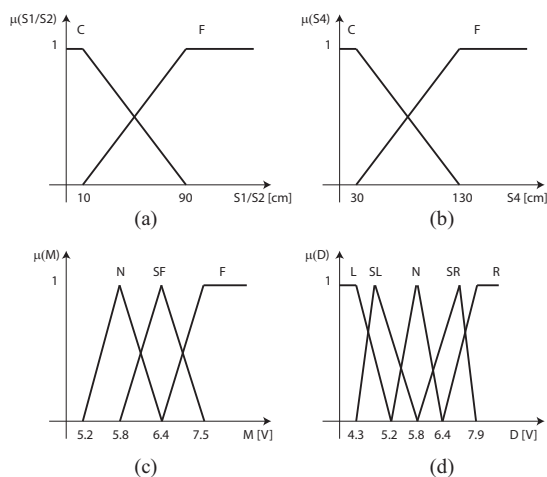


Fig. 4: Definición de los conjuntos difusos. (a) Entradas $S1/S2$ (b) Entrada $S4$ (c) Salida M (d) Salida D .

Tabla 1: Rules for the fuzzy controller with distance inputs

Number	Rules
1	$S1 : C \sqcup S2 : C \sqcup S4 : C \Rightarrow M : N \sqcup D : N$
2	$S1 : C \sqcup S2 : C \sqcup S4 : F \Rightarrow M : MF \sqcup D : N$
3	$S1 : C \sqcup S2 : F \sqcup S4 : C \Rightarrow M : N \sqcup D : L$
4	$S1 : C \sqcup S2 : F \sqcup S4 : F \Rightarrow M : MF \sqcup D : ML$
5	$S1 : F \sqcup S2 : C \sqcup S4 : C \Rightarrow M : N \sqcup D : R$
6	$S1 : F \sqcup S2 : C \sqcup S4 : F \Rightarrow M : MF \sqcup D : MR$
7	$S1 : F \sqcup S2 : F \sqcup S4 : C \Rightarrow M : N \sqcup D : R$
8	$S1 : F \sqcup S2 : F \sqcup S4 : F \Rightarrow M : F \sqcup D : N$

Por el contrario, al aparecer objetos en el camino de la silla el valor promedio disminuye y los conjuntos se estrechan. En la figura 5 se muestra el comportamiento de los conjuntos para las variables $S1/S2$, en sus casos de ajuste mínimo y máximo.

Para calcular el tamaño del universo de discurso se implementó un sistema difuso de tipo SISO (Single-input, Single-Output), que usa la variable de entrada "Spacez la variable de salida "Scale". El factor de escalamiento permite ajustar los parámetros que definen las funciones de pertenencia de acuerdo al espacio disponible. Las reglas de ajuste son las siguientes:

- R1: IF 'space' IS 'small' THEN 'scale' IS 'reduced'
 R2: IF 'space' IS 'large' THEN 'scale' IS 'expanded'

En la figura 6a se muestra el diagrama de bloques para el control de navegación sin el ajuste de parámetros, y en figura 6b se presenta el control de

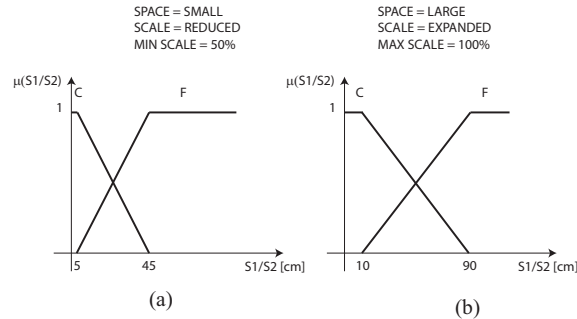


Fig. 5: Límites de escalamiento para los conjuntos de entrada de las variables $S1/S2$. (a) Conjunto reducido (b) Conjunto expandido

navegación junto con el bloque de ajuste SISO y el bloque para calcular el promedio de las distancias de entrada.

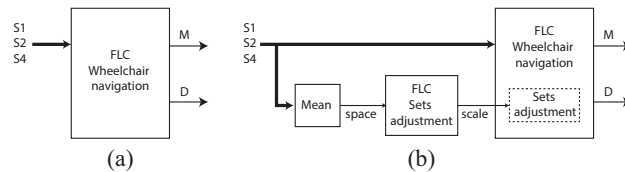


Fig. 6: Diagrama de bloques de controladores. a) Controlador difuso de navegación b) Controlador difuso con universo del discurso ajustable.

2.5. El controlador T2

Con los conjuntos ajustados al universo de discurso mínimo y máximo de la figura 5, se definieron los conjuntos difusos T2. Con ello, los límites para navegar en espacios estrechos y en espacio abiertos se incluyen dentro de la misma función; como consecuencia, la versión con lógica difusa T2 para el controlador ajustable no necesita el bloque SISO descrito para el sistema de T1. Los conjuntos del control T2 se muestran en la figura 7. Solamente se definieron las entradas difusas T2, ya que el tamaño de la huella de incertidumbre para las salidas es muy pequeño y no causa efecto debido a la mecánica de la silla de ruedas. Observe que las huellas de incertidumbre no son simétricas.

3. Pruebas y resultados

Las pruebas experimentales del prototipo se realizaron en los escenarios mostrados en la figura 8. El primero es un pasillo de espacio regular, mientras que

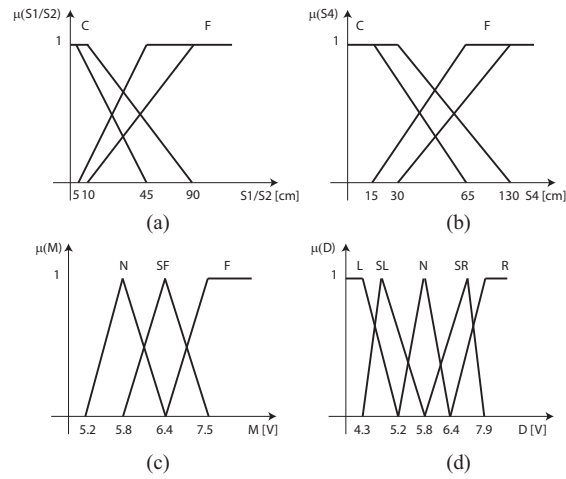


Fig. 7: Definición de conjuntos difusos tipo 2 para el controlador ajustable.

en el segundo las medidas se van estrechando gradualmente. Estos escenarios se crearon usando bloques de un material suave para evitar daños en caso de existir algún choque con los muros. La silla de ruedas debe navegar a través de los escenarios operada primero por el control de navegación T1, luego por el control adaptativo T1 y al final por el control T2. Su velocidad se fijó en 4 m/seg, aproximadamente. Las trayectorias descritas fueron registradas en video y marcadas con puntos mediante el software *VideoPoint Physics Fundamental*.

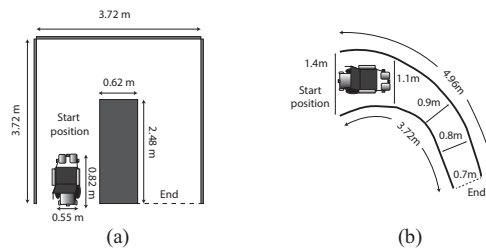


Fig. 8: Escenarios experimentales. (a) Trayectoria con espacio regular (b) Trayectoria con espacio variable

3.1. Trayectoria de espacio regular

Para el caso del escenario 8a se realizaron dos conjuntos de pruebas: primero libre y después con pequeños obstáculos que modifican las condiciones del trayecto. En los resultados mostrados en la figura 9 se observa que las trayectorias

de los casos (a) y (b) son muy parecidas; no así el resultado para el T2 donde la silla se mantuvo más cerca de los muros. Similarmente los casos (d) y (e) son parecidos, pero (f) es distinto.

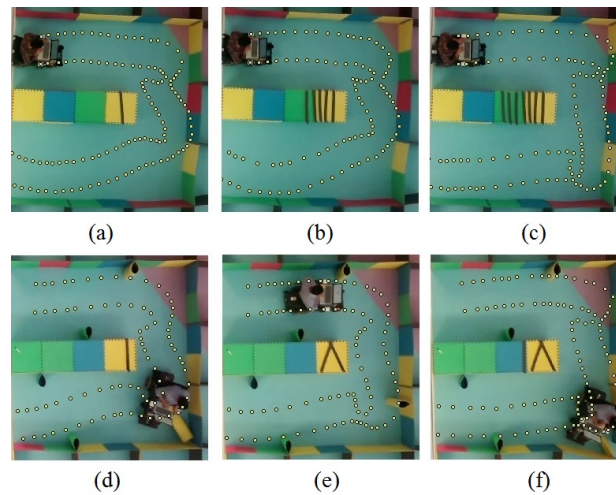


Fig. 9: Trayectorias obtenidas en la prueba con obstáculos y sin obstáculos. a) Controlador de navegación T1. b) Controlador ajustable T1. c) Controlador T2. d) Controlador de navegación T1. e) Controlador ajustable T1. f) Controlador T2.

Usando los datos de $S2$ se calculó la distancia promedio de la silla al muro por el lado izquierdo. También se midieron los tiempos de recorrido en segundos y se registraron los choques que tuvo el prototipo. Los datos encontrados para las pruebas con obstáculos y sin obstáculos se resumen en la tabla 2.

