

Special Issue

# LAJC

---

LATIN-AMERICAN  
JOURNAL OF COMPUTING

---

FACULTAD DE INGENIERÍA DE SISTEMAS  
QUITO - ECUADOR

VIII JORNADAS  
JISIC  
2017

Editor in Chief:  
PhD. Jenny Torres, Escuela Politécnica Nacional, Ecuador

Invited Editor:  
PhD. Marco Benalcázar, Escuela Politécnica Nacional, Ecuador

<http://lajc.epn.edu.ec/>



ESCUELA  
POLITÉCNICA  
NACIONAL



FACULTAD DE INGENIERÍA DE SISTEMAS



# LATIN AMERICAN JOURNAL OF COMPUTING

## LAJC

Vol IV, Issue 3, November 2017

ISSN: 1390-9266

e-ISSN: 1390-9134

Published by:  
Escuela Politécnica Nacional  
Facultad de Ingeniería de Sistemas

Quito – Ecuador

## **LATIN AMERICAN JOURNAL OF COMPUTING – LAJC**

### **Published by:**

Escuela Politécnica Nacional  
Facultad de Ingeniería de Sistemas  
Ecuador

### **Editorial Committee:**

Dra. Jenny Torres, Escuela Politécnica Nacional, Ecuador  
Dr. Edison Loza, Université Grenoble Alpes, France  
Dr. Alex Buitrago, Universidad Externado de Colombia, Colombia  
Dr. Henry Roa, Pontificia Universidad Católica, Ecuador

### **Editor in Chief:**

Dra. Jenny Torres, Escuela Politécnica Nacional, Ecuador

### **Section Editors:**

Ing. Hernán Ordoñez, Escuela Politécnica Nacional, Ecuador  
Dra. Jenny Torres, Escuela Politécnica Nacional, Ecuador  
Dr. Edison Loza, Université Grenoble Alpes, France

### **Mailing Address:**

Escuela Politécnica Nacional, Facultad de Ingeniería de Sistemas  
Ladrón de Guevara E11-253, La Floresta  
Quito-Ecuador, Apartado Postal: 17-01-2759

### **Web Address:**

<http://lajc.epn.edu.ec>

### **E-mail:**

[lajc@epn.edu.ec](mailto:lajc@epn.edu.ec)

### **Frecuency:**

2 issues per year

### **Circulation:**

350

## EDITORIAL

La inteligencia artificial y el aprendizaje de máquina son disciplinas relativamente nuevas. Su nacimiento se remonta al inicio de la segunda mitad del siglo XX, donde ocurrieron dos hitos importantes que marcaron el comienzo de estas disciplinas. En el primer hito, Alan Turing, en su trabajo *Computing Machinery and Intelligence* publicado en 1950, propuso considerar la pregunta “can machines think?”. En este trabajo, Turing aportó varias ideas que actualmente se usan dentro de la inteligencia artificial: test de Turing, aprendizaje de máquina, algoritmos genéticos y aprendizaje por refuerzo. El segundo hito consiste en el uso público del término “inteligencia artificial” en 1956. En este año, John McCarthy realizó una convocatoria para un workshop de verano de 2 meses de duración. Este workshop se desarrolló en el Dartmouth College de Estados Unidos. La propuesta para la convocatoria realizada por McCarthy es la siguiente:

“Proponemos que durante 2 meses, 10 personas realicen un estudio sobre inteligencia artificial durante el verano de 1956 en el Dartmouth College de Hanover, New Hampshire. Este estudio se debe realizar sobre la base de la hipótesis de que cada aspecto del aprendizaje o cualquier otra característica de la inteligencia pueden, en principio, describirse de manera muy precisa de tal modo que un computador los pueda simular. Se intentará encontrar cómo hacer que las máquinas usen el lenguaje, formen abstracciones y conceptos, resuelvan tipos de problemas hasta ahora reservados para los humanos, y que se mejoren a sí mismos. Creemos que se puede lograr un avance significativo en uno o más de estos problemas si un grupo de científicos cuidadosamente seleccionados trabajan juntos durante un verano.”

El workshop de Dartmouth fue el espacio que en 1956 permitió el nacimiento de lo que, desde ese momento hasta la actualidad, se llama inteligencia artificial. Es conveniente remarcar que algunos problemas planteados para el workshop de Dartmouth todavía son objeto de investigación en la actualidad.

Con esta breve historia, pasemos a describir brevemente en qué consiste la inteligencia artificial y el aprendizaje de máquina, o en inglés machine learning. Estas dos disciplinas pertenecen al área de ciencias de la computación, que fusiona la computación con la matemática. La inteligencia artificial, tal como hoy en día se la concibe, se encarga del estudio de agentes que actúan de manera racional; es decir, agentes que producen el mejor resultado o resultado esperado, en términos de alguna función de costo, en ambientes determinísticos o estocásticos, respectivamente. Un agente se define como una entidad que opera de manera autónoma, percibe su ambiente, persiste sobre un tiempo prolongado, se adapta a los cambios, y crea y persigue objetivos.

El aprendizaje de máquina consiste en la creación de algoritmos que mejoran su desempeño en base a ejemplos de entrenamiento. Bajo el paradigma de la programación tradicional, un usuario debe escribir explícitamente las instrucciones que debe ejecutar un algoritmo. En contraste a este hecho, bajo el paradigma de aprendizaje de máquina, un “algoritmo de aprendizaje” en base a ejemplos de entrenamiento genera automáticamente un modelo que es capaz de realizar tareas muy complejas como detectar rostros humanos en imágenes naturales, un problema que resultaría demasiado difícil de resolver usando el paradigma de programación tradicional.

La inteligencia artificial y el aprendizaje de máquina son disciplinas que están muy interrelacionadas y que han permitido generar conocimiento y tecnología que, otrora eran exclusivas de las películas de ciencia ficción. Por citar sólo un par de ejemplos, en este 2017 contamos con aplicaciones capaces de reconocer el tono de voz y las palabras que pronunciamos. También contamos con los primeros prototipos de autos que se conducen a sí mismos. Esto evidencia que la inteligencia artificial y el aprendizaje de máquina son parte del presente y futuro de la humanidad y cada día nos plantean nuevos problemas y desafíos que

requieren la atención de la comunidad científica mundial y, por supuesto, de la comunidad científica del Ecuador. Por esta razón, la Facultad de Sistemas de la Escuela Politécnica Nacional, pone a su vuestra consideración la octava edición de las Jornadas de Ingeniería de Sistemas Informáticos y de Computación (JISIC) 2017 con el tema inteligencia artificial y aprendizaje de máquina.

En esta edición especial de la revista Latin American Journal of Computing (LAJC) se recoge los mejores papers presentados en las JISIC 2017. Es conveniente anotar que todos los trabajos incluidos en esta edición especial son inéditos y pasaron por un riguroso proceso de selección y revisión por parte de pares académicos nacionales e internacionales, quienes en base a su experticia, juzgaron la calidad científica de los papers en el esquema "doble ciego." A nombre del comité científico de las JISIC 2017, y del mío propio, extendemos nuestro más sincero reconocimiento tanto a los revisores como a los autores por su valiosa contribución.

Esperamos que las JISIC 2017, a través de las conferencias magistrales internacionales y nacionales (adjuntos a este editorial), presentación de papers, workshops, talleres y cursos, sean para el Ecuador en este 2017 el equivalente al workshop de Dartmouth para el mundo en 1956. Anhelamos que este evento científico-tecnológico sea el hito que masifique la investigación científica y generación de tecnología en el Ecuador en inteligencia artificial y aprendizaje de máquina.

Sean todos bienvenidos.

\*\*\*

Artificial intelligence and machine learning are new disciplines in the field of computer science. Their birth dates back to the beginning of the second half of the 20th century, when two important milestones marked the start point of these two disciplines. In the first milestone, Alan Turing, in his work "computing machinery and intelligence" published in 1950, posed the question "can machines think?" In this work, Turing contributed several ideas that are currently still used within the field of artificial intelligence: Turing test, machine learning, genetic algorithms and reinforcement learning. The second milestone consists of the public use of the term "artificial intelligence" in 1956. In this year, John McCarthy made a call for a summer workshop of 2 months. This workshop was held at the Dartmouth College in the United States. The proposal for the workshop of Dartmouth is the following:

"We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer."

The workshop of Dartmouth was the space that, in 1956, allowed the birth of what, from that moment up to the present is called artificial intelligence. It should be noted that some problems stated for this workshop are still open for new research.

With this brief history, let's summarize what artificial intelligence and machine learning consist of. These two disciplines belong to the area of computer science, which fuses computing theory with mathematics. Artificial intelligence, as it is conceived today, is in charge of the study of agents who act rationally; that is, agents that produce the best result or expected result, in terms

of some cost function, in deterministic or stochastic environments, respectively. An agent is defined as an entity which operates autonomously, perceives its environment, persists over a prolonged time period, adapts to changes, and creates and pursues goals.

Machine learning consists of creating algorithms that improve their performance based on training examples. Under the paradigm of traditional programming, a user must explicitly write the instructions or code that an algorithm has to execute. In contrast to this fact, under the new paradigm of machine learning, a "learning algorithm" based on training examples automatically generates a model that is capable of performing very complex tasks such as, for example, detecting human-faces in natural images, a problem that would result too difficult to solve using the paradigm of traditional programming.

Artificial intelligence and machine learning are disciplines that are highly interconnected. They have allowed us the discovery of new knowledge and the generation of new technology that was once exclusive to only science fiction movies. To mention just a couple of examples, in this 2017 we have applications and software capable of recognizing the tone of voice and our speech. We also have the first prototypes of self-driving cars. These examples provide evidence that artificial intelligence and machine learning are part of the present and the future of humanity. Every day these disciplines bring new problems and challenges that require the attention of the scientific community of the world. For this reason, the Facultad de Sistemas of the Escuela Politécnica Nacional organized the eighth edition of the conference "Jornadas de Ingeniería de Sistemas Informáticos y de Computación (JISIC) 2017" with the topic of artificial intelligence and machine learning. We hope that this conference contributes significantly to the progress and growth of artificial intelligence and machine learning in Ecuador.

This special issue of the Latin American Journal of Computing (LAJC) contains the best papers presented in the JISIC 2017. It is convenient to remark that all the papers included in this issue are novel and went through a rigorous selection and revision process. In this process, we had reviewers from Ecuador and from the rest of the world. Each paper submitted to the conference was double-blind peer reviewed based on its technical strength and contribution, presentation and quality of the references. On behalf of the scientific committee of the JISIC 2017, and of my own, I extend my most sincere recognition to both the reviewers and the authors for their valuable contribution.

Finally, we hope that the JISIC 2017 will be for Ecuador in 2017 the equivalent of the workshop of Dartmouth for the world in 1956. We also hope that the international and national keynote speakers (abstracts are attached to this editorial), scientific papers, workshops, and courses of the JISIC 2017 motivate our world-wide audience, especially the ones from Ecuador, to do more research and generate significant breakthroughs in the fields of artificial intelligence and machine learning.

You all are very welcome to JISIC 2017.

**Dr. Marco E. Benalcázar**  
**General chair of JISIC 2017**  
**Chair of the Computational Intelligence Society of IEEE – Ecuador**  
**Associate Professor of the Departamento de Informática y Ciencias de la Computación**  
**Escuela Politécnica Nacional**  
**Invited Editor**

**Dr. Ulisses Braga-Neto**

Texas A&amp;M University, USA

**Pattern Recognition for Small-Sample Data in Genomics**

**Abstract:** Modern biomedical applications based on genomic and proteomic expression profiles often produce large amounts of data characterized by a very large number of measurements made on a much smaller number of sample points. This introduces difficult challenges in the application of classification methods to obtain accurate predictive models. In this talk, we discuss our recent work on this topic, highlighting Bayesian approaches to classification of proteomic and metagenomic data, as well as novel approaches for classification and error estimation based on restricted and nonstationary data.

**Dr. Selmer Bringsjord**

Rensselaer Polytechnic Institute, USA

**Can an AI System Have and Communicate in Prose a Sense of Ecuador?**

**Abstract:** Those who have spent much of their lives in Ecuador know what it is like “from the inside” to be there for an extended period of time. We can say that such people feel what it is like to be a person experiencing Ecuador. If they are good creative writers, these people can in addition communicate in prose what it is like to experience aspects of the Ecuadorian life. What this means in the human case is that there is a deep self-consciousness of Ecuador —of its coast, islands, mountains, culture, cuisine, heritage, etc. Some of the literature of Ecuador (which I have been reading in the English) conveys to readers thereof at least some of what it is like to live in Ecuador, to the point that the reader himself/herself experiences a sense of the very place. Parallel points can of course be made about places and writers with which I am personally long acquainted with (e.g., for New York: Winter’s Tale by Helprin, or Seize the Day by Bellow). Now, could an AI ever have such a sense of a place? and have a capacity to convey this sense in literature that it writes? My answer is a double No, for reasons I will explain in the talk. However, as I also explain, we can pursue AI engineering that produces artifacts that, in limited situations, appear to have a sense of place, and perhaps also appear to have the capacity to communicate that sense of place, because of the language they generate.

**Dra. Isabel Passoni**

Universidad Nacional de Mar del Plata, Argentina

**Machine Learning: Applications in the field of Bioengineering**

**Abstract:** The field of Bioengineering will be presented in principle, emphasizing on its interdisciplinary and complex character in function of the diverse scientific communities involved and on its related specialties which include medicine and biology. Different aspects will be highlighted: research, technological development and management of health institutions that are addressed in this domain of engineering. I will provide an overview of Computational Intelligence of Artificial Neural Networks, Fuzzy Logic and Evolutionary Computation that provide tools to the group of techniques of Machine Learning (ML) for the design of models from relevant data in Bioengineering. Designs of Decisions Support Models in Medicine will be presented. They involve diverse sources of information such as signals, images and data processing. Also, it will be shown that the growth of robotics in medical applications is undoubtedly an opportunity for the application of ML techniques. Emphasis will be placed on the opportunity and relevance of addressing new R & D in ML motivated by their potential application in the field of Bioengineering.

## **Latin American Journal of Computing – LAJC**

### **Reviewers**

We are most grateful to the following individuals for their time and commitment to review manuscripts for Latin American Journal of Computing – LAJC in the context of JISIC 2017.

MSc. Andrés Jaramillo, Escuela Politécnica Nacional, Ecuador  
Phd. Antonio Fernández Baldera, Universidad Politécnica de Madrid, España  
Phd. Edison Loza-Aguirre, Escuela Politécnica Nacional, Ecuador  
MSc. Henry Paz, Escuela Politécnica Nacional, Ecuador  
MSc. Iván Carrera, Escuela Politécnica Nacional, Ecuador  
Phd. Josafá Pontes, Escuela Politécnica Nacional, Ecuador  
Phd. José Aguilar, Universidad de los Andes, Venezuela  
Phd. José Lucio, Escuela Politécnica Nacional, Ecuador  
Phd. José Manuel Gómez Soriano, Universidad de Alicante, España  
Phd. Juan Lara, Universidad a Distancia de Madrid, España  
MSc. Julio César Ibarra Fiallo, Universidad San Francisco de Quito, Ecuador  
MSc. Marco Danilo Burbano, Escuela Politécnica Nacional, Ecuador  
MSc. Marco Molina, Escuela Politécnica Nacional, Ecuador  
Phd. Marco Segura, Escuela Politécnica Nacional, Ecuador  
Phd. María Pérez, Escuela Politécnica Nacional, Ecuador  
MSc. Monserrate Intriago, Escuela Politécnica Nacional, Ecuador  
Phd. Myriam Hernández, Escuela Politécnica Nacional, Ecuador  
Phd. Sang Guun Yoo, Escuela Politécnica Nacional, Ecuador  
Phd. Wilbert Aguilar, Universidad de las Fuerzas Armadas-ESPE, Ecuador  
Phd. Yadira Quiñonez, Universidad autónoma de Sinaloa, México



## TABLE OF CONTENTS

<b>Sistema de telemedicina para el monitoreo de la frecuencia cardiaca empleando análisis de componentes independientes en videos</b> <i>Telemedicine system for monitoring heart rate using independent component analysis in videos</i>	
Karla Portilla, Víctor Santos, María Trujillo y Andrés Rosales.....	11-16
<b>Implementación de un Modelo Afetivo para MASOES</b> <i>Implementation of an Affective Model for MASOES</i>	
Saúl Piña, Niriaska Perozo.....	17-22
<b>Vibratory Neurostimulator for Patients with Essential Tremor</b> G Delgado G., Gavilanes J., Mena L., Encalada P.....	23-28
<b>Diseño e Implementación de una Plataforma Experimental de Mano Robótica</b> <i>Design and Implementation of a Robot Hand Experimental Platform</i>	
Washington Caraguay, Marco Sotomayor, Miguel Alonso .....	29-36
<b>First Order Methods for High Resolution Image Denoising</b> David Villacís.....	37-42
<b>Diseño y Construcción de un Escáner 3D e Integración a una Impresora 3D (ingeniería inversa accesible de bajo costo y fuente abierta)</b> <i>Design and Construction of a 3D Scanner and Integration to a 3D Printer (accessible low cost and open source reverse engineering)</i>	
Leombardo Hernández, Andrés Cela.....	43-48
<b>Optimización de un Sistema de Reconocimiento de Idioma Basado en la Representación Vectorial de Fonemas</b> <i>Optimising a Language Recognition System Through Phoneme-Based Vector Representation</i>	
Francisco Charro, Marco Herrera, Nataly Pozo y Andrés Rosales.....	49-54
<b>Dissolved gas content forecasting in power transformers based on Least Square Support Vector Machine (LSSVM)</b> Roberto J. Fiallos .....	55-60
<b>Simulation of evacuation with multi-agents on georeferenced layers with GAMA</b> Zoila Ruiz-Chavez, Jaime Salvador-Meneses .....	61-66
<b>Development of a Prototype for the Quality Control of Bovine Meat Determined by its Organoleptic Characteristics, based on an Automatic Inspection System for Artificial Vision</b> Portero P., Mena B., Gavilanes J.....	67-72
<b>Deep Learning como modelo predictivo para clasificar dígitos manuscritos</b> <i>Deep Learning as a predictive model to classify handwritten digits</i>	
Omar Alexander Ruiz-Vivanco .....	73-78



# Sistema de Telemedicina para el Monitoreo de la Frecuencia Cardiaca Empleando Análisis de Componentes Independientes en Videos

## *Telemedicine System for Monitoring Heart Rate Using Independent Component Analysis in Videos*

Karla Portilla, Víctor Santos, María Trujillo y Andrés Rosales

**Resumen**—El corazón es el órgano encargado de suministrar sangre, oxígeno y nutrientes a través del cuerpo humano, su monitoreo inicial se ve reflejado en el valor de la frecuencia cardiaca, esta medición estima una posible irregularidad dentro del organismo. El control de este signo vital conlleva en la mayoría de casos la ayuda de personal médico o el uso de sensores que necesiten mantener contacto con la piel, estos métodos habituales de monitoreo generan molestias e inconformidad en recién nacidos, adultos mayores o pacientes que sufren de sensibilidad en la piel. Considerando estos inconvenientes se ofrece un sistema no invasivo y sin contacto, basado en el principio de telemedicina, capaz de proporcionar un monitoreo remoto de la frecuencia cardiaca de forma rápida, confiable y económica. Los resultados son comparados con un oxímetro de pulso de una resolución de 1 lpm y precisión de  $\pm 2\%$ , obteniendo errores bajos y aceptables.

**Palabras Clave**— análisis componentes independientes, frecuencia cardiaca, sin contacto, seguimiento facial, separación ciega fuentes, telemedicina.

**Abstract**—Heart is the organ responsible for providing blood, oxygen and nutrients through the human body, their initial monitoring was reflected in heart rate value, this measurement is a possible irregularity within the organism. Control of this vital sign in most cases involves help of medical staff or using sensors that need to maintain contact with the skin, these common monitoring methods generate discomfort and nonconformity in newborns, senior adults or patients suffering from skin sensitivity. According these disadvantages, project offers a non-invasive and non-contact system based on telemedicine principles, which provides a quick, reliable and economical remote heart rate monitor. Results are compared with a pulse oximeter with a resolution of 1 bpm and accuracy of  $\pm 2\%$ , obtaining low and acceptable errors.

**Index Terms**— independent component analysis; heart rate; contactless; face tracking; blind source separation; telemedicine.

Los autores pertenecen al Departamento de Automatización y Control Industrial de la Escuela Politécnica Nacional, Ladrón de Guevara, E11-253, Quito, Ecuador. (e-mail: {karla.portilla, victor.santos, maria.trujillo01, andres.rosales}@epn.edu.ec)

### I. INTRODUCCIÓN

El flujo continuo de sangre está regido por el sistema cardiovascular, sus insuficiencias estructurales o funcionales, son causantes de altos índices de discapacidad y de morbilidad en aproximadamente un tercio de la población mundial [1]. Los signos vitales como: temperatura corporal, frecuencia cardiaca, frecuencia respiratoria y presión arterial [2], proporcionan medidas de las funciones más básicas e importantes del cuerpo humano. Mediante el monitoreo de la frecuencia cardiaca es posible tener una referencia del estado en el que se encuentra el corazón humano. Al existir variaciones fuera del rango normal en reposo, 60 a 100 latidos por minuto [3], esta medición se convierte en un indicador de alteraciones dentro del organismo.

El monitoreo de la frecuencia cardiaca se basa en detectar pulsaciones rítmicas en la arteria explorada [2], para lo cual existen diversos métodos como la medición manual cuyo principal inconveniente se presenta al confundir la frecuencia cardiaca del examinador con la del paciente o aumentar el tiempo de adquisición si el paciente presenta irregularidades cardíacas [4]. Por otro lado, se encuentra el monitoreo digital de forma directa con instrumentos como: relojes con pulsómetro o monitores con o sin banda sobre el pecho; también es posible obtener información de forma indirecta a partir de instrumentos que valoran otros signos vitales tales como: medidores de presión arterial o de saturación de oxígeno en la sangre [3]. Los métodos digitales presentan la característica de usar elementos que mantienen contacto con la piel para adquirir los datos, causando incomodidad para quien los usa, considerando a personas que posean alguna enfermedad, particularmente en la piel.

Actualmente los métodos no invasivos se ven limitados a aplicaciones teóricas que no son comerciales por su alto costo y complejidad como el caso de los sensores térmicos, ópticos o que usen radares de efecto Doppler [5] [6] [7]. La técnica presentada trata de contrarrestar los inconvenientes de los

métodos citados con anterioridad, manteniendo la característica de ser no invasivo, eliminando completamente el contacto con la piel.

Como se aprecia en la Fig. 1 el sistema propuesto para el monitoreo de la frecuencia cardiaca se forma a partir de una cámara de video, un procesador de placa reducida, un computador central y un software computacional.



Fig. 1. Esquema del sistema propuesto.

El sistema implementado, conserva la fiabilidad de los datos obtenidos a la par que genera comodidad en la persona, al incorporar, como se indica en la sección II, algoritmos que adquieran un archivo de video, utilizando un computador de placa reducida y una cámara de video; en la sección III se explica el proceso para la detección de una sección de piel; con una región de interés delimitada se procede a obtener la frecuencia cardíaca mediante un procesamiento digital, basado en cuatro principios: obtención de señales RGB, filtrado, análisis de componentes independientes y transformada rápida de Fourier, como se indica en la sección IV. Los resultados alcanzados con el sistema son analizados y contrastados con técnicas convencionales en la sección V; finalmente se presentan las conclusiones obtenidas en la sección VI.

## II. ADQUISICIÓN DE DATOS

Uno de los elementos que conforman el dispositivo para la adquisición de datos es un computador de placa reducida Raspberry Pi 3 Modelo B, que mantiene un equilibrio entre robustez y versatilidad sumado a mejorar la relación costo-beneficio, en comparación a otro tipo de tarjetas embebidas, al no necesitar software propietario y su capacidad de realizar multiprocesamiento. Este dispositivo cuenta también con: una cámara de video web Microsoft LifeCam 1425 con la capacidad de obtener video en alta definición conservando el enfoque y una pantalla LCD TFT táctil de 3.5 pulgadas.

El sistema propuesto presenta una opción para el monitoreo remoto de la frecuencia cardíaca, por lo que es necesario comunicar el dispositivo de adquisición de datos, mediante tecnologías de la información y comunicación, con el computador central donde se realizará el procesamiento de la imagen, así como un almacenamiento digital de las mediciones realizadas.

Para mantener la eficiencia del sistema es necesario reducir la carga computacional en las etapas de obtención y envío del archivo de video, por esta razón, la resolución de la cámara de video es manipulada en tiempo real utilizando software no propietario para visión por computador (OpenCV 3.2.0) [8], dentro de una interfaz de programación (Python) donde se desarrollaron algoritmos que permitan definir parámetros

importantes como la resolución en la captura de video, la relación o tasa de imágenes adquiridas por cada segundo (fps) o el tipo de codificación del archivo de video para reducir su tamaño en el almacenamiento garantizando su nitidez [9].

Dentro del dispositivo de adquisición de datos se cuenta con una interfaz donde es necesario ingresar el código de registro del usuario a monitorear, para iniciar el envío de información hacia el computador central, el cual se realiza de forma inalámbrica utilizando el módulo BCM43438 incorporado en la Raspberry Pi [10] [11]. El proceso de envío de información se basa en el Protocolo de Control de Transmisión (TCP) utilizando puertas de enlace predeterminadas sobre una arquitectura para el diseño de software tipo cliente-servidor, evitando el procesamiento excesivo en protocolos basados en el modelo OSI al utilizar un protocolo sencillo de solicitud-respuesta, por otra parte se elimina la dependencia del medio físico cableado al vincular a cada uno de los dispositivos mediante direcciones IP y números de puerto (sockets) para identificar a un programa dentro del dispositivo de adquisición de datos y el computador central.

## III. DETECCIÓN Y SEGUIMIENTO DE UNA SECCIÓN DE PIEL

Para la detección y seguimiento de una sección de piel, se requiere determinar aquella que sea accesible como el rostro o las manos y adicionalmente posea características específicas útiles para efectuar un correcto seguimiento, considerando estos requerimientos se seleccionó al rostro como la región de interés.

Con el objetivo de obtener la ubicación de la región de interés, como se muestra en la Fig. 2a, es necesario aplicar el algoritmo propuesto por Paul Viola y Michel Jones [12], el cual establece un método eficiente para la detección inicial de rostros, aplicando elementos básicos basados en el contraste de intensidades luminosas entre zonas rectangulares adyacentes [13]. Este algoritmo presenta un porcentaje de detección muy alto, generando falsos positivos muy bajos, trabaja en tiempo real en al menos dos frames por segundo [14].

La función de Matlab *vision.CascadeObjectDetector* permite la detección de rostros, empleando una arquitectura en cascada, la cual permite construir clasificadores a partir del principio de Viola-Jones, los cuales poseen un porcentaje de detección efectiva del 100% [15].

### A. Rastreador de rasgos faciales KLT

Variaciones de iluminación al identificar el rostro humano pueden crear confusión con objetos, este problema se elimina al aplicar el algoritmo de Kanade-Lucas-Tomasi (KLT), este método realiza de forma conjunta la selección de particularidades, como se observa en la Fig. 2b, así como el seguimiento de estas características de una imagen hacia otra correctamente. Es importante establecer restricciones de brillo y movimientos dentro de la imagen [16].

El proceso de detección y seguimiento de rostro necesita ser aplicado sobre una imagen en escala de grises, de la cual se obtendrá características específicas dentro del rostro, utilizando el método de valor propio mínimo [14], esta acción

es posible realizarla con la función *detectMinEigenFeatures* de Matlab. La ubicación de estos puntos puede variar entre cada uno de los frames presentes en el video, debido a cambios de iluminación o ciertos movimientos articulados, por lo que su adquisición debe ser continua, esto se puede realizar mediante la función *vision.PointTracker* [17].

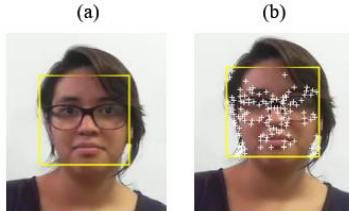


Fig. 2. (a) Detección de rostro aplicando el método de Viola-Jones.  
(b) Selección de particularidades del rostro mediante el algoritmo KLT.

#### IV. OBTENCIÓN DE LA FRECUENCIA CARDIACA

El valor de la frecuencia cardiaca ( $fc$ ) tiene relación directa con los pequeños cambios de color presentes en la piel, estos son generados por la diferencia de volumen sanguíneo ante variaciones de presión sistólica y diastólica dentro del tejido circundante, conocido como ciclo cardiaco con un periodo aproximado de  $1/fc$ , gracias al resultado de estas pequeñas alteraciones es posible estimar el porcentaje de coloración en RGB que tendrá una sección de piel. Este valor tiende a cambiar debido a condiciones ambientales o por actividades físicas que realiza la persona [4].

##### A. Filtrado y normalizado

Con la finalidad de facilitar el tratamiento de las señales, estas se deben representar en el modelo de color RGB debido a que el ojo humano posee la capacidad de distinguir claramente los colores: rojo, verde y azul; además de la facilidad que estos colores presentan, según sus intensidades de luz relativas, para conseguir distintos colores, incluyendo el color negro y blanco. Dentro del procesamiento de imágenes el modelo RGB es útil cuando se establece un mapa de colores específico e individual para cada canal, como se indica en la Fig.3, evitando que sus ejes varíen entre sí.

Estas señales obtenidas dentro de la región de interés pueden contener ruido externo que puede provocar una medición errónea, por lo que deben ser filtradas y normalizadas. El filtro de mediana elimina la mayor cantidad de ruido dentro de una imagen, en especial el de tipo impulso producido por el parpadeo de los ojos, eliminando efectos extraños, su principal ventaja es su desarrollo en el dominio espacial por lo que su aplicación se reduce significativamente, con la desventaja de introducir pequeños errores debido al redondeo [20] [21].

De forma general se puede representar un filtro de mediana de  $n$  elementos como indica (1), donde se define la posición relativa de la mediana como  $x$ . Todos los patrones de tipo puntual o lineal fuera de este rango de dimensiones en pixeles serán eliminados del frame.

$$x = (n + 1) / 2 \quad (1)$$

Para terminar esta etapa los valores atípicos deben ser omitidos, para asegurar la medición final. Se debe entonces aplicar a cada uno de los pixeles dentro del frame, un filtro de normalización, como se presenta en la Fig. 4.

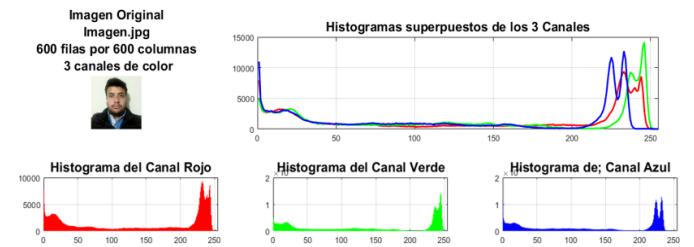


Fig. 3. Obtención de señales RGB y sus histogramas independientes.

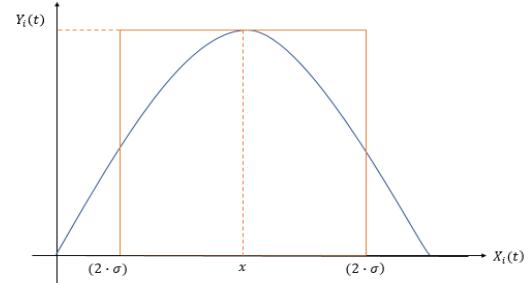


Fig. 4. Normalización aplicada a las señales RGB obtenidas.

Donde  $y_i(t)$  representa el vector de salida con los datos normalizados,  $x_i(t)$  el vector de datos de entrada filtrados y  $\sigma$  la desviación estándar calculada en cada uno de los canales RGB, la cual establece el grado de variabilidad del conjunto de datos.

##### B. Análisis de Componentes Independientes

Establecer el valor de la frecuencia cardiaca conlleva a realizar un análisis de la separación ciega de fuentes, las cuales se basan en la estimación un grupo de señales fuente sin conocer su naturaleza, a partir de un conjunto de mediciones. Particularmente consiste en separar los componentes adyacentes principales de una mezcla de varias señales superpuestas [22]. Dentro del conjunto, las mediciones se encuentran combinadas entre sí, dado que se transmiten en un medio compartido, tal como se aprecia en la Fig. 5.

De forma análoga el diagrama de la Fig. 5 puede ser representado como se indica en (5).

$$S = AF + R \quad (5)$$

Donde  $S$  representa el vector de muestras de las señales obtenidas,  $A$  la matriz de combinaciones con dimensiones  $P \times Q$ ,  $F$  el vector de muestras de las señales fuente y  $R$  el vector de ruido asociado con componentes estadísticamente independientes de las fuentes [23].

Esta dificultad es posible reducirla al aplicar una diagonalización aproximada articular de las matrices propias (JADE), mediante este algoritmo es factible diagonalizar la matriz con los autovalores de las señales combinadas, basado en el cálculo estadístico de alto orden [24] [25]. De esta forma

se encuentra una matriz  $B$  de dimensiones  $Q \times P$ , de forma que, al leer las muestras de las señales obtenidas, se facilite la extracción de señales de salida  $Y$ , similares a las fuentes originales, como se muestra en la Fig. 6.

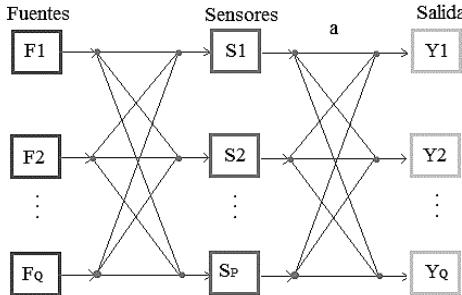


Fig. 5. Representación gráfica de separación ciega de fuentes [18].

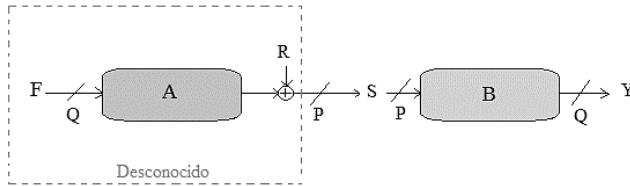


Fig. 6. Diagrama de bloques de separación ciega de fuentes.

Estimando los coeficientes de la matriz de combinaciones  $A$  y las fuentes  $F$  a partir de las muestras obtenidas  $S$  como el resultado de una combinación lineal de las fuentes; considerando ciertas fases para su correcto funcionamiento tales como: un proceso de blanqueamiento de la señal, estimación de sus matrices acumulativas, identificación de una matriz de rotación y finalmente la separación ciega de fuentes [23].

### C. Extracción de la frecuencia cardiaca

Las señales resultantes de JADE se encuentran en el dominio del tiempo por lo que es necesario establecerlas en el dominio frecuencial, esto facilitará el análisis computacional y su interpretación para lograr un estimado de la frecuencia cardiaca, este procedimiento se realiza al aplicar la transformada de Fourier [26].

Para realizar la transformada de Fourier se pueden aplicar tanto la transformada discreta (DFT) como la transformada rápida (FFT), obteniendo resultados muy similares, con la diferencia que en Matlab no es posible realizar la DFT por lo que es necesario incluir una función propia para su desarrollo.

La principal diferencia entre estos dos métodos radica en su tiempo de procesamiento, para el caso de los archivos analizados se cuenta con tres canales de colores cada uno con 1302 datos, al aplicar la DFT el tiempo de procesamiento superó los 4 segundos mientras que al aplicar la FFT el procesamiento se redujo a 0.0007 segundos, debido a la simplificación de operaciones complejas, al aplicar propiedades como simetría y periodicidad, además de realizar un proceso de enventanado rectangular que secciona la señal continua, como se indica en la Fig. 7.

El valor estimado de la frecuencia cardiaca en latidos por

minuto (lpm) está asociado con el pico máximo presente en el espectro de frecuencias obtenido en los canales RGB, para un rango determinado (50 a 200 lpm).

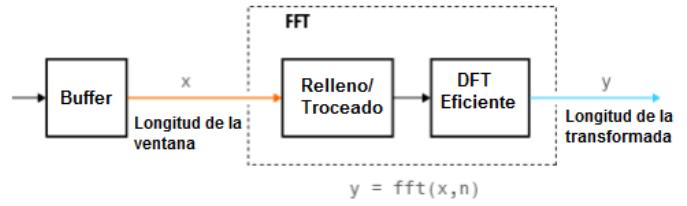


Fig. 7. Diagrama para el cálculo de la transformada rápida de Fourier [28].

## V. RESULTADOS

Para obtener el valor de la frecuencia cardiaca se utilizaron archivos de video obtenidos desde el procesador de placa reducida, empleando una cámara de video con una resolución de 640x480 pixeles (480p) y una duración de 10 segundos. El procesamiento de este archivo fue realizado en un computador central mediante Matlab, en el cual se desarrolló un sistema de visualización y almacenamiento de datos.

En una primera etapa se realizó un análisis tomando en cuenta el tiempo de operación, presentado en la Tabla 1, utilizado en técnicas convencionales como medición manual, oxímetro de pulso y tensiómetro digital en comparación con el monitor de frecuencia cardiaca no invasivo implementado, donde se puede observar que en relación a técnicas convencionales como medición manual y tensiómetro digital el tiempo de operación ha sido reducido evitando largos períodos que pueden causar molestias en el paciente.

Las mediciones de frecuencia cardiaca fueron realizadas en un grupo de 39 personas aleatorias, con diversas tonalidades de piel, en un rango de edad de 21 a 30 años, dentro de una habitación iluminada con luz artificial, por cada participante se realizó una única toma de video. La Tabla 2. indica los resultados obtenidos utilizando el sistema propuesto y técnicas convencionales antes mencionadas. Del mismo modo se incluye el cálculo del error absoluto tomando como referencia la medición obtenida con el oxímetro de pulso, considerando que este dispositivo presenta una resolución de 1 lpm y precisión de  $\pm 2\%$  [27].

Los resultados posteriores a las fases de detección y seguimiento de rostro, extracción de señales RGB, filtrado y normalizado, son sometidos al análisis de las componentes independientes (ICA), este proceso ofrece información útil para la medición.

TABLA I  
PERIODOS APROXIMADOS DE PROCESAMIENTO

Instrumento de medición	Tiempo (segundos)
Medición manual	40
Oxímetro de pulso	8
Tensiómetro digital	28
Monitor no invasivo	24

**TABLA II**  
**MEDICIONES Y ERRORES AL OBTENER LA FRECUENCIA CARDIACA**

Personas	Medición manual*	Oxímetro de pulso*	Tensiómetro digital*	Monitor no invasivo*	Error **
1	64	60	52	61	1.6
2	88	90	87	92	2.2
3	64	63	71	61	3.2
4	72	70	69	67	4.3
5	76	70	71	73	4.3
6	70	71	72	67	5.6
7	68	71	70	67	5.6
8	58	52	53	55	5.7
9	76	86	83	80	6.9

\*La unidad de medición para la frecuencia cardiaca son los latidos por minuto (lpm).

\*\*El error obtenido se expresa en unidades de porcentaje.

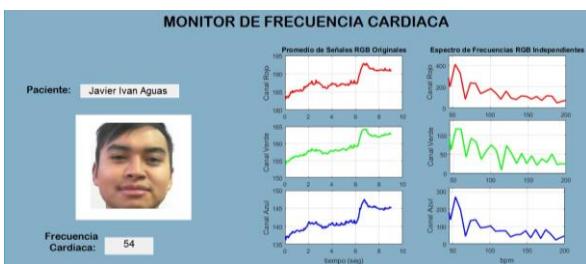


Fig. 8. Pantalla de visualización de la frecuencia obtenida.

A continuación, es necesario analizar el espectro de frecuencias en cada una de las señales RGB que contienen los cambios temporales, determinando el promedio de los valores máximos en los tres canales para obtener el valor de la frecuencia cardiaca de la persona, como se indica en la Fig. 8.

El personal médico tiene acceso a información de cada uno de los usuarios que se encuentren ingresados en el sistema, al introducir el código de registro, es posible visualizar datos personales y el historial mediciones obtenidas, como se muestra en la Fig. 9.

Se pueden presentar variaciones en los resultados obtenidos con el sistema propuesto si se incluyen participantes con edades superiores a 65 años, neonatos o niños hasta 3 años, en los que la frecuencia cardiaca en reposo será inferior a 60 lpm o superior a 100 lpm respectivamente [3].

Considerando las mediciones obtenidas a 9 personas aleatorias, se determinó un intervalo de confianza del 95%, este resultado permite corroborar la efectividad del proyecto, dado que el rango obtenido, como se presenta en la Fig. 10, se encuentra dentro de los límites normales de la frecuencia cardiaca.



Fig. 9. Pantalla de visualización de historial de mediciones.

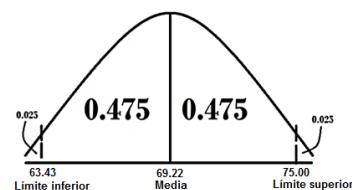


Fig. 10. Intervalo de confianza del 95%

## VI. CONCLUSIONES

El sistema propuesto es una alternativa innovadora que facilita el monitoreo remoto de la frecuencia cardiaca sin contacto con la piel, además ofrece mayor confort en varios pacientes que presentan enfermedades dérmicas, dado que no es necesaria la intervención de una persona externa para esta medición.

Contrastando los resultados obtenidos con el presente proyecto frente a varios métodos convencionales, se puede considerar que el monitor de frecuencia cardiaca es una solución ágil, económica y viable, tomando en cuenta los bajos errores obtenidos en la sección V.

El monitor de frecuencia cardiaca presentado ofrece un servicio de telemedicina económico que proporciona transmisión de datos bajo una misma red, efectivizando los períodos de procesamiento en comparación con sistemas habituales que mantienen contacto con la piel.

Este trabajo constituye la base para proyectos futuros en donde será factible incrementar la cantidad de personas en las que se realice esta medición, a su vez poder agruparlas por edades, etnias y condición médica de manera que se garantice la fiabilidad del mismo y una posible inserción en el mercado.

La técnica propuesta puede ser empleada como referencia inicial en la obtención de otros parámetros fisiológicos importantes, tales como la frecuencia respiratoria y la saturación de oxígeno en la sangre.

## REFERENCIAS

- [1] B. Herreros, F. Bandrés, "Prevención primaria de la arteriosclerosis en la hipertensión arterial," en *Prevención primaria de la arteriosclerosis*, 1ra ed., Madrid , España, ADEMÁS, 2009, pp. 123-125.
- [2] J. Daza, "Evaluación clínico-funcional del sistema cardiovascular," en *Evaluación clínico-funcional del movimiento corporal humano*, 1ra ed., Bogotá, Colombia, Panamericana, 2007, pp. 48-69.
- [3] A. Aguayo, A. Lagos. (2012, Junio). Guía clínica de control de signos vitales, [En Línea]. Disponible en: <http://academico.upv.cl/doctos/KINE-4068/%7B328B1B37-2C2A-4747>

- 8B38-169806A27753%7D/2012/S1/GUIA%20TECNICA%20DE%20CONTROL%20DE%20SIGNOS%20VITALES%20KINE.pdf
- [4] H. Jinich, A. Lifshitz, J. García, M. Ramiro, "Signos Vitales," en *Síntomas y signos cardinales de las enfermedades*, 6ta ed., México D.F., México, El Manual Moderno, 2013, pp. 339-360.
  - [5] R. Janssen *et al.*, "Video-based respiration monitoring with automatic region of interest detection", *Physiological measurement*, vol. 37, no. 1, pp. 100-114, 2015.
  - [6] X. García *et al.*, "Estimating cardiac output. Utility in the clinical practice. Available invasive and non-invasive monitoring", *Medicina Intensiva*, vol. 35, no. 9, pp. 552-561, 2011.
  - [7] C. Bustamante, W. Jara, M. Yapur, "Diseño y construcción de un contador de ritmo cardiaco, basado en un principio óptico", *Revista Tecnológica ESPOL*, vol. 15, no. 1, pp. 72-78, 2002.
  - [8] W. Abaya, J. Basa, M. Sy, A. Abada y E. Dadios, "Low cost smart security camera with night vision capability using Raspberry Pi and OpenCV", *Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2014 Conferencia Internacional en Palawan, 2014, pp. 1-4.
  - [9] A. Kaehler, G. Bradski, "Image Processing," en *Learning OpenCV 3*, 1ra ed., Boston, O'Reilly, 2016, pp. 109-129.
  - [10] A. Pajankar, "Working with Images, Webcams and GUI," en *Raspberry Pi computer vision programming*, 1ra ed., Birmingham, Reino Unido, Packt Publishing, 2015, pp. 30-36.
  - [11] X. Qu, J. Xu Yu y R. Brent, "A mobile TCP socket", Australian National University, Tech Rep. TR-CS-97-08, Abr. 1997.
  - [12] P. Viola y M. Jones, "Rapid object detection using a boosted cascade of simple features", *Proc. IEEE Computer Vision and Pattern Recognition*, 2001.
  - [13] E. Parra, "Aceleración del algoritmo de Viola-Jones mediante rejillas de procesamiento masivamente paralelo en el plano focal", Tesis Ingeniería, Dep. de Electrónica y Electromagnetismo, Univ. Sevilla, Sevilla, 2015.
  - [14] Mathworks. (2017, Septiembre). Detect corners using minimum eigenvalue algorithm and return corner Points object, [En Línea]. Disponible en: <https://www.mathworks.com/help/vision/ref/detectmineigenfeatures.html>
  - [15] G. Mur, "Inferencia de la respuesta afectiva de los espectadores de un video", Tesis Ingeniería, Dept. Sistemas Audiovisuales, Univ. Carlos III de Madrid, Madrid, 2015.
  - [16] F. Bourel, C. Chibelushi y A. Low, "Robust facial feature tracking", *Proc. 11th British Machine Vision Conf.*, vol. 1, pp. 232-241, 2000.
  - [17] Mathworks. (2017, Agosto). Track points in video using Kanade-Lucas-Tomasi (KLT) algorithm, [En Línea]. Disponible en: <https://www.mathworks.com/help/vision/ref/vision.pointtracker-system-object.html>
  - [18] W. Verkruyse, L. Svaasand y J. Nelson, "Remote plethysmographic imaging using ambient light", *Opt. Express*, vol. 16, pp. 21434-21445, 2008.
  - [19] H. Rahman, M. Ahmed y S. Begum, "Non-contact heart rate monitoring using Lab Color Space", *Proc. 13th International Conference on Wearable, Micro & Nano Technologies for Personalized Health (pHealth2016)*, pp. 43-56, 2016.
  - [20] J. Chávez, "Tratamiento digital de imágenes multiespectrales", 2da ed., México D.F., México, UNAM, México, 2010, pp. 273-326.
  - [21] B. Aldalur y M. Santamaría, "Realce de imágenes: filtrado espacial", *Revista de teledetección*, vol. 17, pp. 31-42, 2002.
  - [22] J. Cardoso, "Blind signal separation: statistical principles", *Proc. IEEE*, vol. 9, no. 10, pp. 2009–2025, 1998.
  - [23] J. Cardoso y A. Soloumiac, "Blind beamforming for non-Gaussian signals", *IEEE Proceedings*, pt. F, vol. 140, pp. 362-370, 1993.
  - [24] J. Miettinen, K. Nordhausen y S. Taskinen, "Blind source separation based on joint diagonalization in R: the packages JADE and BSSasymp", *Journal of Statistical Software*, vol 76, 2017, [En Línea]. Disponible en : <https://www.jstatsoft.org/index.php/jss/article/view/v076i02/v76i02.pdf>
  - [25] A. Ziehe, "Blind source separation based on joint diagonalization of matrices with applications in biomedical signal processing", Disertación Ph.D., Dept. Mat., Postdam Univ., Berlín, 2005.
  - [26] L. De la Fraga. (2001, Mayo). La transformada discreta de Fourier y la transformada rápida de Fourier, [En Línea]. Disponible en: [http://www.angelfire.com/falcon/shadow\\_rsv/t\\_senales/tdf.pdf](http://www.angelfire.com/falcon/shadow_rsv/t_senales/tdf.pdf)
  - [27] Equipamiento Médico Integral Peru. (2017, Septiembre). Manual Oxímetro MD300C21C, [En Línea]. Disponible en: [http://labmedical.com.pe/index.php?controller=attachment&id\\_attachme nt=12](http://labmedical.com.pe/index.php?controller=attachment&id_attachme nt=12)
  - [28] Mathworks. (2017). Fast Fourier transform, [En Línea]. Disponible en: [www.mathworks.com/help/matlab/ref/fft.html#buuuyt-6l](http://www.mathworks.com/help/matlab/ref/fft.html#buuuyt-6l).



**Karla Portilla** nació el 1 de septiembre 1993 en Quito, Ecuador. Estudió en la escuela San Francisco de Quito y colegio Manuela Cañizares, bachillerato general único. Estudia en la universidad "Escuela Politécnica Nacional", la carrera de Ingeniería en Electrónica y Control y es miembro de Club de Robótica.

Actualmente se encuentra realizando su proyecto de titulación.



**Víctor Santos** nació en Quito-Ecuador el 9 de octubre de 1991 en Quito, Ecuador. Realizó sus estudios secundarios en la Unidad Educativa Julio María Matovelle, donde obtuvo el bachillerato especialidad Físico Matemático. Estudia en la Escuela Politécnica Nacional la carrera de Ingeniería en Electrónica y Control, donde encuentra desarrollando su proyecto de



**María Trujillo** nació en Riobamba-Ecuador el 15 de febrero de 1990. Realizó sus estudios secundarios en el Colegio Santa Mariana de Jesús de su ciudad natal. Se graduó en la Escuela Politécnica Nacional como Ingeniera en Electrónica y Control en 2014. Obtuvo su título de Magíster en Ingeniería Biomédica en julio del 2016 en la Universidad Politécnica de Madrid. Durante sus estudios de maestría colaboró en el Centro de Tecnología Biomédica de la universidad en el área de bio instrumentación y nanotecnología. Actualmente desempeña el cargo de docente en el Departamento de Automatización y Control Industrial de la Escuela Politécnica Nacional.



**Andrés Rosales** Ingeniero en Electrónica y Control (Escuela Politécnica Nacional – EPN, Ecuador, 2001), y Doctor en Ingeniería en Sistemas de Control (Universidad Nacional de San Juan – UNSJ, Argentina, 2009). Investigador invitado (Universidad de Hannover, Alemania, 2007-2008). Actualmente, es Asesor del Vicerrector de Investigación y Proyección Social de la EPN. Es el Coordinador de Investigación de la Red Ecuatoriana de Universidades y Escuelas Politécnicas para Investigación y Posgrados – REDU. Ocupa el cargo de Profesor Principal del Departamento de Automatización y Control Industrial de la EPN.

# Implementación de un Modelo Afectivo para MASOES

## *Implementation of an Affective Model for MASOES*

Saúl Piña, Niriaska Perozo

**Resumen**—Diferentes autores buscan mejorar la interacción entre agentes inteligentes en sistemas multiagente, un ejemplo es el modelo afectivo de MASOES, el cual a través de la generación de emociones promueve diferentes tipos de comportamientos. Aunque este modelo afectivo ha sido verificado formalmente a nivel de diseño, no ha sido verificado a nivel de implementación. En este trabajo se presenta metodológicamente la implementación realizada para el modelo afectivo de MASOES sobre un sistema multiagente, verificando los requerimientos exigidos a nivel de diseño por esta arquitectura multiagente, con la finalidad de brindar un entorno para la interacción entre los procesos emocionales y las diferentes funciones de un agente. Adicionalmente, se propone el cálculo de la Emoción Social, permitiendo describir el estado emocional colectivo de un grupo de agentes emocionales.

**Palabras Clave**—Sistemas Multiagente, Computación Emocional, Modelo Afectivo, Interacción Emocional, MASOES.

**Abstract**—Different authors seek to improve the interaction between intelligent agents in multi-agent systems, an example is the affective model of MASOES, which through the generation of emotions promotes different types of behaviors. Although this affective model has been formally verified at the design level, it has not been verified at the implementation level. This work presents methodologically the implementation for the affective model of MASOES on a multiagent system, verifying the required at the design level by this multiagent architecture, in order to provide an environment for the interaction between the emotional processes and the different functions of an agent. In addition, we propose the calculation of Social Emotion, allowing to describe the collective emotional state of a group of emotional agents.

**Index Terms**—Multiagent System, Affective Computing, Affective Model, Emotional Interaction, MASOES.

### I. INTRODUCCIÓN

Se sabe que las emociones juegan un papel importante en el desarrollo de los seres humanos para fines sociales o de supervivencia [1], [2]. Un objetivo importante planteado por la comunidad científica es construir sistemas artificiales que exhiban comportamiento emocional, para mejorar la interacción hombre-máquina. Los procesos emocionales se han convertido en un requisito esencial en arquitecturas de agentes cognitivos,

S. Piña (e-mail: sauljabin@gmail.com) y N. Perozo (e-mail: npero@ucla.edu.ve) son parte de la Unidad de Investigación en Inteligencia Artificial, Decanato de Ciencias y Tecnología, Universidad Centroccidental Lisandro Alvarado, Lara, Venezuela.

se espera que el procesamiento afectivo mejore la calidad y la credibilidad de las respuestas emocionales generadas por los agentes.

En la actualidad no se conoce completamente los procesos cerebrales y mentales asociados a las emociones, sin embargo, se realizan esfuerzos para aplicar las teorías existentes en sistemas computacionales. Diferentes autores estudian modelos emocionales en sistemas multiagente, esto, con el objetivo de mejorar la interacción de los agentes y ayudar a la auto-organización y emergencia en dichos sistemas, además, incorporar emociones a agentes inteligentes es de utilidad, debido a que las emociones pueden hacer a los agentes más atractivos y creíbles para que puedan desempeñar un mejor papel en diversos sistemas interactivos que involucren simulación [3]. Un ejemplo es el modelo afectivo de MASOES (“*Multiagent Architecture for Self-Organizing and Emergent Systems*”, en inglés), propuesto en [4], es un modelo afectivo dimensional el cual considera un conjunto de emociones positivas y negativas que permiten generar un cambio dinámico de comportamiento en los agentes a nivel individual (Reactiva, Cognitiva) y colectivo (Imitativa). Este modelo afectivo para MASOES ha sido verificado a nivel de diseño [4], [5], por tal razón, en este trabajo se presenta metodológicamente la implementación realizada para el modelo afectivo de MASOES sobre un sistema multiagente. Dicha implementación puede usarse para realizar simulaciones de formación de comunidades, división de tareas, formación de flujos de peatones, sistemas robóticos, entre otras.

### II. TRABAJOS RELACIONADOS

En [6] se demuestra la aplicabilidad de los sistemas multiagente con modelos emocionales en un problema específico, en este caso sistemas multirobot (a nivel de hardware). La investigación usa como base el modelo MASOES propuesto por [4]. En su investigación proponen una arquitectura con tres niveles, el primer nivel individual proporciona los aspectos relacionados a la conducta y emociones del robot. El segundo nivel es colectivo, soporta los procesos de interacción de los robots. El último nivel, de los procesos de aprendizaje y gestión del conocimiento.

En [7], se propone extender el lenguaje de programación de agentes inteligentes llamado “2APL”, con el objetivo de

integrar emociones en este. 2APL es un lenguaje de programación lógico que fue desarrollado para apoyar la programación de sistemas multiagente. Entre otras cosas, este lenguaje le permite a los agentes recibir eventos y representar información propia, de otros agentes o del entorno. Esta investigación sirve de inspiración para el presente trabajo, ya que guarda similitudes con la arquitectura a nivel individual propuesta en MASOES, en la que se tiene una Base de Conocimiento Conductual, un Modelo Afectivo Dimensional y se procesan eventos, acciones u objetos para generar emociones. Además, sirve como base para justificar el uso del lenguaje de programación lógico *Prolog* en la implementación propuesta y así definir el conocimiento asociado a los agentes, emociones, comportamientos, eventos, acciones u objetos.

### III. ASPECTOS TEÓRICOS

#### A. MASOES

La arquitectura multiagente para sistemas emergentes y auto-organizados llamada MASOES, es una herramienta para el diseño no formal de sistemas, que produzcan un estado auto-organizado el cual emerja de las interacciones locales entre los agentes y de los cambios que se dan en el entorno. En esta arquitectura, cada agente puede cambiar su comportamiento dinámicamente, guiado por su estado emocional, para satisfacer los objetivos del sistema a través de la auto-organización de sus actividades. Ver más detalles en [4], [8], [9].

#### B. Componente Conductual de MASOES

El objeto de estudio del presente trabajo de investigación es el **Componente Conductual**, el cual favorece la adaptación de un agente con su entorno, ya que contiene un modelo afectivo interno que regula su comportamiento. Entre los elementos que lo conforman está el **Configurador Emocional** encargado de manipular las emociones del agente. También, está el **Manejador de Comportamiento**, que se encarga de activar, inhibir y priorizar algunos comportamientos en el agente basado en el estado emocional actual. El conocimiento asociado con la gestión de las emociones y comportamientos es almacenado en la **Base de Conocimiento Conductual**.

#### C. Modelo Afectivo de MASOES

Este modelo considera un conjunto de emociones positivas y negativas, para de esta manera promover un comportamiento individual (Reactivos, Cognitivo) o colectivo (Imitativo) en los agentes. Está representado por un espacio bidimensional, donde el eje *x* es el nivel de Activación, y el eje *y* representa el nivel de satisfacción, también en el intervalo [-1, 1] (Fig. 1).

El modelo afectivo está dividido en cuatro fases [4]: Fase I: Clasificación de las emociones; Fase II: Asociación de las emociones al tipo de comportamiento a través de las reglas en la Tabla I; Fase III: Determinación de la emoción actual; y Fase IV: Determinación del tipo de comportamiento.

### IV. ASPECTOS RELACIONADOS AL DISEÑO DE LA IMPLEMENTACIÓN

#### A. Modelo Afectivo

El modelo afectivo de MASOES está compuesto por 8 emociones, distribuidas en los 4 cuadrantes del plano cartesiano  $R^2$ , las cuales son representadas geométricamente como polígonos regulares o irregulares, como se muestran en la Fig. 2.

Con respecto a la selección de la emoción por parte del modelo afectivo al recibir un estado emocional, básicamente se verifica por cada emoción si el punto en el plano proporcionado por el estado emocional está contenido dentro del polígono de una emoción, posteriormente ésta será asignada al agente.

#### B. Base de Conocimiento Conductual

Está implementada en el lenguaje de programación lógico *Prolog* [10]. El conocimiento gestionado por la BCC puede ser de cuatro tipos: conocimiento sobre *el agente*, *las emociones*, *las reglas de prioridad de los comportamientos* y *los estímulos*. Con respecto al agente, en la Tabla II se observan las cláusulas que definen el conocimiento en relación a sí mismo y a los demás.

La Tabla III, muestra los hechos del conocimiento asociado a los tipos de emociones. La estructura de la regla para definir un tipo de emoción se compone del predicado, en este caso *tipo\_emocion* y los parámetros: emoción y tipo de emoción.

La base de conocimiento conductual también contiene las reglas definidas por MASOES en las que se asocia un estado emocional a un tipo de comportamiento. Así, para las reglas de asociación de comportamiento (Tabla I), se define el conjunto de cláusulas de la Tabla IV, donde E representa la emoción.

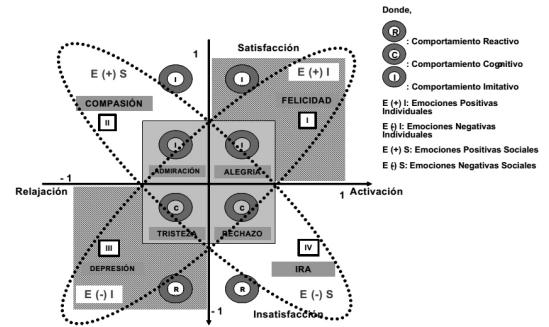


Fig. 1. Modelo Afectivo de MASOES

TABLA I  
REGLAS DE PRIORIZACIÓN DE COMPORTAMIENTOS

<b>Regla 1:</b>	Si el Estado Emocional es Positivo entonces priorizar Comportamiento Imitativo
<b>Regla 2:</b>	Sino Si el Estado Emocional es Ligeramente Negativo entonces priorizar Comportamiento Cognitivo
<b>Regla 3:</b>	Sino Si el Estado Emocional es Altamente Negativo entonces priorizar Comportamiento Reactivo

Para agregar estímulos a la BCC, es necesario que el agente asigne a cada uno de ellos un *Parámetro de Activación* ( $P_a$ ) y un *Parámetro de Satisfacción* ( $P_s$ ), estos valores incrementarán o decrementarán la activación y satisfacción del agente. La estructura para definir un estímulo se compone del predicado *estimulo* y los argumentos: nombre del agente, nombre del estímulo,  $P_a$  y  $P_s$ .

Uno de los aspectos importantes del presente trabajo es que le permite a la persona definir estímulos de manera sencilla y dinámica, además cabe destacar que deben estar asociados al sistema que se está modelando. Es necesario definir para cada estímulo que se está modelando con MASOES, cuanto afectará positivamente o negativamente al agente con respecto a la activación y satisfacción. Un ejemplo de esto puede ser establecido en Wikipedia, diferentes investigaciones sobre la reputación y confiabilidad en esta plataforma [11]–[13], concuerdan que no sólo los usuarios con mayor reputación son los más propensos a colaborar, sino que también generan contenido de mejor calidad. Entonces se pueden definir los siguientes estímulos:

Para el **Aumento de Reputación** (Crear Estímulos Positivos):

**Artículo Sobresaliente:** se podría utilizar como un estímulo para motivar y premiar las contribuciones sobresalientes de los Wikipedistas, por ende este estímulo, incrementaría de gran manera la activación y la satisfacción del agente, se le podrían colocar valores positivos altos a  $P_a$  y  $P_s$ , a fin de promover las emociones positivas y el comportamiento imitativo en los wikipedistas, según el modelo afectivo de MASOES.

Para el **Decremento de Reputación** (Crear Estímulos Negativos):

**Artículo Borrado:** este estímulo favorece las emociones negativas, ya que, este agente puede experimentar desánimo al sentir que su trabajo ha sido eliminado, se le podrían colocar valores negativos altos o moderados a  $P_a$  y  $P_s$ , promoviendo un comportamiento reactivo o cognitivo en los wikipedistas, según el modelo afectivo de MASOES.

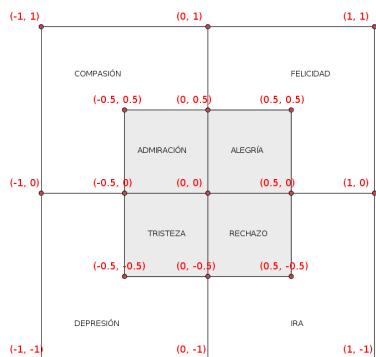


Fig. 2. Polígonos en el Modelo Afectivo

TABLA II  
CONOCIMIENTO RELACIONADO AL AGENTE EN LA BCC

Cláusula	Descripción
yo (agente) .	Definición del agente actual.
otro (A) :- not yo (A) .	Definición de otro agente.
	A representa el nombre del agente

De la misma manera se podría establecer los estímulos en otros sistemas, como puede ser en el desarrollo de software de código abierto, un agente puede experimentar un aumento de su activación y satisfacción si el código que ha propuesto es incluido en la rama principal del proyecto, lo que se entiende como un aumento de reputación, de lo contrario este agente se vería afectado negativamente, disminuyendo su activación y satisfacción. Las redes sociales son otro ejemplo de sistema colectivo, donde se puede definir tanto estímulos positivos como negativos, basados en el contenido creado: un estímulo positivo sería “Me gusta” y uno negativo “No me gusta”.

### C. Configurador Emocional

Este elemento tiene la responsabilidad de evaluar los estímulos y actualizar la emoción del agente. Para esto se definen las siguientes ecuaciones de actualización para la activación y la satisfacción:

$$A'(ag_i) = A_i + P_A \quad (1)$$

Donde  $A'(ag_i)$  representa el nuevo valor de activación del agente  $ag_i$ ,  $A_i$  es la activación actual y  $P_A$  se define como el parámetro de activación obtenido de la BCC, y está comprendido en el intervalo:  $-1 \leq P_A \leq 1$ .

$$S'(ag_i) = S_i + P_S \quad (2)$$

Donde  $S'(ag_i)$  representa el nuevo valor de satisfacción del agente  $ag_i$ ,  $S_i$  es la satisfacción actual y  $P_S$  es el parámetro de satisfacción obtenido de la BCC, y se encuentra en el intervalo:  $-1 \leq P_S \leq 1$ .

Para obtener el estímulo, el configurador emocional hará un cuestionamiento a la BCC:

?- estimulo(nombreAgente, nombreEstimulo, PA, PS) .

Las variables “PA” y “PS”, son las incógnitas a conocer.

TABLA III  
CONOCIMIENTO RELACIONADO A LAS EMOCIONES EN LA BCC

Cláusula	Descripción
tipo_emocion(admiration, positiva) .	La admiración es positiva
tipo_emocion(compasion, positiva) .	La compasión es positiva
tipo_emocion(felicidad, positiva) .	La felicidad es positiva
tipo_emocion(alegría, positiva) .	La alegría es positiva
tipo_emocion(rechazo, ligeramente_negativa) .	El rechazo es ligeramente negativa
tipo_emocion(tristeza, ligeramente_negativa) .	La tristeza es ligeramente negativa
tipo_emocion(ira, altamente_negativa) .	La ira es altamente negativa
tipo_emocion(depresion, altamente_negativa) .	La depresión es altamente negativa

TABLA IV  
CONOCIMIENTO RELACIONADO A LAS COMPORTAMIENTOS EN LA BCC

Cláusula	Descripción
prioridad_comportamiento(E, imitativo) :- tipo_emocion(E, positiva) .	El comportamiento es imitativo si la emoción E es positiva
prioridad_comportamiento(E, cognitivo) :- tipo_emocion(E, ligeramente_negativa) .	El comportamiento es cognitivo si la emoción E es ligeramente negativa
prioridad_comportamiento(E, reactivo) :- tipo_emocion(E, altamente_negativa) .	El comportamiento es reactivo si la emoción E es altamente negativa

#### D. Manejador de Comportamiento

El Manejador de Comportamiento se encarga de actualizar la prioridad del comportamiento a ejecutar por el agente, se basa en las reglas de prioridad definidas en MASOES y el estado emocional actual.

Tiene como entrada la emoción dada por el configurador emocional. Para obtener el tipo de comportamiento asociado a la emoción, este componente consulta la BCC, con el siguiente formato en lenguaje *Prolog*:

```
?- prioridad_comportamiento(nombreEmocion, TIPO).
```

Este cuestionamiento retorna el valor de la variable “TIPO”, que puede ser “imitativo”, “cognitivo” y “reactivo”.

#### E. Cálculo de la Emoción Social

Basado en lo definido en [14], a continuación se expone la propuesta para el cálculo de la *emoción social*, la cual se compone de tres valores: *emoción central*, *distancia máxima con respecto a la emoción central* y *dispersión emocional*.

La **Emoción Social** está representada por un conjunto de tres valores (Ecuación 3).

$$ES(Ag) = \{EC(Ag), m(Ag), \sigma(Ag)\} \quad (3)$$

Donde  $Ag$  representa al grupo de agentes en estudio,  $EC(Ag)$  se refiere a la emoción central exhibida por el grupo de agentes,  $m(Ag)$  es el estado emocional más alejado de la  $EC$ ,  $\sigma(Ag)$  representa la dispersión emocional entorno a la  $EC$ .

La **Emoción Central** dada por la Ecuación 4, se define como la emoción promedio (Ecuaciones 5 y 6) que exhibe un grupo de agentes  $Ag$ .

$$EC(Ag) = (\bar{A}(Ag), \bar{S}(Ag)) \quad (4)$$

Donde  $Ag$  representa al grupo de agentes en estudio,  $\bar{A}$  es el promedio de activación y  $\bar{S}$  el promedio de satisfacción del grupo en estudio.

$$\bar{A}(Ag) = \frac{\sum_{i=1}^n A_i}{n}, \forall ag_i \in Ag \quad (5)$$

$$\bar{S}(Ag) = \frac{\sum_{i=1}^n S_i}{n}, \forall ag_i \in Ag \quad (6)$$

Donde  $Ag$  representa al grupo de agentes en estudio,  $A_i$  es la activación y  $S_i$  la satisfacción del agente  $i$ , para  $1 \leq i \leq n$ .

La **Distancia Máxima** con respecto a la  $EC$  (Ecuación 7), permite saber si existen agentes con estados emocionales muy lejanos o cercanos a la emoción central. Se define como la distancia máxima euclíadiana (Ecuaciones 8 y 9) con respecto a la emoción central.

$$m(Ag) = (m_A(Ag), m_S(Ag)) \quad (7)$$

Donde  $Ag$  representa al grupo de agentes en estudio,  $m_A(Ag)$  es la activación más alejada (máxima activación) y  $m_S(Ag)$  es la satisfacción más alejada (máxima satisfacción).

$$m_A(Ag) = \max \left( \sqrt{(A_i - \bar{A}(Ag))^2} \right), \forall ag_i \in Ag \quad (8)$$

$$m_S(Ag) = \max \left( \sqrt{(S_i - \bar{S}(Ag))^2} \right), \forall ag_i \in Ag \quad (9)$$

Donde  $Ag$  es el grupo de agentes,  $A_i$  es la activación y  $S_i(Ag)$  la satisfacción del agente  $ag_i$ ,  $\bar{A}$  es el promedio de activación y  $\bar{S}(Ag)$  el promedio de satisfacción del grupo en estudio.

Para una mejor comprensión de la diversidad de emociones en el grupo de agentes, surge la **Dispersión Emocional** entorno a la  $EC$  representada por la Ecuación 10 y se define como la desviación estándar con respecto a la emoción central (Ecuaciones 11 y 12).

$$\sigma(Ag) = (\sigma_A(Ag), \sigma_S(Ag)) \quad (10)$$

Donde  $\sigma_A(Ag)$  es la desviación estándar de la activación y  $\sigma_S(Ag)$  es la desviación estándar de la satisfacción del grupo de agentes  $Ag$ .

$$\sigma_A(Ag) = \sqrt{\frac{\sum_{i=1}^n (A_i - \bar{A}(Ag))^2}{n}}, \forall ag_i \in Ag \quad (11)$$

$$\sigma_S(Ag) = \sqrt{\frac{\sum_{i=1}^n (S_i - \bar{S}(Ag))^2}{n}}, \forall ag_i \in Ag \quad (12)$$

Donde  $Ag$  es el grupo de agentes,  $A_i$  es la activación y  $S_i$  la satisfacción del agente  $ag_i$ , para  $1 \leq i \leq n$ ,  $\bar{A}(Ag)$  es el promedio de activación y  $\bar{S}(Ag)$  el promedio de satisfacción del grupo en estudio.

Si  $\sigma(Ag) \gg 0$ , el grupo tiene una alta dispersión emocional, es decir, los miembros del grupo tienen diferentes estados emocionales (muy heterogéneos).

Si  $\sigma(Ag) \cong 0$ , el grupo tiene una dispersión emocional baja, esto significa que los individuos tienen estados emocionales similares (muy homogéneos).

## V. DESCRIPCIÓN DE LA IMPLEMENTACIÓN

Se usó JADE (“Java Agent DEvelopment”, en inglés) [15], para el desarrollo de una herramienta computacional con el diseño propuesto. JADE es uno de los marcos de trabajo más populares para el desarrollo de agentes. Provee bibliotecas de clases para la creación de agentes mediante la herencia y la sobrescritura de comportamientos. Una de las características más importantes de JADE es que cumple con las especificaciones estándar FIPA (“Foundation for Intelligent Physical Agents”, en inglés), las cuales representan una colección de normas que tienen como objetivo promover la interoperabilidad de agentes heterogéneos y los servicios que pueden representar.

Especificamente, la herramienta permite: su uso interactivo a través de interfaces gráficas que se pueden desarrollar, configurar simulaciones con fines de investigación académica, exportar e importar configuraciones, generar gráficas en tiempo real y guardar los resultados de las simulaciones en archivos de texto plano. En la Fig. 3, se muestran las opciones de la interfaz gráfica, provee: 1) Una barra de menú con

opciones generales; 2) Sección para establecer los parámetros de cada agente en la simulación; 3) Información de los estados emocionales del grupo de agentes en tiempo de ejecución de la simulación; y 4) Una sección para controlar la simulación.

## VI. CONCLUSIÓN Y TRABAJOS FUTUROS

En este trabajo se abordó la implementación del modelo afectivo propuesto en MASOES, y por ende su componente conductual, el cual permite generar cambios dinámicos de comportamientos en los agentes emocionales, guiados por su estado emocional. Adicionalmente, se propone el cálculo de la Emoción Social de un grupo de agentes para evaluar las emociones a nivel colectivo, compuesta por un conjunto de tres valores: la Emoción Central, que determina la emoción que tienden a exhibir los agentes; la Distancia Máxima, determina los estados emocionales más alejados del grupo; y la Dispensión Emocional, que define la variación de estados emocionales del grupo, es decir, dicta si los estados emocionales son homegéneos o heterogéneos.

Con respecto a los aportes, este trabajo de investigación representa un marco de trabajo para realizar diversas simulaciones de grupos de agentes emocionales que pueden cambiar dinámicamente su comportamiento, guiado por su estado emocional a nivel individual y colectivo, ya que no sólo provee una implementación del modelo afectivo de MASOES, sino que también, proporciona la implementación del componente conductual completo. Además es importante destacar que la Base de Conocimiento Conductual se desarrolló con el lenguaje de programación *Prolog*, ampliamente utilizado por su versatilidad y facilidad de uso.

Como trabajo futuro, se debería extender este framework, implementando otros componentes de MASOES a nivel individual y colectivo, tales como: los componentes Cognitivo, Reactivo, Social, y la Base de Conocimiento Colectivo, todo esto con la finalidad de completar la implementación de esta arquitectura, y ser usada para modelar e instanciar un sistema emergente y auto-organizado de manera real y no sólo a nivel de diseño. Finalmente, en un próximo trabajo se evaluará la implementación realizada a través de la elaboración de diversos casos de estudios a nivel individual y colectivo, a fin de verificar los resultados obtenidos a nivel de diseño [4].

## REFERENCIAS

- [1] J. Cuevas, “Emociones: calculemos. las promesas de la computación afectiva,” Master’s thesis, Universidad Nacional de Educación a Distancia, Facultad de Filosofía, Madrid, España, 2015.
- [2] L. Rodríguez and F. Ramos, “Computational models of emotions for autonomous agents: major challenges,” *Artificial Intelligence Review*, vol. 43, no. 3, pp. 437–465, 2015.
- [3] H. Jiang, J. Vidal, and M. Huhs, “EBDI: an architecture for emotional agents,” in *Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*. ACM, 2007, pp. 38–40.
- [4] N. Perozo, J. Aguilar, O. Terán, and H. Molina, “An affective model for the multiagent architecture for self-organizing and emergent systems (MASOES),” *Revista Técnica de la Facultad de Ingeniería. Universidad del Zulia*, vol. 35, no. 1, pp. 1–11, 2012.
- [5] ———, “Self-organization and emergence phenomena in wikipedia and free software development using MASOES,” *Publicaciones en Ciencias y Tecnología. Universidad Centroccidental Lisandro Alvarado*, vol. 7, no. 1, pp. 51–72, 2013.
- [6] A. Gil, J. Aguilar, R. Rivas, E. Dapena, and K. Hernandez, “Architecture for multi-robot systems with emergent behavior,” in *Proceedings on the International Conference on Artificial Intelligence (ICAI)*. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2015, pp. 41–47.
- [7] M. Dastani, C. Floor, and J.-J. C. Meyer, “Programming agents with emotions,” in *Emotion Modeling*. Springer, 2014, pp. 57–75.
- [8] N. Perozo, J. Aguilar, and O. Terán, “Proposal for a multiagent architecture for self-organizing systems (ma-sos),” *Intelligence and Security Informatics*, pp. 434–439, 2008.
- [9] N. Perozo, “Modelado multiagente para sistemas emergentes y auto-organizados,” Ph.D. dissertation, Universidad de los Andes, Mérida, Venezuela, 2011.
- [10] (2017, sep) Prolog website. [Online]. Available: <http://www.swi-prolog.org/>
- [11] D. Anthony, S. W. Smith, and T. Williamson, “Reputation and reliability in collective goods the case of the online encyclopedia wikipedia,” *Rationality and Society*, vol. 21, no. 3, pp. 283–306, 2009.
- [12] S. Javanmardi, Y. Ganjisaffar, C. Lopes, and P. Baldi, “User contribution and trust in wikipedia,” in *5th International Conference on Collaborative Computing: Networking, Applications and Worksharing*. IEEE, 2009, pp. 1–6.
- [13] L. De Alfaro, A. Kulshreshtha, I. Pye, and B. T. Adler, “Reputation systems for open collaboration,” *Communications of the ACM*, vol. 54, no. 8, pp. 81–87, 2011.
- [14] J. Rincon, V. Julian, and C. Carrascosa, “Social emotional model,” in *Advances in Practical Applications of Agents, Multi-Agent Systems, and Sustainability: The PAAMS Collection*. Springer, 2015, pp. 199–210.
- [15] (2017, sep) Jade website. [Online]. Available: <http://jade.tilab.com/>

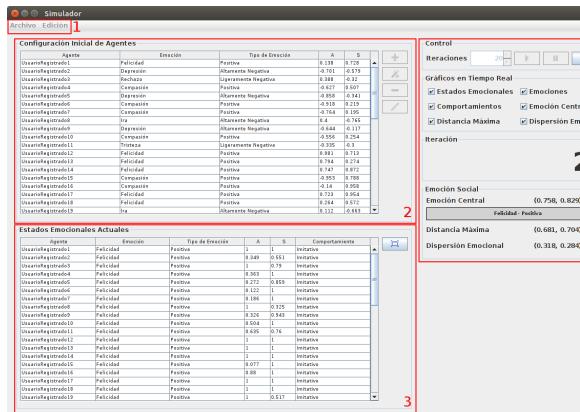


Fig. 3. Interfaz Gráfica Desarrollada para la Configuración de Simulaciones



**Saúl Piña** Es Ingeniero en Informática graduado en 2012 en la Universidad Centroccidental Lisandro Alvarado, Barquisimeto, Venezuela. M.Sc. en Ciencias de la Computación Mención Inteligencia Artificial en 2017 de la Universidad Centroccidental Lisandro Alvarado, Barquisimeto, Venezuela. Actualmente se desempeña como desarrollador de la empresa ThoughtWorks Ecuador.



**Niria Perozo** Es Ingeniero en Informática graduada en 1997 en la Universidad Centroccidental Lisandro Alvarado, Barquisimeto, Venezuela. M.Sc. en Ciencias de la Computación en 2004 de la Universidad de los Andes, Mérida, Venezuela. Doctora en Ciencias Aplicadas (Universidad de los Andes, Mérida, Venezuela) y Neurociencia, Cognición y Comportamiento Colectivo (Universidad Paul Sabatier, Toulouse, Francia) en 2011. Actualmente es Profesora e Investigadora en Inteligencia Artificial en la Universidad Centroccidental Lisandro Alvarado, Venezuela.

# Vibratory Neurostimulator for Patients with Essential Tremor

Delgado G., Gavilanes J., Mena L., Encalada P.

**Abstract**— The present research shows a prototype that seeks the suppression of essential tremor in a neuroprosthesis that emits vibrations to saturate the channels of communication of this pathology, initially presents a general concept of the disease, and the treatments that seek to correct this symptom known as tremor. It was essential to follow the general outlines of the design and development of the device, and subsequently the results obtained and conclusions that validate the hypotheses raised, the authors propose a series of works that validate the system and the hypotheses itself.

**Index Terms**—Essential tremor, neuroprosthesis, Pacini corpuscles, , upper extremities, vibratory actuators.

## I. INTRODUCTION

A tremor is an abnormal and involuntary type of movement characterized by excess movement that may interfere with the individual's ability to perform everyday tasks such as eating, writing, and so on. It can affect various parts of the human body such as the hands, head, torso and legs, [3]. Within the numerous disorders of the body movement, tremors are the most common form, they are a frequent occurrence in elderly people, sometimes being part of a helpless and incapacitated image of old age, a tremor must be assumed as a situation of a disease and not as part of aging. Among tremors, essential tremor is the most frequent, and is produced by specific alterations of the cerebellum, which is where the disorder is generated,[1]. Essential tremor should not be confused with Parkinson's disease, although there may be progression. To date there is no scientific evidence to prove its complete cure, either with pharmacological treatment, surgical or brain stimulation, [13]. The primary goal of all treatments is to maintain the patient's basic functions so that he or she can have as normal and independent a life as possible while reducing the disorder. If we consider that drugs used are not useful in 50% of patients and that the patient's tolerance to drugs decreases with age, along with the impossibility of applying invasive therapies in elderly patients, demonstrates that research and new more

effective therapeutic options are needed to stop the tremor, [18]. That is why this research proposes the development of a system capable of supplying vibratory waves to the sensory system of the human body, which connects with the cerebellum, and with this it aims to reduce the Essential Tremor (E.T.) in the upper extremities of the patient. It should be noted that the system is based on a novel hypothesis and the goal of this work is to develop a platform that allows us to validate and evaluate the most adequate values of amplitude and frequency of mechanical vibration, so that the signal transmitted from the mechanoreceptors to the central nervous system is as effective as possible in suppressing the tremor.