Para la prueba sin obstáculos se puede ver que el controlador T2 mantuvo una distancia promedio de 38.94 cm, la cual es menor que la de los otros controladores y coincide con lo observado en la figura 9. El tiempo del trayecto es mayor para el control T2 y no se registraron choques para ninguna prueba de los controles T1, pero sí del T2.

Por otro lado, en la prueba con obstáculos el control T2 mantuvo una distancia promedio de 39.54 cm, menor que las de los controladores T1. También el tiempo de recorrido fue mayor para el control T2. Finalmente, el control de navegación T1 y el control T2 registraron un choque con el tercer objeto que apareció en el camino, pero el control adaptativo T1 lo libra bien.

3.2. Trayectoria de espacio variable

Observe que el escenario mostrado en la figura 8b es una curva donde el tamaño del pasillo que se forma se va estrechando. Bajo estas condiciones se probó

Tabla 2: Datos registrados para las pruebas realizadas.

Control	dist.(cm)	tiempo(seg)	choques	*dist.(cm)	*tiempo(seg)	*choques
Navegación	47.83	11.78	no	45.68	11.72	si
Adaptativo	48.58	11.5	no	47.26	12.74	no
Tipo 2	38.94	13.3	si	39.54	13.02	si

*Prueba con obstáculos

la adaptabilidad del sistema; primero con el control de navegación T1, luego el control adaptativo T2 y finalmente el control T2. Las trayectorias obtenidas se presentan en la figura 10, donde se puede ver que el controlador adaptativo T1 (figura 10b) describió una trayectoria sin oscilaciones y sin choques, a diferencia de los otros dos casos. El control de navegación T1 ocupó 13.56 segundos para terminar el recorrido, el control ajustable T1 ocupó 15.01 segundos y el T2 en 12.76 segundos.

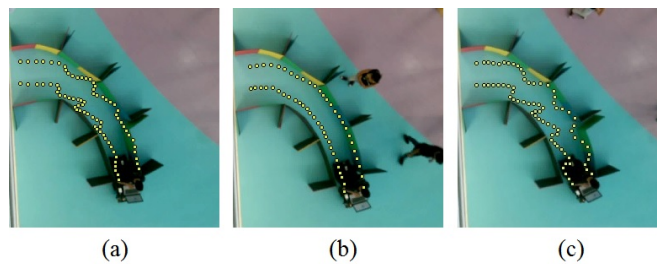


Fig.10: Trayectorias en pasillo estrecho. a) Controlador de navegación T1. b) Controlador ajustable T1. c) Controlador T2.

4. Discusión

4.1. Desempeño

De acuerdo a las pruebas realizadas y con datos obtenidos, se resumen las características de los controladores adaptativos T1 y T2 en la tabla 3. En primer lugar se destaca la estructura usada, ya que a consecuencia de ello, el controlador T1 tiene un tiempo de respuesta menor que el T2. Los bloques de la implementación del sistema de control difuso T2 (ver fig. 1b) hacen que sea mas lenta su ejecución. El tiempo de ejecución del controlador adaptativo T1 consume 1.02 ms y el de T2 consume 6.85 ms.

Para las prueba del pasillo con espacio regular, el controlador T2 se mantuvo siempre más cerca del muro que cualquiera de las versiones T1. Esta característica puede ser útil si se desea que la silla de ruedas siga el contorno de paredes.

El tiempo de recorrido del control T2 siempre fue más grande, pero es razonable si se considera que una trayectoria más cercana al muro exterior representa una distancia ligeramente mayor. Por otro lado, el controlador ajustable T1 no describió una trayectoria muy diferente a la del control de navegación T1 por lo cual no se ven resultados del ajuste de funciones. Sin embargo, en las pruebas donde el espacio del pasillo se reduce gradualmente el control ajustable T1 se adapta muy bien, a diferencia del control T2 que presenta muchas oscilaciones y choques con los muros. En este caso el controlador ajustable es muy superior.

En la práctica, el proceso de ajuste es más simple en el control adaptativo, ya que solo se tienen que cambiar los parámetros de las variables “spacez “scale”. Por otro lado, para ajustar el control tipo 2 se deben modificar las definiciones de las huellas de incertidumbre cambiando los parámetros de las UMF y LMF para las tres variables de distancia.

4.2. Limitaciones

Las condiciones de las pruebas con la silla están sujetas al ambiente controlado del laboratorio. No se consideran otras variables que influyen en la trayectoria descrita por el sistema tales como el tipo de superficie donde se desplaza la silla o su velocidad, tampoco las variaciones en los sensores que pueden ser ocasionadas por los cambios en los ángulos de incidencia de las señales o las propiedades de las superficies de reflexión, ni los cambios de peso en el usuario de la silla.

Por otra parte, la prueba con obstáculos permitió observar colisiones con el tercer objeto de la trayectoria. Este inconveniente se observó con el controlador de navegación sin ajuste y el de T2. Sin embargo, el T1 adaptativo logró evadir bien ese obstáculo. Es importante señalar esta diferencia, ya que se ha observado que el controlador de navegación y el de conjuntos ajustables describen una trayectoria similar, pero éste último demostró ser superior en esta prueba. Las colisiones que se tuvieron en las pruebas indican una debilidad del método de ajuste, sin embargo, muchas serán inevitables en aquellos espacios muy reducidos. Para este caso, se debe asegurar que sean de baja intensidad con el fin de que no puedan dañar al paciente o a la silla de ruedas. Así mismo, se debe contar con otros sensores o sistemas que garanticen mayor seguridad para el sistema.

Tabla 3: Comparativa de algunos aspectos de los controladores adaptativos

Características	Control T1	Control T2
Diseño e implementación	2 controladores difusos	Sistema tipo 2
Tiempo de respuesta	1.02 ms	6.85 ms
Respuesta en espacios reducidos	Se adaptó a las condiciones	Navegó más cerca de los muros
Colisiones con muros	No	Si
Colisiones con obstáculos	No	Si
Ajuste de parámetros	Cálculo automático	Con LMF y UMF

4.3. Trabajo futuro

Se puede aplicar la metodología de los sistemas de control con algún dispositivo de alta velocidad. Sin embargo, el funcionamiento dependerá en gran medida de la velocidad de respuesta de los sensores. No se descarta que este sistema pueda funcionar de mejor manera en un FPGA o un procesador dedicado.

Así mismo, es necesario realizar análisis estadísticos que arrojen más información para comparar el rendimiento de las diferentes implementaciones. Los datos estadísticos pueden ayudar a encontrar características que no se aprecian simplemente con las trayectorias.

Finalmente, sería importante realizar pruebas con pacientes que padezcan alguna discapacidad, ya que hasta el momento se han hecho pruebas con personas que no son usuarios de silla de ruedas.

5. Conclusiones

Se presentaron dos alternativas de controladores difusos que permiten a una silla de ruedas eléctrica navegar y adaptarse a las condiciones de espacio reducido en el escenario. La primera versión fue implementada usando una estructura de universo de discurso variable, cuyo ajuste se realiza por medio de reglas difusas. La segunda alternativa fue el uso de lógica difusa T2, donde la huella de incertidumbre permite incluir los rangos de ajustes mínimo y máximo en la función de pertenencia de manera simultánea. Después de hacer pruebas de validación para las versiones de los controladores implementados, se observó que el control con funciones T2 mantiene a la silla más cerca de los muros del laberinto pero el control ajustable T1 funciona mejor en condiciones donde el espacio se va haciendo estrecho.

Los sistemas T2 se han usado ampliamente y con buenos resultados, pero es muy importante estudiar cuáles son los efectos de la huella de incertidumbre en el sistema para que se comporte adecuadamente. Por otro lado, el controlador de ajuste es muy sencillo y puede ser reemplazable por alguna función lineal de escalamiento, sin embargo, no se descarta que se puedan incluir más reglas para mejorar la función de adaptabilidad.

Agradecimientos Este trabajo fue realizado con el apoyo del Tecnológico de Monterrey Campus Ciudad de México y el CONACYT.

Referencias

1. Simpson, R.C., LoPresti, E.F., Cooper, R.A.: How many people would benefit from a smart wheelchair? *J. Rehabil. Res. Dev.*, 45, 53–71 (2008)
2. Krishnan, R.H., Pugazhenti, S.: Mobility assistive devices and self-transfer robotic systems for elderly, a review. *Intel Serv Robotics*, 7, 37–49, doi:10.1007/s11370-013-0142-6 (2014)