### A. Hypothesis

The description of the pathophysiology of Essential Tremor, which is more accepted by the medical community, shows that the essential tremor could be caused by the anomalous and maintained activation of the Inferior Olivary Nucleus (ION), which can be translated into pathological oscillations in the olive- cerebellum-thalamo-cortical during movement, being able to be responsible for the postural and kinetic tremor observed in the extremities in patients with E.T., [19].

Taking into account that ION can be inhibited by stimulation of the cuneiform nucleus, the first data station of the tactile information, the possibility of continuous Pacini receptor stimulation at high frequencies (vibratory sensitivity using a wide range of frequencies)) could to produce a stimulation of the cuneiform nucleus capable of suppressing ION activity, [20] thereby changing the pathological oscillations of ION and thereby producing a decrease or abolition of postural and kinetic tremor observed in patients with E.T.

This finding is in agreement with a recent work on the restoration of locomotion in rodent models through spinal cord stimulation. Therefore, current evidence supports the idea that stimulation of afferent pathways may alleviate the symptoms of tremors, [21].

## II. THE ESSENTIAL TREMOR

Essential tremor (E.T.) is a neurodegenerative disease [1], that affects mainly adults older than 65, [2]; this disease manifests itself with a postural tremor, which most of the time is symmetrical and involves the hands and forearms [3]

Around 90% of patients present tremor in the upper limbs from the beginning, although it may affect other parts of the anatomy

G. Delgado is a professor at Universidad del Azuay, Ecuador (e-mail: gabrieldelgado@uazuay.edu.ec)

J. Gavilanes is a professor at Escuela Superior Politécnica de Chimborazo, Ecuador (e-mail: javier.gavilanes@espoch.edu.ec)

L. Mena is a professor at Escuela Superior Politécnica del Litoral, Ecuador (e-mail: lismena@espol.edu.ec)

P. Encalada is a professor at Universidad Técnica de Ambato, Ecuador (e-mail: pg.encalada@uta.edu.ec)

such as the head, voice, lower limbs, jaw and, less frequently, the trunk or tongue, [5]. Koller, [6] emphasizes in his studies that patients suffering from essential tremor may be considered handicapped either by physical limitations or by the social disorder that the tremor can cause. Essential tremor is usually less severe when compared to parkinsonian tremor, however, it often causes more social dysfunction, [12].

#### A. Pathophysiology of tremors

The pathophysiological processes involved in the origin of the tremor are mostly speculative; however, neurophysiological studies based on electromyography have been performed on the affected muscles and by means of accelerometry and the application of computer technology have contributed in a relevant way to our knowledge about the physical characteristics of the tremor.

There is a central generator whose neurons discharge spontaneously in a rhythmic way and consequently produce tremor. This has been seen in neurophysiological studies with an intraneuronal record in the lower olive, the ventral thalamus, the subthalamic nucleus and in the inner pallid globe [7].

The mechanical component of the tremor depends on the physical properties of the oscillating limb because its mass and inertia provide a resonance frequency. The frequency range of essential tremor is characterized differently among researchers. Jain, [8] states that the range of the essential tremor is between 5 and 8 Hz while Cooper from 4 to 12 Hz, [9], and Elble from 3 to 11 Hz, [10].

#### B. Treatments

There is no drug capable of solving the problem of each patient, [12]. On the contrary, some drugs may work well in some patients and not be effective in others, also we should point out that many scientific research institutes are trying to find a solution that does not require surgery or medication.

#### C. Technological assistance to the superior members of the body.

Some studies show that, in the long term, medical treatment fails in 85% of cases. g, has influenced the growth of biomechanical load application research as a treatment method for the reduction of pathological tremor, a series of research shows that the increase in inertia or damping of the system could reduce an involuntary shaking movement, [13].

Among the recently investigated tremor reduction systems is the WOTAS (Wearable Orthosis for Tremor Assessment and Suppression) exoskeleton Fig. 1. The WOTAS platform was designed for the elbow and wrist joints, to allow the application of different control strategies on the patient.

### III. NEUROPROSTHESIS CONCEPTUAL DESIGN

For this study, it is necessary to know the corpuscles found in the skin, since it is innervated by different types of independent nerve endings and receptors that recognize stimuli, facilitating skin as a sensorial organ.

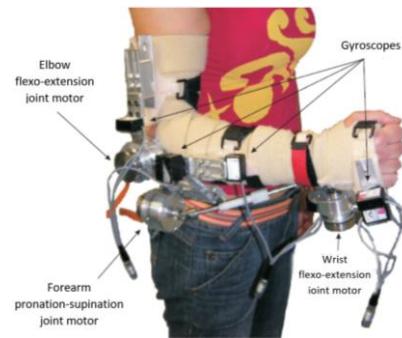


Fig. 1 WOTAS Exoskeleton.

Pacini corpuscles and Meissner corpuscles located in the subcutis are sensitive to mechanical vibrations and deformations, and detect different frequency ranges. Once stimulated, they transmit sensory information through the posterior cord to the cuneiform nucleus, where relay neurons transmit this information to the sensory thalamus and from there to the somatosensory cortex, [14]. The cuneiform nucleus contains GABAergic neurons projecting towards the inferior olfactory nucleus [17], which hypothetically reduces the amplitude of the tremor. Pacini's corpuscles detect a range of frequencies from 80Hz to 450Hz, while Meissner's detect low frequencies between 10Hz and 80Hz.

#### A. System Architecture

The vibration system proposed for the upper limb of a patient consists of four fundamental stages: 1) Stage of Action in charge of generating the oscillatory movement at frequencies perceptible by the corpuscles; 2) Control stage, capable of producing different waves and amplitudes by the user; 3) Monitoring Stage, which will be the agent that registers the trembling movement of the patient and finally: 4) Software Stage, which joins all the stages through a graphic interface (HCI) that facilitates the handling by the operator.

##### 1) Actuation stage

The actuation is the main stage for the operation of the system, since it is responsible for generating the vibrations as close to the dermis as possible so they can be captured by the corpuscles of Pacini. It is also responsible for the selection of the type of control and software for its handling. The devices responsible for transforming energy to vibration are called piezoelectric type actuators, which give a variation of position or movement, proportional to the voltage when a potential difference is applied to them. [15] (Fig. 3), but in fine variations, making this device very sensitive and precise, useful for the system requirements, being able to work in open loop by its linearity.

##### 2) Control architecture

Considering the piezoelectric actuator, it is possible to use a driver for these actuators, which raises the input voltage according to its set gain configuration. In turn, Texas Instruments has a modifiable evaluation module for the driver called DRV8662EVM. These characteristics and their reduced size make it suitable for this work, so that the DRV8662 chip will allow designing and building a new card to control the vibratory actuators in a right way.

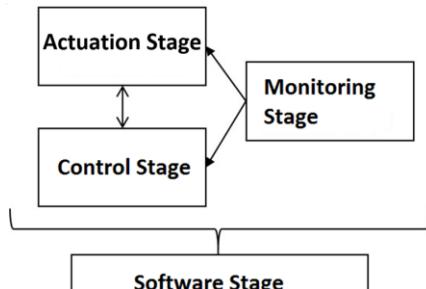


Fig. 2 System Architecture

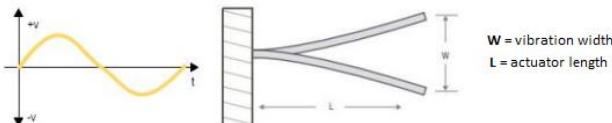


Fig. 3 Voltage Vs. Piezoelectric Actuator Displacement

### 3) Monitoring stage

The monitoring stage is covered by IMU inertial sensors of the company TECHNAID, which facilitates monitoring both the patient's tremor and the effect of the piezoelectric actuators on the patient, from which only two will be used in the monitoring of the patient's activity with E.T.

Essential tremor typically affects both upper extremities, particularly hands and forearms. It is always accompanied by a rhythmic tune of discharge from the motor units that forces the body to go into oscillation.

According to studies and experiments of some experts (Koller, Jainda, Cooper, Elble, Deuschl) it is concluded that the frequency of essential tremor differs in each patient, however, the range varies from 3 to 15Hz. Therefore, using software a band pass filter has been designed for the reading of only the values where the E.T is present. In such a way that it is possible to evaluate the usability of the prototype in patients.

The stages analyzed synthesize the conceptual design of the system, as shown in Fig. 4.

## IV. DEFINITION OF VALIDATION PROTOCOL

In Table I some experimental protocols of system validation are proposed, as well as the description of the process.

A clinical validation of the concept should also be carried out where typical neurological examinations are considered as:

- Resting arm in the lap
- Keeping both arms extended
- Movement of the finger towards the nose.

Each one of these tasks has 60 seconds monitoring time, while the patient is sitting comfortably.

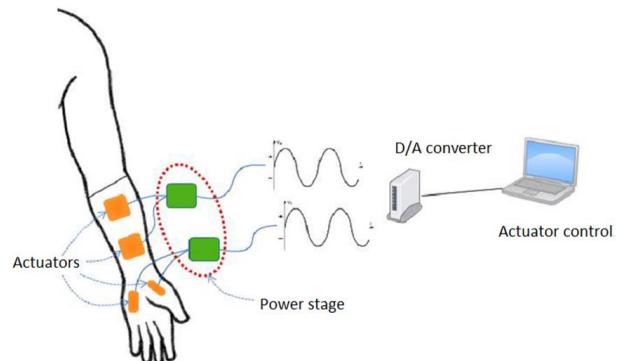


Fig. 4 Detailed conceptual design.

The garment will work in two ways to validate the process:

- Simulated Mode. - Frequencies generation for therapeutic purposes lower than 60Hz
- Verum Mode. - Therapeutic frequencies generation between 60-400 Hz, where the exact frequency is given by the experimental procedures named in Table 1.

In each session, 3 people will be present:

- Patient
- System operator
- Medical specialist

During the trial neither the patient nor the physician will know when the system is in simulated or Verum mode, with the purpose of avoiding the placebo effect of the system on the patient [16].

The complete development of the neuroprosthesis was performed, with each one of its stages. It is important to point out that this is a prototype capable of analyzing the effects produced by the vibration on the mechanoreceptors of the patient. Fig. 5 shows a photograph of the attachment of the vibratory actuators on the surface of the patient's arm, a garment was used for holding the entire palm of the hand as shown in the figure.

During the placing of the neuroprosthesis on a patient with E.T, protocols were carried out to validate the hypothesis, a measurement was established with inertial sensors placed strategically on the wrist and the patient's hand; Fig. 6 shows the patient's tremor signal in the time domain.

TABLE I  
EXPERIMENTAL PROTOCOLS

Number	Experiment	Description
1	Operation of actuators in the hand	The frequency and amplitude are chosen, which can be varied to evaluate the patient's reaction.
2	Operation of actuators in the forearm.	Like method 1, frequency, amplitude and time of operation are varied, also applying frequency sweep.
3	Operation of actuators in the hand and forearm at the same frequency.	The actuators operate at the same frequency in two different parts of the arm.
4	Operation of actuators in the hand and in the forearm at a different frequency	The actuators operate at different frequencies in two different parts of the arm.

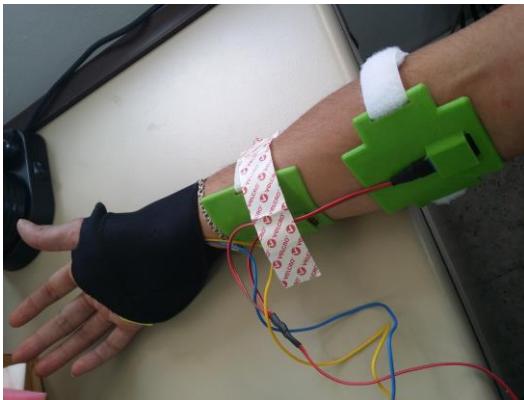


Fig. 5 Attachment of vibratory actuators.

## V. PRESENTATION OF RESULTS

After receiving and validating the E.T, a vibration frequency signal is sent in Crescendum mode (Ramp Up), from 70 to 350 Hz in 60 seconds; Fig. 7 takes the E.T. data while the patient receives that vibration. At the end, the patient enters a rest period of 30 seconds without monitoring the activity of Essential Tremor (E.T) and after this time the patient's data is taken again in 60 seconds to perform an overlap between E.T Preprocess and E.T Postprocess (Fig. 8).

As in the case analyzed, multiple vibration frequencies were sent to the patient, the next figures (Fig. 7 to Fig. 12) show the most significant frequency graphs for the final analysis on the patient.

In order to show a correct final analysis of the process, the patient's essential tremor graphs are superimposed, before and after the process, obtaining a reduction of the tremor in the patient of 61.43%.

In Fig. 13 the analysis is better appreciated, where the Preprocess is in blue and the Postprocess is in orange.

## VI. CONCLUSIONS

During the development of this work, a prototype of a neuroprosthesis was built, which was capable of supplying vibrations on the skin of a patient with Essential Tremor, for this purpose the best components were selected in the corresponding stages:

**Actuation Stage:** The piezoelectric vibration actuators have a very wide range of operation, which was suitable to obtain frequencies necessary to be detected by the Pacini corpuscles, which have a direct connection with the inferior olivary nucleus that controls the motor functions and generates the tremor.

**Control Stage:** Printable materials with 3D technology allowed to design and make subjections with better transfer of the vibration towards the skin of the patient, without being a nuisance.

**Monitoring Stage:** From inertial sensors, it was possible to obtain a quantitative feedback of the effect that the vibration produces on the patient.

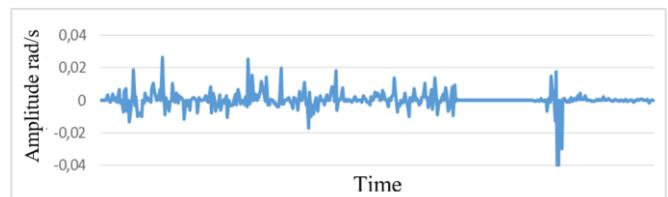


Fig. 6 Initial Tremor Preprocess of the patient in 60 seconds



Fig. 7 Patient trembling in 60 seconds at 200Hz.

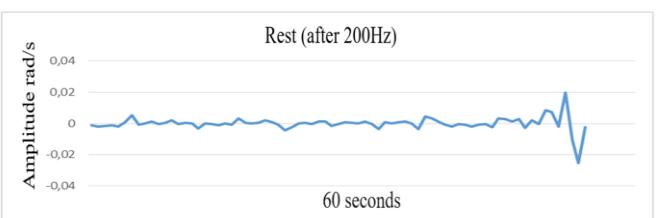


Fig. 8 Patient trembling in 60 seconds at rest.

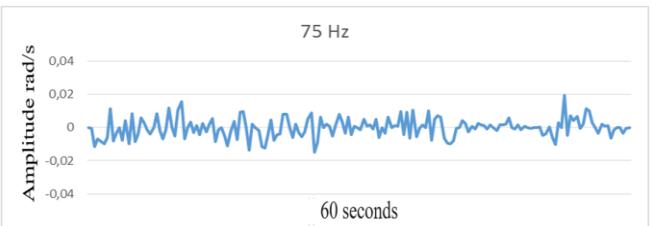


Fig. 9 Patient trembling in 60 seconds at 75Hz.

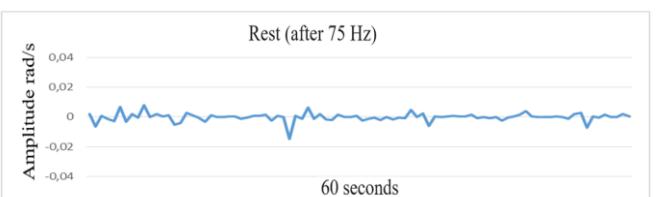


Fig. 10 Patient trembling in 60 seconds at rest.

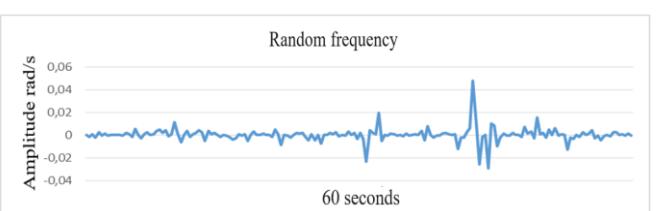


Fig. 11 Patient trembling in 60 seconds at random frequency



Fig. 12 Patient trembling in 60 seconds during postprocess

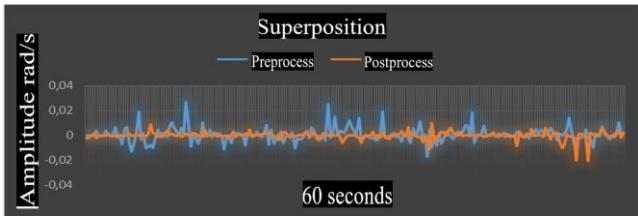


Fig. 13. Patient's tremor in 60 seconds during Pre-processing and Post-processing

**Software Stage:** An experimental protocol was presented for the validation of the neuroprosthesis capable of varying its frequency in an automatic way to observe the effects that said frequency exerts on the patient.

**Validation stage:** The experimental protocol gave us a small sample of what the presented hypothesis of suppression of tremor in a patient with E.T can validate, reducing 61.43% of the tremor.

## VII. FUTURE WORK

Conduct research on different vibration frequencies on patients with E.T and obtain which are the best for each person.

Design and build flexible piezoelectric actuators capable of being placed in clothes to improve the patient's comfort.

Carry out an analysis of the neuroprosthesis for a larger group of patients in order to obtain an accurate validation of the hypothesis, since it is still premature to validate it with a single patient.

## REFERENCES

- [1] U. Bonuccelli, "Essential tremor is a neurodegenerative disease," *J. Neural Transm.*, vol. 119, no. 11, pp. 1383-7; discussion 1373, Nov. 2012.
- [2] E. D. Louis, "Essential tremor," *Lancet Neurol.*, vol. 4, no. 2, pp. 100-10, Feb. 2005.
- [3] G. Deuschl, P. Bain, and M. Brin, "Consensus statement of the Movement Disorder Society on Tremor. Ad Hoc Scientific Committee." *Mov. Disord.*, vol. 13 Suppl 3, pp. 2-23, Jan. 1998.
- [4] E. D. Louis and M. S. Okun, "It is time to remove the 'benign' from the essential tremor label," *Parkinsonism Relat. Disord.*, vol. 17, no. 7, pp. 516-20, Aug. 2011.
- [5] K. L. Busenbark, J. Nash, S. Nash, J. P. Hubble, and W. C. Koller, "Is essential tremor benign?", *Neurology*, vol. 41, no. 12, pp. 1982-3, Dec. 1991.
- [6] Elble, Roger J., and William C. Koller. *Tremor*. Johns Hopkins University Press, 1990.
- [7] S. E. Hua, F. a Lenz, T. a Zirh, S. G. Reich, and P. M. Dougherty, "Thalamic neuronal activity correlated with essential tremor." *J. Neurol. Neurosurg. Psychiatry*, vol. 64, no. 2, pp. 273-6, Feb. 1998.
- [8] S.S. Jain y S.C. Kirshblum. Movement disorders, including tremors. In A. Joel y J.B. DeLisa, editors, *Rehabilitation Medicine: Principles and practice*, volume 2. Lippincott Company: Philadelphia, 1993.
- [9] G. Cooper y R. Rodnitzky. The many forms of tremor. Precise classification guides selection of therapy. *Postgraduate Medicine online*, July 2000.
- [10] R.J. Elble. Essential tremor frequency decreases with time. *Neurology*, 2000.
- [11] Marsden CD, Origins of normal and pathologic tremor. En: Findley LJ, Capildeo R (Eds). (1984). *Movement Disorders: Tremor*. New York: Oxford University Press, pp 37-84. En: Fahn S, Jankovic J (Eds). (2007).
- [12] J. Benito-León and E. D. Louis, "Essential tremor: emerging views of a common disorder," *Nat. Clin. Pract. Neurol.*, vol. 2, no. 12, Dec. 2006.
- [13] Kotovsky J and Rosen MJ. A wearable tremor suppression orthosis, *J Rehab Res Develop*, 1998.

- [14] E. Kandel, J. Schwartz, and T. Jessell, *Principles of Neural Science*. 2000.
- [15] CURIE, Jacques; CURIE, Pierre. Développement, par pression, de l'électricité polaire dans les cristaux hémièdres à faces inclinées. *Comptes Rendus*, 1880, vol. 91, p. 294-295.
- [16] E. Rocon, J. M. Belda-Lois, A. F. Ruiz, M. Manto, J. C. Moreno, and J. L. Pons, "Design and validation of a rehabilitation robotic exoskeleton for tremor assessment and suppression," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 3, pp. 367-378, Sep. 2007.
- [17] A. R. Gibson, K. M. Horn, and M. Pong, "Inhibitory control of olivary discharge. *Ann. N. Y. Acad. Sci.*, vol. 978, pp. 219-31, Dec. 2002.
- [18] J. Benito-León, "Essential tremor: one of the most common neurodegenerative diseases?" *Neuroepidemiology*, vol. 36, no. 2, pp. 77-8, Jan. 2011.
- [19] K. E. Zeuner and G. Deuschl, An update on tremors., *Curr. Opin. Neurol.*, vol. 25, no. 4, pp. 475-82, Aug. 2012.
- [20] P. Geborek, H. Jörntell, and F. Bengtsson, Stimulation within the cuneate nucleus suppresses synaptic activation of climbing fibers., *Front. Neural Circuits*, vol. 6, no. January, p. 120, Jan. 2012.
- [21] R. Fuentes, P. Petersson, and M. A. L. Nicolelis, Restoration of locomotive function in Parkinson's disease by spinal cord stimulation: mechanistic approach., *Eur. J. Neurosci.*, vol. 32, no. 7, pp. 11008, Oct. 2010.



**Delgado G.** was born in Ecuador, Cuenca City, in 1990. He received the B.S. in Electronics Engineering from Universidad del Azuay, Cuenca - Ecuador in 2014 and the MSc. degree in Automation and Robotics from University Polytechnic of Madrid, Spain in 2015. In 2014 -2015 He was a Research Assistant in Group of Neural and Cognitive Engineering from Centre for Automation and Robotics (CSIC-UPM) in Spanish National Research Council. Since 2016, He has been professor in Universidad del Azuay, Cuenca - Ecuador and he has been Research. His research interests include, artificial intelligence, artificial vision, manipulators robot, mobile robots, machine learning and control systems.



**Gavilanes J.** was born in Ecuador, Santo Domingo City, in 1987. He received the B.S. in electronics engineering from Escuela Superior Politécnica del Chimborazo, Riobamba Ecuador in 2012 and the MSc. degree in automation and robotics from University Polytechnic of Madrid, Spain in 2015. In 2014 -2015 He was a Research Assistant in Field & Service Robotics Group from Centre for Automation and Robotics (CSIC-UPM) in Spanish National Research Council. He collaborated in the Tiramisu European project. Since 2016, He has been professor in Escuela Superior Politécnica del Chimborazo, Riobamba Ecuador and he has been Research Assistant with Biomechanics Studies Research Group. His research interests include artificial vision, biomechanics, mobile robots and additive manufacturing.



**Mena L.** was born in Ecuador, Ambato City, in 1987. She received the B.S. in electronics engineering from Salesian Polytechnic University, Quito Ecuador in 2012 and the MSc. degree in automation and robotics from University Polytechnic of Madrid, Spain in 2015.

In 2014 -2015 she was a Research Assistant in Field & Service Robotics Group from Centre for Automation and Robotics (CSIC-UPM) in Spanish National Research Council. She collaborated in the Tiramisu European project. Since 2016, she has been professor with the Electrical and Computer Engineering Department, Escuela Superior Politécnica del Litoral, Guayaquil Ecuador and she has been Research Assistant with Vision and Robotics Center. Her research interests include locomotion systems, mobile robots and additive manufacturing.



**Encalada P.** was born in Ecuador, Ambato City, in 1986. He received the M.S.c degrees in Automatic and Robotic from the University Politehnica of Madrid, España, in 2016 and the degree in engineering of Electronic and Instrumentation from ESPE, Ecuador, in 2012.

From 2015 to 2016, he was a Research Assistant with the CAR of the University Politehnica of Madrid.

Since 2013, he has been an Professor with the Electrical and Comunication and Industrial Engineering Department of the Universidad Técnica de Ambato.

He is the author of one chapter of a Robocity 2016 and his research interests include robotic systems, machine learning and control systems.

# Diseño e Implementación de una Plataforma Experimental de Mano Robótica

## *Design and Implementation of a Robot Hand Experimental Platform*

Washington Caraguay, Marco Sotomayor, Miguel Alonso

**Resumen**—En el presente artículo, se describe el diseño e implementación de una plataforma experimental de mano robótica con la finalidad de crear un entorno real para pruebas de señales que hayan sido tratadas con diferentes técnicas aplicadas o dispositivos utilizados en la extracción de señales bioeléctricas. Mecánicamente la plataforma dispone de una mano de cinco grados de libertad, uno para cada dedo. El movimiento de cada dedo lo genera un servomotor, el mismo que recibe la señal desde una tarjeta multiplexora electrónica. La recepción de señales desde el exterior es tanto alámbrica como inalámbrica. Para comprobación del movimiento de manera individual de los dedos, se ha provisto de un joystick, un teclado y un dispositivo bluetooth. El movimiento de las articulaciones en cada dedo se logra al tensar un hilo nylon desde la polea ubicada en el eje del servomotor. Se utilizó señales PWM para la comprobación del movimiento de los dedos.

**Abstract**—In this article, we describe the design and implementation of an experimental robotic hand platform in order to create a real environment to tests signals that have been treated with different applied techniques or devices used in the extraction of bioelectrical signals. Mechanically the platform has a hand of 5 degrees of freedom, one for each finger. The movement of each finger is generated by a servomotor, which receives the signal from an electronic distribution board. The reception of signals from the outside is both wired and wireless. To check the individual movement of the fingers, a joystick, a keyboard and Bluetooth device have been provided. The movement of the joints on each finger is achieved by tightening a nylon thread from the pulley located on the servomotor shaft. PWM signals were used to check the movement of the fingers.

**Index Terms**—robotic hand, experimental platform, test signal, degrees of freedom, PWM signals.

### I. INTRODUCCIÓN

Este trabajo forma parte del proyecto de investigación “Inclusión educativa y laboral de personas con capacidades especiales mediante la ayuda de Tecnología Asistiva”, el mismo que es financiado por el Centro de Investigaciones de la Universidad Espíritu Santo-Ecuador.

Washington Caraguay, Universidad Espíritu Santo-Ecuador (e-mail: wcaraguay@uees.edu.ec).

Marco Sotomayor, Universidad Espíritu Santo-Ecuador (e-mail: mvinicio@uees.edu.ec).

Miguel Alonso, Universidad Espíritu Santo-Ecuador (e-mail: aalonso@uees.edu.ec).

N los últimos registros del Consejo Nacional para la Igualdad de Discapacidades (CONADIS), en Ecuador se encuentran registradas 199.284 personas con discapacidad física [1]. Dentro de este grupo se encuentran aquellas personas que han sufrido amputaciones tanto de brazos como de manos. La imposibilidad que representa para una persona realizar acciones como tocar, dejar y recoger objetos, entre otras acciones es evidente debido a la falta del miembro superior. Sin embargo, del total de personas registradas, solamente 31.924 se encuentran incluidas laboralmente.

A nivel mundial, existen avances científicos considerables en el diseño de prótesis de manos robotizadas con la finalidad de lograr aprovechar señales bioeléctricas propias del cuerpo humano. Varios métodos para el control de las prótesis, así como diferentes enfoques en la obtención y tratamiento de las señales han sido considerados, entre los cuales se estudian algoritmos genéticos como entradas para un clasificador simple [2], y técnicas de visión por computador para controlar la posición del brazo [3]. Carrozza et. al, propone movimientos de la mano controlados por los movimientos de los dedos del pie, demostrando una alta efectividad en el control [4]. Además, se ha investigado el uso de técnicas como Fast Orthogonal Search (FOS), para la estimación de la fuerza de la muñeca inducida por el codo por medio de señales EMG [5]. Castenelli et. al, 2008 aborda la fuerza necesaria que debería ser aplicada en el cierre de la mano, para lo cual aplica técnicas de machine learning tales como redes neuronales y support vector machine [6]. Harada et.al, 2010, discuten el diseño de una mano robótica, la cual utiliza redes neuronales para el reconocimiento de la operación de cada dedo, diferenciando como opera cada dedo de la mano de manera independiente [7]. Las manos robotizadas multiagarré son de gran interés en la comunidad científica y los esfuerzos en su desarrollo han logrado resultados satisfactorios [8], sin embargo la aparición de nuevos materiales y técnicas de inteligencia artificial, motivan a experimentar con diferentes enfoques.

En el Ecuador, varias universidades están contribuyendo con incrementar el estado del arte en el desarrollo de manos robotizadas. A continuación se exponen algunos desarrollos:

En los años 2009, en la Escuela Politécnica Nacional, se diseñó e implementó una mano robótica capaz de emular el movimiento de una mano humana, controlada por un guante con sensores a través de comandos desde un computador. La mano provista de cinco dedos, con tres grados de libertad cada

uno, emula los movimientos de flexión de los dedos, cuyas dimensiones son muy parecidas a las de una mano humana. Al guante de mando se le acoplaron quince sensores fotoeléctricos, uno en cada articulación que conforman los dedos de la mano humana. Se implementó un controlador PID para cada articulación, el mismo que toma los datos proporcionados por el guante o el computador, los compara con los valores de posición de la mano, ejecuta el algoritmo de control y un microcontrolador se encarga de transformar las señales a PWM. Pruebas realizadas mostraron que la diferencia entre el ángulo real de las articulaciones de los dedos y aquellas en las que se posiciona la mano robótica, difieren con un error de aproximadamente  $\pm 1.55\%$  cuando es controlado por el guante, y  $\pm 0.46$  cuando es controlado por el computador [9].

En el año 2012, en la Universidad Politécnica Salesiana de la ciudad de Cuenca, se desarrolló un prototipo de mano robótica activada por señales electromiográficas (EMG). El prototipo dispone de 4 grados de libertad y consiste en el movimiento de 3 dedos y rotación de la muñeca, comandado por pulsos recibidos de señales EMG. Cada dedo tiene un sensor fuerza para controlar la presión ejercida por la mano sobre los objetos. Las pruebas de funcionalidad realizadas fueron exitosas. Sin embargo, el usuario debe aprender a coordinar sus movimientos [10]. En el año 2014, estudiantes de la Escuela de Electrónica y Telecomunicaciones de la Universidad Técnica Particular de Loja, desarrollan el prototipo “Hand of Hope, Una mano de Esperanza”, el mismo que trata de un Proyecto para reproducir prótesis de extremidades superiores a precios reducidos. La prótesis desarrollada hace uso de bioseñales para su control y fue confeccionada con biopolímeros plásticos [11]. Al año siguiente, la Escuela Superior Politécnica del Ejército, diseña una prótesis biónica de mano derecha de siete grados de libertad, utilizando materiales inteligentes y control mioeléctrico, adaptada para varios patrones de sujeción. El objetivo principal de este proyecto era permitir realizar movimientos básicos de agarre para la ejecución de tareas cotidianas a una persona con amputación transradial debajo del codo.

Mediante el análisis de la biomecánica de la mano humana se determina sus parámetros principales, los cuales son utilizados para el diseño de la estructura mecánica del dispositivo, asegurando el desempeño funcional al simular los movimientos naturales de la mano [12]. Posteriormente, se le aplicaron algoritmos de control y se realizaron pruebas de movimiento, carga, agarre y presión, validando el modelo mediante pruebas estadísticas de independencia de Chi cuadrado [13]. En el mismo año, la Universidad Católica de Cuenca en conjunto con la Secretaría Nacional de Educación Superior, Ciencia, Tecnología e Innovación dentro del marco del Proyecto Prometeo, desarrollaron una prótesis biomecánica de mano derecho para niños mayores de ocho años y adolescentes.

Con un peso de hasta 200 gramos, el dispositivo protésico funciona con un sistema de reconocimiento de voz que permite realizar tres funciones importantes: presión cilíndrica, pinza fina y la extensión del índice para permitir el contacto con pantallas táctiles. Morfológicamente emula una palma de mano articulada a sus cinco dedos, trabaja con unas barras de

mecanismo lateral que permiten realizar las funciones de flexión y extensión de todos los dedos de acuerdo a los ángulos naturales de la mano, exceptuando el pulgar. No utilizan resortes ni hilos de tensión, reduciendo su peso significativamente [14].

En el año 2016, en la Universidad Técnica del Norte, como trabajo de titulación se presenta la construcción de una mano robótica enfocado al control de movimiento de los dedos. En este trabajo, se analiza la sujeción de forma puntual o pinza entre pulgar e índice, y apertura y cierre total de la mano. Se planteó un diseño con apariencias aproximadas a una mano real. El movimiento de los dedos se hace a través de hilos que hacen la función de tendones. El control total de la estructura se realiza a través de un guante, que contiene tres sensores de flexión para emular el movimiento del pulgar, el dedo índice y tres dedos restantes. Además contiene un acelerómetro que simula el movimiento de pronosupinación de la muñeca. Así la estructura ejecuta el mismo movimiento que realiza el guante [15].

Con el marco referencial expuesto, el presente trabajo es un módulo del Proyecto de Investigación “Desarrollo de Tecnología Asistiva para la Inclusión Laboral y Educativa de personas con Capacidades Especiales”, cuyo objetivo es utilizar diferentes técnicas en el tratamiento de señales bioléctricas y a su vez desarrollo de dispositivos como prototipos de brazos o manos robóticas. Por tanto, se propone el diseño de un banco de pruebas experimental de mano robótica de carácter multifuncional y mecánicamente desacoplada entre sus dedos pero acopladas entre sus falanges, generando un entorno real de comprobación de señales que hayan sido filtradas previamente y así contribuir con el estado del arte de la tecnología asistiva en el tratamiento de señales aplicadas a prótesis de mano robotizadas.

## II. MATERIALES Y MÉTODOS

En [16], se sostiene que las simulaciones son importantes durante las primeras etapas del diseño de algoritmos de control. Sin embargo, los estudios de simulación generalmente son incompletos debido a que se desprecian aspectos prácticos, de ahí que los resultados de simulación sean de valor limitado. Por otro lado, la validación experimental de algoritmos de control de señales, asegura su éxito potencial en el mundo real de las aplicaciones.

### A. Estructura de la Plataforma Experimental de Mano Robótica

Estructuralmente, La plataforma experimental se compone de cuatro módulos: mando, control, el prototipo mecánico de mano y un regulador de voltaje. En el módulo de mando se dispone de 3 modos: un Joystick para pruebas independientes de abre y cierre de la mano, una tarjeta electrónica Bluetooth para la recepción de señales inalámbricas que provienen de dispositivos móviles y un banco de pulsadores para el movimiento desacoplado de cada dedo. También se dispone de un switch selector del modo a considerar.

El módulo de control consta de un dispositivo programable que multiplexa cada una de las señales recibidas direccionalmente hacia los respectivos servos, este microcontrolador puede ser remplazado fácilmente y

reprogramado en tareas de acondicionamiento de señales. El prototipo mecánico de mano robótica recibe las respectivas señales desde los servos y dependiendo de la anchura del pulso de la señal PWM entregada a los servos, ejecuta el movimiento de los dedos. Para su funcionamiento la plataforma experimental de Mano Robótica dispone de voltajes regulados de +5 Vcc., y +6.5Vcc. El diagrama funcional de bloques, se muestra en la Fig. 1.

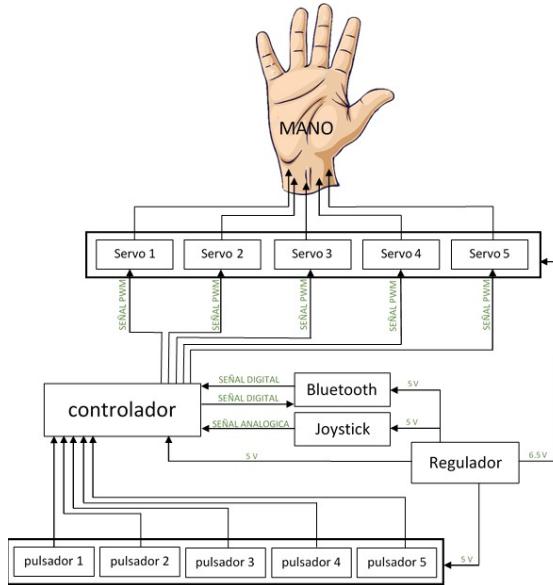


Fig. 1. Diagrama funcional de bloques de la Plataforma Experimental de Mano Robótica.

Morfológicamente, el prototipo de mano, consiste de cinco servomotores, uno para cada dedo y para la transmisión de movimiento para cada una de las articulaciones de los dedos, se hacen uso de hilos que hacen las veces de tendones. Cada dedo se compone de tres eslabones seriales, los mismos que se mueven conforme lo hace el servomotor ubicado en la base de la plataforma.

Se utilizó la impresión 3D para el desarrollo del prototipo de mano, el material que se utilizó para su fabricación es plástico ABS. La Fig. 2, muestra la plataforma implementada.

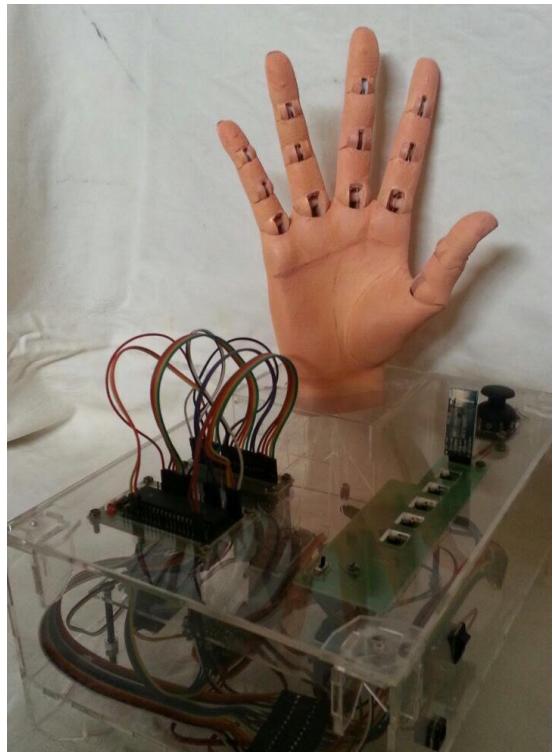


Fig. 2. Plataforma Experimental de Mano Robótica.

#### B. Modelo Cinemático del prototipo de mano robótica

Con fines de lograr la predicción y reproducción del comportamiento de la mano en su movimiento, se ha desarrollado haciendo uso del software MATLAB una simulación, para lo cual se requiere de conocer el modelo exacto del sistema. A continuación se describe la metodología utilizada.

En [16], [17], [18], concuerdan que para resolver el modelo cinemático de un robot manipulador existen dos problemas fundamentales: el modelo cinemático directo y el modelo cinemático inverso. Por otra parte la cinemática trata de relacionar las velocidades del movimiento de las articulaciones y las del extremo, esta relación se conoce como modelo diferencial representado por la matriz Jacobiana. El prototipo de mano que dispone la plataforma experimental, se compone de cinco dedos o miembros, cuatro de ellos poseen tres eslabones seriales, a diferencia del pulgar que solo tiene dos. Todos ellos de configuración articular Rotacional-Rotacional-Rotacional (RRR) y muy semejantes a un robot manipulador serial. Es de mencionar que el funcionamiento biomecánico de cada uno de los dedos de la mano es similar a un robot manipulador serial con características fuertemente desacopladas entre ellos y acoplados entre sus falanges. Por tanto, se expone el análisis cinemático de un miembro de tres eslabones seriales con tres articulaciones. La Fig. 3, expone el esquema de uno de los dedos del prototipo.

La solución al problema cinemático directo está dado por relaciones matemáticas (1), que permiten conocer cuál es la posición y orientación que adopta el extremo o la punta del dedo, cuando cada una de las variables que fijan la posición u orientación de sus articulaciones toma valores determinados.

$$\begin{aligned}x &= f_x(q_1, q_2, q_3). \\y &= f_y(q_1, q_2, q_3). \\z &= f_z(q_1, q_2, q_3). \\&\phi = f_\alpha(q_1, q_2, q_3). \\&\theta = f_\beta(q_1, q_2, q_3). \\&\psi = f_\gamma(q_1, q_2, q_3).\end{aligned}\quad (1)$$

Para representar la posición y orientación se ha considerado coordenadas cartesianas ( $x, y, z$ ) y ángulos de Euler ( $\phi, \theta, \psi$ ). Donde  $q_1, q_2, q_3$ , son los desplazamientos angulares de las articulaciones 1, 2 y 3 respectivamente.

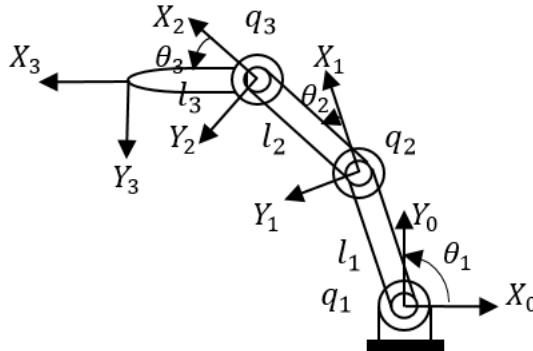


Fig. 3. Esquema de uno de los dedos de tres eslabones seriales del prototipo

Para resolver el problema cinemático directo del prototipo de mano, se ha hecho uso de la representación Denavit – Hartenberg (DH), [19], [20], [21], [17]. Donde las siguientes relaciones para una sucesión de rotaciones y traslaciones, permiten relacionar el sistema de referencia  $i - 1$  con el sistema del elemento  $i$ :

- Rotación alrededor del eje  $z_{i-1}$  un ángulo  $\theta_i$ .
- Traslación a lo largo de  $z_{i-1}$  una distancia  $d_i$ .
- Traslación a lo largo de  $x_i$  una distancia  $a_i$ .
- Rotación alrededor del eje  $x_i$  un ángulo  $\alpha_i$ .

Los cuatro parámetros ( $\theta_i, d_i, a_i, \alpha_i$ ), dependen solo de las características geométricas de cada eslabón y de las articulaciones que le unen con el anterior y siguiente, como muestra la Fig.4. Una vez determinados los parámetros DH, las matrices de transformación homogénea permiten una representación conjunta de la posición y orientación. La relación entre matrices de transformación homogénea y el método DH está determinada por (2):

$${}^{i-1}A_i = \text{Rot}_z(\theta_i)T(0,0, d_i)T(a_i, 0, 0) \text{Rot}_x(\alpha_i). \quad (2)$$

Desarrollando las operaciones matriciales correspondientes se obtiene la expresión matricial (3), que permite calcular las relaciones entre los eslabones consecutivos del robot de manera inmediata.

$${}^{i-1}A_i = \begin{bmatrix} \cos \theta_i & -\cos \alpha_i \sin \theta_i & \sin \alpha_i \sin \theta_i & a_i \cos \theta_i \\ \sin \theta_i & \cos \alpha_i \cos \theta_i & -\sin \alpha_i \cos \theta_i & a_i \sin \theta_i \\ 0 & \sin \alpha_i & \cos \alpha_i & d_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3).$$

Aplicando el método DH a uno de los dedos del prototipo de mano, que muestra la Fig. 3, se obtiene los parámetros expuestos en la Tabla1.

Considerando los parámetros DH y la ecuación (2) y (3) se logra la matriz de transformación homogénea (4) que relaciona la base del robot con el extremo final del dedo, resolviendo completamente el problema de la cinemática directa del dedo.

$${}^0A_3 = \begin{bmatrix} C\theta_{123} & -S\theta_{123} & 0 & l_1C\theta_1 + l_2C\theta_{12} + l_3C\theta_{123} \\ S\theta_{123} & C\theta_{123} & 0 & l_1S\theta_1 + l_2S\theta_{12} + l_3S\theta_{123} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

Tabla 1. Parámetros obtenidos al aplicar el método Denavit-Hartenberg.

$q_i$	$\theta_i$	$d_i$	$a_i$	$\alpha_i$
$q_1$	$\theta_1$	0	$l_1$	0
$q_2$	$\theta_2$	0	$l_2$	0
$q_3$	$\theta_3$	0	$l_3$	0

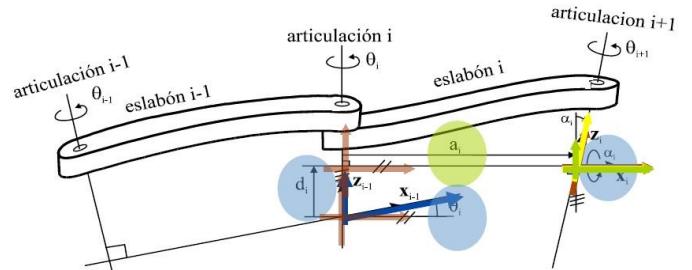


Fig. 4. Parámetros de Denavit – Hartenberg para un eslabón rotacional, [22] .

Donde el vector de posición final para el extremo del dedo, queda identificado por (5).

$$\begin{aligned}p_x &= l_1C\theta_1 + l_2C\theta_{12} + l_3C\theta_{123} \\p_y &= l_1S\theta_1 + l_2S\theta_{12} + l_3S\theta_{123} \\p_z &= 0\end{aligned}\quad (5)$$

Una vez que se ha determinado el modelo cinemático directo, se determina el modelo cinemático inverso del dedo de la plataforma experimental, con base en la siguiente relación (6).

$$\begin{bmatrix} q_1 \\ q_2 \\ q_3 \\ q_4 \\ q_5 \\ q_6 \end{bmatrix} = f^{-1}(x, y, z, \phi, \theta, \psi). \quad (6)$$

Con base en la geometría del miembro en análisis, las ecuaciones (7) y (8), dan solución al problema cinemático inverso.

$$k = \frac{p_x^2 + p_y^2 - l_1^2 - l_2^2}{2l_1l_2}$$

$$\theta_2 = \text{Cos}^{-1} k \quad (7)$$

$$C\theta_1 = \frac{p_x(l_1 + l_2 C\theta_2) + p_y l_2 S\theta_2}{l_1^2 + l_2^2 + 2l_1 l_2 C\theta_2}$$

$$S\theta_1 = \frac{-p_x l_2 S\theta_2 + p_y(l_1 + l_2 C\theta_2)}{l_1^2 + l_2^2 + 2l_1 l_2 C\theta_2}$$

$$\theta_1 = \text{Atan2}(S\theta_1, C\theta_1) \quad (8)$$

La orientación del extremo del dedo queda definido por (9):

$$\varphi = \theta_1 + \theta_2 + \theta_3 \quad (9)$$

A partir de la información contenida en las matrices de transformación homogénea, que definen el modelo cinemático, se obtiene la Jacobiana “J”,(10) y (11), utilizando el método de propagación de velocidades, muy común en [17], [18].

$$\begin{bmatrix} v_x \\ v_y \\ \omega_z \end{bmatrix} = J \begin{bmatrix} \dot{\theta}_1 \\ \dot{\theta}_2 \\ \dot{\theta}_3 \end{bmatrix} \quad (10)$$

$$J = [J_1 \ J_2 \ J_3] \quad (11)$$

Dónde:

$$J_1 = \begin{bmatrix} -(l_1 S\theta_1 + l_2 S\theta_{12} + l_3 S\theta_{123}) \\ (l_1 C\theta_1 + l_2 C\theta_{12} + l_3 C\theta_{123}) \\ 1 \end{bmatrix}$$

$$J_2 = \begin{bmatrix} -(l_2 S\theta_{12} + l_3 S\theta_{123}) \\ (l_2 C\theta_{12} + l_3 C\theta_{123}) \\ 1 \end{bmatrix}$$

$$J_3 = \begin{bmatrix} -(l_3 S\theta_{123}) \\ l_3 C\theta_{123} \\ 1 \end{bmatrix}$$

Físicamente, cada columna de  $J$  representa la dirección del vector velocidad tanto lineal como angular del extremo originado por el giro de su correspondiente articulación.

### III. DISCUSIÓN DE RESULTADOS.

El prototipo de mano robótica fue sometido a pruebas de funcionalidad mediante señales PWM, evidenciando el movimiento de los dedos, como también el abre y cierre de la mano, como muestra la Fig. 5. Mediante una aplicación móvil se comprobó la funcionalidad del módulo inalámbrico en la

transmisión y recepción de señales de prueba así como la respuesta en el movimiento de los dedos y falanges.



Fig. 5. Pruebas de Funcionalidad de Mano Robótica.

Los modelos matemáticos expuestos fueron simulados. La Fig. 6 muestra la simulación de la cinemática directa de uno de los dedos del prototipo.

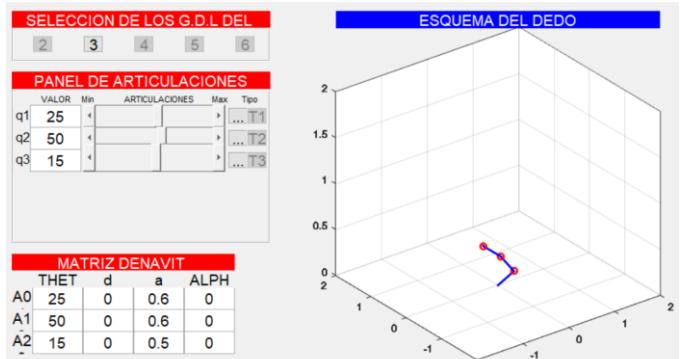


Fig. 6. Simulación de la Cinemática Directa de un dedo de la Plataforma Experimental de Mano Robótica.

### IV. CONCLUSIONES Y TRABAJOS FUTUROS

En el presente trabajo se diseñó e implementó una plataforma experimental de mano robótica cuyo objetivo es servir como banco de pruebas de arquitectura abierta para futuras investigaciones en el tratamiento de señales bioeléctricas, sean estas electromiográficas o electroencefalográficas. Para este propósito se implementaron cuatro módulos: una interfaz de mando, un módulo de control, el prototipo mecánico de mano y la regulación de voltaje. El

sistema dispone de un Joystick para pruebas de abre y cierre de la mano, una tarjeta electrónica Bluetooth para la recepción de señales inalámbricas y un banco de pulsadores para el movimiento individual de cada dedo. Los resultados demuestran que la plataforma de mano robótica exhibe la funcionalidad esperada, comprobándose el movimiento de las articulaciones de cada dedo y el abre y cierre de la mano. En trabajos futuros es deseable comprobar esquemas de control basados en técnicas de machine learning o redes neuronales que identifiquen señales o expresiones corporales que sean reproducidas en un entorno real.

## V. BIBLIOGRAFÍA

- [1] Consejo Nacional para la Igualdad de Discapacidades, «INFORMACIÓN ESTADÍSTICA DE PERSONAS CON DISCAPACIDAD LABORALMENTE ACTIVAS,» 17 Agosto 2017. [En línea]. Available: <http://www.consejodiscapacidades.gob.ec/estadistica/index.html>. [Último acceso: Septiembre 2017].
- [2] D. Peleg, E. Braiman, E. Yom-Tov y G. Inbar, «Classification of Finger Activation for Use in a Robotic Prosthesis Arm,» *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 10, nº 4, pp. 290-293, 2002.
- [3] D. Hann, D. kuschner y Y.-f. Wang, «Upper Limb Position Sensing: A Machine Vision Approach,» de *2nd International IEEE EMBS Conference on Neural Engineering*, Arlington, Virginia, 2005.
- [4] M. C. Carroza, A. Persichetti, C. Laschi, F. Vecchi, R. Lazzarini, V. Tamburrelli, P. Vacalebri y P. Darío, «A Novel Wearable Interface for Robotic Hand Prostheses,» de *9th International Conference on Rehabilitation Robotics*, Chicago, Illinois, 2005.
- [5] F. Mobasser, J. M. Eklund y K. Hashtrudi-Zaad, «Estimation of Elbow-Induced Wrist Force With EMG,» *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, vol. 54, nº 4, pp. 683-693, 2007.
- [6] C. Castellini, P. Van der Smagt y G. Hirzinger, «Surface EMG for Force Control of Mechanical Hands,» de *IEEE International Conference on Robotics and Automation*, Pasadena, California, 2008.
- [7] A. Harada, T. Nakakuki, M. Hikita y C. Ishii, «Robot Finger Design for Myoelectric Prosthetic Hand and,» de *International Conference on Automation and Logistics*, Hong Kong and Macau, 2010.
- [8] J. T. Belter, M. T. Leddy y K. D. Gemmell Jr., «Comparative Clinical Evaluation of the Yale Multigrasp Hand,» de *6th IEEE RAS/EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, UTown, Singapore, 2016.
- [9] D. I. Pilaquinga, «Diseño y Construcción de una Mano Robótica Controlada Mediante un Guante Sensorizado,» Quito, 2009.
- [10] J. Collahuazo y E. Ordoñez, «Design and Construction of a Robot Hand Activated by Electromyographic Signals,» de *2012 IEEE International Symposium on Robotic and Sensors Environments*, 2012.
- [11] C. A. Calderón Córdova, «Una Mano de Esperanza para Todos,» *Perspectivas de Investigación*, vol. 1, nº 1, p. 3, 2014.
- [12] M. B. Monar Naranjo y L. A. Murillo Mantilla, «Diseño y Construcción de una Prótesis Biónica de Mano de 7 Grados de Libertad Utilizando Materiales Inteligentes y Control Mioeléctrico Adaptada para Varios Patrones de Sujeción,» Latacunga, 2015.
- [13] D. F. Morales Reinoso, «Diseño e Implementación del Sistema de Control de una Prótesis Biónica de 7 Grados de Libertad Utilizando Materiales Inteligentes y Control Mioeléctrico Adaptada para varios Patrones de Sujeción,» Latacunga, 2016.
- [14] B. Gámez, C. Flores, F. Cabrera y J. Cabrera, «Design of a biomechanics prosthesis for child,» *Ingeniería UC*, vol. 23, nº 1, pp. 58-66, 2016.
- [15] H. Pozo, «Construcción de una Mano Robótica Enfocada al Movimiento de los Dedos,» Trabajo de Titulación, Ibarra, Ecuador, 2016.
- [16] F. Reyes Cortez, Robótica. Control de Robots Manipuladores, C.V. México: Alfa/Marcombo, 2011.
- [17] A. Barrientos, L. Peñín y C. Balaguer, Fundamentos de Robótica, Madrid, España: Mc Graw-Hill, 2007.
- [18] L.-W. T. Tsai, Robot Analysis, New York: John Wiley , 199.
- [19] J. Denavit y R. Hartenberg, «A kinematic notation for lower-pair mechanisms,» *ASME Journal of Applied Mechanics*, nº 1955, pp. 215-221, 1955.
- [20] R. Paul, Robot Manipulators: Mathematics, Programming, and Control, Massachusetts,: The MIT Press, 1981.
- [21] J. Craig, Robótica, México: Prentice Hall, 2006.
- [22] F. Reyes y R. Kelly., «On parameter identification of robot manipulators,» de *Proceedings of International Conference on Robotics and Automation*, Albuquerque, New Mexico, 1997.



**Washington Caraguay**, es Ingeniero en Electrónica y Telecomunicaciones por la Escuela Superior Politécnica del Litoral y Máster en Automática y Robótica por la Universidad Politécnica de Madrid. Con una reconocida y demostrada trayectoria técnica, académica y científica en Instituciones de Educación Superior. Actualmente, se desempeña como Profesor de Física y Robótica en la Facultad de Ingeniería en Sistemas y Telecomunicaciones de la Universidad Espíritu Santo, Ecuador. Colabora con el Centro de Investigaciones de la UEES como Director del Proyecto: Inclusión Educativa y Laboral de Personas con Discapacidad Visual, Auditiva y Motriz en las Extremidades Superiores mediante el Diseño y Aplicación de Tecnología Asistiva. Como compensación y agradecimientos a la vida, comparte su tiempo en actividades de enseñanza y capacitación los fines de semana impartiendo cursos de formación técnica en Robótica y Nuevas Tecnologías.



**Marco Sotomayor Sánchez**, es Máster en Tecnologías de la Información por la Universidad de Griffith, Australia. Ha sido miembro del equipo de investigación del Departamento de Inteligencia Artificial y

Tecnologías Semánticas de la Escuela de Tecnologías de la Información y Comunicación de la Universidad de Griffith bajo la dirección del matemático Dr. Kewen Wang, autoridad en áreas de Inteligencia Artificial como razonamiento monótonico y Web Semántica. Ha colaborado en CSIRO, agencia del gobierno de Australia para investigación científica, en el Departamento de e-Health, bajo la supervisión del Dr. Michael Lawley, en el proyecto de estudio de modularidad en ontologías médicas, en particular SNOMED CT. Actualmente se desempeña como Profesor y Director del proyecto de investigación “Framework de análisis geoespacial predictivo de patrones criminales”.



**Miguel Alonso** es estudiante de la carrera de Ingeniería en Telecomunicaciones en la Universidad de Especialidades Espíritu Santo. Actualmente, se encuentra desarrollando el proyecto de titulación: “Diseño e Implementación de Plataforma experimental de mano robótica” el mismo que busca contribuir a

las tecnologías inclusivas de las problemáticas sociales de movilidad y pérdida de miembros superiores. Como parte de los puntales de su vida es instructor y facilitador en temas de tecnología, programas ofimáticos y redes de datos en centros de capacitación.



# First Order Methods for High Resolution Image Denoising

David Villacís

**Abstract**—In this paper we are interested in comparing the performance of some of the most relevant first order non-smooth optimization methods applied to the Rudin, Osher and Fatemi (ROF) Image Denoising Model and a Primal-Dual Chambolle-Pock Image Denoising Model. Because of the properties of the resulting numerical schemes it is possible to handle these computations pixelwise, allowing implementations based on parallel paradigms which are helpful in the context of high resolution imaging.

**Index Terms**—Image Denoising, High Resolution, Parallel Computing, First Order Optimization Methods, Non-smooth optimization methods.

## I. INTRODUCTION

Image Denoising is one of the most revisited image analysis tasks, it consists of removing noise from a damaged original image. According to [1, p. 145] noise can be added in an image due to problems in the image adquisition processes such as: malfunctioning pixels in the camera sensors, faulty memory locations in hardware, analog-to-digital conversion, a noisy transmission channel, etc. Traditional techniques used to solve this problem include the use of filters [2], Wavelets [3], and lately in [4] the use of variational models have been popularized due to the elegant way of enforcing properties in the image without the need of single them out explicitly. We are interested in addressing the efficiency of some relevant variational models relying on conjugate duality and saddle point formulations and their application on parallel computing paradigms.

In this work we will consider an image as a  $n_1 \times n_2$  pixel matrix, and for simplicity on the mathematical treatment we will map this matrix as a vector  $x \in \mathbb{R}^n$ , where  $n = n_1 \cdot n_2$ . Thus a pixel in location  $(i, j)$  is the element  $x_{i+n_1(j-1)}$  of the vector  $x$ .

We will consider the image denoise problem as an inverse problem where the observed image is defined in (1).

$$f = T(x) + \eta, \quad (1)$$

where  $T$  is a non-linear operator and  $\eta$  is an additive type of noise. In order to solve this problem we can use a Tikhonov regularization of the problem in (2).

$$f = \phi(x) + \lambda R(x), \quad (2)$$

where  $\phi$  is a data fidelity term that is selected according to the type of noise present in the image and  $R$  is called regularizer and promotes certain properties in the solutions obtained.

David Villacís is with the Research Center on Mathematical Modelling (MODEMAT), Escuela Politécnica Nacional, Quito, Ecuador

## II. IMAGE DENOISING MODELS

### A. Total Variation Regularization (TV)

The idea of using Total Variation (TV) as a regularizer is to induce sparsity on the gradients of an image, this particular property favours piecewise constant images with sparse edges. Introducing the discrete gradient operator  $\mathbb{K} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  with  $m = 2 \times n$  and  $x \in \mathbb{R}^n$  a given image, the discrete Total Variation (TV) is defined by (3).

$$\|\mathbb{K}x\|_{p,1} = \sum_{j=1}^n \|(\mathbb{K}x)_j\|_p = \sum_{j=1}^n ((\mathbb{K}x)_j^p + (\mathbb{K}x)_{n+j}^p)^{\frac{1}{p}} \quad (3)$$

The  $p$  parameter is used to realize anisotropic ( $p = 1$ ) or isotropic ( $p = 2$ ), the latest is usually prefered since it does not exhibit a grind bias.

$$\|\mathbb{K}x\|_{2,1} = \sum_{j=1}^n \|(\mathbb{K}x)_j\|_2 = \sum_{j=1}^n \sqrt{(\mathbb{K}x)_j^2 + (\mathbb{K}x)_{n+j}^2} \quad (4)$$

Figure 1 shows the impact of using this regularizer on the denoising model.

### B. Rudin-Osher-Fatemi (ROF) Model

The TV denoising variational model was first introduced by Rudin, Osher and Fatemi in their seminal paper [5]. They proposed a model that contains a data fidelity term for gaussian noise and a regularizer according to the following model:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|x - f\|_2^2 + \lambda \|\mathbb{K}x\|_{2,1}, \quad (5)$$

where  $f \in \mathbb{R}^n$  is the input corrupted image and  $\lambda$  is the Tichkonov regularization parameter.

In order to handle the regularization term for this model, it is a usual practice to make use of its Fenchel-Rockafellar dual version:

$$\min_{y \in \mathbb{R}^m} \|f - \mathbb{K}^\top y\|_2^2 + \delta_{B_\lambda}(y) \quad (6)$$

Where  $\delta$  is the indicator function of the  $\lambda$ -radius ball  $B_\lambda(0)$ , and  $y \in \mathbb{R}^m$  is the dual variable.

In order to approach this problem numerically it is helpful to use a saddle point formulation of (5):

$$\min_x \max_y \frac{1}{2} \|x - f\|_2^2 + \langle \mathbb{K}x, y \rangle - \delta_{B_\lambda}(y) \quad (7)$$

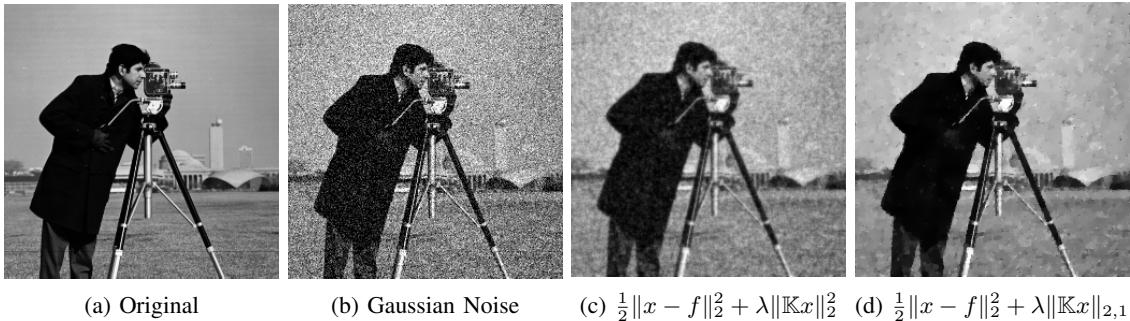


Fig. 1: Image denoising examples using Forward-Backward method for  $\|\mathbb{K}x\|_2^2$  (1c) and  $\|\mathbb{K}x\|_{2,1}$  (1d) regularizers. We can see that the use of Total Variation (4) promotes sharper edges inside the image. The original image has been modified with additive gaussian noise with mean 0 and variance 0.2.

### C. TV- $l_1$ Model

The denoising task is very sensitive to the noise type appearing in the image, as a matter of fact, the ROF Model presented previously works for gaussian distributed noise as noted in the work presented by Chan et al.[6]. When it comes to impulse distributed noise, seminal work by Nikolova [7] shows that using a  $l_1$  data fidelity term gives us better results in the output image; this effect is illustrated in Figure 2. This model was described as follows:

$$\min_{x \in \mathbb{R}^n} \|x - f\|_1 + \lambda \|\mathbb{K}x\|_{2,1}, \quad (8)$$

where  $f \in \mathbb{R}^n$  is the input corrupted image and  $\lambda$  is the Tichonov regularization parameter. It is not hard to see that this model contains two non-differentiable terms, which yields a harder problem to be solved. In order to derive numerical methods for this model, let us formulate the corresponding saddle-point problem:

$$\min_x \max_y \|x - f\|_1 + \langle \mathbb{K}x, y \rangle - \delta_{B_\lambda}(y) \quad (9)$$

### III. NUMERICAL TREATMENT

In this section we will present the numerical methods used to solve (5) and (8) along with their corresponding algorithms. Several analytical tools such as Conjugate Duality and Proximal Operators will be used to derive formulations more suitable for a parallel computation environment.

#### A. ROF Model

1) *Forward-Backward Splitting*: The conjugate dual ROF Model (6) presented in section II has a particular structure that can be exploited, it is the smoothness of the data fidelity term  $\|f - \mathbb{K}^\top y\|_2^2$ , which in this case is differentiable. Therefore, we can use several splitting numerical mechanisms to find the solution. Let us first analyze the Forward-Backward splitting. Taking in consideration this specific model,  $f(y) = \delta_{B_\lambda}(y)$  the characteristic function of the  $\lambda$ -ball, and  $\partial f$  the convex subdifferential of  $f$  we can describe the optimality condition

as:

$$\begin{aligned} 0 &\in \nabla g(x) + \partial f(x), \\ 0 &\in \mathbb{K}(f - \mathbb{K}^\top y) + \partial f(y), \\ \partial f(y) &\ni -\mathbb{K}(f - \mathbb{K}^\top y), \\ \tau^{-1}y + \partial f(y) &\ni \tau^{-1}y - \mathbb{K}(f - \mathbb{K}^\top y), \\ y &= (I + \tau\partial f)^{-1}(y - \tau\mathbb{K}(f - \mathbb{K}^\top y)). \end{aligned}$$

We will name  $(I + \tau\partial f)^{-1}$  as  $\text{prox}_{\tau\partial f}$ , the proximal operator for  $f$ . The proximal operator provides the unique optimizer of a Moreau-Yosida (MY) regularization of a function  $f$ :

$$f_{MY}(x) = f(x) + \frac{1}{2\tau} \|\bar{x} - x\|_2^2. \quad (10)$$

In [8] it is shown that (10) has a unique minimizer and the resulting function is proper, lower-semicontinuous and strongly convex. Moreover, its minimizer solves the following optimization problem:

$$\text{prox}_{\tau\partial f}(x) = \arg\min_{\bar{x}} f(x) + \frac{1}{2\tau} \|\bar{x} - x\|_2^2 \quad (11)$$

Leading to the following numerical iteration:

$$y_{k+1} = \text{prox}_{\tau\partial f}(y_k - \tau\mathbb{K}(f - \mathbb{K}^\top y_k)). \quad (12)$$

This iteration requires the calculation of the proximal operator of  $\delta_{B_\lambda}(y)$ , for this application in particular this operator corresponds to the pixel-wise projection of  $y$  onto the  $\lambda$ -radius ball.

$$[\text{proj}_{B_\lambda}(y)]_i = \frac{y_i}{\max\{1, \lambda^{-1}\|y\|_2\}}, \quad \forall i = 1, \dots, n \quad (13)$$

Now, working on the dual optimization problem exclusively we can make use of Algorithm 1 to obtain the optimal dual value and retrieve its primal  $x_k = f - \mathbb{K}^\top y_k$ .

2) *Chambolle-Pock Method*: This method handles the saddle point formulation of the problem (7). We can find in this formulation that it contains a primal part and a dual part. To tackle this problem this method proposes two proximal steps, one for the primal part and one for the dual part, and an interpolation step that in the case of the ROF model yields the numerical scheme presented in Algorithm 2.

As a preliminary experiment, in Figure 3 we can see the convergence properties of the two methods presented in this section. Indeed we can see that the Chambolle-Pock method

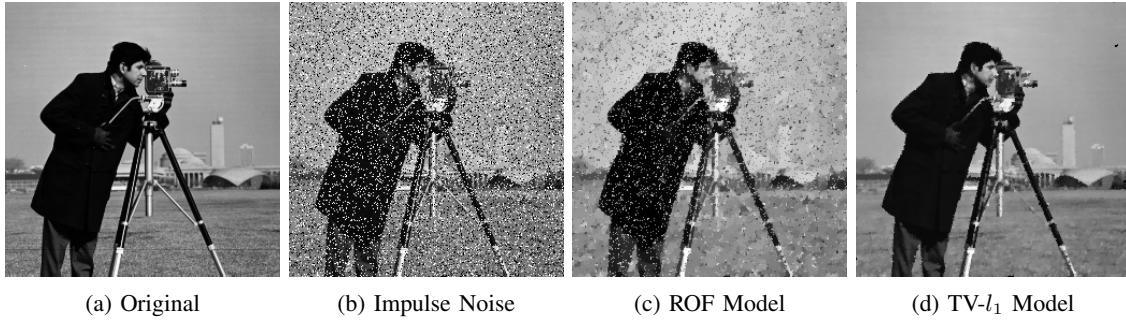


Fig. 2: Image denoising examples using Chambolle-Pock method for ROF Model (2c) and TV- $l_1$  model (2d) regularizers with  $\lambda = 0.2$ . We can see that the ROF model cannot tackle the image denoising task appropriately when it is presented with impulse noise.

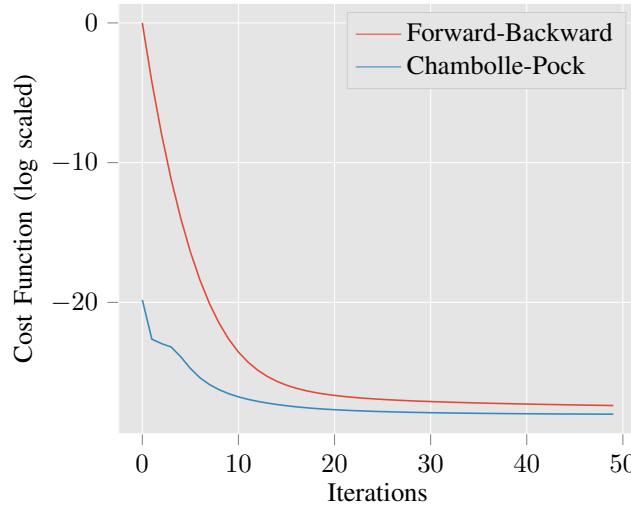


Fig. 3: ROF Cost Function evolution for Forward-Backward and Chambolle-Pock Methods.

---

**Algorithm 1** Forward-Backward Splitting for ROF Denoising

---

- 1: Choose  $\lambda > 0$ ,  $\tau > 0$ ,  $k = 0$ .
  - 2: Set number of iterations  $niter$ .
  - 3: **while**  $k < niter$  **do**
  - 4:   Calculate  $\nabla g(y_k) = \mathbb{K}(f - \tau \mathbb{K}^\top y_k)$
  - 5:   Perform projection onto the  $\lambda$ -ball  

$$y_{k+1} = \text{proj}_{B_\lambda}(y_k - \tau \nabla g(y_k))$$
  - 6:    $k = k + 1$
  - 7: **end while**
  - 8: Retrieve primal value  $x_{k+1} = f - \mathbb{K}^\top y_{k+1}$ .
- 

presents better convergence properties thanks to the use of the extra information provided by the proximal operator.

### B. TV- $l_1$ Model

1) *Chambolle-Pock Method:* For this model we will make use of a saddle point formulation of the TV- $l_1$  model (9). In order to apply the Chambolle-Pock Method we need to derive the form of the proximal operator for  $g(x) = \|x - f\|_1$ .

$$\text{prox}_{\tau \partial g} = \underset{\bar{x}}{\operatorname{argmin}} \|x - f\|_1 + \frac{1}{2\tau} \|\bar{x} - x\|^2. \quad (14)$$

---

**Algorithm 2** Chambolle-Pock Method for ROF Denoising

---

- 1: Choose  $\lambda > 0$ ,  $\tau, \sigma > 0$ ,  $k = 0$ .
  - 2: Set number of iterations  $niter$ .
  - 3: **while**  $k < niter$  **do**
  - 4:   Calculate the proximal step for the primal function  

$$x_{k+1} = \frac{x_k + \tau(f - \mathbb{K}^\top y_k)}{\tau + 1}$$
  - 5:   Calculate the interpolation step  $\bar{x}_{k+1} = 2x_{k+1} - x_k$
  - 6:   Perform projection onto the  $\lambda$ -ball  

$$y_{k+1} = \text{proj}_{B_\lambda}(y_k + \sigma \mathbb{K} \bar{x}_{k+1})$$
  - 7:    $k = k + 1$
  - 8: **end while**
- 

which, using the procedure explained in Appendix A, yields:

$$\bar{x}_i = \text{sign}(x_i - f_i) \max(|x_i - f_i| - \tau, 0), \quad \forall i = 1, \dots, n. \quad (15)$$

Therefore, we can make use of the numerical scheme presented in Algorithm 3. Let us remark the fact that every operation is performed pixel-wise.

**Algorithm 3** Chambolle-Pock Method for TV- $l_1$  Denoising

```

1: Choose  $\lambda > 0$ ,  $\tau, \sigma > 0$ ,  $k = 0$ .
2: Set number of iterations  $niter$ .
3: while  $k < niter$  do
4:   Calculate the auxiliary value:  $\hat{x}_{k+1} = x_k - \tau(\mathbb{K}^\top y_k)$ 
5:   Calculate the proximal step for the primal function
      
$$x_{k+1} = f + sign(\hat{x}_{k+1} - f) \max(|\hat{x}_{k+1} - f| - \tau, 0)$$

6:   Calculate the interpolation step  $\bar{x}_{k+1} = 2x_{k+1} - x_k$ 
7:   Perform projection onto the  $\lambda$ -ball
      
$$y_{k+1} = proj_{B_\lambda}(y_k + \sigma \mathbb{K} \bar{x}_{k+1})$$

8:    $k = k + 1$ 
9: end while

```

**IV. NUMERICAL EXPERIMENTS**

For the numerical experiments presented in this work we will be using both a set testing images with synthetic noise and the dataset provided in [9], this dataset provides real noise generated for high resolution images of playing cards. This dataset was obtained using a PhaseOne XF medium format camera equipped with an achromatic IQ260 digital back and a PhaseOne Digita AF 120mm F4 lens. The equipment used generated 16 bit TIFF images of 8964x6716 pixels. The images were shot with an ISO of 3200 and a histogram approximately spanning a quarter of the full dynamic range, yielding an image that contains noise generated from round-off errors from the digital-analog conversor, photon counting noise and electronic noise.

Since all operations for the numerical methods described previously are applied pixel-wise, we can make use of parallel computation efectively. In this work, we used CUDA parallel programming [10] to implement the update for the numerical schemes presented for the chambolle-pock method for both ROF and TV- $l_1$  denoising problems.

For the CPU computations, all methods described above were implemented using python Numpy [11] numerical libraries, and all the GPU computations were coded using PyCUDA [12] with kernels written in C.

All the presented experiments where executed in the Mode-Mat HPC Cluster, this facility provided us with Xeon Phi processors for the CPU computations and NVIDIA Tesla K80 GPU coprocessors for the GPU computations.

**A. ROF Model**

We tested the Forward-Backward (FB) and Chambolle-Pock (CP) primal-dual methods against a set of images: Circle (106x106 px), Cameraman (256x256 px) and Lena (512x512 px) using both the CPU and GPU implementations. The results are presented in Table I, this synthetic images presented additive gaussian noise with 0 mean and 0.2 variance. In Table II the results over the real noise dataset with the following set of images: Playing Cards 1 (1280x720 px - HD Resolution), Playing Cards 2 (1920x1080 px - FullHD Resolution), Playing

TABLE I: ROF Model Processing Time

Image	FB CPU	CP CPU	CP GPU
	time (s)	time (s)	time (s)
Circle	0.0877 $\pm$ 0.008	0.0791 $\pm$ 0.001	<b>0.0261 <math>\pm</math> 0.004</b>
Cameraman	0.4499 $\pm$ 0.008	0.4064 $\pm$ 0.001	<b>0.0269 <math>\pm</math> 0.001</b>
Lena	1.7019 $\pm$ 0.007	1.6732 $\pm$ 0.069	<b>0.0358 <math>\pm</math> 0.001</b>

TABLE II: ROF Model Processing Time

Image	FB CPU	CP CPU	CP GPU
	time (s)	time (s)	time (s)
Playing Cards 1	7.89 $\pm$ 0.12	7.26 $\pm$ 0.02	<b>0.067 <math>\pm</math> 0.01</b>
Playing Cards 2	24.40 $\pm$ 0.06	21.60 $\pm$ 0.16	<b>0.118 <math>\pm</math> 0.01</b>
Playing Cards 3	585.04 $\pm$ 0.98	511.36 $\pm$ 0.81	<b>1.717 <math>\pm</math> 0.01</b>

TABLE III: TV- $l_1$  Model Processing Time - Synthetic Noise

Image	CP CPU	CP GPU
	time (s)	time(s)
Circle	0.094254 $\pm$ 0.000323	<b>0.025320 <math>\pm</math> 0.000736</b>
Cameraman	0.456685 $\pm$ 0.000841	<b>0.026760 <math>\pm</math> 0.000779</b>
Lena	1.802704 $\pm$ 0.003430	<b>0.037916 <math>\pm</math> 0.000587</b>

TABLE IV: TV- $l_1$  Model Processing Time - Real Noise

Image	CP CPU	CP GPU
	time (s)	time(s)
Playing Cards 1	8.23238 $\pm$ 0.030	<b>0.067131 <math>\pm</math> 0.001167</b>
Playing Cards 2	24.16407 $\pm$ 0.051	<b>0.128384 <math>\pm</math> 0.007352</b>
Playing Cards 3	599.70860 $\pm$ 0.754	<b>1.895479 <math>\pm</math> 0.003539</b>

TABLE V: ROF Model Processing Time Low CPU

Image	FB CPU LOW	CP CPU LOW	CP GPU
	time (s)	time (s)	time (s)
Circle	0.180 $\pm$ 0.02	0.139 $\pm$ 0.01	<b>0.0261 <math>\pm</math> 0.004</b>
Cameraman	1.395 $\pm$ 0.10	0.872 $\pm$ 0.05	<b>0.0269 <math>\pm</math> 0.001</b>
Lena	5.971 $\pm$ 0.01	5.358 $\pm$ 0.1	<b>0.0358 <math>\pm</math> 0.001</b>

TABLE VI: TV- $l_1$  Model Processing Time Low CPU

Image	CP CPU LOW	CP GPU
	time (s)	time (s)
Circle	0.16 $\pm$ 0.01	<b>0.025 <math>\pm</math> 0.001</b>
Cameraman	0.935 $\pm$ 0.3	<b>0.027 <math>\pm</math> 0.001</b>
Lena	6.06 $\pm$ 0.1	<b>0.038 <math>\pm</math> 0.001</b>

Cards 3 (7680x4320 px - UltraHD Resolution).

For the experiments, in the synthetic dataset we used 100 runs, the mean and standard deviation values are reported in the corresponding tables. In the case of the real noise dataset we used 10 trials with the same values reported.

**B. TV- $l_1$  Model**

In Table III and Table IV we tested Chambolle-Pock (CP) CPU and GPU implementations against the same set of images described in the previous experiment. The synthetic images generated presented additive impulse noise.

For comparison purposes in Tables V and Table VI the results of the paralell version of the algorithms is compared with the performance using a desktop computer using a traditional Intel Core i5 processor.

**V. CONCLUSION**

In this work we can see that the ROF and TV- $l_1$  image denoising models can be formulated using a saddle point formulation. This formulation allows Chambolle-Pock method

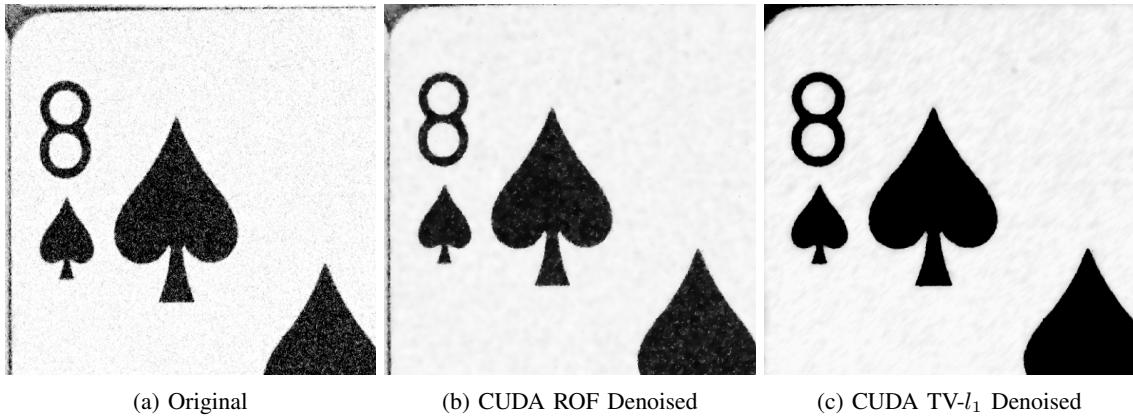


Fig. 4: Image denoising examples using a CUDA Implementation Chambolle-Pock method for ROF Model (4b) and TV- $l_1$  model (4c) regularizers with  $\lambda = 0.2$ . This image patch corresponds to an UltraHD Resolution Image, the execution time for the algorithm was 1.89548 seconds.

to implement an algorithm to find a solution that presents some separability properties, since the proximal operators obtained can be implemented pixelwise. Therefore, we can make use of parallel computation. In particular we used CUDA parallel computation which yields dramatic speedups for finding solutions when compared with pure serial implementations.