3. Simpson, R.C.: Smart wheelchairs: A literature review. *J Rehabil Res Dev.*, 42, 423–436 (2005)
4. Urdiales, C.: Collaborative Assistive Robot for Mobility Enhancement (CARMEN) - The bare necessities: assisted. Intelligent Systems Reference Library, Springer, Berlin (2012)
5. Wang, Y., Wang, S.: Identification of directional intention for a wheelchair robot using a distance fuzzy reasoning algorithm. In: 2014 International Conference on Advanced Mechatronic Systems (ICAMEchS), pp. 260–263, doi:10.1109/ICAMEchS.2014.6911661 (2014)
6. Wu, B.-F., Jen, C.-L., Tsou, T.-Y., Li, W.-F., Tseng, P.-Y.: Accompanist detection and following for wheelchair robots with fuzzy controller. In: 2012 International Conference on Advanced Mechatronic Systems (ICAMEchS), pp. 638–643 (2012)
7. Zadeh, L.A.: Fuzzy sets. *Information and Control*, 8, 338–353, doi:10.1016/S0019-9958(65)90241-X (1965)
8. Klabi, I., Masmoudi, M.S., Masmoudi, M.: Advanced user interfaces for intelligent wheelchair system. In: 2014 1st International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), pp. 130–136. doi:10.1109/ATSIP.2014.6834592 (2014)
9. Bayar, V., Akar, B., Yayan, U., Yavuz, H.S., Yazici, A.: Fuzzy logic based design of classical behaviors for mobile robots in ROS middleware. In: 2014 IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA) Proceedings, pp. 162–169, doi:10.1109/INISTA.2014.6873613 (2014)
10. Zal, F., Chen, T.-S., Chi, S.-W., Kuo, C.-H.: Fuzzy controller based subsumption behavior architecture for autonomous robotic wheelchair. In: 2013 International Conference on Advanced Robotics and Intelligent Systems (ARIS), pp. 158–163, doi:10.1109/ARIS.2013.6573552 (2013)
11. Tyagi, V., Gupta, N.K., Tyagi, P.K.: Smart wheelchair using fuzzy inference system. In: 2013 IEEE Global Humanitarian Technology Conference: South Asia Satellite (GHTC-SAS), pp. 175–180 doi:10.1109/GHTC-SAS.2013.6629911 (2013)
12. Cardoso Bissoli, A.L., de Lima, L.E.M., Ardizzon Lovatte, M.: Automatic path generation and reactive behavior for a wheel chair. In: 2013 IEEE International Conference on Mechatronics and Automation (ICMA), pp. 525–532, doi:10.1109/ICMA.2013.6617972 (2013)
13. Moslehi, H.R., Gu, J.: Design and development of fuzzy logic operated smart motorized wheelchair. In: 2011 24th Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 869–873, doi:10.1109/CCECE.2011.6030581 (2011)
14. Liu, G., Yao, M., Zhang, L., Zhang, C.: Fuzzy Controller for Obstacle Avoidance in Electric Wheelchair with Ultrasonic Sensors. In: 2011 International Symposium on Computer Science and Society (ISCCS), pp. 71–74, doi:10.1109/ISCCS.2011.27 (2011)
15. Njah, M., Jallouli, M.: Wheelchair obstacle avoidance based on fuzzy controller and ultrasonic sensors. In: 2013 International Conference on Computer Applications Technology (ICCAT), pp. 1–5, doi:10.1109/ICCAT.2013.6522062 (2013)
16. Wang, Y., Deng, H., Chen, Z.: Adaptive fuzzy logic controller with rule-based changeable universe of discourse for a nonlinear MIMO system. In: 5th International Conference on Intelligent Systems Design and Applications, ISDA '05, pp. 8–13, doi:10.1109/ISDA.2005.18 (2005)
17. Wang, Y.: Adaptive fuzzy logic controller with changeable universe of discourse using neural fuzzy networks. In: Proceedings of 2005 International Conference on Machine Learning and Cybernetics, pp. 893–898, Vol. 2, doi:10.1109/ICMLC.2005.1527070 (2005)

18. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning—I. *Information Sciences*, 8, 199–249, doi:10.1016/0020-0255(75)90036-5 (1975)
19. Karnik, N.N., Mendel, J.M., Liang, Q.: Type-2 fuzzy logic systems. *IEEE Transactions on Fuzzy Systems*, 7, 643–658, doi:10.1109/91.811231 (1999)
20. Mendel, J.M., John, R.I.: Type-2 fuzzy sets made simple. *IEEE Transactions on Fuzzy Systems*, 10, 117–127, doi:10.1109/91.995115 (2002)
21. Dereli, T., Baykasoglu, A., Altun, K., Durmusoglu, A., Türksen, I.B.: Industrial applications of type-2 fuzzy sets and systems: A concise review. *Computers in Industry, Fuzziness in Industry and Applications*, 62, 125–137, doi:10.1016/j.compind.2010.10.006 (2011)
22. Wu, D., Mendel, J.M.: Designing practical interval type-2 fuzzy logic systems made simple. In: *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pp. 800–807, doi:10.1109/FUZZ-IEEE.2014.6891534 (2014)

Enriquecimiento de señal para activación de frenos ABS en automóviles

José Vázquez Quezada¹, Juan Villegas-Cortez¹, Iván Vázquez-Álvarez¹, Carlos Avilés-Cruz¹, Ismael Osuna-Galán², Yolanda Pérez-Pimentel²

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

² Universidad Politécnica de Chiapas, Tuxtla Gtz., Chiapas, México

jose.pp.vq@gmail.com, {juanvc, iva, caviles}@azc.uam.mx,
{iosuna, ypimentel}@upchiapas.edu.mx

Resumen. El sistema de frenos ABS es ampliamente usado en los automóviles de uso doméstico por su alta eficiencia y mayor aporte de seguridad, además con la masificación de su uso, el costo de éste se ha reducido. Actualmente la activación de este sistema de frenos depende en gran medida de la aplicación abrupta del pedal por parte del conductor del vehículo, por lo que el funcionamiento de éste depende principalmente de las aptitudes del conductor. En este documento se presenta una propuesta de mejoramiento del sistema de frenos ABS, considerando un enriquecimiento de la señal de activación, usando visión por computadora del espacio de rodamiento (camino transitado), tomando la señal proporcionada por el módulo de visión, a fin de tener activaciones más rápidas e independientes del conductor, y antes de que los sensores detecten que se está produciendo un bloqueo desde los neumáticos. Aquí se muestra el detalle del prototipo de mezclado-unión de las señales (computadora del automóvil + la incorporación de visión), así como la implementación simulada en el laboratorio con resultados prometedores.

Palabras clave: procesamiento digital de señales, FPGA, ABS.

1. Introducción

Actualmente el Sistema de Frenos Antibloqueo (Antilock Brake System, ABS), se utiliza en situaciones donde se debe detener un vehículo que viaja a una velocidad considerable sin perder maniobrabilidad del automóvil durante el proceso de frenado [1], e.g. en una carretera resbaladiza por condiciones climáticas, factores humanos o condiciones de la superficie misma. El sistema de frenos ABS permite que se mantenga el control del vehículo en el frenado sin perder el arrastre o tracción, evitando un derrape o amarre, incluso en pistas cubiertas de hielo.

El funcionamiento es simple, cada rueda tiene un sensor de velocidad angular (generalmente un sensor magnético), cuando la velocidad angular de una de

las ruedas cambia muy rápido con respecto a las demás se detecta una posible condición de bloqueo, entonces las válvulas presurizan, o liberan presión, en dicho neumático según lo necesite para evitar el amarre y poder frenar y a su vez evadir obstáculos de manera controlada si es preciso mientras se detiene. Aunque hay varios modelos de ABS [2], [3], todos constan de cuatro componentes básicos que se mencionan continuación y se muestran en la figura 1.

1. Sensores de velocidad: Cada rueda tiene un sensor que estima la velocidad angular de cada rueda, con lo cual es posible determinar cuando alguna rueda está a punto de bloquearse.
2. Barras de frenado (válvulas): Línea del líquido de frenos controlado por el ABS. Presuriza o libera presión en el neumático que lo requiera.
3. Unidad hidráulica (bomba): Recupera la presión que se liberó en los frenos mediante las válvulas.
4. Computadora ABS: Recibe las señales de los sensores de velocidad de las ruedas, con esta información opera las válvulas [4].

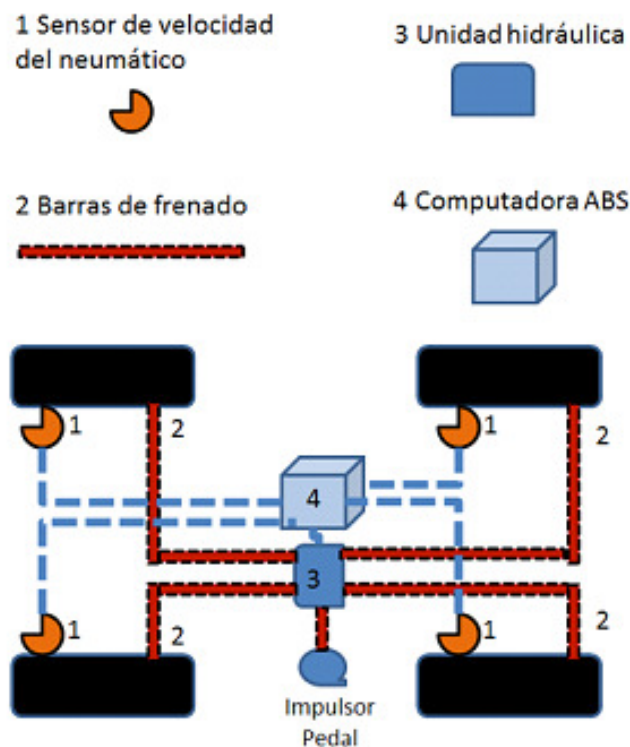


Fig. 1. Diagrama principal de un ABS típico [3].

Una FPGA (Field Programmable Gate Array) es un dispositivo semiconductor que contiene campos de matrices de compuertas programables. Se caracterizan por altas densidades de compuertas, alto rendimiento y un número muy grande de entradas/salidas definidas por el usuario, tienen un esquema de interconexión flexible y su entorno de diseño similar a una matriz de puertas, no se limita a la típica matriz AND-OR [5]. Además de contar con una matriz interna configurable de relojes lógicos. La FPGA tiene la ventaja de ser re-programable (lo que añade una enorme flexibilidad al flujo de diseño), su costo de desarrollo y adquisición son bajos y el tiempo de desarrollo es también menor. El proyecto consiste en tomar la señal de activación del ABS y enriquecerla con la señal de coeficiente de fricción adquirida en otro módulo, usando visión por computadora mismo que aquí no se detalla pero que se simula su señal, para que pueda ser activado el ABS tomando en cuenta el tipo de superficie, de esta manera estamos conjuntando a través de la FPGA el sistema de visión con el sistema ABS.

2. Desarrollo

En este proyecto se busca generar una señal enriquecida que active el sistema ABS de un automóvil incorporándole una señal de coeficiente de fricción, calculado a partir del procesamiento de imágenes usando lógica difusa, lo cual enriquece la señal para la activación del freno.

La figura 2 muestra la metodología del proyecto. En el inciso (A), puede observarse el comportamiento normal que sigue la activación del freno ABS. En el inciso (B), puede observarse cómo se integra la integración de la señal desde módulo de visión proveniente de la cámara, la cuál se ingresa al módulo de lógica difusa, donde se determina el coeficiente de fricción, éste coeficiente de fricción será añadido a la señal proveniente del escáner del sistema de frenado ABS para enriquecerla [6], [7]. El coeficiente de fricción tiene dos posibles valores, 0 y 1. El 0 significa que el valor del coeficiente de fricción es mayor al límite (es decir, no hay deslizamiento) y por lo tanto, no se requiere activar el sistema ABS, en caso contrario su valor será de 1.

Para el proyecto se han considerado tres posibles escenarios, los cuáles van acorde a los tipos de terrenos de desplazamiento del vehículo en la ciudad: (i) suelo seco y luz de día, (ii) suelo húmedo y luz de día, y (iii) suelo seco y luz baja.

Realizada la adquisición y análisis del video, el sistema propone el cálculo de un coeficiente de fricción asociado a la superficie de rodamiento. Este cálculo es estimado a partir del análisis de la propia imagen adquirida del video [8], y de esta forma se puede obtener una señal adicional, que complementa o coadyuva a la señal de la computadora del sistema ABS ante condiciones no supervisadas por el conductor del vehículo.

Las variables de entrada seleccionadas fueron la velocidad (v), la entropía (e) y la luminosidad (l); con éstas variables se configuró un sistema MISO (Multiple Input Single Output), estableciendo los conjuntos difusos y reglas de

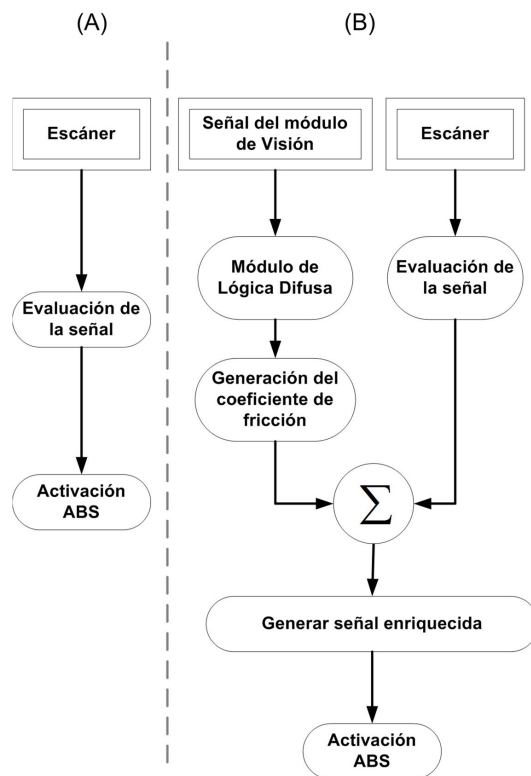


Fig. 2. Metodología del proyecto. (A) Funcionamiento de ABS. (B)Funcionamiento de ABS con señal Enriquecida.

inferencia que permitieran su implementación real. Las características-variables consideradas (v, e, I) fueron extraídas usando imágenes cuadro a cuadro del video obtenido por la cámara, para identificar las tres condiciones de superficie de rodamiento consideradas, (i, ii, iii). Del procesamiento de cada imagen se obtienen los valores de las tres variables (v, e, I), que se ingresan al sistema difuso, y de acuerdo a un conjunto de reglas de inferencia se obtiene el coeficiente de fricción.

Combinando el coeficiente de fricción que da una señal con valor lógico de 0 ó 1 y las señales de activación procedentes del escáner, obtenemos una señal enriquecida. Esto se hará utilizando la siguiente tabla de verdad mostrada en la tabla de verdad mostrada en la figura 3, donde se muestra el flujo de datos como señales que ingresan al dispositivo FTFS, se procesan y entregan como salida la activación o no del freno ABS.

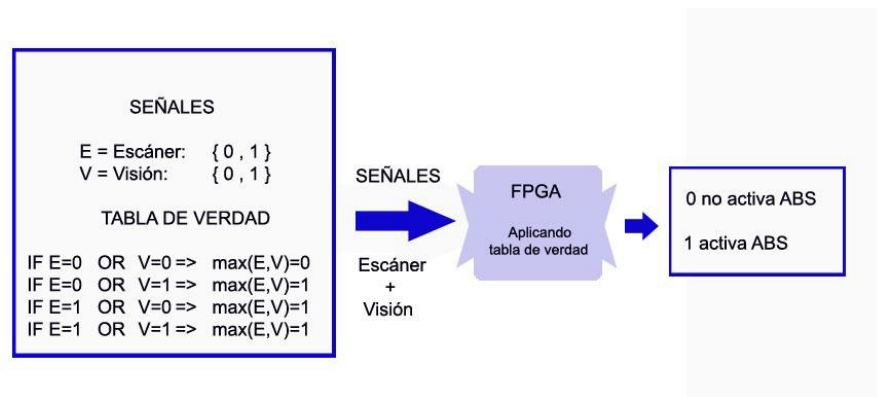


Fig. 3. Tabla de verdad de las señales de activación del ABS.

Se tienen dos variables de entrada, las imágenes de la cámara de video y la señal propia del sistema ABS. El objetivo del procesamiento de la señal de la cámara es ayudar al conductor en el dado caso de que no haya considerado correctamente la superficie sobre la cuál se encuentra y en caso de ser necesaria la activación del freno ABS, activarlo aunque el conductor no haya reaccionado a tiempo, con lo que se pretende evitar accidentes.

Otro aspecto tomado en cuenta es la velocidad de adquisición de señales. La lógica que se siguió es que a menor velocidad del vehículo la ayuda del sistema de activación ABS con señal enriquecida es menor. Por lo tanto, la velocidad de adquisición es directamente proporcional a la velocidad del vehículo. La tabla 1 muestra las velocidades de procesamiento establecidas.

Se considera que el tipo de superficie está asociado a un peso en las reglas de activación. En la Tabla 2, se aprecia el peso de defuzzificación asignado dependiendo el tipo de superficie/camino que se recorre. De igual manera, dentro

Tabla 1. Velocidad de procesamiento de la señal.

Velocidad longitudinal del vehículo (km/h)	Valor del reloj	Frecuencia de procesamiento (Hz)
1 - 19	250	4
20 - 49	125	8
50 - 79	62	16
80 - 119 Km.	31	32
120 o más	16	62

del procesamiento se incluye la eliminación de ruido por parte de la adquisición de imágenes por video, con la finalidad de eliminar los falsos positivos que pueden ser provocados por efectos de luz y sombras, para ello se evalúan 10 cuadros o imágenes consecutivas para evitar que una sombra o un reflejo puede ocasionar un 1. Esto significa que por cada segundo se obtienen tres coeficientes de fricción que son enviados a la señal de activación ABS. El tiempo que se toma para el muestreo de las señales es determinado dependiendo la velocidad del automóvil cada .015 segundos alrededor de 66 tomas por segundo, lo cual es suficiente para tomar las muestras necesarias para el cálculo del coeficiente de fricción.

Tabla 2. Tipos de superficie y pesos asignados.

Tipo de superficie representada	Peso $w(x)$
Superficie resbaladiza (aceite o hielo)	0.8
Superficie mojada	0.5
Superficie seca	0.3

Se realizaron pruebas de funcionamiento del sistema, se modificaron las variables de velocidad del automóvil para verificar la asignación proporcional en la velocidad de procesamiento de la señal y la variable de activación de cámara mediante la captura de imágenes asociadas a diferentes tipos de suelo.

La simulación en el freno ABS fue realizada en laboratorios de Inteco Inc., controlado por Matlab, dicho sistema permite probar el algoritmo en un prototipo de una rueda. Este prototipo base el cual es usado para simular la llegada de las señales de activación del freno ABS. Se realizaron diversas pruebas aleatorias de la señal del pedal y la señal que se recibirá de la cámara para activar el freno, obteniendo resultados satisfactorios ya que la activación manual coincidió con los obtenidos por el sistema, que se discutiran a continuación.

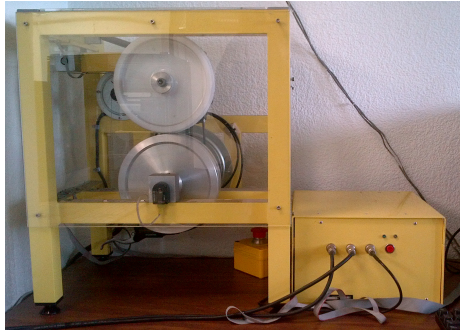


Fig. 4. Laboratorio ABS de Inteco Inc.