#### APPENDIX A

##### $\|x - f\|_1$ PROXIMAL OPERATOR

We can formulate the proximal operator as:

$$\operatorname{argmin}_{\bar{x}} \|x - f\|_1 + \frac{1}{2\tau} \|\bar{x} - x\| \quad (16)$$

We know that  $\bar{x}$  is a minimizer argument of (16) if and only if it satisfies the following optimization criteria:

$$\begin{aligned} 0 &\in \partial \|\bar{x} - f\|_1 + (\bar{x} - x), \\ x &\in \tau \partial \|\bar{x} - f\|_1 + \bar{x} \end{aligned}$$

We know that the  $l_1$ -norm is separable, therefore we can analyze this equation component-wise. Let us take the case  $\bar{x}_i \neq 0$  then  $\partial \|\bar{x}_i - f_i\| = \operatorname{sign}(\bar{x}_i - f_i)$ , then:

$$\begin{aligned} x_i &= \tau \operatorname{sign}(\bar{x}_i - f_i) + x_i, \\ \bar{x}_i &= x_i - \tau \operatorname{sign}(\bar{x}_i - f_i), \\ \bar{x}_i - f_i &= x_i - f_i - \tau \operatorname{sign}(\bar{x}_i - f_i) \end{aligned}$$

If  $\bar{x}_i - f_i > 0$  then  $x_i - f_i < -\tau$  and  $\bar{x}_i - f_i < 0$  then  $x_i - f_i > \tau$ . Hence,  $|x_i - f_i| > \tau$  and  $\operatorname{sign}(\bar{x}_i - f_i) = \operatorname{sign}(x_i - f_i)$ . If  $\bar{x}_i - f_i = 0$  then,

$$\begin{aligned} 0 &\in f_i - x_i + \tau[-1, 1], \\ x_i - f_i &\in [-\tau, \tau], \\ |x_i - f_i| &\leq \tau \end{aligned}$$

Therefore, the proximal operator can be written as

$$\begin{aligned} \operatorname{prox}_{\tau \partial g}(x) &= \bar{x}_i - f_i, \\ &= \begin{cases} 0 & \text{if } |x_i - f_i| \leq \tau, \\ x_i - f_i - \tau \operatorname{sign}(x_i - f_i) & \text{if } |x_i - f_i| > \tau \end{cases} \end{aligned}$$

This condition can be written in a more compact form:

$$\operatorname{prox}_{\tau \partial g}(x) = \max\{|x_i - f_i| - \tau, 0\} \operatorname{sign}(x_i - f_i) \quad (17)$$

#### REFERENCES

- [1] T. Chan and J. Shen, *Image Processing And Analysis: Variational, Pde, Wavelet, And Stochastic Methods*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2005.
- [2] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. The MIT Press, 1964.
- [3] A. Chambolle, R. A. DeVore, N. Y. Lee, and B. J. Lucier, “Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage,” *IEEE Transactions on Image Processing*, vol. 7, no. 3, pp. 319–335, 1998.
- [4] L. I. Rudin and S. Osher, “Total variation based image restoration with free local constraints,” *Image Processing, 1994. Proceedings. ICIP-94, IEEE International Conference*, vol. 60, 1994.
- [5] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.
- [6] R. H. Chan, C.-W. Ho, and M. Nikolova, “Salt and Pepper Noise Removal by Median Type Noise Detectors and Detail-Preserving Regularization,” *IEEE Transactions on Image processing*, vol. 14, no. 10, pp. 1479–1485, 2005.
- [7] M. Nikolova, “A Variational Approach to Remove Outliers and Impulse Noise,” in *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, 2004, pp. 99–120.
- [8] J. Peypouquet, *Convex Optimization in Normed Spaces*. Springer, 2015.
- [9] S. K. S. S. Villacis David, Helenius Teemu, “Photographic dataset: playing cards,” 2017.
- [10] T. M. John Cheng, Max Grossman, *Professional CUDA C Programming*, 1st ed. Wrox, 2014.
- [11] S. C. C. Stfan van der Walt and G. Varoquaux, “Numpy: Open source numerical tools for Python,” 2011. [Online]. Available: <http://aip.scitation.org/doi/abs/10.1109/MCSE.2011.37>
- [12] Y. L. B. C. P. I. A. F. Andreas Klckner, Nicolas Pinto, “Pycuda and pyopencl: A scripting-based approach to gpu run-time code generation,” 2009. [Online]. Available: <https://arxiv.org/abs/0911.3456>



**David Villacís** is a PhD student in Applied Mathematics from Escuela Politécnica Nacional. M.Sc. in Computer Science from University of Birmingham, UK. Currently working in variational image processing problems and optimization methods for machine learning.



# Diseño y Construcción de un Escáner 3D e Integración a una Impresora 3D (ingeniería inversa accesible de bajo costo y fuente abierta)

## *Design and Construction of a 3D Scanner and Integration to a 3D Printer (accessible low cost and open source reverse engineering)*

Leomundo Hernández, Andrés Cela

**Resumen—** El equipo construido posee un escáner 3D con principio en la triangulación láser, integrado a una impresora 3D en base a manufactura por deposición fundida. Se ha utilizado el mismo hardware de la impresora para integrar los periféricos del escáner: láseres y plataforma rotatoria; la cámara se conecta directamente a un computador que posee un software desarrollado para manejar las funcionalidades del escáner, que crea una nube de puntos del modelo, y debido a la integración de herramientas de postprocesado (reconstrucción de la superficie) y laminado (código-g), se genera automáticamente un archivo listo para la interpretación en la impresora 3D, cerrando el ciclo de reproducción de un modelo. Por su bajo costo y fuente abierta este prototipo hace accesible la realización de ingeniería inversa de un objeto, mediante la adquisición digital (escáner 3D) y permite su reproducción total (impresión 3D).

**Palabras Clave—** escáner 3D, impresión 3D, ingeniería inversa, reconstrucción de superficies, visión por computador

**Abstract—** Built equipment has a 3D scanner with a principle in laser triangulation, integrated to a 3D printer based on fused deposition modeling. The same printer hardware has been used to integrate the peripherals of the scanner: lasers and rotating platform, the camera connects directly to a computer that has software developed to handle the functionalities of the scanner, which obtains a point cloud, and due to integration of postprocessing (surface reconstruction) and laminate tools (g-code), a file is automatically generated ready for interpretation in the 3D printer, closing the reproduction cycle of a model. Due to its low cost and open source, this prototype made accessible performing reverse engineer of an object by means of digital acquisition (3D scanner) and allows its total reproduction (3D printing).

**Index Terms —** 3D scanner, 3D printer, reverse engineer, surface reconstruction, computer vision.

Leomundo Hernández es estudiante de Ingeniería en Electrónica y Control de la Escuela Politécnica Nacional (EPN), Quito – Ecuador. (e-mail: leomundo.hernandez@epn.edu.ec)

## I. INTRODUCCIÓN

### A. Impresión 3D

La impresión 3D es un proceso de fabricación aditiva, que utilizando un modelo digital tridimensional CAD, puede reproducirlo físicamente de forma automatizada mediante equipos CNC acondicionadas para este fin.

Hace 30 años aproximadamente los métodos de impresión 3D por deposición fueron monopolizados por pocas empresas (Stratasys® y 3D Systems®) que los patentaron y encarecieron, y por consiguiente limitaron la I+D por parte de otros desarrolladores o fabricantes, por recelo a infringir las leyes de propiedad intelectual.

En el transcurso de esta última década y sobre todo en el año 2009 y 2015 las patentes más insignes caducaron (US5174931 [1], US5569349 [2], US5637169 [3], US5639070 [4]), lo que ha generado un desarrollo exponencial de las impresoras 3D con una orientación hacia el código abierto principalmente, después de que un proyecto comunitario llamado RepRap en el año 2007 se impuso la meta de crear la primera impresora 3D autoreponible, que debió esperar hasta el 2009 para difundir su trabajo legalmente [5, pp. 77-80,186].

### B. Escáner 3D e Ingeniería Inversa

Un modelo digital tridimensional (3D) es una representación numérica de un objeto real. En general existen dos grandes grupos: los modelos volumétricos, los cuales describen las propiedades internas del objeto, mientras que en los modelos superficiales se representa la apariencia visual de los objetos [6, p. 1].

Si se requiere producir un objeto mediante CNC y no se dispone del CAD, es necesario aplicar ingeniería inversa al componente. El escáner 3D es una herramienta para poder digitalizar el objeto.

En el campo de la digitalización tridimensional, cabe destacar proyectos como FabScan, Freels y Horus, entre otros,

Andrés Cela es profesor titular del Departamento de Automatización y Control Industrial de la Escuela Politécnica Nacional (EPN), Quito – Ecuador. (e-mail: andres.cela@epn.edu.ec)

bajo licencia Creative Commons. Las aplicaciones desarrolladas para los escáneres 3D, se han limitado a la generación de archivos que contienen nubes de puntos que carecen de información (malla poligonal y posterior código-g) para poder imprimirlas. FabScan se destaca por integrar una herramienta de ejecución en segundo plano para la reconstrucción de la superficie, resultando un modelo de vértices y triángulos. Pero hasta el momento ningún software abierto ha adicionado una herramienta para el laminado de los modelos, y dar como resultado un archivo a punto (código-g), para la impresión 3D.

### C. Equipos Multifuncionales

Existen no más de tres equipos comerciales que han desarrollado soluciones integrales, como por ejemplo AIO Robotics® con Zeus 3D printer, y XYZprinting® con da Vinci 1.0 AiO bajo arquitecturas cerradas.

### D. Problema

Hasta el momento no se ha encontrado una solución paquetizada tanto en hardware como en software, capaz de realizar la reproducción tridimensional (escaneo e impresión 3D) de un modelo, bajo el marco de bajo costo y fuente abierta.

### E. Propuesta

El equipo desarrollado, realiza una integración de hardware y de software. Por una parte los periféricos del escáner son conectados a la misma tarjeta de control de la impresora, y se los aloja dentro de la misma carcasa. El firmware de la tarjeta de control es modificado y adaptado para manejar estos nuevos elementos. Se desarrolla el software, que obtiene la nube de puntos del modelo, y con el uso de herramientas externas de postprocesamiento y laminado que se ejecutan en segundo plano, se realiza la generación automática de un archivo listo para la impresión 3D.

## II. DISEÑO, CONSTRUCCIÓN E INTEGRACIÓN DEL HARDWARE

### A. Construcción de la Impresora 3D

En este proyecto se parte desde una impresora 3D ya desarrollada por la comunidad RepRap. El mecanismo seleccionado es el CORE XY [7]. Adicionalmente se realiza las modificaciones necesarias para que todos los materiales sean accesibles localmente.

### B. Integración del Escáner 3D

#### 1) Funcionamiento Conceptual - Escáner 3D

Un láser lineal se proyecta sobre el modelo a ser digitalizado, y la cámara localizada a un ángulo determinado, captura la deformación del láser sobre la superficie del objeto, en base a la triangulación del láser-cámara-silueta, se puede obtener la información de profundidad. El proceso se repite mientras el objeto gira mediante una plataforma rotatoria; el resultado final es una sucesión de siluetas ubicadas en el espacio tridimensional correspondientes al objeto digitalizado en forma de nube de puntos [6, p. 27].

## 2) Diseño Mecánico

### a) Movimiento

Se ha realizado el diseño de una plataforma rotatoria con transmisión de movimiento mediante piñones en relación de reducción y un rodamiento para la base. Piñones y soportes han sido impresos en 3D.

### b) Caja/ Estructura

La clase de armazón seleccionado, es de tipo cerrada, para contar con los siguientes beneficios: (1) aislamiento a los cambios bruscos de temperatura ambiental, principalmente en la impresión de plásticos altamente sensibles a las fluctuaciones de temperatura, como el ABS, (2) entorno de luz controlada para posterior integración del proceso de escaneo 3D.



Fig. 1. Equipo Multifuncional – Construcción.

## 3) Diseño Electrónico

Los elementos electrónicos son los especificados por la comunidad RepRap para impresión 3D.

Las tarjetas electrónicas de la impresora 3D (RAMPS 1.4 y Arduino Mega 2560), reciben directamente todos los periféricos correspondientes al escáner 3D. Los 2 láseres lineales se conectan a las salidas disponibles para servomotores, y el motor a pasos de la plataforma rotatoria, se conecta a la segunda salida destinada a extrusión.

Se realiza las modificaciones del firmware CNC (Marlin) para que puedan integrarse totalmente como un solo equipo. Se debió modificar las librerías de servomotores, y adicionalmente inhibir los permisivos de temperatura para que el motor del escáner pueda rotar.

Tanto la impresora 3D y la cámara son conectadas a un computador, desde donde se procesarán las imágenes capturadas y se coordinará las acciones de encendido del láser y rotación de la plataforma.

La Fig. 1 presenta la implementación final del equipo multifuncional.

### C. Procura de materiales

Materiales mecánicos, tales como ejes y caja son de fabricación nacional (Ecuador). Todos los componentes electrónicos son comercializados localmente, pero son importados desde países asiáticos, principalmente desde China.

#### III. DISEÑO DEL SOFTWARE

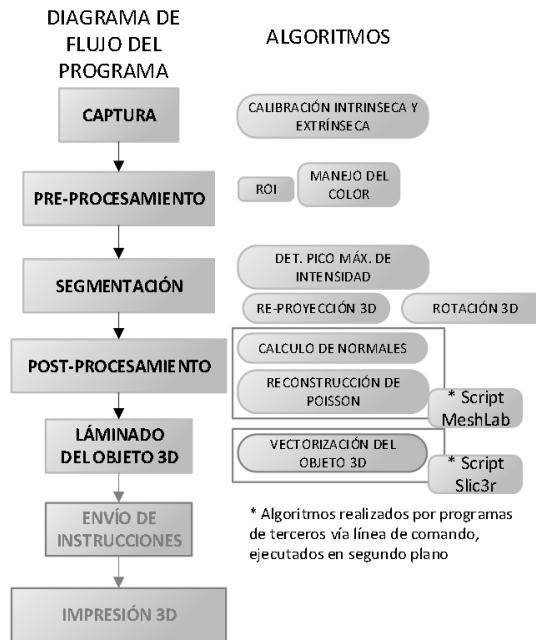


Fig. 2. Diagramas de Flujo: Programa, Algoritmos.

El software desarrollado, nombrado *Escan3r*, se lo ha implementado mediante el lenguaje de programación Python 2.7 (e interfaz gráfica wxPython) y la librería de visión por computador OpenCV 2.4. En la Fig. 2 se muestra el diagrama de flujo del programa y sus algoritmos.

#### A. Captura

##### 1) Adquisición de imágenes

Este proceso se realiza mediante la librería de visión artificial OpenCV. Para fines de pre-procesamiento se capturan dos imágenes en la misma posición, con y sin láser lineal proyectado.

##### 2) Calibración intrínseca y extrínseca

Cuando se realiza la captura de imágenes mediante una cámara, se pierde la información de profundidad del espacio tridimensional del mundo real, por lo cual es importante la obtención de los parámetros intrínsecos y extrínsecos de la cámara para realizar la reconstrucción de la escena 3D con las imágenes obtenidas. El modelo matemático de la cámara pinhole se indica en (1) y se desarrolla matricialmente en (2) [8, p. 374] [9].

$$s \begin{bmatrix} m' \end{bmatrix} = A[R|t]M' \quad (1)$$

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (2)$$

donde:

$X, Y, Z$	Coordenadas de un punto 3D en el espacio del mundo real.
$u, v$	Coordenadas de la proyección del punto en pixeles (punto en la imagen 2D).
$A$	Matriz de la cámara (matriz de parámetros intrínsecos).
$c_x, c_y$	Punto principal que usualmente es el centro de la imagen.
$f_x, f_y$	Distancias focales expresadas en unidades de pixel.
$[R t]$	Matriz de rotación y traslación (parámetros extrínsecos).
$A[R t]$	Matriz de proyección.

Los parámetros intrínsecos (matriz  $A$ ), presentan las características propias de la cámara (generalmente fijas), los parámetros extrínsecos ( $[R|t]$ ) representan la ubicación espacial de la cámara respecto al sistema de coordenadas del mundo real. Estos parámetros se obtienen mediante la captura y procesamiento de patrones conocidos, en este proyecto se utilizó el tablero de ajedrez y la librería OpenCV [8, pp. 381-384].

#### B. Pre-procesamiento

##### 1) ROI

En la configuración de la aplicación del escáner 3D, se incluye parámetros que definen el ROI, encerrando el objeto en un rectángulo y proporcionando desde la adquisición una referencia del tamaño de la imagen objetivo, desechariendo los pixeles no deseados por estar fuera del objeto.

##### 2) Manejo del Color

Se ha optado por la estrategia de eliminar el fondo de la imagen, haciendo la diferencia del canal rojo de la imagen con el láser proyectado y la imagen normal en escala de grises.

#### C. Segmentación

##### 1) Detección del pico máximo de intensidad

La detección del pico máximo de intensidad entre los pixeles que conforman el ancho de la línea láser, ocurre en el pixel más brillante, pero se calcula a nivel de subpixel para disponer de la mayor resolución posible en el eje horizontal de la nube de puntos. En [10], se realiza una evaluación de los estimadores de pico máximo de intensidad. El método seleccionado para este proyecto es el Centro de Masa mostrado en (3).

$$CM_j = \frac{\sum_i^w v_j \cdot i}{\sum_i^w v_j} \quad (3)$$

donde:

$i, j$  Índice de columna y fila de la imagen.

$v$  Fila del pixel de la imagen

$w$  Ancho de la imagen (en pixeles)

##### 2) Re-proyección de la escena 3D

En la triangulación, la imagen con el perfil láser debe ser reproyectado a las coordenadas del mundo real, para lo cual la transformada de perspectiva tiene que realizar un proceso inverso, empleando la matriz inversa  $A^{-1}$ , para emitir un rayo

desde el centro de la cámara ( $c_x, c_y$ ), que pase por las coordenadas ( $u, v$ ) del perfil en la imagen hacia la escena 3D.

Luego se procede a realizar el cálculo de la intersección entre el rayo proyectado, con el plano formado por el láser lineal. Las coordenadas resultantes, son el perfil capturado en la imagen 2D, reprojectado a coordenadas 3D del mundo real. En la ecuación (4) [11, p. 76], se ha despejado las coordenadas cartesianas reproyectadas en base a parámetros conocidos.

$$X_l = \frac{d}{d \cdot v} v \quad (4)$$

donde:

- $X_l$  Coordenadas cartesianas reproyectada (silueta - nube de puntos).
- $d$  Distancia mínima del plano hacia el origen de coordenadas (centro de la cámara).
- $v$  Vector de dirección de la línea de reproducción (dado por:  $A^{-1} \times m'$ ).

### 3) Rotación y apilamiento de perfiles 3D

Para realizar el apilamiento de los perfiles obtenidos en el espacio tridimensional, se realiza la rotación en el espacio tridimensional en base al eje de la plataforma rotatoria que corresponde al eje "y" de la nube de puntos. La ecuación (5), representada matricialmente en (6), ha sido aplicada para realizar la rotación del perfil en un ángulo  $\theta$ , correspondiente a la rotación realizada por la plataforma. Las iteraciones de captura de los perfiles multiplicadas por el ángulo  $\theta$ , deben sumar 360°.

$$p' = R_y(\theta)p \quad (5)$$

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & 0 & \sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ -\sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (6)$$

donde:

- $x, y, z$  Coordenadas del punto  $p$
- $x', y', z'$  Nuevas coordenadas del punto rotado  $p'$
- $\theta$  Ángulo de rotación
- $R_y(\theta)$  Matriz de rotación en el eje y, en función de  $\theta$

### D. Post-Procesamiento

Una vez obtenida la nube de puntos, se procede con el cálculo de la superficie (malla de polígonos), mediante la Reconstrucción de Poisson [12].

#### 1) Macro-algoritmo del script de reconstrucción

Como se detalla en el diagrama de flujos, Fig. 2, el post-procesamiento se realiza mediante el uso de un script, donde se listan todos los parámetros e instrucciones que deben de ejecutarse en el programa Meshlabserver que corre en segundo plano la reconstrucción, vía línea de comandos.

En la Tabla 1, se muestra el macro-algoritmo que contienen el script de reconstrucción de superficies.

En la Fig. 3, se muestra el procesos de reconstrucción, de la nube de puntos, literal b), variando el número de muestreo y profundidad de reconstrucción, en los literales d), e) y f).

### E. Laminado

Conseguida la superficie a partir de la nube de puntos del objeto escaneado en formato .stl, es necesario un último procesamiento previo a la impresión, llamado laminado. Se debe ingresar un archivo de parámetros pre-configurado con los valores de maquinado (diámetro de la boquilla, altura de capa, ancho de perímetro, porcentaje de relleno y temperatura de extrusión). El resultado del laminado de un objeto es la sucesión de instrucciones de maquinado (movimiento de ejes, calentamiento de boquillas, etc.) llamado código-g, una capa sobre otra, en base al proceso de manufactura de deposición fundida. Nuevamente se recurre a scripts para detallar las configuraciones de laminado que se ejecutan en segundo plano por medio del software slic3r, dando como resultado un archivo listo para impresión.

TABLA I  
MACRO-ALGORITMO: RECONSTRUCCIÓN DE SUPERFICIES

Entradas: nube de puntos, número de muestras, n.º de vecindario, profundidad de reconstrucción

Procedimiento:

1. Muestreo (Opcional)
  - 1.1. Número de muestras de la nube de puntos, mediante el algoritmo "Poisson-disk"
2. Reconstrucción de Poisson
  - 2.1. Cálculo de normales con el número de vecindario del conjunto de muestras.
  - 2.2. Reconstrucción de Poisson, en base a profundidad de reconstrucción.

Salida: archivo.stl (malla de polígonos)

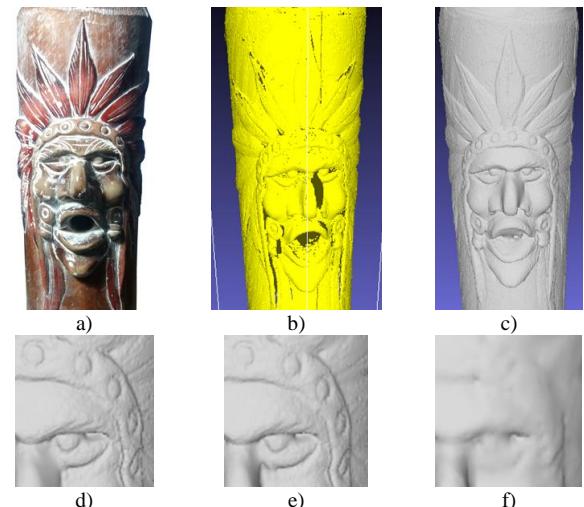


Fig. 3. Reconstrucción de nube de puntos. a) Modelo Original, b) Nube de puntos, c) Reconstrucción de superficie, d), e) y f) Comparativa de reconstrucción de superficie, con diferente configuración de parámetros.

## IV. METODOLOGÍA

A continuación se lista la metodología para la ejecución de ensayos y posterior tratamiento de datos:

(1) Se consideraron 3 objetos de prueba, con diferentes morfológicas y materiales. (2) Los objetos son escaneados en 3D, para obtener la nube de puntos. (3) El software procesa automáticamente las nubes de puntos. Se realiza la reconstrucción (mallado poligonal) y laminado (obtención del código-g). (4) Los modelos digitales son impresos en 3D.

(5) Se realizan 5 mediciones aleatorias de diferentes detalles, tanto del objeto original como del modelo reproducido. (6) Las medidas se repiten 3 veces, y se considera su promedio para posterior análisis.

## V. RESULTADOS Y DISCUSIÓN

En la Fig. 4 se encuentran 2 de los 3 objetos de prueba (“Pájaro” y “Busto”), donde se aprecia el modelo original y su reproducción.

En la Tabla 2, se han ordenado y presentado los datos de las mediciones de los 3 objetos de prueba. En las últimas columnas se calcula el error absoluto en (mm), y el error relativo porcentual. Para fines de visualización y comparación en la gráfica, los errores se muestran en valor absoluto.

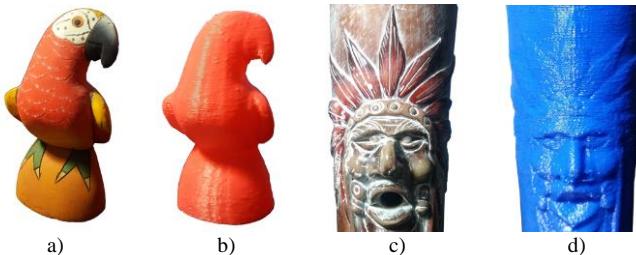


Fig. 4. Pruebas – Proceso completo de reproducción. a) y b) Objeto de prueba y reproducción: “Pájaro”, c) y d) Objeto de prueba y reproducción: “Busto”.

TABLA II  
COMPARACIÓN DIMENSIONAL

Muestra	Medida modelo (mm)	Medida Imp. 3D (mm)	Error (mm)	Error (%)
<b>Objeto de Prueba 1</b>				
1	89,52	90,7	1,18	1,32
2	78,51	77,48	1,03	1,31
3	63,44	64,51	1,07	1,69
4	11,73	12,58	0,85	7,25
5	7,77	9	1,23	15,83
<b>Objeto de Prueba 2</b>				
6	110,77	111,98	1,21	1,09
7	53,25	54,6	1,35	2,54
8	29,85	30,73	0,88	2,95
9	24,15	23,06	1,09	4,51
10	5,6	4,75	0,85	15,18
<b>Objeto de Prueba 3</b>				
11	45,32	46,53	1,21	2,67
12	36,78	37,88	1,10	2,99
13	22,41	21,34	1,07	4,77
14	18,24	17,18	1,06	5,81
15	5,22	4,46	0,76	14,56

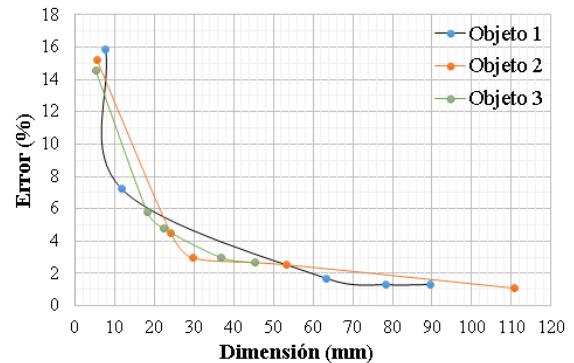


Fig. 5. Error porcentual en función de las dimensiones del modelo.

En la Fig. 5, se grafica el error porcentual en función de las dimensiones obtenidas en las diferentes mediciones, de los 3 objetos de prueba.

### A. Análisis de resultados

El error máximo en las medidas generales de los objetos es de hasta  $\pm 2.5\%$ , experimentando en mediciones de 50 a 100 mm. En la contrastación dimensional de detalles de 5 a 10mm se observan errores de hasta el  $\pm 15\%$ .

Detalles menores a 1 mm se van suavizando en la manipulación del modelo 3D obtenido. La primera desviación se encuentra en la adquisición, ya que la línea láser tiene un grosor de aproximadamente 0.4 mm. La segunda ocurre principalmente en el proceso de reconstrucción de superficies. No ha sido posible comprobar el grado de desviación en el proceso de laminado, ya que no se ha encontrado un software que reconstruya el código-g y permita realizar mediciones digitales. Finalmente la impresión 3D también inserta un error dimensional, de aproximadamente  $\pm 0.25$  mm.

El proceso de reconstrucción superficial tiende a llenar los agujeros o cavidades más profundas del modelo, porque la nube de puntos carece de información en estas áreas por la falta de detección de la línea láser, ya que la luz se va perdiendo en estas zonas.

El proceso totalmente automático para obtener el archivo de código-g (a punto para impresión 3D) del objeto digitalizado, tiene un éxito del 75%, esto se debe a que los parámetros pre-configurados en los algoritmos de reconstrucción superficial no son universales para todos los tipos de morfologías de las piezas. Principalmente se ha debido re-configurar el vecindario de cálculo de normales. Para piezas con morfologías difusas, no ha sido posible realizar este proceso automático, o puntualmente su calidad no ha sido aceptable, y se ha debido realizar varios escaneos e integrarlos manualmente por software externo.

### B. Limitaciones y Futuros Trabajos

El algoritmo de escaneo permite actualmente el barrido del láser por todo el objeto una sola vez, se plantea a futuro la implementación del algoritmo de *Iteración de Puntos Cercanos* (*por sus siglas en inglés, ICP* [13]) para realizar la alineación y registro de varios escaneos de la misma pieza desde varias posiciones para acceder a las cavidades no alcanzadas en un solo barrido.

Actualmente, por medio del escáner 3D se obtiene la forma de las piezas y sus medidas para utilizar el archivo .stl generado

como molde para redibujar el objeto en una aplicación CAD. Se podría implementar el reconocimiento de formas 3D (básicas: prismas, cilindros, conos y esferas o desde ShapeNet [14]) en los modelos obtenidos, para mayor facilidad en el análisis y modificación en aplicaciones nuevas, rediseños o mejoras en el campo de la ingeniería inversa.

## VI. CONCLUSIONES

El equipo multifuncional cumple con la expectativa inicial de ser una solución integral para la reproducción de objetos 3D en el campo del prototipado rápido.

El error absoluto en el proceso de reproducción tiende a 1 mm, lo cual no es aceptable para aplicaciones de mecánica de precisión, pero puede ser adecuado para la reproducción de accesorios donde este error se pueda desestimar.

Los procesos automáticos de reconstrucción de superficie y laminado, aunque no han sido efectivos al 100%, son viables y se pueden mejorar con técnicas dinámicas de calibración automática de parámetros dependiendo del objeto a escanearse.

Los modelos digitalizados en 3D pueden ser utilizados para modificación o mejora como parte de procesos de ingeniería inversa.

## VII. REFERENCIAS

- [1] T. A. Almquist, «Method of and apparatus for making a three-dimensional product by stereolithography». USA Patente US5174931 A, 29 12 1992.
- [2] T. A. Almquist, «Thermal stereolithography». USA Patente US 5569349 A, 1996 10 29.
- [3] C. W. Hull, P. F. Jacobs, K. A. Schmidt, D. R. Smalley y W. A. Vinson, «Method of building three dimensional objects with sheets». USA Patente US 5637169 A, 10 06 1997.
- [4] C. R. Deckard, «Method for producing parts by selective sintering». Patente US5639070 A, 17 07 1997.
- [5] B. v. d. Berg, S. v. d. Hof y E. Kosta, 3D Printing Legal, Philosophical and Economic Dimensions, vol. 26, The Hague: Asser Press, Springer, 2016.
- [6] F. Bellocchio, 3D Surface Reconstruction, Multi-Scale Hierarchical Approaches, 1 ed., Crema/Milano: Springer Science, 2013, pp. 1-7; 21-42.
- [7] RepRap, «CoreXY,» RepRap, 31 03 2016. [En línea]. Available: <http://reprap.org/wiki/CoreXY>. [Último acceso: 23 10 2017].
- [8] G. Bradski y A. Kaehler, Learning OpenCV, Sebastopol: O'Reilly, 2008.
- [9] M. Breier, P. Möller, W. Li, M. Bosling, T. Pretz y D. Merhof, «Accurate laser triangulation using a perpendicular camera setup to assess the height profile of PCBs,» 2015 IEEE International Conference on Industrial Technology (ICIT), pp. 1613-1618, 2015.
- [10] J. Forest, J. Salvi, E. Cabruja y C. Pous, «Laser stripe peak detector for 3D scanners. A FIR filter approach,» Universitat de Girona - Computer Vision and Robotics Lab, 2004.
- [11] K. Torneslev, «3D scanning using multibeam laser,» IMM, Lyngby, 2005.
- [12] M. Kazhdan, M. Bolitho y H. Hoppe, «Poisson Surface Reconstruction,» Eurographics Symposium on Geometry Processing, p. 10, 2006.
- [13] P. Besl y N. D. McKay, «A method for registration of 3-D shapes,» IEEE Trans. on Pattern Analysis and Machine Intelligence, Los Alamitos, CA, USA, 1992.
- [14] A. X. Chang, T. Funkhouser y L. Guibas, «ShapeNet: An Information-Rich 3D Model Repository,» Princeton Vision & Robotics, 2015.



**Leombardo Hernández**, realizó sus estudios secundarios en el Instituto Tecnológico Superior “Central Técnico”, donde obtuvo el título de Bachiller Técnico Industrial, Especialidad: Electrónica. Actualmente es estudiante (egresado) de la carrera de Ingeniería en Electrónica y Control en la Escuela Politécnica Nacional.



**Andrés Cela**, estudió su grado en la Escuela Politécnica Nacional obteniendo el título de Ingeniero en Electrónica y Control, después de dos años de trabajar como docente en la misma Universidad realizó los estudios de Maestría en la Universidad de Alcalá obteniendo el título de Master en Sistemas Electrónicos Avanzados. Posteriormente realizó un año de investigación en la Universidad Carlos III de Madrid en temas relacionados con la Visión Artificial y Control Fuzzy. A partir de sus trabajos de investigación ha publicado varios artículos en las revistas IEEE, Politécnica, Sensors, entre otras. Se ha desempeñado también como Coordinador de Posgrados de la FIEE. Actualmente desempeña el cargo de jefe del área de circuitos eléctricos de la misma facultad.

# Optimización de un Sistema de Reconocimiento de Idioma Basado en la Representación Vectorial de Fonemas

## *Optimising a Language Recognition System Through Phoneme-Based Vector Representation*

Francisco Charro, Marco Herrera, Nataly Pozo y Andrés Rosales

**Resumen —** En este artículo se analiza la representación vectorial de fonemas como una alternativa para mejorar un sistema de reconocimiento de idioma (LID). Se estudian las arquitecturas, CBOW (Continuous Bag-of-Words) y Skip-gram propuestas por Mikolov. Estos modelos permiten predecir palabras dentro de un contexto mediante la generación de vectores n-dimensionales. En este trabajo se analizará la aplicación de estos modelos en unidades fonéticas más pequeñas o n-gramas.

**Palabras Clave —** Representación Vectorial, Reconocimiento de Idioma, Skip-gram, n-gramas, embeddings.

**Abstract -** This article analyzes vector representation of phonemes as an alternative to improve a language identification system (LID). CBOW (Continuous Bag-of-Words) and Skip-gram architectures proposed by Mikolov are studied. These models allow predicting words within a context by generating n-dimensional vectors. In this work we will analyze the application of these models in smaller phonetic units or n-grams.

**Keywords -** Vector Representation, Language Recognition, Skip-gram, n-grams, embeddings.

### I. INTRODUCCIÓN

El reconocimiento de idioma es una tarea implícita en la mayoría de sistemas informáticos que interactúan con los seres humanos, sin embargo el mejoramiento de su rendimiento sigue en desarrollo. Se han probado varios enfoques, los más básicos desde la semántica formal con modelos simbólicos que siguen reglas fijas, tienen un procesamiento rápido pero ignoran las sutilezas del lenguaje y el contexto que al no ser considerados resulta en un sinfín de extensiones y alternativas [1]. Una opción válida a esta técnica es la semántica distribucional, que convierte el lenguaje en un modelo matemático, cambiando de un enfoque simbólico a uno estadístico, y de esta manera se pueden estudiar las relaciones

Los autores pertenecen al Departamento de Automatización y Control Industrial de la Escuela Politécnica Nacional, Ladrón de Guevara, E11-253, Quito, Ecuador. (e-mail: {francisco.charro, marco.herrera, nataly.pozov, andres.rosales}@epn.edu.ec)

entre distribución y significación aumentando la probabilidad de éxito [2]. Los modelos estadísticos actuales basan su función de distribución en la representación vectorial de palabras, definen la distancia entre ellas en un espacio vectorial, y construyen un modelo de predicción en el que se puede pronosticar la siguiente palabra que va a aparecer en base a la palabra actual y el contexto en el que se circunda [1].

La aportación de este documento está dirigido a estudiar los efectos de la representación vectorial a nivel fonético o poli-fonético (varios fonemas unidos en una sola unidad) en la identificación de idioma.

El sistema LID implementado se presenta en la Figura 1. Consta de 2 etapas principales, el Front-End donde se extraen los parámetros principales de cada idioma, en este caso la representación vectorial fonética, y el Back-End que modela los idiomas a reconocer con técnicas basadas en i-Vectores. Estos procesos se los realiza tanto para los datos de entrenamiento como para los evaluados. Utiliza un clasificador basado en regresión logística y máxima verosimilitud estadística.

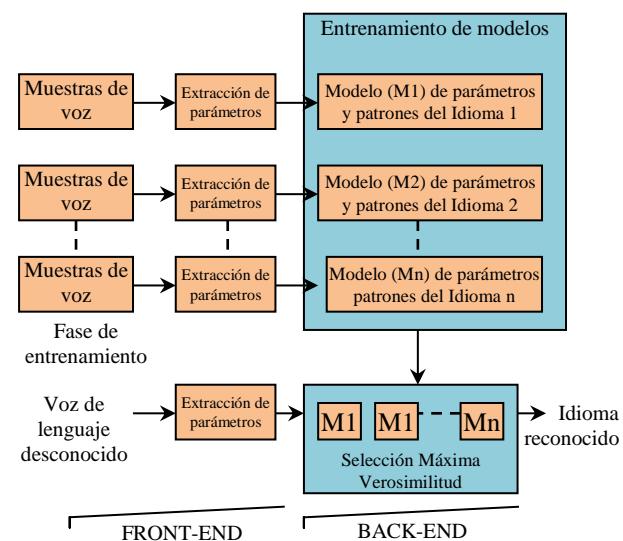


Figura 1. Diagrama básico de un sistema LID

La técnica de modelado con la que se comprobó la calidad de la representación vectorial es conocida como PRLM (modelado de idioma con reconocimiento fonético) en paralelo o (PPRLM) [3], implementada en investigaciones anteriores por el Grupo de Tecnología del Habla (GTH) de la Universidad Politécnica de Madrid (UPM) a la que se tuvo acceso para la experimentación. Este sistema permite también modelar la frecuencia de ocurrencia de las secuencias de fonemas en cada uno de los idiomas a reconocer lo que permite centrarse en la representación vectorial únicamente [4].

Los modelos se crean a partir de las secuencias de fonemas generadas con reconocedores existentes y así se obtiene una secuencia diferente para cada reconocedor. Por lo tanto la calidad y tamaño de la base de datos es un factor importante.

El presente trabajo se ha verificado sobre la base de datos KALAKA-3 desarrollada para la competición europea de reconocimiento de idioma Albayzin LRE 2012 [5] en la condición plenty-closed, es decir se poseen datos de audio etiquetados en 6 idiomas (vasco, catalán, inglés, gallego, portugués y español) para el entrenamiento (plenty) y no admite otros diferentes para la evaluación (closed). Se tienen además 3 reconocedores de fonemas que utilizan Modelos Ocultos de Markov (HMM's) (húngaro, ruso y checo) que generan un conjunto de secuencias de fonemas para cada uno de los 6 idiomas (en el presente caso es posible obtener, 3 HMM's  $\times$  6 idiomas = 18 distribuciones de idioma o PRLM's). Durante el reconocimiento, la muestra a ser evaluada es procesada por cada PRLM y sus salidas son promediadas en el dominio logarítmico como sistemas independientes, la distribución de los datos a evaluar se compara con los modelos y se generan scores basados en las verosimilitudes.

## II. DESCRIPCIÓN DEL SISTEMA

El sistema implementado está basado en un sistema PPRLM donde se tienen seis modelos correspondientes a los seis idiomas de la base de datos KALAKA-3 por cada reconocedor de fonemas disponible o HMM (Húngaro, Ruso o Checo).

En la Figura 2 se muestra el modelo planteado, tomando como ejemplo el HMM de Húngaro y el idioma Vasco. Este sistema corresponde al que presentó mejores métricas, pero es necesario explicar las pruebas previas hasta alcanzar este modelo propuesto. A partir de ahora, llamaremos "embeddings" [6] a la representación vectorial de los fonemas.