3. Análisis de resultados

En la figura 5, se muestra un ejemplo del estado de las señales para saber cuál debe ser el resultado esperado en la señal enriquecida, como se describe a continuación:

- (a) Señal de freno en 0, señal de cámara en 0, señal enriquecida en 0.
- (b) Señal de freno en 0, señal de cámara en 1, señal enriquecida en 1.
- (c) Señal de freno en 1, señal de cámara en 0, señal enriquecida en 1.
- (d) Señal de freno en 1, señal de cámara en 1, señal enriquecida en 1

Con el diagrama que se mostró anteriormente en la figura 2, en el Device-Driver del Abs Main se obtiene la señal enriquecida que se estaba buscando. A continuación se muestran algunas de las gráficas que resumen el estándar de más de 100 corridas en ellas se puede ver una acotación promedio de los resultados obtenidos del procesamiento de estas dos señales. En la imagen se tiene primero la señal del pedal, segundo la señal de la cámara y tercero la señal enriquecida producida por las dos anteriores.

En la gráfica se aprecian las señales que activan el freno ABS cuando está en 1 y lo desactiva si estas en 0, el eje de las X está delimitada la gráfica con el tiempo que transcurre en la simulación que es de (1000 milisegundos) y respecto al eje Y se tiene el valor de la señal enviada al freno desde la cámara, el pedal y la señal enriquecida que es de 0 ó 1 pero se toma en la gráfica de -0.5 a 1.5 para poder apreciar de manera clara la señal.

Con esto se comprobó que la señal enriquecida se obtiene eficientemente. Como se muestra en la figura 5, de arriba hacia abajo se tiene la señal del pedal, señal de cámara y señal enriquecida.

La gráfica muestra la media de activaciones al simular una superficie con charcos o semi-mojada se puede ver como es más activa la señal de la cámara, también se realizó para una superficie casi perfecta, tomando como referencia una carretera sin baches, desniveles, sin basura o charcos en ella y para una superficie normal (tomando como normal una calle sin baches de nuestra ciudad).

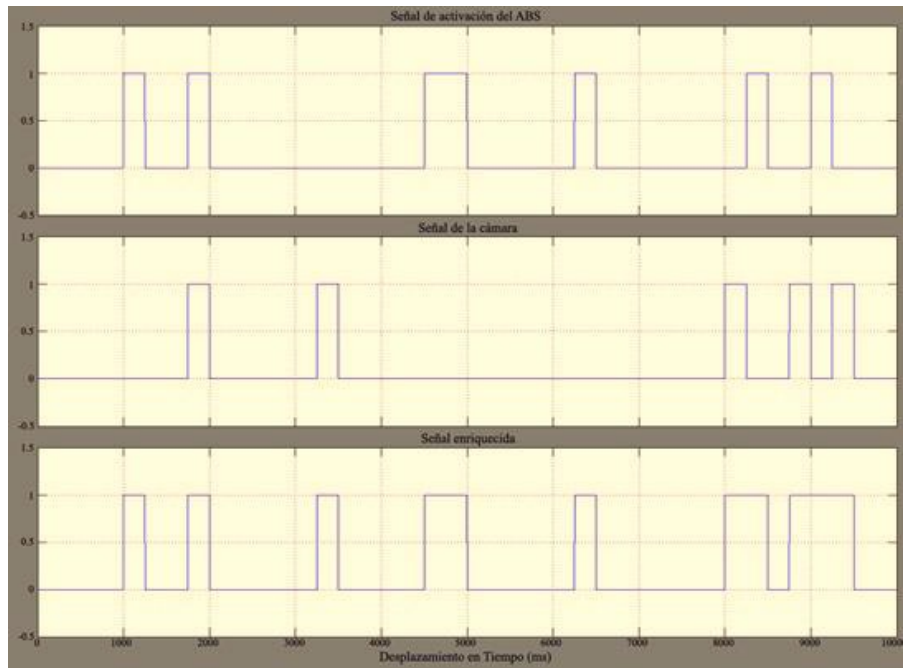


Fig. 5. Señal enriquecida de activación del ABS

Con lo que se puede afirmar que se obtuvo la señal enriquecida y su desempeño es factible y seguro de acuerdo a lo esperado para activar el freno del automóvil. Dejando una gran apertura a la implementación de muchas más señales que ayuden al control de seguridad del automóvil.

4. Conclusiones y trabajo futuro

En este trabajo se presentó una propuesta de mejora del desempeño del sistema ABS, mostrando cómo adicionar una señal generada por un sistema de visión por computadora externo, con la finalidad de ayudar a enriquecer de manera visual y mecánica la activación de éste; lo que se comprueba al integrar dicha información es la mejora del tiempo de respuesta de activación del freno ABS, ya que en lugar de esperar a que se inicie el proceso de frenado por parte del conductor y se presente una situación cercana al bloqueo, en su lugar se detecta una situación de riesgo, dadas las condiciones de la superficie en que se circula. Las condiciones de activación obtenidas por el sistema de visión han sido simuladas efectivamente, y con estas condiciones se ha validado el funcionamiento del sistema propuesto.

Un resultado obtenido es que puede prolongar la vida útil del freno reduciendo sus costos de mantenimiento, siendo un extra al aumento de su nivel de seguridad.

También se puede hacer mención de la posibilidad de anexar más señales al sistema propuesto, y esto da pie a la incorporación de otras opciones diversas de señales para activar el freno, e.g. cámaras de visión traseras y laterales que se trabajan en sistemas para estacionar vehículos, o más sensores colocados en los neumáticos entre muchas otras. Acorde con los resultados mostrados se obtuvo el desempeño esperado en la simulación con la instrumentación de laboratorio.

Referencias

1. Nieto, J.W.: Diseño y adaptación de un sistema de seguridad active para estacionamiento vehicular y monitoreo continuo. Technical report, Escuela Politecnica del Ejercito - Ecuador (2012)
2. Albert, M.P.: Frenos ABS. Alfaomega (1995)
3. BOSCH: Automotive Handbook. Bentley Publishers (2007)
4. Maosheng, Y.: Road surface condition detection and identification and vehicle anti-skid control. Master's thesis, Cleveland State University (2008)
5. López, F.J.P.: Implementación de control por modos deslizantes para sistema abs. Proyecto termina división de ciencias básicas e ingeniería, Universidad Autónoma Metropolitana Unidad Azcapotzalco (2010)
6. Wang, J., Song, C., Jin, L.: Modeling and simulation of automotive four-channel hydraulic abs based on amesim and simulink/stateflow. In: Intelligent Systems and Applications (ISA), 2010 2nd International Workshop on. (2010) 1–4
7. Hernández, J.J.O.: Aplicación de control por modelo inverso a sistema abs. Proyecto terminal, división de ciencias básicas e ingeniería, Universidad Autónoma Metropolitana Unidad Azcapotzalco (2011)
8. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA (1992)

Implementación de un control multiresolución empleando un dispositivo háptico

E. Mendoza-Monjaraz, J. A. Cruz-Tolentino, A. Jarillo-Silva, J.
Pacheco-Mendoza

Universidad de la Sierra Sur,
Miahuatlán de Porfirio Díaz, Oaxaca, México

{eufra_mendoza}@hotmail.com, {jacruz, ajarillo, jpacheco}@unsis.edu.mx
<http://www.unsis.edu.mx>

Resumen. En este trabajo de investigación se realiza la implementación de un control multiresolución sobre un dispositivo háptico Novint Falcon, mismo que se somete a una tarea de regulación basada en seguimiento, empleando polinomios de quinto orden para validar de forma experimental al control multiresolución, con la finalidad de realizar una comparación con otro tipo de control. Los resultados son comparados con un control PID discreto, mediante una medida de tendencia central y una medida de dispersión. Con ellos se busca ratificar y aprovechar las bondades que posee este tipo de control en sistemas robóticos, especialmente en un sistema háptico. Los resultados obtenidos permiten verificar y determinar dichos beneficios que el control multiresolución provee durante el seguimiento de trayectoria planificada, en comparación con controles lineales clásicos.

Palabras clave: control multiresolución, sistema háptico, control wavelet.

1. Introducción

En años recientes S. Parvez [1] desarrolló un nuevo controlador, el cual se basa en el análisis de la señal del error, mismo que es descompuesto en diferentes niveles de resolución, el cual nombró PID wavelet. El PID wavelet, nombrado así por tener ciertas similitudes con el control PID clásico (Proporcional-Integral-Derivativo), considera la señal de error (diferencia entre la señal de referencia y la señal de salida del sistema) para procesarla (obtener la derivada, la integral y una parte proporcional de la señal de error) y entregar una señal de control en función del error y una serie de ganancias previamente calculadas. Por otra parte el control PID wavelet no contempla solamente tres términos, sino que es posible obtener tantos como se requiera dependiendo del nivel de resolución al que se someta la señal de error. Por otra parte, los controles multiresolución (categoría en el que recaen no solo el PID wavelet sino una gran variedad de controles que utilizan técnicas de análisis multiresolución basada en la teoría wavelet, e.g. control wavelet auto-sintonizado con una wavenet [2], control PD-Wavelet+I [3],

control P-wavelet+I [4], control de modos deslizantes+filtro wavelet [5], PID wavenet [6]) permiten un mejor análisis de la señal de error, al descomponer a diversos niveles de escala-frecuencia dicha señal. Es por ello que uno de los beneficios que han brindado estos controles es la inmunidad al ruido, lo que permite entregar una señal de control más suave que aquella entregada en otro tipo de control sin la necesidad de una etapa de filtrado y que a su vez garantiza prolongar la vida útil de los actuadores de la planta debido al libre *chattering*. Después de su aparición y gracias a estos beneficios ha sido implementado para el control de diversos sistemas como en [7], [8], [9], [10] y [11].