### A. FRONT-END

#### a. Creación de las listas de entrenamiento

Las listas de entrenamiento se crean a partir de los datos pre-adquiridos por los reconocedores fonéticos disponibles de KALAKA-3. La lista de entrenamiento consta de 4656 archivos y se los separa en dos bloques, el bloque inferior (train\_DOWN, con menor cantidad de elementos) para la generación de los embeddings y el bloque superior (train\_UP) se utilizará en la creación de los i-Vectores de entrenamiento de cada modelo PRLM, esta división es necesaria para evitar el sobrentrenamiento del sistema. Las listas empleadas para desarrollo constan de 458 elementos y las de evaluación de 941 elementos.

Los parámetros a definir en esta etapa son: 1) el porcentaje de cada bloque superior e inferior del entrenamiento y 2) el

número de n-gramas en los que se agruparán los fonemas de cada archivo de audio.

#### b. Entrenamiento de los embeddings

Se utilizan las listas de entrenamiento "train\_DOWN" para la generación de embeddings a través del toolkit Wor2Vec propuesto por Mikolov [7]. Este software genera una representación vectorial de un poli-fonema utilizando el modelo CBOW o Skip-gram. En esta experimentación se probaron las dos configuraciones, y se obtuvieron vectores para bi-gramas, tri-gramas o cuatri-gramas hasta obtener la mejor "calidad" de los vectores.

Generamos una nueva lista de todos los n-gramas contenidos en las listas train\_DOWN (en total seis listas, una por idioma), se cuenta el número de apariciones de los n-gramas presentes, se ordenan por frecuencia de aparición y se calcula para cada uno, un vector de la dimensión requerida según el modelo seleccionado (CBOW/Skip-Gram) y sus parámetros.

Estas listas contienen todo el vocabulario de fonemas para cada uno de los 6 idiomas, representados vectorialmente para un HMM determinado.

#### a. Representación vectorial de fonemas o embeddings

Para generar las listas de la secuencia de n-gramas de cada archivo de entrenamiento (train\_UP), desarrollo y evaluación, se realiza un mapeo de estos archivos con la lista de embeddings para construir los pares fonema-vector que son los que nos interesan. Si un fonema no aparece en la lista de embeddings se le asigna un vector "missing grams" de valor cero. Al final se tendrán todos los archivos de entrenamiento, desarrollo y evaluación mapeados con la lista de embeddings.

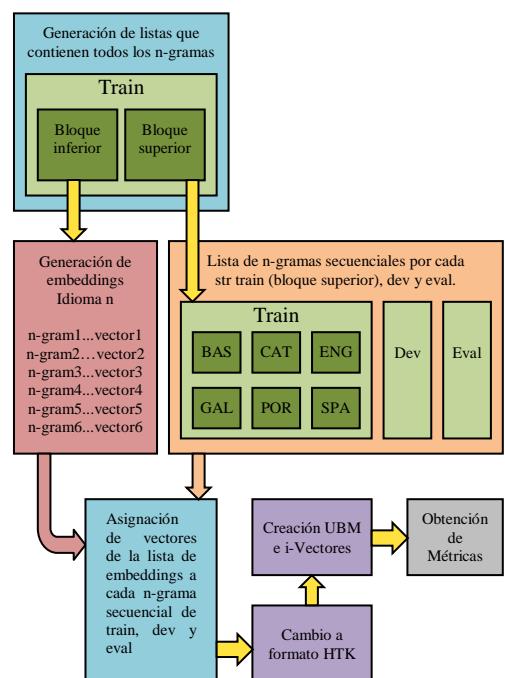


Figura 2. Diagrama de bloques del sistema LID implementado

A estas nuevas listas que contienen la información de la secuencia de n-gramas y sus embeddings se agrega información de contexto recalculando su valor con información de sus vecinos. En esta nueva y definitiva representación los vectores se calculan ponderando con un 50% del valor del vector actual, un 25% de su vector anterior y un 25% del vector siguiente. El tiempo de cálculo aumenta pero las métricas mejoran considerablemente.

Para la generación de i-Vectores mediante el fusionador disponible, se requiere cambiar los datos a formato binario HTK mediante la función de Matlab writehtk.m. El contenido de estos archivos de extensión .fea es el mismo que el de las listas que contienen los vectores con contexto.

### B. BACK-END

La experimentación contempla la creación de modelos independientes de idiomas o UBM (Universal Background Model) que son modelos de mezclas gaussianas (GMM) entrenados con una gran cantidad de datos. Se crearon UBM's para cada uno de los 6 idiomas a reconocer utilizando uno de los 3 HMMs disponibles, se modifican los parámetros de los algoritmos CBOW o Skip-gram hasta alcanzar los valores óptimos. En la experimentación final se crearon los seis sistemas correspondientes a cada idioma.

El UBM genera un modelo de Gaussianas y luego se entrena los i-Vectores que representan la información de los ficheros de entrenamiento, desarrollo y evaluación. A partir de estos i-Vectores se obtienen las métricas en el sistema de fusión.

## III. RESULTADOS DE LAS PRUEBAS

Las métricas utilizadas para evaluar el rendimiento del sistema son: la Función de Costo de Detección (Fact), y la Función de Costo de Decisión Promedio (Cavg) que miden el costo de decisión provocado por las tasas de error de falsa aceptación y de falso rechazo generadas en la evaluación [8].

La experimentación partió de un sistema inicial al que se le han modificado sus parámetros uno a uno hasta encontrar una configuración óptima.

Este sistema inicial tiene 2 tipos de parámetros, los propios del sistema LID y los correspondientes a la generación de los embeddings.

### A. SISTEMA INICIAL

En el sistema inicial se utilizó un solo UBM del HMM de Húngaro [9] con los vectores de parámetros de los seis idiomas para entrenar los i-Vectores. En primer lugar se crearon las listas para los 6 idiomas, que contienen todos los poli-fonemas entrenados y su representación vectorial ordenados por frecuencia de aparición y de dimensión N establecida manualmente. Seguidamente se generaron las listas de poli-fonemas ordenados secuencialmente para cada archivo del set de entrenamiento, desarrollo y evaluación. Se realizó la búsqueda de cada fonema en las 6 listas de embeddings y a continuación se concatenaron todos los vectores en uno solo de dimensión 6N. Si el fonema no existía en la lista de embeddings de algún idioma se le asignó el vector de "missing grams" vector de la misma dimensión N con un valor cercano a cero.

La experimentación partió de la configuración descrita en la Tabla 1 para los parámetros de los embeddings y de la configuración de la Tabla 2 para los parámetros del sistema

LID. En la Tabla 1 se muestra un sistema de bi-gramas que utiliza a Skip-gram (SG) como el modelo utilizado, con una ventana de 8, muestreo negativo de 25, 10 iteraciones y la lista de entrenamiento dividida (el 40% se utilizó en la generación de los embeddings y el 60% como set de entrenamiento para la generación de los i-Vectores). La Tabla 2 muestra el tamaño del vector de cada lista de los embeddings generados en este caso 40, el vector concatenado total fue de 240, el número de gaussianas entrenadas de 128 y la técnica de reducción de dimensionalidad PCA (Principal Components Analysis) de dimensión 100.

"CW/SG" es el modelo utilizado CBOW o Skip-gram, "W" es la ventana de datos del modelo, "NS" el valor para muestreo negativo e "Iter" el número de iteraciones para el algoritmo de Back-Propagation que recalcula los pesos.

Los resultados de la Tabla 4, muestran que un mayor número de gaussianas mejora las métricas por tener un espacio gaussiano más preciso, lo cual era esperado [10]. Por otro lado la Tabla 3 y 5 muestran que PCA no mejora el sistema y aumenta el espacio de almacenamiento requerido por los vectores.

Una vez analizados estos primeros resultados se continuó con otro enfoque para la creación del UBM e i-Vectores, se decidió probar sistemas individuales de cada idioma y fusionarlos, este punto de vista permitió el desarrollo de un nuevo sistema que fue el definitivo. El sistema inicial anterior sin embargo, permitió extraer conclusiones importantes como el número de gaussianas y la pertinencia de PCA

N-gramas	CW/SG	W	NS	Iter	Lista entrenamiento
2	SG	8	25	10	60-40

Tabla 1. Parámetros iniciales para los embeddings

Vector	Vector total	Gauss	Red. Dimensión
40	240	128	PCA 100

Tabla 2. Parámetros iniciales LID

EXP. N°	Gauss	Vector	Vector total	PCA	Fact	Cavg
1	128	40	240	100	0,5859	0,2665
2	128	40	240	120	0,5830	0,2801
3	128	100	600	100	0,6419	0,2797

Tabla 3. Pruebas de tamaño del vector con PCA

EXP. N°	Gauss	Fact	Cavg
1a	16	0,7045	0,3037
1b	32	0,6318	0,2860
1c	64	0,5942	0,2731
1	128	0,5859	0,2665
1d	256	0,5644	0,2615
1e	512	0,5288	0,2469

Tabla 4. Pruebas para el número de gaussianas

EXP. N°	Gauss	Vector	Vector total	PCA	Fact	Cavg
4	32	40	240	120	0,6340	0,2876
5	32	20	120	NO	0,6132	0,2762

Tabla 5. Experimentación sin PCA

EXP. N°	Gauss	Vector	Vector total	Fact	Cavg
6	32	20	20	0,6879	0,3058

Tabla 6. Experimento inicial para un sistema de un solo idioma

EXP.Nº	Gauss	Vector	Vector total	PCA	Fact	Cavg
7	32	50	50	No	0,7295	0,3281
8	32	10	10	No	0,7727	0,3357

Tabla 7. Pruebas de tamaño del vector sin PCA

EXP.Nº	N-gramas	Vector	Vector total	PCA	Fact	Cavg
9	3	50	50	No	0,7873	0,3445

Tabla 8. Pruebas con trigramas

EXP. N°	Gauss	Vector	Contex	Fact	Cavg
10	32	20	Yes	0,5946	0,2760

Tabla 9. Pruebas agregando información de contexto

EXP. N°	Gauss	Vector	Contex	Fact	Cavg
11	32	10	Yes	0,5784	0,2703
12	32	8	Yes	0,6469	0,2959

Tabla 10. Pruebas para el tamaño del vector con contexto

EXP. N°	Gauss	Vector	PCA	Contex	Fact	Cavg
13	32	10	No	Yes	0,6139	0,2786

Tabla 11. Prueba con la lista de entrenamiento completa

#### A. SISTEMA FINAL

Para el experimento N°6, se creó un sistema utilizando la plantilla de embeddings obtenida para el idioma Vasco similar al experimento N°5, este sistema es el presentado en la Figura 1. Se utilizó un tamaño de vector de 20 para entrenar directamente el UBM y los i-Vectores. Este nuevo sistema aportó métricas inferiores en un 10%, pero se consideraron mejores desde la perspectiva de que se trató de un solo sistema para un idioma y aún quedaban otros 5 para fusionarlos.

##### a) Tamaño del vector

Para el experimento N°7 se utilizó un tamaño de vector de 50 y para el experimento N°8 el tamaño fue de 10.

##### b) N-gramas

Las métricas del experimento N°9 fueron inferiores, el sistema de bi-gramas es mejor que el de tri-gramas en un 7,34% en Fact y en 4,76% en Cavg. Revisamos el mapeo de fonemas en el sistema de tri-gramas y nos percatamos de un número elevado de “missing-grams”. Continuamos utilizando al experimento N°6 con bi-gramas como experimento base y dadas las características de los sistemas LID y las propias de los embeddings, introducimos información de contexto para verificar su aporte.

##### c) Incorporación del contexto

Se introdujo información de contexto en los embeddings, realizando operaciones algebraicas entre cada elemento de los vectores para recalcularlos, utilizando su propia información, la del embedding anterior y del siguiente. La fórmula aplicada para cada vector de la secuencia fonética fue:

$$\text{Vect}'_{\text{act}} = 0,5 * \text{Vect}_{\text{act}} + 0,25 * \text{Vect}_{\text{ant}} + 0,25 * \text{Vect}_{\text{sig}}$$

Donde:  $\text{Vect}'_{\text{act}}$  es el vector recalculado,  $\text{Vect}_{\text{act}}$  es el vector actual,  $\text{Vect}_{\text{ant}}$  es el vector anterior y  $\text{Vect}_{\text{sig}}$  es el vector siguiente.

Los resultados obtenidos fueron los presentados en la Tabla 9.

#### d) Tamaño del vector con contexto

N-gramas	Listas %	CW/SG	W	NS	Iter	Vector	Gauss	Fact	Cavg
2	60-40	SG	8	25	10	10	32	0,5784	0,2703

Tabla 12. Resumen de los parámetros del experimento N°11

EXP. N°	CW/SG	W	NS	Iter	Fact	Cavg
14	CW	8	25	10	0,6869	0,3204
15	SG	5	25	10	0,6591	0,2955
16	SG	10	25	10	0,5761	0,2623
17	SG	8	10	10	0,6353	0,2833
18	SG	8	35	10	0,5877	0,2808
19	SG	8	25	25	0,6142	0,2801

Tabla 13. Pruebas con diferentes configuraciones en W2V

n gram	Listas %	CW/ SG	W	NS	Iter	Vector	Gauss	Fact	Cavg
2	60-40	SG	10	25	10	10	32	0,5761	0,2623

Tabla 14. Parámetros finales con mejores resultados

En la Tabla 10 se muestran los resultados para varios tamaños de vector, los mejores resultados se obtuvieron para vectores de dimensión 10 (experimento N°11). Este tamaño fue el definitivo para las siguientes experimentaciones, y concluimos con la determinación de los parámetros del sistema LID.

#### e) Tamaño del set de entrenamiento dedicado a los embeddings

Se repitió el experimento N°11 variando el set de entrenamiento, utilizando las listas completas tanto para entrenar el sistema como para generar los embeddings, es decir, sin bloques superior e inferior.

Los resultados fueron ligeramente inferiores, además que el tiempo de procesamiento incrementó considerablemente, esto se debe a que el sistema se sobreentrena. Por tanto, se eligió trabajar a partir de este punto con listas de entrenamiento divididas 60%-40%. Las siguientes pruebas se realizaron para establecer la mejor configuración de los modelos CBOW y Skip-gram.

#### f) Parámetros CBOW/Skip-gram

La Tabla 12 muestra un resumen completo de los parámetros óptimos obtenidos hasta el momento, corresponden al experimento N°11.

Realizamos los cambios de un parámetro a la vez, manteniendo invariables los demás. Se muestran en la Tabla 13 los valores de Fact y Cavg para esta experimentación (se resalta el parámetro modificado).

#### B. EXTRACCIÓN DE PARÁMETROS ÓPTIMOS

De las pruebas del apartado anterior se obtuvieron los parámetros que proporcionan los mejores resultados. Se observó un cambio positivo únicamente en el tamaño de la ventana, los demás parámetros no aportaron y, por tal motivo, los parámetros finales seleccionados fueron:

Estas métricas se han obtenido para el HMM (Húngaro), un idioma (Vasco) y para 32 gaussianas.

El sistema final propuesto consistió en realizar la fusión de los seis sistemas correspondientes a cada uno de los seis idiomas

y con 512 gaussianas con la mejor configuración que se ha mostrado en la Tabla 14.

EXP. Nº	Idioma	Fact	Cavg
20	Vasco	0,5125	0,2438
21	Catalán	0,4969	0,2440
22	Inglés	0,5077	0,2451
23	Gallego	0,5195	0,2514
24	Portugués	0,5416	0,2623
25	Español	0,5257	0,2544
26	Fusión	0,3754	0,1870

Tabla 15. Métricas para los 6 idiomas y su fusión

EXP. Nº	SISTEMA 1	SISTEMA 2	Fact	Cavg	Improv. Fact	Improv. Cavg
27	None	Embedding Vasco	0,5125	0,2438	NA	NA
28	PPRLM 3-gram Hung	None	0,3284	0,1716	NA	NA
29	PPRLM 3-gram Hung	Embedding Vasco	0,2917	0,1553	11,2%	9,5%
30	PPRLM 3-gram Hung	6 Embeddings	0,2764	0,1493	15,8%	12,9%

Tabla 16. Fusión de los sistemas LID Embeddings-PPRLM

Los datos obtenidos para cada sistema de un idioma (experimentos Nº20 al Nº25) reflejan métricas que están dentro de un mismo rango, los valores de Fact y Cavg son relativamente elevados en comparación con otros sistemas LID.

El experimento Nº26 es la fusión de los 6 sistemas de cada idioma. Claramente se observaron mejoras en las métricas finales producto de la comparación entre la fusión de los 6 sistemas con el mejor de ellos, en este caso Vasco.

#### IV. RESULTADOS DE LA FUSIÓN

La experimentación final realizada, mostrada en la Tabla 16, fue la fusión de estos sistemas de embeddings con otros de diferente naturaleza y verificar su aportación.

La fusión se ejecutó sobre un sistema PPRLM basado en un modelo de 3-gramas con suavizado Witten Bell existente. El suavizado permite evitar las probabilidades cero de los n-gramas no vistos en el texto, redistribuyendo estas probabilidades y compensando a los n-gramas menos frecuentes o no vistos. A diferencia de otras técnicas de suavizado, utiliza como información adicional el contexto.

El experimento Nº27 replica los resultados obtenidos en el experimento Nº20 sin fusiones con otros sistemas, en el experimento Nº28 están únicamente los resultados para el sistema PPRLM para 3-gramas que ha presentado los mejores resultados.

Para el experimento Nº29 se han fusionado el sistema PPRLM y el sistema de embeddings con mejores métricas, en este caso Vasco. Presentaron una mejora respecto al sistema PPRLM de 11,2% en Fact y 9,5% en Cavg.

El último experimento Nº30 muestra los resultados obtenidos para la fusión del sistema PPRLM con los 6 sistemas de embeddings, la mejora fue de 15,8% en Fact y 12,9% en Cavg.

Estos resultados demuestran la aportación positiva de los sistemas de embeddings cuando se combinan con otros sistemas.

#### V. CONCLUSIONES

La metodología secuencial de los experimentos permitió cubrir un amplio rango de las posibilidades sobre la utilización de la representación vectorial a nivel fonético.

Los modelos propuestos Continuous Bag of Words y Skip-gram han sido desarrollados inicialmente para la representación de unidades fonéticas complejas como palabras, sin embargo se ha propuesto emplear unidades más pequeñas como n-gramas que, de acuerdo a las pruebas realizadas con la base de datos disponible, presentó mejores resultados con unidades más pequeñas (2 gramas) porque el vocabulario generado incluye la mayoría de combinaciones posibles. Los poli-fonemas de mayor dimensión fueron más difíciles de abarcar durante el entrenamiento de los embeddings y produjeron más “missing-grams” o fonemas no encontrados, que en el modelo se reemplazaron con un vector de valor cero. Esta observación se tradujo en el aumento de la tasa de error. Una alternativa para realizar pruebas con n-gramas mayores podría ser la utilización de una base de datos mayor para la generación de los embeddings.

Los datos del vector contienen información muy dispersa para las unidades fonéticas empleadas, por lo que la tendencia de reducir el tamaño del vector generó mejores resultados y determinó el mínimo permitido por el reconocedor (dim. 10).

La configuración del modelo de un solo idioma, se la aplicó a los otros 5 para disponer de seis sistemas que al final fueron fusionados para obtener las métricas finales.

El sistema propuesto por sí solo está por debajo de otros sistemas LID, pero su fusión con otras técnicas puede contribuir significativamente como se ha demostrado.

#### REFERENCIAS

- [1] E. Ambikairajah, H. Li, L. Wang, B. Yin, and V. Sethu. “Language Identification: A tutorial”. IEEE Circuits and Systems Magazine, pages 82-108. May 2011.
- [2] E. Singer, P. A. Torres-Carrasquillo, T. P. Gleason, W. M. Campbell, and D. A. Reynolds. “Acoustic, phonetic, and discriminative approaches to automatic language identification”. Interspeech , 2003.
- [3] C. Salamea, L.F. D'Haro, R. de Córdoba, M. A., Caraballo “Incorporación de n-gramas discriminativos para mejorar un reconocedor de idioma fonotáctico basado en i-vectores”, Procesamiento del Lenguaje Natural, Revista nº 51, páginas 145-152, 2013.
- [4] M. A. Zissman et al. “Comparison of four approaches to automatic language identification of telephone speech”. IEEE Transactions on Speech and Audio Processing, pages 31 - 44, 1996.
- [5] L. J. Rodriguez-Fuentes, N. Brummer, M. Penagarikano, A. Varona, G. Bordel, and M. Diez. “The Albayzin 2012 language recognition evaluation”. In Interspeech , pages 1497 - 1501, 2013.
- [6] S. Lai, K. Liu, L. Xu and J. Zhao. “How to Generate a Good Word Embedding”, National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences, China, July 2015.
- [7] T. Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. “Distributed Representations of Words and Phrases and their Compositionality”. In Proceedings of NIPS, 2013.
- [8] M. Díez, A. Varona, M. Peñagarikano, L. J. Rodríguez-Fuentes, and G. Bordel. “On the use of pone log-likelihood ratios as features in spoken language recognition”. In Slt, pages 274-279, 2012.
- [9] P. Schwarz, “Phoneme Recognition based on Long Temporal Context”, PhD Thesis. Brno University of Technology, 2009.
- [10] D. A. Reynolds. “A Gaussian mixture modeling approach to text independent Speaker identification”. Ph.D. thesis, Georgia Inst. of Technol., 1992.



**Francisco Charro**, Ingeniero en Electrónica y Telecomunicaciones por la Escuela Politécnica Nacional (EPN) 2006, Diplomado Superior en Plataformas Operativas e Internetworking por la Escuela Politécnica Nacional 2008, Máster Universitario en Sistemas Electrónicos por la Universidad Politécnica de Madrid (UPM) 2015. Docente a tiempo completo en la Escuela Politécnica Nacional y a tiempo parcial en la Universidad de las Américas desde 2016.



**Marco Herrera**, Ingeniero en Electrónica y de Control por la Escuela Politécnica Nacional (EPN) 2009 y Master Universitario en Automática y Robótica por la Universidad Politécnica de Madrid (UPM), España, en 2014. De 2010 a 2012, docente en la Universidad Internacional del Ecuador, Quito, Ecuador. Desde 2014 es docente en el Departamento de Automatización y Control Industrial de la Escuela Politécnica Nacional. Su interés de investigación incluye control óptimo, control de modo deslizante y modelado borroso.



**Nately Pozo**, Ingeniera en Electrónica y Control en la Escuela Politécnica Nacional. A partir de entonces se trabajó como Técnico Docente Politécnico (DACI - EPN). En el año 2015- 2016 estudió en la Universidad de Newcastle – Reino Unido, en donde obtuvo el título de M.Sc. Electrical Power. Actualmente se desempeña en el cargo de Profesora a Tiempo Completo en el Departamento de Automatización y Control Industrial de la EPN. Sus áreas de interés son: Electrónica de Potencia, Conversores Estáticos, Energía Renovable, Redes Eléctricas Inteligentes, Diseño y Control de Máquinas.



**Andrés Rosales**, Profesor Principal del Departamento de Automatización y Control Industrial. Ingeniero en Electrónica y Control y Doctor en Ingeniería en Sistemas de Control (Universidad Nacional de San Juan – UNSJ, Argentina, 2009). Investigador invitado (Universidad de Hannover, Alemania, 2007-2008). Actualmente, Director de Investigación y Proyección Social de la EPN. Fue Coordinador General Técnico del CEAACES (2014-2015), y Editor de la Revista Politécnica de la EPN (2012-2014).

# Dissolved gas content forecasting in power transformers based on Least Square Support Vector Machine (LSSVM)

Roberto J. Fiallos

**Abstract**—Taking into account the chaotic characteristic of gas production within power transformers, a Least Square Support Vector Machine (LSSVM) model is implemented to forecast dissolved gas content based on historical chromatography samples. Additionally, an extending approach is developed with a correlation between oil temperature and Dissolved Gas Analysis (DGA), where a multi-input LSSVM is trained with the utilization of DGA and temperature datasets. The obtained DGA prediction from the extending model illustrates more accurate results, and the previous algorithm uncertainties are reduced.

A favourable correlation between hydrogen, methane, ethane, ethylene, and acetylene and oil temperature is achieved by the application of the proposed multi-input model.

**Index Terms**—Dissolved gas analysis (DGA), Gas chromatography, machine learning, Least Square Support Vector Machine (LSSVM).

## I. INTRODUCTION

Power transformers constitute one of the most important equipment in an electrical power system. These assets are generally efficient, reliable, and capital intensive, with an expected service life of 40 years or more.

Thermal or electrical stress contributes to insulating system deterioration within power transformers. Mineral oil and/or paper degradation is associated with abnormal functionality and possible incipient faults in the equipment, consequently, different types of hydrocarbons and carbon oxides are produced.

The composition of the gas dissolved in mineral insulating oils can be analyzed by the application of a diagnostic tool called Dissolved Gas Analysis –DGA, which detects and evaluates internal failures and their development trends.

A correct interpretation of DGA results is required to forecast and prevent failures with significant accuracy. References [1] and [2] explain concepts regarding power transformers insulating system composition, the degradation process of mineral oil and cellulose, the effects of operating conditions on gas production, and procedures utilized to detect and analyze possible failures.

The amount of available DGA data has a significant impact on the accuracy of the final results. Data analytic methods for power transformers involve amounts of data without existing formula or equation to correlate variables. As a result, machine learning algorithms have been used to diagnose and forecast dissolved gas concentration levels in power transformers, which are based on learning information directly from past DGA data and adapting their performance for future predictions.

Consequently, this project aims to predict dissolved gas content trends applying real chromatography data. A specific objective refers to obtaining high accuracy in the forecast values, where the randomness behaviour of the DGA data must be reduced by the application of processing techniques.

A Least Square Support Vector Machine (LSSVM) is implemented and validated. Finally, considering the influence of the operating conditions in the dissolved gas content into the power transformer, a correlation between oil temperature and DGA is also proposed to improve the predictions.

Motivated by the above-mentioned difficulties, a Least Square Support Vector Machine model (LSSVM) for DGA data predictions is constructed in this project, where historical real DGA data obtained from the industrial sector is used for training and testing the proposed algorithm. As part of the present work, a pre-processing stage is used to reduce the randomness DGA behaviour, which in addition to the LSSVM capabilities contribute to obtaining more accurate predictions.

As mentioned before, gas content changes are hugely affected by power transformer operating conditions, thereby a correlation between dissolved gas content and oil temperature is included as an extending approach of this project. The construction of a multi-input LSSVM model is developed with the application of DGA and oil temperature data in the training period. The main goal of the second proposed algorithm is to increase the accuracy of the forecasting DGA values.

## II. DISSOLVED GAS ANALYSIS METHODS

All transformers generate gases of some amount at normal operating conditions. Occasionally, this generation can lead to severe faults within the transformers. A dissolved gas analysis, which is the most common type of transformer monitoring can provide important data to increase the availability of power transformers. This analysis is based on chromatography methods, where oil samples are analyzed in laboratories. A number of gases (hydrogen, methane, ethane, ethylene, acetylene), and the relationship between each other help to identify the type of faults at an early stage [1].

### A. Key gas method

The method is dependent on the gas released at various temperatures of oil and cellular (paper) decomposition due to faults. The fault is determined by calculating the relative proportions of the gases. These significant gases are known as ‘key gases’. The four general fault types are described by [1], [3].

### B. Ratio Method

The ratio method is a technique which involves the calculation of key gas ratios and comparing these ratios to a suggested limit. Some of the most commonly used techniques are Doernenburg ratios and Rogers's ratios. The Doernenburg method is one of the effective diagnostic tools available but is less used due to its complexity. In this method, the concentration of one of the principle gas needs to be two times the other gases to be possible to calculate the ratios. The Rogers ration method is an advanced form of Doernenburg method and has almost same principle. But the requirement of needing significant concentration of principle gases is not there. The faults are chosen accordingly with the gases and the ratios [4].

### C. Duval's Triangle

This is one of the most preferred and also a highly-recognized method in IEC guidelines used for the gas analysis. It is recommended for its supreme accuracy in determining the faults. The advantage of this method is that it requires only 3 gases to analyze all types of potential faults within the transformer. The 3 gases are methane ( $\text{CH}_4$ ), acetylene ( $\text{C}_2\text{H}_2$ ), and ethylene ( $\text{C}_2\text{H}_4$ ). The construction of the triangle is in such a way that one calculates the total accumulated amount of three key gases and divides each gas by the total of the three gases and the percentage associated with each gas is found. The arrived values are plotted on a triangle as in the figure to arrive at a diagnosis [2]. Figure 1 illustrates the relative percentages of the 3 gases, which are plotted on each side of the triangle from 0% to 100% [5]. According to the relationship between the 3 gases, the diagnosis can be obtained from the fault zones (Table 1).

### III. LEAST SQUARE SUPPORT VECTOR MACHINE ALGORITHM (LSSVM)

Least square support vector machine (LSSVM) requires a reduced quantity of data to predict the future time series. "Based on the available time series, network internal parameters are tuned using an appropriate tuning algorithm" [6]. LSSVM is a reformulation of the traditional SVM, and it is more suitable to solve the regression problems [7]. Basically, LSSVM approach refers to solving a set of linear equations, due

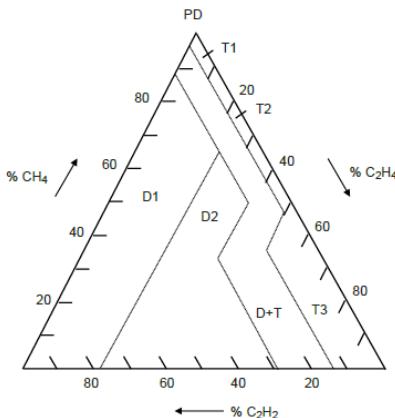


Figure 1 Duval's Triangle [2]

Table 1 Duval's triangle fault zones [2]

Code	Fault zone
T1	Low-temperature thermal fault ( $T < 300^\circ\text{C}$ )
T2	mid temperature thermal fault ( $300^\circ\text{C} \text{ to } 700^\circ\text{C}$ )
T3	High-temperature thermal fault ( $T > 700^\circ\text{C}$ )
D1	discharges of low energy
D2	discharges of high energy
D+T	mix of thermal and electrical faults
PD	partial discharges

to equality instead of inequality constraints in the problem formulation [8].

Given a training dataset  $\{x_k, y_k\}$ , where  $x_k \in R^m$  is the input data, and  $y_k \in R$  is the corresponding output data. In literature [7], a linear equation of higher-dimensional feature space is defined as:

$$f(x) = w^T \varphi(x) + b \quad (1)$$

where  $\varphi(\cdot)$  is a nonlinear mapping of data from input space into a higher-dimensional feature space. The optimization problem can be described by the following equations:

$$\min J(w, e) = \frac{1}{2} w^T w + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2 \quad (2)$$

Subject to  $y_k = w^T \varphi(x_k) + b + e_k$ ,  $k = 1, 2, \dots, N$ , where  $w \in R^m$  error variable  $e_k \in R$ , and  $b$  is bias.  $J$  is the loss function, and  $\gamma$  is an adjustable constant [8]. The Lagrangian function is defined according to the optimal function (2):

$$L(w, \alpha_k, b, e_k) = J + \sum_{k=1}^N \alpha_k [y_k - w^T \varphi(x_k) - b - e_k] \quad (3)$$

Regarding to equation (3),  $\alpha_k$  represents the Lagrange multipliers which also support vector  $\alpha_k \in R$  [8]. According to the linear KKT system [6], the first-order derivatives of  $L$  are:

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{k=1}^N \alpha_k \varphi(x_k) \quad (4)$$

$$\frac{\partial L}{\partial \alpha_k} = 0 \rightarrow w^T \varphi(x_k) + b + e_k - y_k = 0 \quad (5)$$

$$\frac{\partial L}{\partial b} = 0 \rightarrow \sum_{k=1}^N \alpha_k = 0 \quad (6)$$

$$\frac{\partial L}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k \quad (7)$$

After eliminating  $w$  and  $e_k$ , matrix equation (8) is gotten [7], which is a set of nonlinear equations to be solved in  $\alpha$  and  $b$ . These implicitly omissions correspond to creating an  $\varepsilon$ -insensitive zone in the underlying cost function, which is clear from the condition for optimality equations (4), (5), (6), and (7) [9].

$$\begin{bmatrix} 0 & \eta_l^T \\ s & \Omega + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ \alpha_k \end{bmatrix} = \begin{bmatrix} 0 \\ y_k \end{bmatrix} \quad (8)$$

where  $\Omega = K(x, x_k)$ ,  $I$  is the identity matrix,  $\eta_l^T = [1 \dots 1]$ ,  $s = [1 \dots 1]^T$ .

Therefore, the resulting LSSVM model for function estimation becomes [9].

$$f(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b \quad (9)$$

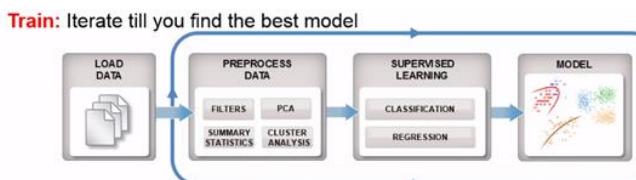
The parameters  $\alpha_k$  and  $b$  in equation (9) represent the solution to the linear system. Similar to the standard SVM,  $K(x, x_k)$  is the kernel function, and in the case of function estimation, RBF kernels can be applied. RBF kernel function has an advantage in comparison with SVM, because it has only two additional tuning hyper-parameters ( $Y, \sigma$ ).

#### IV. DISSOLVED GAS CONTENT PREDICTION BASED ON LSSVM – METHODOLOGY AND VALIDATION

Handling the available data and finding the right algorithm refer the most important challenge in machine learning techniques. Firstly, pre-processing techniques must be applied to attenuate the randomness behaviour of the dissolved gases content samples before using this data as an input of the LSSVM model.

Figure 2 describes the workflow implemented with LSSVM model. The first step in the training and testing periods refers to the pre-process of dissolved gases content data.

According to [7], the pre-processing techniques are required to attenuate the stochastic characteristics of the time sequence data and regularized its performance, in this manner carry out a reasonable prediction to a certain extend. Two different techniques are applied as data preparation before the machine learning algorithm: removing outliers from the series (smoothed) and data normalization. Both techniques have been chosen considering the performance of the LSSVM in the testing stage. Practical DGA sample data are a sequence of random observations taken over different periods of time, because of this some gas content measurements differ significantly in magnitude. The performance of the first technique is shown in Figure 3, which is applied to the ethane dataset. It can be noticed how the original data is rescaled in order to eliminate anomalies and avoid possible inconsistencies in the training period of the machine learning. A data normalization is also applied as a pre-processing method. The dataset is normalized in the range [0, 1] according to equation (10), before applying the DGA samples as an input of the LSSVM.



**Predict:** Integrate trained models into applications



Figure 2 Machine learning workflow [10]

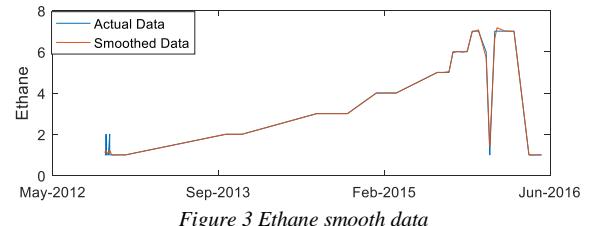


Figure 3 Ethane smooth data

$$y(t) = \frac{x(t)}{\max(x(1), x(2), \dots, x(n))} \quad (10)$$

where  $= 0, 1, \dots, n$ ,  $x(t)$  represent the DGA time-series dataset, while  $y(t)$  is the normalized parameter [6].

The LSSVM forecasting model is implemented based on the MATLAB LS-SVMLab toolbox. In this work, LSSVM is applied as time-series prediction (function estimation) based on the pre-process data, as describes

4. The available data corresponds to 52 samples and it is split into the training and testing datasets. The training dataset corresponds to 75% of the total data (39 samples), and it is applied to build the LSSVM model, while the remaining 25% (13 samples) is used to verify the model performance.

##### A. Training the model

The machine learning algorithm is built and trained using the optimal  $(Y, \sigma)$  hyper-parameters combination, and the total training set. In the proposed model, a robust training function is applied as training function, which is more suitable in the case of data containing non-Gaussian noise or outliers [11]. This specific function improves the final model performance, as a result of the support values influence corresponding to noise and outliers is decreased, and their corresponding large errors can be avoided.

The LSSVM algorithm requires an input training dataset and their corresponding output training dataset in order to build the model. The raw DGA data is used as input, and the results of the pre-process methods (smoothed or normalized) data is the corresponding output dataset in the training period.

After the iterative process, the model performance has the sufficient accuracy in order to starts with the testing stage, where the average error is collected and calculated, and the new parameter combination is replaced. The iterative process is repeated until approach the stopping criteria [6]. Considering the best cross-validation performance and the minimized error, the optimal  $(Y, \sigma)$  hyper-parameters are chosen [9].

##### B. Testing model

The trained model must be integrated into the prediction application, substituting the training dataset by the testing dataset into the LSSVM model, and in this way, the estimation values can be obtained [6]. In order to verify the LSSVM forecasting model performance, a testing dataset is applied to see how it will respond with unknown data. Using the remaining 25% of the available DGA data, the proposed algorithm is validated against the actual values from the corresponding analysis, which error is used to verify the LSSVM model forecasting ability.

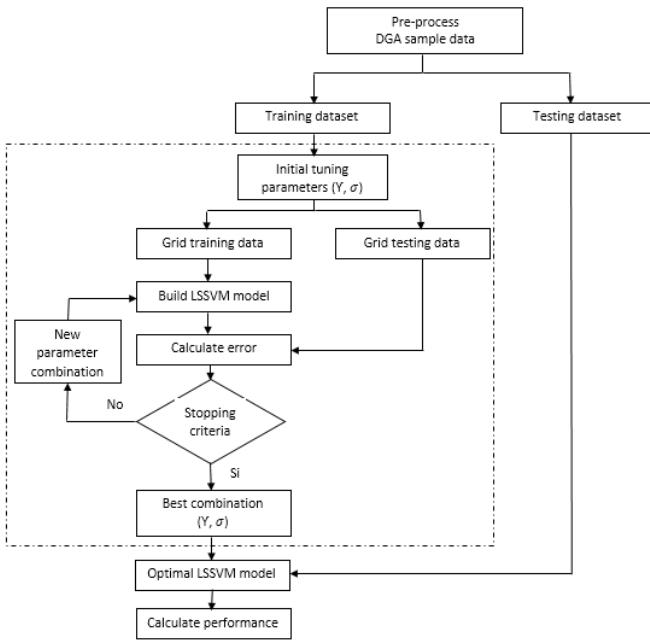


Figure 4 LSSVM flowchart of the proposed model [6].

### C. Validation of DGA forecasting data based on LSSVM

The dataset contains chromatography samples of three power transformers collected between October 2012 and May 2016, where the gas concentration values of hydrogen (H<sub>2</sub>), oxygen (O<sub>2</sub>), nitrogen (N<sub>2</sub>), carbon monoxide (CO), methane (CH<sub>4</sub>), carbon dioxide (CO<sub>2</sub>), ethylene (C<sub>2</sub>H<sub>4</sub>), ethane (C<sub>2</sub>H<sub>6</sub>), and acetylene (C<sub>2</sub>H<sub>2</sub>) have been included.

The DGA forecasting values based on LSSVM model for hydrogen data are shown in Figure 5, where the superposition curves of actual and predicted DGA dataset is demonstrated. Clearly, it can be noticed that a high performance has been achieved with the LSSVM forecasting values for the different gases. Regarding Figure 6, this describes the LSSVM results in the environment of the training data [11], and the corresponding obtained optimal hyper-parameters ( $Y, \sigma$ ).

An excellent correlation between the raw DGA sample and the corresponding pre-process data is defined by the applied machine learning. Most of the gases illustrate a well-defined function correlation between the raw data and the pre-process

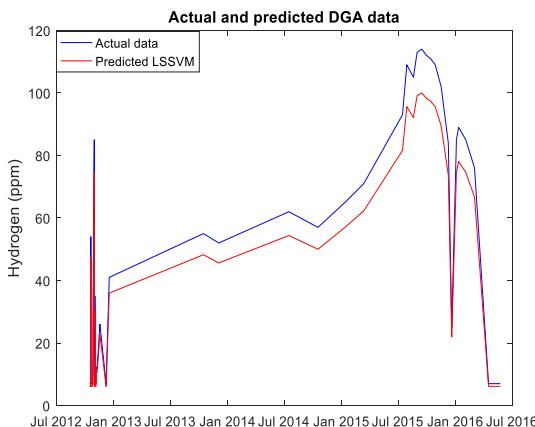


Figure 5 Actual and predicted DGA data using LSSVM model – hydrogen

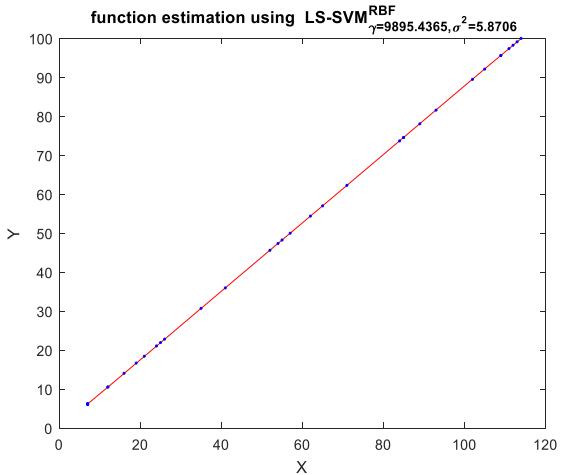


Figure 6 LSSVM performances in the training period

output in the training stage, where the regulation optimal parameter ( $Y$ ) is obtained through the solution of the linear LSSVM, and the optimal kernel parameter ( $\sigma$ ) is determined by choosing the midrange of values through an iterative process until its error is minimized [7].

The LSSVM forecasted results are validated in comparison to the actual corresponding data. The predicted values of H<sub>2</sub>, O<sub>2</sub>, CO, CO<sub>2</sub>, C<sub>2</sub>H<sub>4</sub>, C<sub>2</sub>H<sub>6</sub>, and C<sub>2</sub>H<sub>2</sub> demonstrate a low mean absolute percentage error (MAPE), which varies in the range 0.1154 - 19.9 %.

Consequently, the developed LSSVM algorithm demonstrates an effective performance in the training and testing periods, where a relatively small of datasets have been applied. In some cases the pre-process stage and the robust functions, which are applied in the training stage of the proposed model, apparently cannot avoid the randomness characteristic completely. For instance, nitrogen (N<sub>2</sub>) and methane (CH<sub>4</sub>) forecasting values show higher rates of MAPE (35.76% and 36.48% respectively) in comparison with the rest of gases.

Hence, it can be considered that these two cases do not provide enough effectiveness in the prediction of future trends of gas production, and some misinterpretations can be generated.

In overall, a high performance is achieved by the LSSVM model. Its effectiveness is demonstrated by the similar trends described in Figure 5, and the corresponding error rates in each case. Owing to the possible issues that can be produced by an over/underestimation of gas dissolved content, an alternative approach will be developed in the next section, where some operating conditions of the power transformer will be included in order to improve the gas prediction.

## V. CORRELATION BETWEEN GAS CONCENTRATION LEVELS AND OPERATING CONDITIONS

As mentioned before and according to [2], any gas formation results from a stress of some kind (thermal or electric) inside the equipment, additionally, the authors in [12] state that the insulation system quality depends on the mechanical (temperature, vibrations) and electrical (voltage levels, loading) influences. Hence, evaluate the status of the power transformer considering further parameters such as environmental

influences and operating conditions can contribute to a much better interpretation and understanding of the dissolved gas production changes, and the subsequently malfunction detection that eventually can lead to failure.

Available data can be used to analyze the condition of power transformers and planning maintenance task based on real asset's condition instead of every certain period of time, which benefits might not have a significant impact on the transformer performance.

#### A. Temperature influences on DGA

Reference [2] states that the different gases require specific temperatures in order to accumulate as a stable recombination product. It means that gases are formed depending on fault type and temperature. For instance, gases such as acetylene require temperatures of at least 800°C to 1200°C to show an accumulative trend [2], ethylene with hot spots between 150°C and 1000°C [13], and carbon oxides (CO and CO<sub>2</sub>) can be formed at temperatures higher than 105°C.

Indeed, temperature are strongly linked with gas production and consequently with abnormal oil and paper degradation into the power transformer. In literature [4], results obtained from transformers with different features and operating conditions demonstrate that the solid insulator life-cycle at high loading and/or high operational temperatures decreases as result of abnormal circumstances. For these reasons, temperature measurements corresponding to the same intervals when the DGA samples were obtained have been included to build a multi-dimensional algorithm and improving the accuracy of gas concentration level predictions.

#### B. Multi-input LSSVM model

The applied LS-SVMLab toolbox has a multi-dimensional capability in order to include additional input and/or output variables. The work presented in [12] refers to the arising of nonlinear behaviour in time-series predictions conducted by aggregating data from multiple sources.

As mentioned in [9], the LSSVM multilayer network is first trained on subsets of data  $D1, \dots, Dm$ , and followed by a nonlinear combination, which is determined by solving a parametric optimization problem. In fact, this implementation also uses a kernel function (RBF) to obtaining the hyper-parameters through an iterative process (10-fold cross-validation) in the training period.

The multi-input LSSVM algorithm is first trained on subsets of data  $D1, D2, \dots, Dm$ , and the combination of the models is determined by solving a parametric optimization problem, which provides an interacting cooperation for the ensemble and in this way realize collective intelligence [9].

In this work, multiple input and output values are used in the training of the novel LSSVM algorithm, where the input vectors refer to the actual DGA and temperature data, while the corresponding normalized values are applied as output vectors.

As documented in [14], the machine learning method automatically correlates particular patterns of oil temperature with historical DGA data in the training period. The novel LSSVM model is built based on the input datasets and the hyper-parameters ( $Y, \sigma$ ) obtained from the training stage.

The multi-input LSSVM algorithm performance is tested by the application of unknown DGA data. In this case, the input

testing dataset contains DGA samples as a unique variable to establish the expected gas concentration level changes into the power transformer. Similar to [11], the proposed LSSVM model treats the DGA sample date as an event interval, while the temperature values contain several measurements for the corresponding DGA interval. Owing to this difference between the amount of DGA data and temperature measurements in every interval used in the present work, cumulative statistics should be applied to characterize the temperature contribution to the dissolved gas concentration level changes.

#### C. Multi-input LSSVM model validation

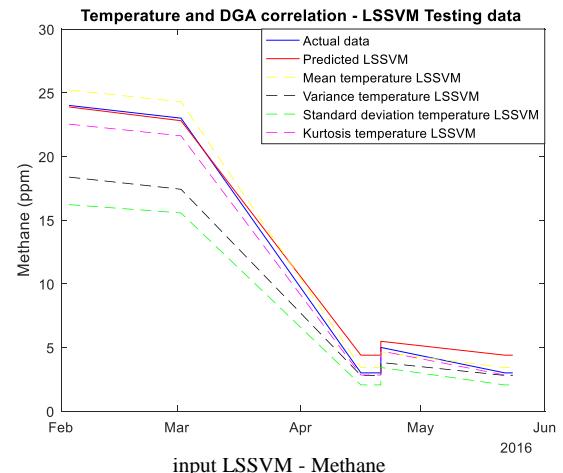
The actual DGA data and the characterize vector obtained from oil temperature measurements by the application of cumulative statistics are applied as input values for the training period of the multi-input LSSVM model. The same pre-process methodology used in the first LSSVM algorithm is applied to attenuate the randomness behaviour of DGA data and the representative feature vectors of temperature for each descriptive statistics method (mean, variance, standard deviation, and kurtosis). The normalized values of both datasets are applied as the corresponding output of the novel model training.

The multi-input LSSVM algorithm is built based on the obtained hyper-parameters ( $Y, \sigma$ ), the DGA samples and the characterize temperature values throughout the training period for each used descriptive statistics method.

The performance of the novel model is verified through the substitution of the training datasets by unknown testing data. In this stage, the algorithm is tested by the application of unused DGA data as unique input. The multi-input LSSVM model responses are validated in comparison to the actual DGA and the values predicted by the earlier LSSVM model for every gas included in this work.

As can be seen in Figure 7, many experiments with the inclusion of temperature patterns in the training period of the novel LSSVM model have been conducted to identify the best performance. Some of the new forecasts indicate improvements in accuracy, others have not shown important changes, while in some cases the values are less accurate in comparison to the earlier predictions obtained by the first LSSVM model.

Figure 7 Temperature and DGA correlation - Testing values based on multi-



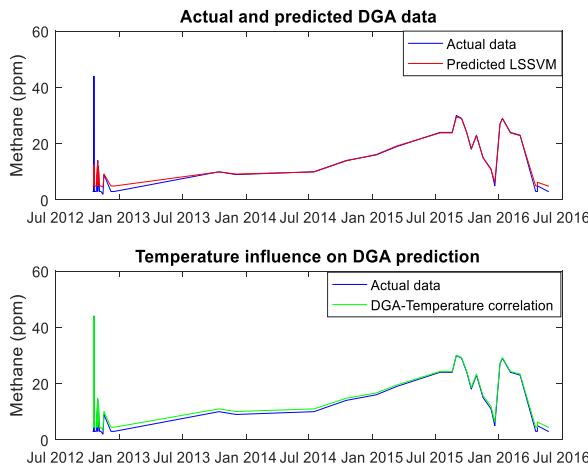


Figure 8 Methane time-series comparison between DGA forecasting model and temperature correlation model

Taking into consideration the mean absolute percentage error (MAPE) calculated values, evidently, the forecasting values of hydrogen ( $H_2$ ), methane ( $CH_4$ ), and ethane ( $C_2H_6$ ) have been improved mainly by the application of kurtosis as feature extraction method of the temperature measurements. Comparing the results obtained from the first and second models, the mean absolute percentage error of hydrogen predictions is improved from 12.16% to 5.87%, methane MAPE decreases from 36.48% to 5.69%, whereas ethane shows a drop from 7.7% to 5.65% in its corresponding MAPE.

Obviously, the correlation between temperature patterns and hydrogen, methane, and ethane has an important influence on the corresponding forecasting values. One of the least accurate predictions obtained by the first proposed LSSVM model represents methane content, whose uncertainty is reduced by 30% with the inclusion of kurtosis as feature extraction of the temperature measurements into the novel model as Figure 8 indicates.

To sum up, the novel LSSVM model developed with the correlation between DGA and oil temperature patterns indicates the influence of this operating conditions of the power transformer in certain gas concentration behaviour, whose relationships must be interpreted bearing in mind the International Standards referred in [1] and [2].

## VI. CONCLUSIONS

Considering the least accurate results obtained from the first LSSVM proposed model, a multi-input novel LSSVM model is developed with the inclusion of oil temperature patterns with the main aim of improving the previous predictions values.

Temperature condition has a significant impact in dissolved gas content changes taking into account the improvements in the predictions obtained with the application of this novel model. Therefore, as part of the extending approach, a correlation between DGA and oil temperature is found by the application of multi-input LSSVM model. According to the obtained results, the most notable improvements in the forecasting values are obtained with the application of kurtosis as feature extraction of temperature due to it reflects the changes between the temperature measurements. An excellent correlation between oil temperature and hydrogen, methane,

ethane, ethylene and acetylene is reached in the three utilized power transformers. Some of the obtained relationships can be linked to the interpretation of the gas analysis. In contrast, the nitrogen ( $N_2$ ) forecasting values show a considerable inaccuracy, whose MAPE value is around 35.76%. The application of oil temperature in the training model utilized in the multi-input LSSVM model developed as extending approach has not affected either (56.93%) and can be considered as a weakness of the proposed models. Considering the interpretations of DGA given by [2], oxygen and nitrogen are found in oil as a result of contact with atmospheric air, which can be stated as the reason for the lack of correlation with oil temperature.

In conclusion, the obtained forecasting DGA data contain important accuracy for the three power transformers used in this project, and it can bring sufficient certainty to applying further analysis.

## REFERENCES

- [1] IEEE Power / Energy Society, *Guide for the Interpretation of Gases in Oil-Immersed Transformers*, New York: IEEE, 2008.
- [2] IEC, *Mineral oil-filled electrical equipment in service - Guidance on the interpretation of dissolved and free gases analysis*, IEC, 2015.
- [3] B. Stewart, *Asset Management-Transformer condition monitoring*, Glasgow: University of Strathclyde, 2017.
- [4] L. Hamrick, *Dissolved Gas Analysis for Transformers*, Tennessee: ESCO Energy Services, 2009.
- [5] M. Duval, *Practicing DGA*, Qualitrol Company, 2015.
- [6] A. Melit, A. Massi Pavan y M. Benghanem, *Least squares support vector machine for short-term prediction of meteorological time series*, Springer, 2012.
- [7] R. J. Liao, J. P. Bian, L. J. Yang, S. Grzybowski, Y. Y. Wang and J. Li, *Forecasting dissolved gases content in power transformer oil based on weakening buffer operator and least square support vector machine-Markov*, vol. 6, THe Institution of Engineering and Technology, 2012, pp. 142-151.
- [8] J.-Y. Zhu, B. Ren, H.-X. Zhang and Z.-T. Deng, *Time series prediction via new support vector machines*, Beijing, 2002, pp. 364-366.
- [9] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor y J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific Publishing, 2002.
- [10] S. Prasanna, “MathWorks,” 2015. [Online]. Available: [https://uk.mathworks.com/videos/machine-learning-with-matlab-100694.html?elqsid=1497965925222&potential\\_use=Student](https://uk.mathworks.com/videos/machine-learning-with-matlab-100694.html?elqsid=1497965925222&potential_use=Student).
- [11] K. De Brabanter, P. Karsmakers, F. Ojeda, C. Alzate, J. De Brabanter, K. Pelckmans, B. De Moor, J. Vandewalle and J. A. Suykens, *LS-SVMlab Toolbox User's Guide*, Version 1.8 ed., 2011.
- [12] C. Voros, B. Németh, S. Szabó y I. Berta, *Diagnostic evalution of power transformers by an expert system*, Vols. %1 de %2306-310, IEEE, 2012.
- [13] M. Duval, *Dissolved gas analysis: It can save your transformer*, vol. 5, IEEE, 1989.
- [14] Y. Kim, A. Goyal y T. Kumar, *Predictive modeling of dissolved gas concentration in oil-immersed substation transformers*, 2016 the 4th IEEE International COnference on Smart Energy Grid Engineering ed., IEEE, 2016, pp. 261-267.



**Roberto J. Fiallos** graduated from the National Polytechnic Institute in Electrical Engineering in 2012. He obtained the MSc degree in Electrical Power Engineering with Business from the University of Strathclyde in 2017. He works for the electrical transmission and distribution industry in projects related to substations, overhead lines, and SCADA. His research interests include Condition Assessment of High Voltage Equipment, Power Systems Optimization and Machine Learning Applications.

# Simulation of evacuation with multi-agents on georeferenced layers with GAMA

Zoila Ruiz-Chavez, Jaime Salvador-Meneses

**Abstract**—This work deals with agent-based modeling within a geographical environment, it reviews concepts on Multi-agents and Geographic Information Systems (GIS). It is oriented to perform a simulation that considers aspects of human behavior (through agents) during an evacuation, considering real restrictions of the environment in which they are developed (GIS layers). This simulation is done in the GAMA<sup>1</sup> platform, which allows us to easily implement models based on agents and geographic layers. It allows us to work with the attributes of the layers and to define constraints based on them.

**Index Terms**—GAMA, Agents, Multi-Agents, GIS, Complex Systems.

## I. INTRODUCTION

Today the growth of geographic information management is evident. The management of spatially referenced information and their topological relationships, is undoubtedly the main feature that defines Geographic Information Systems (GIS) [1]. Thanks to the type of data they manage, we can represent reality in a simplified way to perform different types of analysis.

On the other hand, we find Complex Systems, whose main characteristic is to be unpredictable, because they are composed of different elements that interact with each other, they also evolve dynamically [2].

GAMA is a solution that allows modeling or constructing simulations based on spatially referenced multiagent systems; the model or simulation considers the different elements that make up the complex system and whose medium in which they develop are georeferenced thematic layers containing information to be integrated into the model and to be able to perform calculations or establish restrictions [3].

GIS are currently used in different areas [4] as well as complex systems [5], both techniques are generally used separately. This project intends to use the two techniques and experimentally prove the benefits obtained by using these two techniques simultaneously. Modeling or constructing simulations based on spatially referenced multi-agent systems is an alternative that tries to give realism to the environment in which agents operate to obtain results as close to reality as possible.

The methodology used in this work complements the traditional studies allowing to analyze data in a more dynamic way, modeling the behavior of complex systems on spatial layers that allow visualizing their behavior in an environment with real characteristics and obtain results closer to reality.

It is intended to clearly define the relationships between the different elements that make up the complex system to

be modeled and what effects they have if some component is altered or external elements are introduced.

This document is organized as follows. Section II introduces aspects on simulations based on multi-space agents. Section III presents an application model in the GAMA platform. Finally, in section IV some conclusions and opportunities for future work are presented.

## II. GEOREFERENCED MULTIAGENTS SIMULATIONS

In this section we review simulations with spatial multiagents and describe the two systems used to create them.

Computer models are gaining ground in various areas and sectors, being able to represent phenomena through models is a real challenge today. Integrating multi-agents systems with geographic information systems allows better simulation of phenomena or specific activities whose behavior is complex. In a simulation, to be able to consider spatiotemporal restrictions with multi-agents, potentiates any proposed application [6]. To be able to implement simultaneous strategies using simulations based on spatial agents in a study, allow to face challenges to model complex systems that include specific characteristics like the real form of terrain, slopes, sense of mobilization of vehicles, alleys, among others. In this model, the agents consider the distances to the so-called safe places depending on their current location and the access to roads through which they will be mobilized.

Among the studies analyzed, we find simulations to model the traffic flow in evacuations [7], socio-ecological models [8], models of human behavior due to climate change [9].

Several studies propose simulations that consider complex systems, but the biggest limitation they have found has been the absence of tools that integrate environments with real characteristics, that allows them to create dynamic and realistic environments [10]. Typically, most platforms solve modeling with multi-agent systems separately to solve the complex problems that arise and those involving spatial information, and it is not usual to find a platform that integrates a complete solution that allows to execute both tasks simultaneously.

The open source GAMA platform (GIS and Agent-based Modeling Architecture) is oriented to the modeling and simulation of multi-agent systems with complex environments [11].

### A. Geographic Information Systems

Today the search for facilitating the visualization of factors that intervene in a study, and to include in the study the spatial positioning, has allowed the improvement and popularity of tools oriented to the management of space information [12]. Being able to evaluate more accurately any study that can be

<sup>1</sup>GIS and Agent-based Modeling Architecture

represented with its spatial location is more relevant. Every study related to spatial information tries to project from the point of view of the geographic information systems, all the elements that constitute a particular problem and to be able to obtain efficient and safe designs [13].

The use of geographic information systems based on open source software, proprietary or mixed, in the applications gives us advantages such as the discretization of the information that is stored in geographic databases [14], which contain, on the one hand, information concerning the geographical location of the elements involved in the study and, on the other hand, the descriptive information of them. The first would be the geometry that represents the object, georeferenced geometric form, and the second attributes associated with the object of various types.

### B. Multi-agents Systems

Currently, multi-agent systems are an expanding area with applications in various areas such as artificial intelligence, numerical simulations and distributed systems, among others, making important contributions in solving problems, where traditional methods have not provided sufficient satisfactory solutions, thanks to the natural form of distributed problem solving [15].

Multi-agent Systems (MAS) are composed of multiple agents that are able to meet goals, whether they are intelligent or not. The design of a MAS must aim to coordinate, through some procedure, the behavior of autonomous agents that, with their beliefs, desires and intentions, are able to construct the sequences of actions that they must take to solve the problem, which can be an individual problem (that is, specific to each agent) or global (common to the whole set of agents) [16].

### C. GAMA Platform

This platform<sup>2</sup> allows the representation of features associated with complex systems and environments based on GIS layers. It has its own development language, called GAML, that allows the construction, coupling and reuse of complex models with different agent architectures, real environments and abstraction levels [17].

This platform is oriented not only to professionals with solid programming knowledge, its graphical interface allows to build models through wizards that automatically create the code associated with the generated graphic model. Agents in the GAML language are defined as species (those that provide a set of attributes, actions, behaviors and properties), while the grid is the medium where they develop (this may be spatial layers) [18]. It is possible to generate species of generic type, parent or child that inherit properties or qualities of another species, very similar to what happens in the object oriented design paradigm. GAMA allows us to work with vectorial layers, be these polylines, polygons or points. Not only allows you to visualize the vector layers, but also allows you to interact with the attributes of each one.

<sup>2</sup><http://gama-platform.org/>

### III. EVACUATION MODEL

Today there are many studies that are based on current techniques to optimize the results obtained by simulations of various kinds [19]. Traditional modeling techniques have become increasingly imprecise because they do not contemplate within their calculations the changing factor of Complex Systems or the special features of their environment. The studies related to evacuations in case of seismic risks at urban level are very common, however in most of them the predominant factor in reality is not taken into account: human behavior [20].

Humans, for modeling purposes, should be considered within the Complex Systems field because of their changing characteristics according to the environment, external factors and how they interact with each other. There are few studies oriented to the simulation of evacuations taking into account people and mobility according to the location at the time of an emergency [8].

In this work a model of evacuation to safe sites is generated that verifies the routes of evacuation and the time of response to an emergency, using the GAMA tool. The model consists of three important parts:

- Agents Modeling.
- Environment analysis and definition.
- Agents and environment interactions.

#### A. Agents modeling

The relevant factor to take into account within the model are people, so at this stage we design the different types of agents to create and their specific parameters. First, we create a global species that has common characteristics to all agents that represent people and who can inherit these characteristics to specific species such as:

- Pedestrians: People whose environment at the general level are streets and do not have a defined location schedule. Its basic characteristics are: mobility without fixed schedules, minimum and maximum speed.
- Employees: People who at fixed hours move to buildings considered workplaces and stay in those places during well-defined days. They have characteristics such as: mobility with schedules defined by ranges, job location, minimum and maximum speed.

#### B. Environment analysis and definition

People can occupy different types of environments within the model which are represented by spatial layers, which can be of three types:

- Streets: Vector layer that contains streets with spatial reference and own characteristics like: length, positioning, orientation, connectivity, among others. It represents the environment where people move and thanks to the characteristics of the spatial layer, you can calculate the maximum capacity, limitations to access certain places. Thus you can also perform calculations to find the best alternative route or routes to get from any location to destinations points.

- Buildings: Polygonal vector layer that are associated with the environment, represent workplaces or educational establishments. Here people will be accommodated in certain days.
- Safe places: Layer depicting safe environments for people in emergencies.

Buildings and safe places are defined by the attribute of the vectorial layer. In the GAML language the vectorial layers are also declared as another kind of species allowing to work with their attributes.

### C. Agents and environment interactions

Simulations are now considered an effective method to prevent further complications in the event of natural disasters or terrorist attacks [21]. In case of emergencies, the evacuation is the first step to follow, either in an orderly way or by simple human reflection. For this reason, it is important to consider how the agents interact with each other and with the environment in case of an emergency. The characteristics and functions of agents must be clearly defined under certain restrictions. Performing simulations that according to defined schedules reflect urban problems such as traffic, are of vital importance to obtain optimal results in the simulation.

We define two agents types:

- The "employee" agent type that moves to the buildings for the time set in their restrictions.
- The "pedestrian" agent type that moves over the streets with a minimum speed for certain time ranges.

After defining the agents and the environment in which they operate, it remains to introduce the functions of relationship between agents and the environment. Some function examples are:

- Location: Allows to define the location of the agents on the vectorial layers.
- Displacement: Functionality that allows people to move through the street layers.
- Emergency: In case of activating this functionality, people move only to safe places. Carry out a search for the best route to the nearest point of safe places.
- Restrictions: Restrictions are defined both at the level of people and of vectorial layers, such as streets without pedestrian access (simulates the presence of meshes, doors that limit access), dead ends (owned by the vector layer).

When running the model, the agents move to their places of permanence (it depends on the agent type).

When the emergency is activated, all the agents move to safe places using the maximum speed defined in their restrictions. The safe place to which each agent is mobilized depends on the value of the distance calculated using the *simple\_clustering\_by\_distance*<sup>3</sup> function. Also, the estimated time for this displacement is defined within the model. With this, we can define the percentage of people who can reach a safe place.

Figure 1. Initial status



The initial status is shown in Figure 1. Red buildings represent workplaces or educational establishments, while green buildings are safe places. Pedestrians can be seen on the streets and employees inside buildings.

In the Figure 2 we can see the final state of the simulation. It shows the people who reach a safe site, the accommodation capacity of safe places and people who by the estimated time for evacuation failed to reach or worse still did not come out of buildings.

The simulation environment has the following layers (maps):

- Buildings *Redificacion.shp*
- Streets *Rcalles.shp*
- Zones *Rzona.shp*

These layers allow us to have different levels of interaction of the environment within the model, each of them represents a geographical vector data with their respective attributes.

For the location of the agents within the simulation, the *use\_building* attribute is used, this attribute belongs to the building layer (buildings or safe places). If the agent type is *employee*, its initial location will be inside a building.

Pedestrian agents are located on the streets layer and are mobilized at random. As a restriction in the model, it is necessary that all the agents will move to the safe places only in case of emergency.

The initial parameters of the simulation are:

- Number of employees: 500
- Number of pedestrians: 100
- Minimum speed: 1.0
- Maximum speed: 5.0

<sup>3</sup><https://gama-platform.github.io/wiki/Operators>

Figure 2. Final status



In order to choose the safe place to which the agents will move in the model, the closest distance from the current location of each agent to the safe-type constructions is sought through the *simple\_clustering\_by\_distance* function. At the end of the simulation, the percentage of employees and pedestrians who managed to evacuate in a given time range is calculated.

#### IV. CONCLUSIONS

Among the conclusions that we could draw from this work we have:

The ability to implement a simulation with MAS and GIS allows you to visualize specific problems with real restrictions: take a dead end, not be able to find an optimal route to arrive at a defined time to safe places, the ability to evacuate buildings through the doors access, among others.

When making models or simulations related with geographical characteristics, it is more appropriate to use properties of the geographic data. Therefore, we obtain better results by decomposing the environment into different levels that allow a better interaction with its attributes.

GAMA represents a useful tool when making models that implement the two systems (MAS and GIS). It allows exploiting the full potential of both systems to solve problems related to the modeling and simulation of spatially referenced complex systems.

Simulations of this type should be considered as an aid in decision making in the urban part in case of emergencies.

As future work is intended to carry out a deeper simulation contemplating vehicular traffic and rescue personnel in order to have a complete idea of what happened in case of an evacuation: the response time of the rescue groups, the problem caused by traffic jams and other .

#### ACKNOWLEDGMENT

The authors would like to thank to Universidad Central del Ecuador and its initiative called *Proyectos Semilla* for the support during the writing of this paper. This work has been supported with Universidad Central del Ecuador funds.

#### REFERENCES

- [1] A. Carrara and F. Guzzetti, *Geographical information systems in assessing natural hazards*. Springer Science & Business Media, 2013, vol. 5.
- [2] H. Yu, Z. Shen, C. Leung, C. Miao, and V. R. Lesser, "A survey of multi-agent trust management systems," *IEEE Access*, vol. 1, pp. 35–50, 2013.
- [3] E. Macatulad and A. Blanco, "3dgis-based multi-agent geosimulation and visualization of building evacuation using gama platform," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 40, no. 2, p. 87, 2014.
- [4] M. Tahri, M. Hakdaoui, and M. Maanan, "The evaluation of solar farm locations applying geographic information system and multi-criteria decision-making methods: Case study in southern morocco," *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 1354–1362, 2015.
- [5] S. Li and W. A. Mackaness, "A multi-agent-based, semantic-driven system for decision support in epidemic management," *Health informatics journal*, vol. 21, no. 3, pp. 195–208, 2015.
- [6] J. Izquierdo, I. Montalvo, R. Pérez-García, and D. Ayala-Cabrera, "Multi-agent simulation of hydraulic transient equations in pressurized systems," *Journal of Computing in Civil Engineering*, vol. 30, no. 4, p. 04015071, 2015.
- [7] X. Chen and F. Zhan, "Agent-based modeling and simulation of urban evacuation: relative effectiveness of simultaneous and staged evacuation strategies," in *Agent-Based Modeling and Simulation*. Springer, 2014, pp. 78–96.
- [8] T. Filatova, P. H. Verburg, D. C. Parker, and C. A. Stannard, "Spatial agent-based models for socio-ecological systems: challenges and prospects," *Environmental modelling & software*, vol. 45, pp. 1–7, 2013.
- [9] M. Pons, P. A. Johnson, M. Rosas, and E. Jover, "A georeferenced agent-based model to analyze the climate change impacts on ski tourism at a regional scale," *International Journal of Geographical Information Science*, vol. 28, no. 12, pp. 2474–2494, 2014.
- [10] A. Drogoul, E. Amouroux, P. Caillou, B. Gaudou, A. Grignard, N. Marilleau, P. Taillandier, M. Vavasseur, D.-A. Vo, and J.-D. Zucker, "Gama: multi-level and complex environment for agent-based models and simulations," in *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2013, pp. 1361–1362.
- [11] D.-A. Vo, A. Drogoul, J.-D. Zucker, and T.-V. Ho, "A modelling language to represent and specify emerging structures in agent-based model," *Principles and Practice of Multi-Agent Systems*, pp. 212–227, 2012.
- [12] R. Debnath and A. N. Amin, "A geographic information system-based logical urban growth model for predicting spatial growth of an urban area," *Environment and Planning B: Planning and Design*, vol. 43, no. 3, pp. 580–597, 2016.
- [13] A. Minelli, C. Tissot, M. Rouan, and M. Le Tixerant,

- “Simulation of marine activities by coupling geographical information system and agent based model: improvements and technical achievements,” *PeerJ Preprints*, vol. 4, p. e2231v2, 2016.
- [14] D. J. Peuquet, “It’s about time: A conceptual framework for the representation of temporal dynamics in geographic information systems,” *Annals of the Association of american Geographers*, vol. 84, no. 3, pp. 441–461, 1994.
  - [15] C. M. Macal, “Everything you need to know about agent-based modelling and simulation,” *Journal of Simulation*, vol. 10, no. 2, pp. 144–156, 2016.
  - [16] L. Hernández, C. Baladron, J. M. Aguiar, B. Carro, A. Sanchez-Esguevillas, J. Lloret, D. Chinarro, J. J. Gomez-Sanz, and D. Cook, “A multi-agent system architecture for smart grid management and forecasting of energy demand in virtual power plants,” *IEEE Communications Magazine*, vol. 51, no. 1, pp. 106–113, 2013.
  - [17] C. Le Page, G. Abrami, N. Becu, P. Bommel, B. Bonté, F. Bousquet, B. Gaudou, J. P. Müller, and P. Taillandier, “Multi-platform training sessions to teach agent-based simulation,” 2015.
  - [18] P. Taillandier, D.-A. Vo, E. Amouroux, and A. Drogoul, “Gama: a simulation platform that integrates geographical information data, agent-based modeling and multi-scale control,” *Principles and Practice of Multi-Agent Systems*, pp. 242–258, 2012.
  - [19] D. Prunetti, A. Muzy, E. Innocenti, and X. Pieri, “Utility-based multi-agent system with spatial interactions: The case of virtual estate development,” *Computational Economics*, vol. 43, no. 3, p. 271, 2014.
  - [20] T. Takabatake, T. Shibayama, M. Esteban, H. Ishii, and G. Hamano, “Simulated tsunami evacuation behavior of local residents and visitors in kamakura, japan,” *International Journal of Disaster Risk Reduction*, vol. 23, pp. 1–14, 2017.
  - [21] L. Ma, B. Chen, S. Qiu, Z. Li, and X. Qiu, “Agent-based modeling of emergency evacuation in a railway station square under sarin terrorist attack,” *International Journal of Modeling, Simulation, and Scientific Computing*, p. 1750022, 2016.



**Zoila Ruiz-Chavez** received her Master’s Degree in Lógica Computación e Inteligencia Artificial from Universidad de Sevilla (Spain, 2014) and a Master’s Degree in Sistemas de Información Geográfica Aplicada a la Conservación y Desarrollo Sostenible from Universidad Central del Ecuador (Ecuador, 2012). She graduated as Ingeniera Informática in Universidad Central del Ecuador (Ecuador, 2000). She is currently a Professor at Facultad de Ingeniería, Ciencias Físicas y Matemática of the Universidad Central del Ecuador. Her main research interests include Machine Learning, Agent based systems, and Geographic Information Systems.



**Jaime Salvador-Meneses** received his Master’s Degree in Lógica Computación e Inteligencia Artificial from Universidad de Sevilla (Spain, 2014) and a Master’s Degree in Sistemas de Información Geográfica Aplicada a la Conservación y Desarrollo Sostenible from Universidad Central del Ecuador (Ecuador, 2012). He graduated as Ingeniero Matemático and Ingeniero Informático in Universidad Central del Ecuador (Ecuador, 2001). He is currently a Professor at Facultad de Ingeniería, Ciencias Físicas y Matemática of the Universidad Central del Ecuador. His main research interests include parallel computing in GPU, Machine Learning and Geographic Information Systems.



# Development of a Prototype for the Quality Control of Bovine Meat Determined by its Organoleptic Characteristics, based on an Automatic Inspection System for Artificial Vision

Portero P., Mena B., Gavilanes J.

**Abstract— Abstract—** The world beef market is in full growth and with unsatisfied demands and prices up and these depend very much on their quality when arriving at the final consumer, while the national market, our country also has a future potential in the meat industry. In the present investigation is to evaluate the sensory characteristics of color and texture by means of the presence of a bovine meat sample in the system whose final result will be the acceptance or negation of its organoleptic quality. The system is constituted by a transport band in charge of taking the sample to a capsule where both the Lighting System (LS) and the Computer Vision System (CVS) are located, and when leaving there it will be automatically classified by an actuator, guaranteeing less manipulation and estimation of time by human intervention throughout the process while providing greater security to the final consumer about the meat that is entering their home.

**Index terms—** Artificial intelligence, automation systems, quality control, color space, organoleptic characteristics.

## I. INTRODUCTION

Since the middle of the twentieth century the consumption of beef has been increasing, because people consider it as an essential product for their food, for this reason their production has increased, bringing with it factors that can affect their quality, so the industry (Blandino, 2005) within its project "The Beef Industry in Central America: Situation and Perspectives" defines quality as "the whole characteristics of a product or service that meets the explicit or implicit wishes of the consumer."

However, at present the tool that allows to evaluate characteristics of quality as the color and the texture of quantitative way using a simple digital image that generates a non-destructive, fast, precise, objective evaluation and that avoids the human inspection that in many occasions tends to be expensive and imprecise, are the Computer Vision Systems (SVC) as it allows to replace human manipulation using the robust algorithms used in image processing [1]. Consequently, this has led several researchers to perform work involving

Portero P., Mena B., Gavilanes J. works at Escuela Superior Politécnica del Chimborazo (e-mail: paolaportero2013@gmail.com, bellisiss@gmail.com, Javier.gavilanes@espoch.edu.ec)

SVCs, such as the Advanced Engineering Research and Development Center (CIDIA), which in 2008 developed a Meat Sorting Device that determines the quality of beef through the use of industrial digital cameras evaluating the color, the thickness of the subcutaneous meat and the marbling.

## II. QUALITY CONTROL IN THE MEAT INDUSTRY

Quality control in the meat industry, and especially in beef, was born in 1996 from the appearance of European cases on Bovine Spongiform Encephalopathy and the correlation that existed between this disease and humans. the main markets for this product, demanded systems and tools that would satisfy the safety of food and meat in particular to protect the safety of its consumers.

## III. ARTIFICIAL VISION VS. HUMAN VISION

The human being in tiny amounts of time has the capacity to acquire, process and assimilate information from the outside world, which may or may not be visible in its entirety and in varying lighting conditions [3], but the industry constantly uses the quality control and most are the employees who take care of it.

Where these processes usually have several hours of duration, so the individual tends to become tired or distracted generating erroneous results in the evaluation [4].

In contrast, with Artificial Vision (AV) and with less or equal assimilation times can be emulated variety of elemental vision processes obtaining a minimum of error. Concluding that AV systems can be an important factor in the economic development of human considering that even in several actions its improvement is necessary [3].

## IV. COMPUTER VISION MEASYSTEM

A CVS is technically divided into two systems: the perceptual system and the actuation system. The first consists of lighting systems, image capture and signal acquisition in the computer. The second uses a processing algorithm that provides information that helps us in making decisions and that follows a system of action responsible for making judgments on the outside of the system [5].

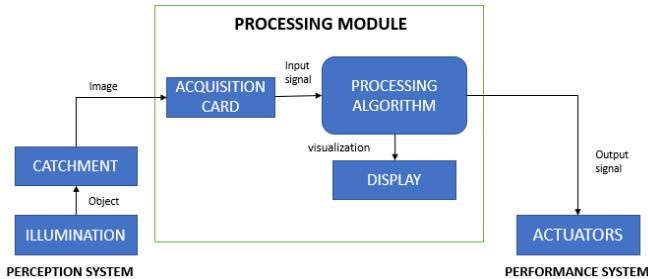


Fig. 1 Technical composition of an Artificial Vision System

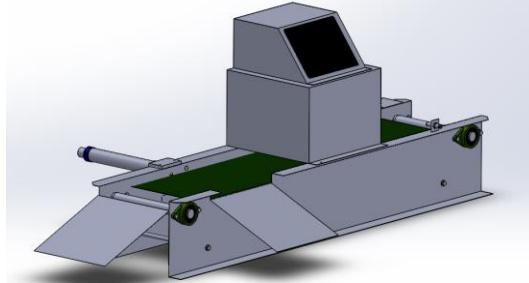


Fig. 2 Structure design

## V. DEVELOPMENT AND IMPLEMENTATION

Due to the fact of working with a food product and that the quality of this depends partly on the little or much manipulation that suffers, it has been considered to use a conveyor belt for the whole process its material is of stainless steel and is constituted by sensors, actuators and the SVA.

The control of the devices and the processing of the algorithm of the SVA is realized through a Raspberry Pi 3B (RPi), using the library of Open CV that allows the processing of images in real time, with more than 500 functions.

Figure 4 describes the operation of the system. Where, the initial sensors determine the existence of the product to start with the conveyor belt until arriving at the capsule, where, it will be integrated the SI and a sensor that allow to execute the algorithm of vision, once finished the analysis with the algorithm is executed again the conveyor belt to a sorting point where it will pass only the qualified product, otherwise it will be rejected to another location by the cylinder drive.

An essential part of a good SVA is the SI, due to the work done with color and texture differentiations, the lateral lighting technique has been selected, which allows highlighting of reliefs, edges, fissures and textures.

For the design, two lamps (L1 and L2) of 6W, 6000K with an external 20° angle were considered, to obtain a better distribution of light in the encapsulation of the vision system. Figure 5.



Fig. 3 Block diagram of prototype operation

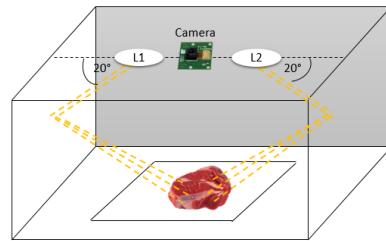


Fig. 4 LS Desing

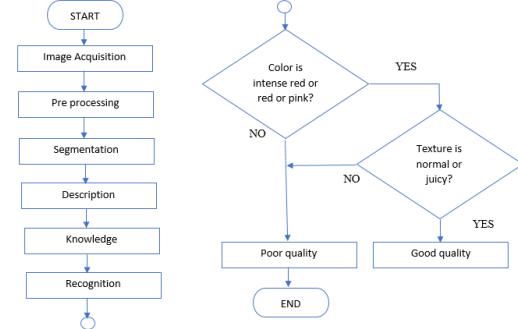


Fig. 5 AV Algorithm Flowchart

The SVA is based on the obtaining of organoleptic parameters of quality specifically of color and texture, using the CIELab and HSL color spaces, where the former uses the methodology (AMSA, 1992) together with the sensorial methodology allowing to obtain the characteristic of the color in an objective way based on the parameters a, b and L [6] while the second allows to obtain the brightness of the sample through the parameter L. For the characterization of the texture was taken as a guide the parameters extracted from the Department of Agriculture of the United States (USDA) and color was taken as a guide the meat color patterns developed by the Faculty of Veterinary Medicine and Animal Science of the National Autonomous University of Mexico (UNAM). However, beef scores and grades were classified into 5 scales. (Table V); information extracted from the document of the meat processing module of Ing. Tania Parra MSc. Magister in Livestock Industries Mention Industries of the Meat and teacher of meat processes of the Polytechnic School of Chimborazo.

The methodology used for the development of the vision algorithm is appreciated in the block diagram in Figure 6 and the function of each of the stages is detailed below.