El artículo está seccionado de la siguiente manera: en la Sección 1 se tiene una breve introducción sobre los controladores multiresolución; en la Sección 2 se presenta la arquitectura del control multiresolución para un sistema SISO y para un sistema MIMO; mientras que en la Sección 3 se da a conocer al sistema háptico Novint Falcon donde se implementa el control; en la Sección 4 se da a conocer las estrategias de control empleadas, así como también la plataforma empleada y los resultados de las pruebas experimentales; en la Sección 5 se discuten los resultados y en la Sección 6 se dan las conclusiones obtenidas del trabajo de investigación; en la última Sección se dan las referencias empleadas en este trabajo de investigación.

2. Control multiresolución

En esencia, un control multiresolución toma la señal de error, mediante técnicas de análisis multiresolución, descompone esta señal en diversas señales en el espacio de escala-frecuencia, estas señales posteriormente son escaladas y sumadas para obtener la señal de control.

$$u(k) = \sum_{i=1}^{N+1} K_i * e_i(k), \quad (1)$$

donde e_i es el i -ésimo componente de la señal de error a un nivel i de resolución en el espacio escala-frecuencia, K_i es la ganancia correspondiente al i -ésimo componente de la señal de error y N es el nivel de resolución al que se somete a la señal de error $e(k)$. De acuerdo al nivel de resolución N que se tenga siempre se tendrá $N + 1$ componentes de la señal de error $e(k)$.

Actualmente, todos los controladores multiresolución emplean el análisis multiresolución basados en la teoría wavelet para su diseño y únicamente requiere la señal de error. Ésta técnica siempre dará un componente de la señal de error de alta escala - baja frecuencia, un componente de baja escala - alta frecuencia y $N - 1$ componentes de mediana escala - mediana frecuencia.

Los componentes de la señal de error a un nivel N de descomposición son obtenidos mediante la siguiente ecuación:

$$e(k) = \sum_{j=-\infty}^{\infty} c_{N,j} \phi_{N,j}(k) + \sum_{i=1}^{N-1} \sum_{j=-\infty}^{\infty} d_{i,j} \psi_{i,j}(k) + \sum_{j=-\infty}^{\infty} d_{N,j} \psi_{N,j}(k), \quad (2)$$

donde: el primer término del lado derecho de la ecuación representa al componente de alta escala - baja frecuencia ($e_1(k)$), el segundo término a los componentes de mediana escala- mediana frecuencia ($e_2(k), \dots, e_N(k)$) y el último término al componente de baja escala - alta frecuencia ($e_{N+1}(k)$); las funciones ϕ (función de escalamiento) y ψ (función wavelet) son funciones base wavelet que permiten generar subespacios ortonormales V y W en $L_2(\mathbb{R})$; c y d son los coeficientes de aproximación y los coeficientes de detalle en cada i -ésimo nivel de descomposición, respectivamente, y son calculados por [12]:

$$c_{i,j} = \langle e(k), \phi_{i,j}(k) \rangle = \sum_{j=-\infty}^{\infty} e(k) \overline{\phi_{i,j}(k)}, \quad (3)$$

$$d_{i,j} = \langle e(k), \psi_{i,j}(k) \rangle = \sum_{j=-\infty}^{\infty} e(k) \overline{\psi_{i,j}(k)}. \quad (4)$$

El control multiresolución para un sistema de múltiples entradas y múltiples salidas (MIMO) se representa de la siguiente manera. Sea $U(k) \in \mathbb{R}^m$ el vector de control que contiene las señales de control para el sistema MIMO, $Em(k) \in \mathbb{R}^{m \times (N+1)}$ es la matriz que contiene los componentes de la señal del error correspondiente a cada salida, $K(k) \in \mathbb{R}^{(N+1) \times m}$ es la matriz que contiene las ganancias del controlador, N es el nivel de descomposición y m es el número de entradas y salidas del sistema MIMO. El vector de control se calcula como:

$$U_i(k) = \sum_{j=0}^{N+1} Em_{i,j}(k) * K_{j,i}, \quad i = 1, 2, \dots, m. \quad (5)$$

En la Figura 1 se muestra el esquema básico de un sistema control multiresolución para sistemas MIMO, donde Y_{ref} y Y es la referencia y la salida del sistema, respectivamente; e y Em es la señal de error y los componentes de la misma, respectivamente; K son las ganancias del control que posteriormente son multiplicadas Em para obtener la señal de control U ; P representa las perturbaciones externas que afectan al sistema y R representa el ruido introducido a los sensores.

El control multiresolución posee estabilidad en lazo cerrado como se demostró en [13], sin embargo actualmente no existen métodos de sintonización. Una alternativa para sintonizarlo es empleando redes neuronales artificiales como en [4] y en [14].

3. Dispositivo háptico Novint Falcon

El dispositivo Novint Falcon [15] es del tipo paralelogramo, usa una variante de 3 grados de libertad $[x, y, z]$ de la configuración original del robot Delta, la cual es llamada traslacional, muy similar al modelo de Tsai [16]. Tiene incorporados efectores finales que pueden removerse fácilmente cuando el dispositivo está en

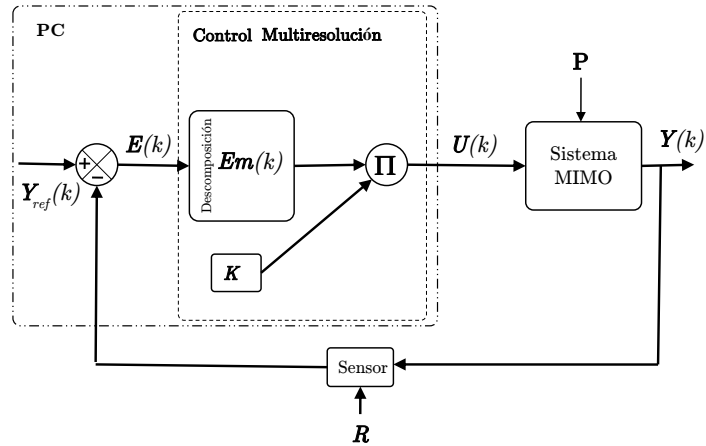


Fig. 1: Esquema básico de un sistema de control multiresolución para un sistema MIMO

uso y puede seguir trabajando de manera normal. Ejemplos de estos efectores son: una pistola y un sujetador esférico (Ver Figura 2). El primero se usa constantemente en videojuegos y el segundo sirve más como apuntador.



Fig. 2: Efectores del dispositivo Novint Falcon.

El dispositivo cuenta con las siguientes características:

- El área de trabajo es aproximadamente de 10.16 cm x 10.16 cm x 10.16 cm (Ver Figura 3).
- Las capacidades de fuerza son mayores a 0.90 kilogramos.
- Suministra 400 dpi (puntos por pulgada) en resolución de posición cartesiana.
- Interfaz de comunicación USB 2.0.
- 1.81 kilogramos de peso.
- Consume 30 watts, 110v-240v, 50Hz-60Hz.
- Usa 3 motores Mabuchi RS-555PH-15 con codificadores ópticos de 320 líneas por revolución.

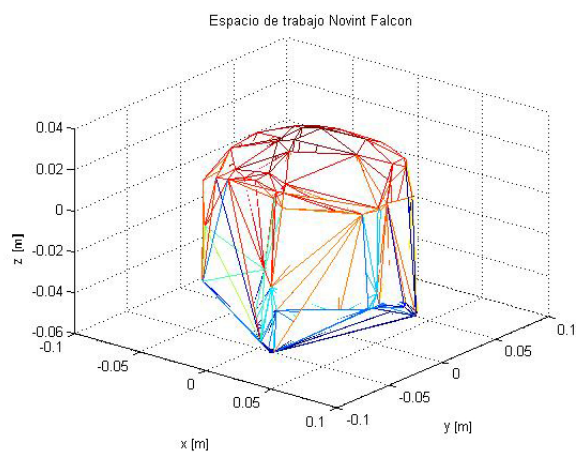


Fig. 3: Área de espacio de trabajo medida tomando puntos aleatorios.

El dispositivo se comunica con la computadora a través de la interfaz USB 2.0, la computadora envía y recibe señales de control las cuales son interpretadas por el sistema de control del dispositivo y enviado a los motores. Estos, a su vez, cuentan con codificadores ópticos para la lectura de la posición angular. El desarrollo de la aplicación se realiza bajo el lenguaje C++ con ayuda del API en su versión 2.1.3 del HDAL, el cual otorga al usuario programador dos visiones, una a bajo nivel y otra a alto nivel para su uso.

La HDAL (Capa de abstracción del dispositivo háptico) es el modelo de arquitectura con el que está implementado el dispositivo Novint Falcon el cual consiste en un sistema de capas que se comunican con el software de aplicación.

La aplicación está conformada de dos principales componentes: la simulación gráfica compuesta de ambientes virtuales y el componente de simulación háptica que contiene la lógica de control del dispositivo. La aplicación se comunica con la HDAL a través de su API, a la cual se le envía como parámetro una función que es llamada a cada lectura de los servos (actuadores), con lo que se proporciona la posición y la fuerza del dispositivo en tiempo real (Figura 4).



Fig. 4: Estructura de la HDAL del dispositivo Novint Falcon.

4. Resultados experimentales

Los resultados experimentales se basan en la aplicación de dos estrategias de control al dispositivo háptico Novint Falcon. Para ello se generan consignas de movimiento basadas en la planificación de una tarea que consta del seguimiento de dos trayectorias.

T1: una trayectoria desde el punto inicial P_0 hasta el punto P_1 , es una trayectoria de referencia y_{ref} generada por el polinomio ξ , como

$$\xi(k) = 10(k/k_b)^3 - 15(k/k_b)^4 + 6(k/k_b)^5, \quad (6)$$

donde k_b es el tiempo de convergencia, que para este caso es de 1 segundo, por lo tanto la trayectoria de referencia es

$$Y_{ref}(k) = P_0 + (P_1 - P_0) * \xi(k), \quad (7)$$

T2: consta de una circunferencia como trayectoria de referencia Y_{ref} exclusivamente sobre el plano X-Y, iniciada en el punto P_1 y terminada sobre el mismo punto, las ecuaciones paramétricas que la definen son:

$$\begin{aligned} Y_{x_{ref}}(k) &= r \cos(2\pi(k - k_b)/T) \\ Y_{y_{ref}}(k) &= r \sin(2\pi(k - k_b)/T) \\ Y_{z_{ref}}(k) &= 0, \end{aligned} \quad (8)$$

donde T es el tiempo de duración de la trayectoria y r es el radio de la circunferencia, que para este caso es $T = 10$ segundos y $r = 2$ cm.

4.1. Estrategias de control

Las estrategias de control aplicadas al dispositivo háptico son: el control multiresolución y el control PID discreto. Para el control multiresolución se utiliza la ecuación (5). De acuerdo a la realización de una serie de pruebas empleando

diferentes niveles de descomposición (N), se determinó utilizar $N = 5$, ya que se obtiene una mejor resolución y una señal de control suave, suprimiendo la señal de ruido. Además se utilizan los coeficientes de filtro Daubechies de soporte compacto de orden 2 dados en [17]:

Tabla 1: Coeficientes del filtro Daubechies de orden 2

n	Coeficientes del filtro pasa bajas
0	0.48296291
1	0.83651630
2	0.22414386
3	-0.12940952

El PID discreto se rige por la siguiente ecuación [18]:

$$U_i^{PID}(k) = U_i^{PID}(k-1) + K_{P,i}^{PID} [e(k) - e(k-1)] + K_{I,i}^{PID} e(k) + K_{D,i}^{PID} [e(k) - 2e(k-1) + e(k-2)], \tag{9}$$

donde $U_i^{PID}(k) \in \mathbb{R}$ es el i -ésima señal de control del PID, $K_{P,i} \in \mathbb{R}$ es la i -ésima ganancia para la parte proporcional, $K_{I,i} \in \mathbb{R}$ es la i -ésima ganancia para la parte integral y $K_{D,i} \in \mathbb{R}$ es la i -ésima ganancia para la parte derivativa.

Las ganancias de ambas leyes de control son obtenidas de manera experimental para el sistema no lineal Novint Falcon y son dadas en el Cuadro 2 y 3.

Tabla 2: Valores de las ganancias del control multiresolución.

Ganancias del control multiresolución	K_1	K_2	K_3	K_4	K_5	K_6
Ganancias para U_1	30	10	0	0	0	0
Ganancias para U_2	30	10	0	0	0	0
Ganancias para U_3	30	10	0	0	0	0

Tabla 3: Valores de las ganancias para el PID discreto.

Ganancias del PID discreto	$K_{P,i}$	$K_{I,i}$	$K_{D,i}$
Ganancias para U_1	12.5	0	4.5
Ganancias para U_2	12.5	0	4.5
Ganancias para U_3	12.5	0	4.5

4.2. Plataforma experimental

La plataforma experimental está constituida por dos elementos importantes: uno corresponde al dispositivo Novint Falcon y el segundo al de un equipo de cómputo que tiene las siguientes características.

- Procesador AMD Turion x2 a 2.00 GHz.
- 2.00 Gb en RAM DDR II 667 mhz.
- Sistema operativo windows 7 service pack 1.
- Tarjeta de vídeo ATI Radeon x1250 con 128 MB de memoria compartida.
- Controlador ATI SB600-OCHI USB 2.0.

4.3. Resultados

En los cuadros 4 y 5, se dan a conocer los resultados obtenidos que corresponden a la media ($\bar{\mu}$) y la desviación estándar (σ) de la señal de error en los tres ejes para cada uno de los controles, respectivamente.

Tabla 4: Media de la señal de error

Eje	Control Multiresolución	PID discreto	unidad
x	0.0027	0.0029	cm.
y	0.0074	0.0121	cm.
z	0.0425	0.0494	cm.

Tabla 5: Desviación estándar de la señal de error

Eje	Control Multiresolución	PID discreto	unidad
x	0.0043	0.0047	cm.
y	0.0030	0.0041	cm.
z	0.0059	0.0073	cm.

En la Figura 5 se presenta el desempeño del efector final del dispositivo háptico Novint Falcon durante el seguimiento de la tarea T1-T2 empleando el control multiresolución. En la Figura 6 es posible observar el comportamiento del error de posición y sus componentes obtenidas mediante la técnica de multiresolución, donde el componente principal de mayor escala pero de baja frecuencia (e_{x1}), es representado por el primer término de la ecuación 2, y los componentes de mediana escala y mediana frecuencia (e_{x2}, \dots, e_{x5}), son representado por el segundo término de la ecuación 2 y al término de menor escala y mayor frecuencia

(e_{x6}), es representado por el tercer término de la ecuación 2. Estos términos son escalados para calcular la señal de control U_1 .

Además se realiza un acercamiento en el intervalo [1.8, 1.9] con la finalidad de percibir con mayor claridad los componentes del error.

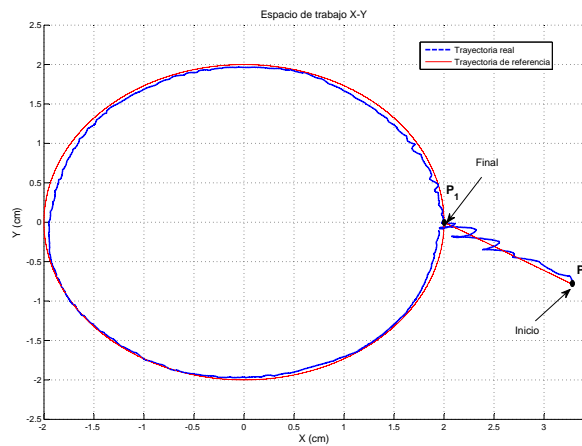


Fig. 5: Trayectoria del sistema háptico en el espacio de trabajo

En la Figura 7 se muestra la señal de control generada por el control multiresolución U_1 y sobre la misma figura se realiza una ampliación en el intervalo de tiempo de [2, 3.2], donde se observa cambios suaves de la señal de control. Finalmente en la Figura 8, se observa la señal de control PID discreto U_1^{PID} y sobre la misma se realiza una ampliación en el intervalo de tiempo de [2, 3.2], en el cual se aprecia la presencia de una señal que oscila a una alta frecuencia en comparación con el control multiresolución.

5. Discusión

A partir de los datos obtenidos de la media de la señal de error y su dispersión en los tres ejes de ambos controles, el sistema presenta robustez y estabilidad en cada instante de tiempo manteniendo en algunas ocasiones un error nulo para el control multiresolución, como se observa de manera gráfica en la Figura 5, donde el efector final del dispositivo Novint Falcon trata de seguir de manera suave, evitando cambios repentinos en la ejecución de la tarea T1-T2. Sin embargo se aprecian pequeñas y constantes oscilaciones durante la mayor parte del seguimiento, esto es provocado por las dinámicas (gravedad, tribología, Coriolis,

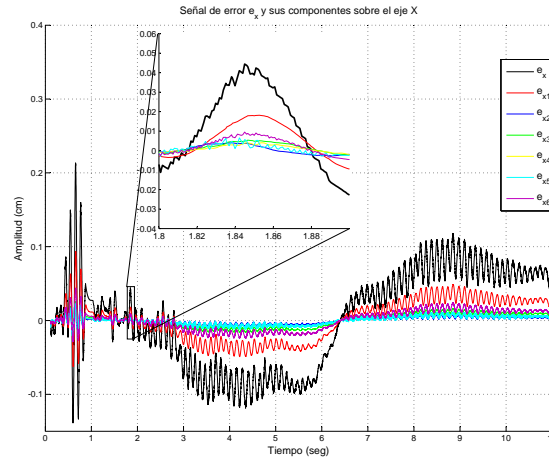


Fig. 6: Señal de error e_x y sus componentes a un nivel de descomposición de $N=5$

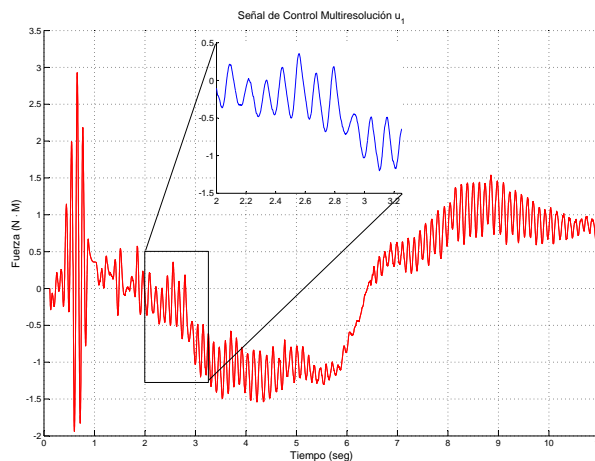


Fig. 7: Señal de control U_1 generada por el control multiresolución

y centrípetas) del dispositivo, mismas que no logran ser compensadas por el control, el cual maneja ganancias constantes durante toda la tarea.

La Figura 6 nos ayuda a seleccionar de alguna manera las ganancias del control multiresolución, ya que los términos que contiene altas frecuencias pueden ser escaladas con ganancias muy pequeñas y en algunos casos nulas, para evitar introducir el ruido al dispositivo háptico por medio de la señal de control, por

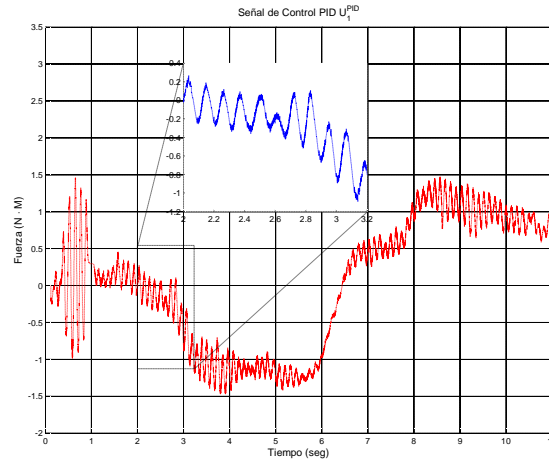


Fig. 8: Señal de control U_1^{PID} generada por el control PID discreto

tal razón las ganancias K_3 , K_4 , K_5 y K_6 en las tres señales de control dadas en el Cuadro 2 son cero. Por lo tanto, se obtiene una señal de control suave debido a proceso de selección de las ganancias de anular los componentes de la señal de error de alta frecuencias, eliminando señales no deseadas (chattering) que pueden causar efectos en los mecanismos del sistema. Mientras que en la Figura 8 se muestra como la señal de control del PID contiene todavía ruido proveniente de la señal de error y amplificado por la parte derivativa del mismo.

6. Conclusiones

El control multiresolución presenta ventajas sobre la gran mayoría de los controles clásicos, ya que su comportamiento tiene la característica de ser suave (sin chattering). Ésto se debe a que muestra inmunidad al ruido, mismo que es introducido por los sensores ópticos del dispositivo, actuadores y la estructura dinámica del dispositivo. También, porque a diferencia de la mayoría de los sistemas de control emplean una etapa de filtrado para cancelar los componentes de alta frecuencia de la señal de error, mientras que en el control multiresolución esta etapa no es requerida. Así, esas señales no son eliminadas, sino estudiadas de mejor manera para elegir cuáles pueden ser incorporadas y cuáles no al sistema, de esta forma no se pierde información. Además, en el control multiresolución no se requiere del conocimiento de las velocidades lineales o articulares para el diseño del mismo, lo que minimiza el cálculo matemático en el dispositivo, mismo que requiere un procesamiento adicional. Sin embargo, se carecen de técnicas o procedimientos para el cálculo ideal de sus ganancias. Una opción sería el uso de redes neuronales, que permitan auto-sintonizar las ganancias en toda la

ejecución de la tarea y sobre todo el espacio de trabajo del dispositivo háptico. En consecuencia un trabajo futuro es la implementación de redes neuronales que permitan sintonizar las ganancias del control multiresolución en tiempo real, realizando además pruebas experimentales ante perturbaciones externas.

Referencias

1. Parvez, S., Gao, Z.: A Wavelet-Based Multiresolution PID Controller. *IEEE Transactions on Industry Applications*. 41(2), 537–543 (2005)
2. Cruz, J. A., Ramos, L. R., Espejel, M. A.: A Self-Tuning of a Wavelet PID Controller. pp. 73–78, IEEE Press, Puebla, México (2010)
3. Parvez, S.: *Advanced Control Techniques for Motion Control Problem*. Cleveland State University [Ph. D. Dissertation], Cleveland, Ohio, USA (2003)
4. Cruz, J. A.: *Diseño y Aplicaciones de un Control PID Wavelet*. Universidad Autónoma del Estado de Hidalgo [Tesis de Grado de Maestría], Pachuca, Hidalgo, México (2009)
5. Jarillo, A., Domínguez, O. A., Cruz, J. A., Ramos, L. E., Mijangos, T. S.: Design and Experimental Result of a Wavelet Filter for High Performance in Trajectory of Robotc System. *International Journal of Scientific and Engineering Research*, 3(10), 1300–1306 (2012)
6. Sedighzadeh, M., Rezazadeh, A.: Adaptive PID Control of Wind Energy Conversion Systems Using RASPI Mother Wavelet Basis Function Networks. In: *Proceedings of World Academy of Science, Engineering and Technology*, 27, pp. 269–273 (2008)
7. Khan, M. A., Azizur Rahman, M.: Real-Time Implementation of a Wavelet Based Speed Controller for Induction Motor Drives. In: *Canadian Conference on Electrical and Computer Engineering*, pp. 546–549, IEEE Press, Vancouver, Canada (2007)
8. Khan, M. A., Hinchey, M. J., Azizur Rahman, M.: Implementation of Wavelet Controller for Battery Storage System of Hybrid Electric Vehicle. In: *IEEE Industry Applications Society Annual Meeting*, pp. 1–8, IEEE Press, Houston, Texas (2009)
9. Cruz, J. A., Jarillo, A., Ramos, L. E.: Control de un Robot Manipulador Usando Control Wavelet: Experimental. In: *VI Semana Nacional de Ingeniería*, pp. 33–42, Oaxaca, México (2010)
10. Jeevan, L. J., Malik, V.: A Wavelet Based Multi-Resolution Controller. In: *Journal of Emerging Trends in Computing and Information Sciences*, pp. 17–21 (2011)
11. Yao, C. Zhang, G.: A Multi-Resolution PID Controller based on the wavelet transform. In: *Advanced Materials Research*, pp. 632–636 (2012)
12. Goswami, J. C., Chan, A. K.: *Fundamentals of Wavelet: Theory, Algorithms and Applications*. John Wiley and Sons, New Jersey, USA (2010)
13. Roldan, H., Itzá, B. A., Ramos, L. E.: Análisis de estabilidad de un control PID wavelet. *Revista Internacional de sistemas Computacionales y Electrónicos*, 3(6), 27–32, Hidalgo, México (2011)
14. Khan, M. A., Azizur Rahman, M. A.: A Novel Neuro-Wavelet Based Self-Tuned Wavelet Controller for IPM Motor Drives. In: *IEEE Industry Applications Society Annual Meeting*, 2008, pp. 1–8. IEEE Press, Edmonton, Canada (2008)
15. Novint Technologies, Inc.: <http://www.novint.com>
16. Tsai, L. W.: A Parallel Manipulator with Only Translational Degrees of Freedom. In: *ASME Design Engineering Technical Conferences* (1996)
17. Daubechies, I.: *Ten Lectures on Wavelets*. SIAM, Philadelphia, USA (1992)
18. Visioli, A.: *Practical PID Control*. Springer-London, England (2006)

Reviewing Committee

Ricardo Acevedo	Ivan Adrian Lopez Sanchez
Moisés Alencastre Miranda	Antonio Marin-Hernandez
Roberto Alonso Rodriguez	Lourdes Martínez
Joanna Alvarado-Uribe	Martinez Medina
Gustavo Arroyo Figueroa	Miguel Angel Medina Perez
Christian Arzate	Efrén Mezura-Montes
Ivonne Maricela Ávila Mora	Sabino Miranda-Jiménez
Jorge Bautista López	Daniela Moctezuma
Ivo Buzon	Raul Monroy
Maria Victoria Carreras	Jaime Mora-Vargas
Felix Castro Espinoza	Saturnino Job Morales Escobar
Noé Alejandro Castro Sánchez	Lourdes Muñoz Gómez
Bárbara Cervantes	Antonio Neme
Jair Cervantes	Alberto Oliart Ros
Efren Chavez Ochoa	Mauricio Osorio
Gustavo Delgado Reyes	Elvia Palacios
Sofía N. Galicia-Haro	Hiram Eredín Ponce Espinosa
Natalia Garcia	Carlos Pérez Leguizamo
Alexander Gelbukh	Maricela Quintana López
David Gonzalez	Carlos A. Reyes-García
Miguel Gonzalez-Mendoza	Carlos Alberto Rojas Hernández
Hugo Gustavo González	Rafael Rojas Hernández
Hernández Mario Graff	Dafne Rosso
Fernando Gudiño	Oscar S. Siordia
Pedro Guevara López	Grigori Sidorov
Yasmin Hernandez	Abraham Sánchez López
Neil Hernandez Gress	Israel Tabarez Paz
Oscar Herrera	Eric Sadit Tellez
Rodolfo Ibarra	Nestor Velasco-Bermeo
Yulia Ledeneva	Elio Villaseñor
Asdrubal Lopez Chau	Alisa Zhila
Juan Carlos Lopez Pimentel	

Impreso en los Talleres Gráficos
de la Dirección de Publicaciones
del Instituto Politécnico Nacional
Tresguerras 27, Centro Histórico, México, D.F.
mayo de 2015
Printing 500 / Edición 500 ejemplares