### A. Image Acquisition

Technique that consists in digitizing the signal delivered by the camera of the RPi.

Let an image be a two-dimensional function  $f(x, y)$ , the digitization consists of decomposing this function into an  $M \times N$  matrix, which is no more than a grid of pixels

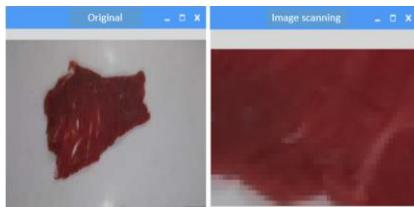


Fig. 6 Original (Left) and image scanning (Right)

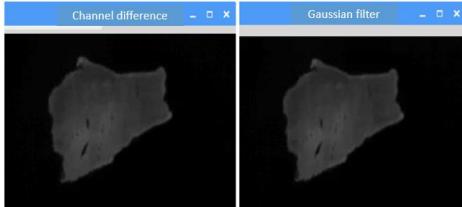


Fig. 7 Channel difference (left), Gaussian filter (right)

### B. Pre processing

Once the Acquisition is made, the pre-processing is performed to obtain an improvement of the digitized image.

The RGB image is converted to grayscale, followed by this the Gaussian Filter is applied in order to eliminate the noise that occurs during the capture, producing a smoother or smoother more uniform through the image than the filter of the half.

### C. Segmentation

Next, the segmentation process in charge of decomposing the image into regions is executed, in order to extract only the section that constitutes the area of the sample. From the binarization of the image.

Binarization consists of zeroing all the pixels smaller than a threshold value and one of those equal or greater, the final image being constituted by a set of ones and zeros, which is nothing more than the representation of the black and white image. The threshold in most cases is fixed, contrary to this study in which the color of the samples is continuously changed, for that reason uses the OTSU thresholding technique, which obtains a binary image (black and white) by means of an automatic threshold from a grayscale or bimodal image.

To find the value of the threshold, Otsu makes use of the histogram of a bimodal image obtaining an approximate one half of the two peaks representing the color black and white respectively. The automatic threshold value proves to be very useful for this segmentation, however, within the analysis of the images it has been considered to perform a variation (add a constant to the final threshold value) of the final value, allowing to eliminate fragments that are not of interest to image.

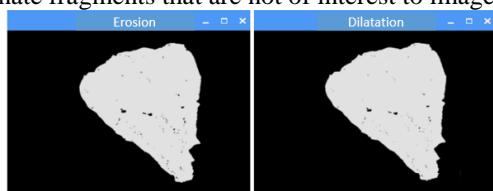


Fig. Morphological operations

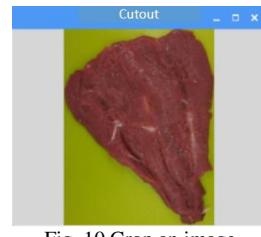


Fig. 10 Crop an image

Following this step morphological operations are executed, which are those in charge of exalting the geometry and shape of the objects, it is fundamental to use them prior to the extraction of characteristics [3]. For this case erosion is used followed by dilation.

Erosion removes pixels at the borders of the image, makes it smaller and dilation adds pixels at the borders of the image, makes it larger. These operations have been implemented in the event that veins are present, torn at the edges of the sample.

The last step of the segmentation is to extract or to cut the section of interest determining by means of the functions of OpenCV the coordinates of the pixels that conform the outline of the sample.

### D. Description

Five scales have been taken into account both in color and texture as can be seen in the following tables.

### E. Knowledge

The parameters of color and texture are obtained for later evaluation and to determine if the meat is of good or poor quality. To know the color variations that the image presents the conversion of RGB color space to CIELAB is performed which is a chromatic model frequently used to better describe the colors perceived by the human eye as it does not depend on any device resulting more objective to constantly describe more precisely the same color.

OpenCV through the use of its functions allows to obtain the maximum and minimum values of each parameter corresponding to the resulting color space. The color parameters representing each color are shown below.

TABLE I  
COLOR SCALE CONSIDERED

Very intense red	Intense Red	Red	Pink	Very light pink

TABLE II  
TEXTURES SCALE CONSIDERED

Very Juicy	Juicy	Normal	Dry	Very Dry

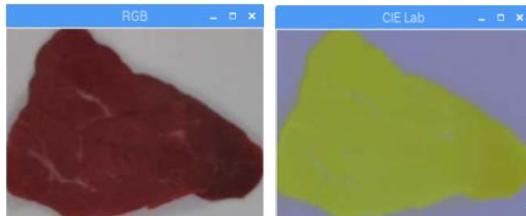


Fig. 11 RGB image (left), CIELab image (right)

TABLE III  
DESCRIPTION OF CIELAB PARAMETERS

Colors	CIE LAB PARAMETERS					
	Minimum Values			Maximum Values		
	L	A	B	L	a	B
Very intense red	19	136	131	38	145	137
Intense red	36	143	133	57	155	137
Red	43	152	136	97	159	142
Pink	59	156	132	99	166	149
Very light pink	45	145	133	78	157	142

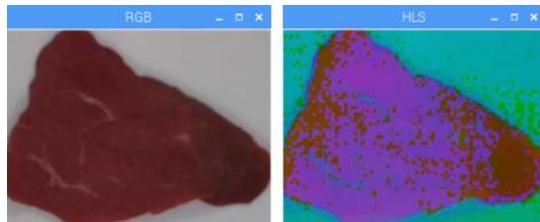


Fig. 12 RGB image (left), HLS image (right)

To determine the colors that will represent the juiciness within the texture algorithm, HLS is used because it allows to have a better appreciation of the brightness. In this case allows to differentiate the amount of water on the meat, as they will vary their values depending on how juicy or dry this is.

The texture parameters are obtained in the same way as the color parameters.

#### F. Recognition

Beef does not present a uniform color in its entirety, that is why the color algorithm is responsible for segmenting the meat by its color from very intense red to very light pink.

The segmentation is done by means of a mask, this being a binary image where its pixels are placed to one as long as the value of each of the pixels of the CIELab space are within the preset range for each color. The multiplication between the matrix of the mask and the matrix of the CIELab color space is then performed, resulting in an image containing only one color, which belongs to the 5 working scales.

TABLE IV  
DESCRIPTION OF HLS PARAMETER

HLS PARAMETERS					
Minimum Values			Maximum Values		
H	L	S	H	L	S
175	54	51	180	121	149

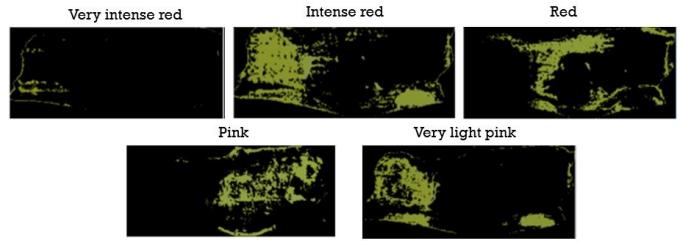


Fig. 13 Color filters on an image

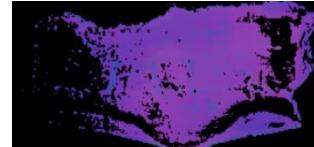


Fig. 14 Representation of juiciness

Meat color determination is performed by comparing between the areas of the resulting 5 images, and color will be defined by the image having the largest area.

As for the texture, the same procedure is executed as for the color, but only once and the juiciness will be determined by comparing the total area of the sample with the area of the image resulting from the multiplication between the mask and the HLS image.

And the quality will depend on the percentage yielded (Very dry: 30%, dry: 45%, normal: 50%, juicy: 55%, very juicy: 70%). To know if meat is juicy or not, the pixels of the mask add juiciness and compared with the total area of the meat.

#### G. Make decisions

The quality of the meat is considered according to its color and texture by means of the following conditions:

- RFN: Red meat, firm and juicy.
- PSE: Pale flesh, soft and juicy.
- DFD: Dark, firm and dry meat.

Where the first RFN describes quality meat and is the desirable condition for its ability to retain water, for its bright red color and for its firm texture. The two remaining conditions were considered to be of poor quality.

Within the algorithm, to comply with these conditions the colors that are accepted are: intense red, red and pink and for juicy and normal texture

## VI. TESTS AND RESULTS

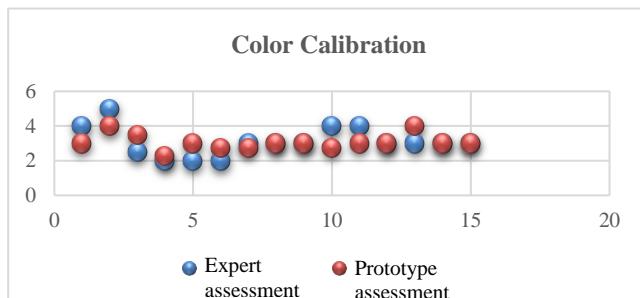
A comparative analysis of the machine is carried out with respect to a human sensory evaluation panel, to evaluate times, color variations and texture of each sample. For this purpose, 5 semi-expert judges were selected, who evaluated the characteristics of 15 samples in two replicates, scoring in 5 ranks (Table 5), based on the sensorial methodology where the weights of each judge were weighted by each sample to tabulate the data. While for the time it is considered the joint commitment of each judge for the 30 samples valued reaching an average time of 01:09:56.

TABLE V  
SCALES OF COLOR AND TEXTURE AND ITS WEIGHTING

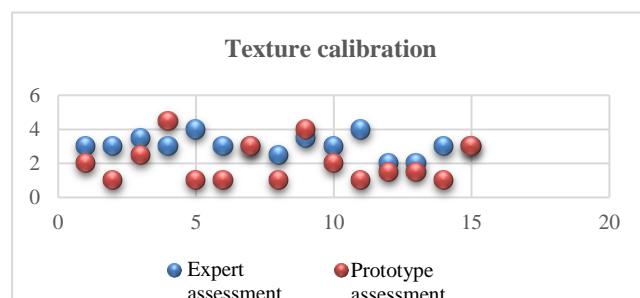
WEIGHTING	COLOR	TEXTURE
5	Very intense red	Very Juicy
4	Intense red	Juicy
3	Red	Normal
2	Pink	Dry
1	Very light pink	Very Dry

It should be added that the evaluation carried out by the prototype used the same samples and previous repetitions, taking as a guide the methodology (AMSA, 1992) and the sensory methodology from which only the values of the CIELab parameters of the color pattern will be taken. While for the texture the percentage levels of the L parameter of the HLS color space shown in the total area of the sample were taken.

After a proper calibration of parameters, better results were obtained as can be seen in the following graphs. For the analysis of this study, the t-dependent test for paired samples was used because it is desired to observe the behavior of the prototype against the recovery performed by the judges taking into account that the sample has been evaluated twice and that it is desired to verify significant differences in the behavior of the prototype. Using the SPSS statistical software, we finally obtained a correlation significance of 0.595 in color without significant differences, while for the texture a correlation significance of 0.77 was obtained, there being a significant difference between the two valuations.



Gráf. 1 Color calibration



Gráf. 2 Texture calibration

## VII. CONCLUSIONS

This device has provided an objective non-invasive method for assessing color and texture of meat, in intervals shorter times with respect to a panel of selected sensory judges for evaluating these organoleptic characteristics of beef, substituting this panel evaluation by a single user, responsible for management and interpretation of results easily, thanks to the friendly interface incorporated

## VIII. FUTURE WORKS

This device has provided a non-invasive objective method for assessing color and texture of the meat, in smaller time intervals with respect to a panel of sensorial judges generating a saving of approximately 67% and with an 18.384% error in the evaluation. Being able to replace the evaluating panel by a single user, the same one that is in charge of the handling and interpretation of the results of a simple way, thanks to the friendly interface incorporated.

The implemented system proved to be optimal. However, being a prototype has a range of error that can be corrected by future research.

## IX. REFERENCES

- [1] BLANDINO HERRERA, Luis Javier. La Industria de la carne bovina en centro América: Situación y Perspectivas [On Line], Costa Rica: Gráfica Litho Offset, (221-0121), 2005. [Query: December 15, 2016]. Available in: [https://books.google.com.ec/books?id=md8OAQAAIAAJ&pg=PA40&dq=calidad%20de%20carne%20bovina&hl=es-419&sa=X&redir\\_esc=y#v=onepage&q=calidad%20de%20carne%20bovina&f=false](https://books.google.com.ec/books?id=md8OAQAAIAAJ&pg=PA40&dq=calidad%20de%20carne%20bovina&hl=es-419&sa=X&redir_esc=y#v=onepage&q=calidad%20de%20carne%20bovina&f=false)
- [2] VALDEZ, G et al. "Acortando la brecha digital para la trazabilidad sanitaria: el problema de la transferencia tecnológica en la ganadería sonorense, caso SITAGAN". Estudios Sociales: Revista de Investigación científica [On Line], 2011, 19(37), pp.141-174, [Query: January 20, 2017]. ISSN-e 0188-4557. Available in: <https://dialnet.unirioja.es/servlet/articulo?codigo=3424029>
- [3] SOMOLINOS, J. Avances en Robótica y Visión por Computador [On Line Cuenca: Ediciones de la Universidad de Castilla-La Mancha, 2002. [Query: March 25, 2017] Available in:[https://books.google.com.ec/books?id=VXwzEIngQC&pg=PA31&so=urce=gbs\\_toc\\_r&cad=4#v=onepage&q&f=false](https://books.google.com.ec/books?id=VXwzEIngQC&pg=PA31&so=urce=gbs_toc_r&cad=4#v=onepage&q&f=false)
- [4] LOZANO MANTILLA, German & ORDUZ ROFRIGUEZ, Jairo. Diseño de un Sistema de Visión Artificial para la Revisión del Nivel de Llenado de Bebidas Embotelladas [On Line] (tesis). (Ingeniería) Universidad Autónoma del Caribe. Facultad de Ingenierías. Ingeniería Mecatrónica. Colombia 2015. p.16. [Query: April 18, 2017]. Available in: <http://repositorio.uac.edu.co/bitstream/handle/11619/1371/TMCT%200013C.pdf?sequence=1&isAllowed=y>
- [5] GROOVER, Mikell. Robótica Industrial. [On Line]. España: MICGRAW-HILL, 1989. [Query: April 2, 2017]. Available in: <https://www.casadellibro.com/libro-robotica-industrial/9788476153024/155000>
- [6] BRAÑA, Diego, et al. "Manual de Análisis de Calidad en Muestras de Carne", Centro Nacional de Investigación Disciplinaria en Fisiología y Mejoramiento Animal Folleto técnico, ISBN:978-607-425-612-3, No. 11,2011, México, pp.21-32



**Portero P.** was born in Ecuador, Ambato City, in 1993. She received the B.S. in Electronic Engineering in Control and Industrial Networks from Escuela Superior Politécnica del Chimborazo, Riobamba Ecuador in 2017. In 2013 - 2014 She was treasurer in IEEE-ESPOCH Student Branch. In 2014 - 2015 She was president of the affinity group WIE-ESPOCH.



**Mena B.** was born in Ecuador, Latacunga City, in 1992. She received the B.S. Electronic Engineering in Control and Industrial Networks from Escuela Superior Politécnica del Chimborazo, Riobamba Ecuador in 2017. From 2012 to 2016, she was a member IEEE. In the 2013 she was president of the affinity group WIE-ESPOCH and the 2015 she was president of the Student branch IEEE-ESPOCH.



**Gavilanes J.** was born in Ecuador, Santo Domingo City, in 1987. He received the B.S. in electronics engineering from Escuela Superior Politécnica del Chimborazo, Riobamba Ecuador in 2012 and the MSc. degree in automation and robotics from University Polytechnic of Madrid, Spain in 2015.

In 2014 -2015 He was a Research Assistant in Field & Service Robotics Group from Centre for Automation and Robotics (CSIC-UPM) in Spanish National Research Council. He collaborated in the Tiramisu European project.

Since 2016, He has been professor in Escuela Superior Politécnica del Chimborazo, Riobamba Ecuador and he has been Research Assistant with Biomechanics Studies Research Group. His research interests include artificial vision, biomechanics, mobile robots and additive manufacturing.

# Deep Learning como modelo predictivo para clasificar dígitos manuscritos

## *Deep Learning as a predictive model to classify handwritten digits*

Omar Alexander Ruiz-Vivanco

**Resumen**—En este trabajo se presentan los resultados de aplicar modelos de predicción Deep Learning para identificar el dígito de una imagen que contiene un número escrito a mano de la base de datos MNIST, este conjunto de dataset se obtiene de la competencia de Kaggle: Digit Recognizer. Como primer paso se utilizan técnicas de preprocesamiento de imágenes las cuales se enfocan en obtener una imagen lo más nítida posible y la reducción de tamaño de la misma, objetivos que se logran con técnicas de umbralización por el Método de Otsu, transformada de Wavelet de Haar y el Análisis de sus Componentes Principales (ACP), para obtener como resultado un conjunto de nuevos dataset para ser evaluados. En el siguiente paso se toman estos dataset y se aplican los modelos Deep Learning MxNET y H2o ejecutados en el lenguaje estadístico R, obteniendo varias predicciones. Por último se envió a la competencia Digit Recognizer la mejor de las predicciones obtenidas en el experimento, y el resultado de la evaluación de la misma fue de 99,129 % de predicción.

**Palabras Claves**—Inteligencia Artificial, Deep Learning, Dígitos manuscritos, Método Otsu, Wavelet Haar.

**Abstract**—In this research work, the results of applying Deep Learning prediction models to identify the digit of an image, that contains a handwritten number of the MNIST database, are presented. This set of dataset was acquired from the competition of Kaggle: Digit Recognizer. The following process was applied: First, image preprocessing techniques were used, which focus on obtaining a pretty clear image and to reduce the size of the same, these objectives that are achieved with Otsu Method, transformed from Haar Wavelet and the Principal Component Analysis (PCA), thus obtaining as a result, one set of new dataset to be evaluated. Second, the Deep Learning MxNET and H2o models, which were executed in the statistical language R, were applied to these datasets obtained, this way, several predictions were acquired. Finally, the best obtained predictions in the experiment were sent to the Digit Recognizer competition, and the results of this evaluation scored 99,129% of prediction.

**Index Terms**—Artificial Intelligence, Deep Learning, Handwritten digits, Otsu Method, Wavelet Haar.

### I. INTRODUCCIÓN

La investigación se fundamenta en la competición de Kaggle denominada Digit Recognizer [1], cuya finalidad es el reconocimiento de dígitos escritos a mano. Kaggle es una plataforma en donde las empresas y los investigadores publican datos y estadísticas con el objetivo de brindar a participantes de cualquier parte del mundo la posibilidad de producir los mejores modelos predictivos a través de competiciones en donde se presentan un sinnúmero de estrategias para

el mismo problema [2]. Los dataset para esta competición en particular son tomados de MNIST [3] que es una gran base de datos de dígitos escritos a mano de acceso público y usada ampliamente para entrenamiento y pruebas en el campo de aprendizaje automático [4].

De acuerdo con [3] sacar conclusiones razonables a partir de experimentos de aprendizaje requiere que el resultado sea independiente de la elección de los conjuntos de entrenamiento y de prueba. La base de datos MNIST tiene 70.000 registros particionada en dos dataset, el primero con 60.000 ejemplos utilizados como dataset de entrenamiento (training) y el segundo con 10.000 ejemplos utilizados como dataset de prueba (test). Esta partición es estándar en la definición de MNIST [3] y utilizada en ese formato por la mayoría de trabajos que han experimentado con el dataset [5], [6], [7], [8], [13]. La base de datos MNIST se elaboró con la colaboración de estudiantes de secundaria y empleados de la oficina del censo mostrando diferencias en los trazos, siendo los dígitos formado por el segundo grupo más limpio y más fácil de reconocer, MNIST es la mezcla aleatoria de ambos [3], [4].

La competencia Digit Recognizer [1] propone la utilización de la base de datos MNIST con una partición diferente, los registros se han mezclado y dividido con 42.000 elementos para formar el dataset de entrenamiento y 28.000 elementos para el dataset de prueba. Esta partición fué la que se tomó para la experimentación en el presente trabajo, siendo mejor en el sentido de que al mezclar la partición original de MNIST obliga a que en el aprendizaje de los dígitos se tomen nuevos elementos al igual que en el test. La imagen del dígito manuscrito está formada por 784 columnas con valores entre 0 para el color blanco y 255 para el color negro que representan los píxeles que conforman la imagen de 28x28 en escala de grises. El dataset de entrenamiento es el único que contiene una columna adicional en cada fila de elementos que etiqueta el dígito de la imagen. En la Figura 1 se muestran 100 dígitos tomados al azar del dataset de entrenamiento de Kaggle, el dígito manuscrito de color negro y la etiqueta de color rojo.

Otras bases de datos utilizadas para el reconocimiento de patrones con varias publicaciones han sido CENPARMI (Centre for Pattern Recognition and Machine Intelligence) [10] fundado en 1972 por los profesores del Departamento de Ciencias de la Computación de la Universidad de Concordia en Canadá, cuyos miembros se dedican a la investigación multidisciplinaria del reconocimiento de escritura a mano y

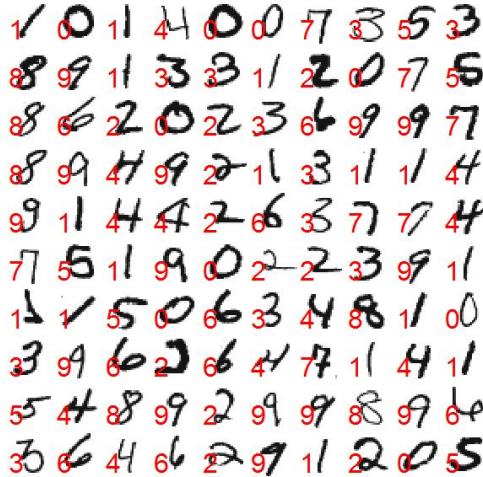


Figura 1. Dígitos del dataset de entrenamiento de Kaggle

verificación de firma, análisis de imágenes y procesamiento de documentos, combinando diferentes técnicas de clasificación. **CEDAR** [11] imágenes de caracteres japoneses creada en la Universidad Estatal de Nueva York en Buffalo desde 1978, estas imágenes se extrajeron de una variedad de fuentes de documentos, incluidos libros, faxes, revistas, impresoras láser, revistas y periódicos. Sin embargo MNIST es la base de datos de mayor utilización en experimentos de aprendizaje máquina especializada en patrones de dígitos escritos a mano.

La principal contribución de este trabajo es la combinación de las técnicas de preprocessamiento de imágenes con los modelos de predicción, trabajos previos no realizan un análisis de preprocessamiento de imágenes como en [12] donde se utiliza una arquitectura de redes neuronales, en [7] se compara el rendimiento de varios algoritmos clasificadores en una base de datos estándar de dígitos manuscritos, en [8] se crea un clasificador neuronal Limited Receptive Area (LIRA), en [13] se utiliza retro-propagación en línea para perceptrones lisos de múltiples capas.

Tomando en consideración lo mencionado se estima que la combinación de técnicas previas para el tratamiento de imágenes y los modelos predictivos, según los resultados del presente trabajo, se obtienen predicciones comparables con los principales artículos científicos escritos sobre el tema, lo que justifica por si mismo la razón de dar conocer esta investigación que es la síntesis del trabajo de tesis *Análisis predictivo para clasificar dígitos escritos a mano utilizando la base de datos MNIST* en [14] del autor.

## II. TÉCNICAS Y MODELOS

Las utilidades que se obtienen hoy en día del reconocimiento automático de dígitos a partir de una imagen pueden ser muy variadas, lo importante es desarrollar una técnica eficiente a partir de datos de entrenamiento. El aprendizaje de máquina (machine learning) consiste en “enseñar”, a un computador a reconocer dígitos escritos a mano por distintas personas en donde factores como intensidad del trazo, curvatura, posición de la mano, etc., generan múltiples variaciones del dibujo lo que constituye el problema a resolver.

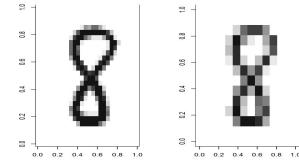


Figura 2. Reducción del dígito 8 dataset Kaggle

En esta sección se mencionan las técnicas de preprocessamiento de imágenes utilizadas para mejorar la calidad de las mismas. En la minería de datos se extrae información de un conjunto de datos para transformarla en una estructura más comprensible y manejable. La técnica a utilizar es el aprendizaje supervisado en donde se deduce una función a partir de datos de entrenamiento para obtener una etiqueta de clase [15], es decir la clasificación de dígitos del 0 al 9. Esta tarea se lleva a cabo realizando un análisis de los datos iniciales para encontrar las técnicas de preprocessamiento más adecuadas para luego aplicar el modelo predictivo que maximice el grado de acierto en la predicción de los dígitos. Estas técnicas utilizadas son Wavelet Haar [20] para reducir el tamaño de la imagen. Umbralización por Otsu [16] para obtener una imagen más nítida y por último PCA [21] para el análisis de los componentes principales.

### II-A. Material y métodos

**II-A1. Análisis del dataset:** Los dígitos del dataset contienen varias formas de escritura con diferentes trazos, curvatura, forma, etc., es decir diferencias muy particulares de cada escritor. Con el objetivo de encontrar tendencias en la escritura de los dígitos se analiza el nivel de intensidad de cada uno, añadiendo esta nueva característica al dataset mediante la obtención de la media. Esta característica de intensidad de trazo puede ser útil al momento de realizar el análisis predictivo. Con esta técnica se obtiene el dataset llamado: “train\_kaggle\_int”.

**II-A2. Reducción de la imagen:** La reducción de la dimensión de los datos se logra aplicando la transformada de Wavelet Haar, esta función transforma cualquier secuencia 1

$$(a_0, a_1, \dots, a_{2n}, a_{2n+1}), \quad (1)$$

de cualquier longitud en una secuencia de dos componentes vectoriales 2

$$((a_0, a_1), \dots, (a_{2n}, a_{2n+1})), \quad (2)$$

Si se multiplica por la derecha cada vector con la matriz  $H2$ , se obtiene el resultado 3

$$((s_0, d_0), \dots, (s_n, d_n)), \quad (3)$$

de una etapa de la transformada rápida de wavelet de Haar. La función devuelve las secuencias s y d y se utiliza la secuencia s para continuar con los cálculos [20]. En el dataset train de Kaggle se utiliza dos etapas de cálculo equivalentes a tomar cuatro píxeles continuos y promediar su valor para obtener un nuevo píxel que reemplace a los anteriores. El objetivo es reducir la imagen de cada dígito manuscrito de 28x28 píxeles

a una imagen de 14x14 píxeles. Para lo cuál se utiliza el principio de Wavelet de Haar. En R la función ht, de la librería binhf [22], proporciona la transformada. El Algoritmo 1 es utilizado para lograr la reducción del dataset original de 784 columnas a 196.

#### Algoritmo 1 Transformada Wavelet de Haar

**Entrada:** Seleccionar matriz original (28x28)

**Salida:** Matriz Reducida (14x14)

```

1: para  $i = 1$  hasta 14 veces hacer
2:   para  $i = 1$  hasta 14 veces hacer
3:     vector = matriz original[i, j]
4:     haar = funcion wavelet haar(vector)
5:     haar = funcion wavelet haar(haar$)
6:     matriz reducida[i, j] = funcion redondear(haar$)
7:   fin para
8: fin para
9: devolver matriz reducida

```

El resultado de esta fase devuelve un dataset con una reducción a 196 columnas más la columna de etiqueta, Figura 2. Se obtiene la reducción del dataset llamado: "train\_kaggle\_r".

**II-A3. Binarización de la imagen:** La binarización es el método mediante el cuál se convierte los píxeles de una imagen de varios valores a únicamente dos (negro: 255 y blanco: 0). Para lo cual se establece un umbral sobre el cual todos los valores se convierten en 255 y por debajo de este en 0. El método de Otsu [16], [17] encuentra el umbral óptimo (threshold) maximizando la varianza entre clases (between-class variance) mediante una búsqueda exhaustiva. La umbralización se utiliza cuando la diferencia entre la imagen y el fondo existe una clara diferencia [18].

Se fundamenta en la similitud entre los píxeles que forman el objeto principal y sus diferencias respecto al resto de objetos. La escena debería tener un fondo uniforme y objetos parecidos. A continuación se describe la fórmula de la técnica de Otsu basada en estadísticas sobre el histograma unidimensional de una imagen [19].

$$T = \max(\sigma^2) \quad (4)$$

siendo

$$\sigma^2 = w_B(\mu_B - \mu)^2 + w_F(\mu_F - \mu)^2 \quad (5)$$

$$w_k = \sum_{i=0}^k p_i \quad (6)$$

$$\mu_k = \sum_{i=0}^k i.p_i \quad (7)$$

$$\mu_c = \frac{\mu_k}{w_k} \quad (8)$$

Donde  $p_i$  es la probabilidad de aparición de un determinado nivel  $i$ . Se aplica la binarización, Figura 3, al dataset train original de 784 columnas, y al dataset train reducido de 196 columnas, además se aumenta la columna de intensidad a ambos para obtener un total de 4 dataset como resultado de esta fase: "train\_kaggle\_bo", "train\_kaggle\_br", "train\_kaggle\_boi", y "train\_kaggle\_bri".

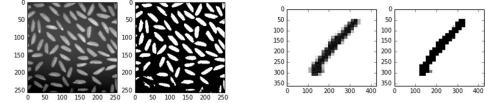


Figura 3. Binarización: Izq: grano de arroz, Der: dígito 1 dataset Kaggle

**II-A4. Extracción de nuevas variables:** El análisis de componentes principales es una técnica de transformación con el objetivo de identificar los principales factores que explican las diferencias en los datos y una descripción más comprimida de sus correlaciones. Constituye una herramienta para extraer tendencias generales de un conjunto de datos. PCA ordena los componentes principales por su significado ofreciendo una excelente base para la reducción de la dimensión de datos multidimensionales [21]. Este proceso identifica un grupo de variables ficticias que se forman de la combinación de las anteriores observadas. De esta forma se sintetizan los datos y se relacionan entre sí, sin recurrir a ninguna hipótesis previa sobre el significado de cada factor inicial. Los componentes principales obtenidos mediante el proceso de cálculo de raíces y vectores característicos de una matriz simétrica contienen la mayoría de la varianza observada, evitando la información redundante. Para esto las variables deben ser incorreladas entre sí y se expresan como combinación lineal de las demás variables que han sido observadas. El porcentaje de mayor varianza mostrada en cada una de estas componentes se traduce en que las mismas contiene mayor cantidad de información. Se aplica la reducción por PCA para las 64 primeras componentes de los dataset: train kaggle bo y train kaggle br describiendo el 99.21 % y el 99.28 % de las características respectivamente. Dataset resultantes de este proceso: "train\_kaggle\_pca\_br", "train\_kaggle\_pca\_bo".

**II-A5. Modelo predictivo:** El Deep Learning [9] forma parte de un conjunto amplio de métodos de aprendizaje automático los cuales se basan en aprender representaciones de datos, mediante una serie de operaciones matemáticas sobre una lista de números, dando como resultado otra lista de números. Por ejemplo el reconocimiento de imágenes, se codifica como una lista de números, por tanto, la red neuronal recibiría tantos números en su entrada como píxeles tienen las imágenes o el caso de imágenes en color tres por cada píxel. Su salida bastaría con un solo número cercano a 1.0 para determinar si se trata de un rostro (figura, letra, dígito, etc.) o cercano a 0.0 si no reconoce un rostro, los valores intermedios se interpretan como inseguridad o probabilidad.

La arquitectura de las redes de neuronas [23] se basa en la conexión que tienen las neuronas en cada capa con la siguiente, estas conexiones tienen asociado un número llamado peso. La principal operación dentro de las redes consiste en multiplicar los valores de una neurona por los pesos de sus conexiones salientes, luego cada neurona de la siguiente capa suma todos los input que recibe de las conexiones entrantes. En cada capa se utiliza una función de activación que se encarga de mantener los números producidos por cada neurona dentro de un rango razonable, números reales entre 0 y 1. La función más habitual es la sigmoidal que es no lineal, es decir que si se grafican los valores de entrada y de salida el resultado no sería una línea.

Tabla I  
RESULTADOS DEL PREPROCESAMIENTO

Preprocesamiento del dataset train de Kaggle			
Tratamiento	Dataset Resultante	Dimensión	Tiempo (seg.)
ninguno	train_kaggle	42000 x 785	0,00
int	train_kaggle_int	42000 x 786	4,57
red	train_kaggle_r	42000 x 197	110,17
bin	train_kaggle_bo	42000 x 785	288,19
red + bin	train_kaggle_br	42000 x 197	162,91
bin + int	train_train_kaggle_boi	42000 x 786	292,76
red + bin + int	train_kaggle_bri	42000 x 198	167,48
bin + pca	train_kaggle_pca_bo	42000 x 65	315,68
red + bin + pca	train_kaggle_pca_br	42000 x 65	164,96

bin = binarización (umbralización por método de Otsu)

red = reducción (transformada Wavelet de Haar)

int = intensidad (aumento de nueva característica)

pca = análisis de componentes principales

Cuando se nombra a una red neuronal se lo hace así: n-layer. La capa de salida se toma para representar los resultados de clase, 10 en el caso de los dígitos manuscritos de Kaggle. Se dimensiona la red neuronal por el número de neuronas o por el número de parámetros.

### III. EXPERIMENTACIÓN Y RESULTADOS

Se procura que la experimentación con los modelos predictivos sea en las mismas condiciones, para esto se fija la misma semilla<sup>1</sup> en todas las pruebas y se realiza el mismo procedimiento. Teniendo en cuenta que el dataset test proporcionado por Kaggle no contiene la columna de etiqueta se hace difícil la comprobación de la predicción realizada en la experimentación. Se crea por lo tanto una muestra a partir del dataset train para obtener nuevos conjuntos de training y testing que contengan ambos la etiqueta del dígito a predecir. Para estas pruebas en particular se obtienen un subconjunto de 10.000 dígitos para testing y otro de 32.000 dígitos para training.

#### III-A. Experimentos con los modelos predictivos

**III-A1. Preprocesamiento de datos:** En el transcurso del presente trabajo se obtuvieron algunos dataset con diferente tipo de tratamiento previo. La Tabla I muestra el resumen del preprocesamiento realizado con el dataset train original de Kaggle. Se puede identificar el tratamiento realizado, el nombre del dataset resultante con la descripción de sus dimensiones y el tiempo empleado en el proceso expresado en segundos.

**III-A2. Análisis predictivo con los modelos:** Tres son los modelos predictivos utilizados para la experimentación, el primero h2o con su implementación de Deep Learning ofrece escalabilidad y rapidez con configuraciones que permiten el multiprocesamiento, es decir el uso de dos o más procesadores o CPUs, y entre otras, la posibilidad de especificar el tamaño

<sup>1</sup>En R se utiliza la función set.seed( ) para obtener números aleatorios reproducibles, el argumento es un número entero cualquiera.

Tabla II  
RESULTADOS DE LA EXPERIMENTACIÓN DEL ANÁLISIS PREDICTIVO

Experimentación del modelo predictivo con los dataset obtenidos			
Dataset	Accuracy (%)	T. (seg.)	T. Total (seg.)
Modelo h2o			
train_kaggle	98,41	22371,26	22371,26
train_kaggle_bo	98,22	18331,55	18619,74
train_kaggle_br	97,59	26224,96	26387,87
train_kaggle_boi	98,09	15515,57	15808,33
train_kaggle_bri	97,53	27115,42	27282,90
train_kaggle_pca_bo	98,49	15027,91	15343,59
train_kaggle_pca_br	97,68	18342,16	18507,12
Modelo MXNet			
train_kaggle	98,04	4291,57	4291,57
train_kaggle_bo	98,82	4875,65	5163,84
train_kaggle_br	97,15	879,83	1042,74
train_kaggle_pca_bo	81,89	416,20	731,88
train_kaggle_pca_br	85,60	413,43	578,39
Modelo RN 2l			
train_kaggle	93,55	806,36	806,36
train_kaggle_bo	94,33	1025,34	1313,53
train_kaggle_br	91,12	369,55	532,46

de memoria, aspectos importantes que hacen de esta herramienta una de las más utilizadas en problemas de clasificación y con suficiente documentación disponible en [24]. La versión y configuración utilizada para estas pruebas permiten el uso de máximo 4 unidades de procesamiento con un tamaño de memoria de 12 GB, pero las restricciones de hardware de la máquina utilizada permite un máximo de 2 CPUs y un máximo de 10,67 GB, lo que hace que los tiempos para este modelo no sean de los mejores presentados.

El segundo modelo MXNet [25] dispone de una flexible y eficiente librería para Deep Learning, se ejecuta en las CPU o GPU4, en los clústeres, servidores, ordenadores de sobremesa, o teléfonos móviles [26]. Este modelo evidencia los mejores resultados en tiempo de ejecución con el mismo hardware. Y por último el tercer modelo es una implementación sencilla codificada en R de una red neuronal 2-layer en la que se pretende realizar un ensayo didáctico enfocado en el conocimiento más que en obtener los mejores resultados. La Tabla II muestra el resumen de la experimentación en donde se identifica el modelo predictivo, el dataset insumo para los cálculos, la exactitud de predicción, el tiempo empleado y el tiempo total que consolida los tiempos globales tomando en cuenta el tiempo de preproceso.

#### III-B. La competencia de Kaggle Digit Recognizer

El insumo para la presentación en la competencia de Kaggle se obtiene utilizando el modelo y preproceso con el mejor resultado, esto es el modelo MXNet con el dataset binarizado que obtuvo el registro de 98,82 % con un tiempo de 5163,84 segundos. La Tabla III muestra el resultado de la participación en la competencia, alcanzando un 99,129 % de efectividad en la predicción de dígitos manuscritos de la competencia. Participación realizada el 30 de agosto de 2016.

Tabla III  
RESULTADOS DE LA COMPETENCIA DE KAGGLE: DIGIT RECOGNIZER

Tratamiento	Tiempo (seg.)	Modelo	Accuracy (%)	Tiempo (seg.)
bin	288,19	MXNet	99,129	6940,212

bin = binarización (umbralización por método de Otsu)

#### IV. CONCLUSIONES

- El modelo h2o y el modelo MXNet alcanzan similares resultados sin embargo el segundo tiene un coste computacional mucho más bajo, constituyéndose en el modelo con los mejores resultados de la presente investigación.
- La técnica de añadir una nueva característica al dataset mediante la obtención de la media de la intensidad no mejora la efectividad predictiva para ninguno de los dos dataset probados.
- Binarizar la imagen constituye el mejor preproceso de la presente investigación y se evidencia en las imágenes producidas, siendo estas más nítidas que las originales.
- La reducción no mejora los resultados obtenidos con el dataset original esto puede ser comprensible desde el hecho que visualmente se nota la diferencia en las imágenes producidas por cada dataset.
- El porcentaje de efectividad predictiva (accuracy) es alto para toda la experimentación. El mejor resultado global de las pruebas es de 98,82 %, y la calificación obtenida en la participación de la competencia Digit Recognizer es de 99,129 % de efectividad de predicción para datos reales. Esta calificación da como resultado un margen de error del 0,871 %.
- La utilización del método de Otsu combinado con el modelo MXNet presenta el mejor porcentaje de predicción de este trabajo.

#### AGRADECIMIENTOS

El autor agradece al Dr. Iñaki Inza y al Dr. Yosu Yurramendi director de tesis y responsable, respectivamente, del Máster Universitario en Ingeniería Computacional y Sistemas Inteligentes de la Universidad del País Vasco / Euskal Herriko Unibertsitatea (UPV/EHU), por su apoyo en la consecución de los objetivos académicos planteados en la realización del máster.

#### REFERENCIAS

- [1] Kaggle, (2016). Digit Recognizer. *Learn computer vision fundamentals with the famous MNIST data*. Recuperado de: <https://www.kaggle.com/c/digit-recognizer>.
- [2] Kaggle, (2016). The Home of Data Science & Machine Learning. *Kaggle helps you learn, work, and play*. Recuperado de: <https://www.kaggle.com/>.
- [3] Y. LeCun and C. Cortes and C. Burges, (2015). THE MNIST DATABASE of handwritten digits. Recuperado de: <http://yann.lecun.com/exdb/mnist/>.
- [4] L. Deng. (2012). The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Processing Magazine*, 29, pp. 141142.
- [5] Y. LeCun, L. Bottou, Y. Bengio, y P. Haffner. (1998). Gradient-based learning applied to document recognition. *IEEE*, 86(11), pp. 2278-2324.
- [6] P. Y. Simard, D. Steinkraus, y J. C. Platt. (2003). Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis. *Document Analysis and Recognition. IEEE Computer Society Washington, DC, USA*, pp. 958.

- [7] L. Bottou, C. Cortes y J.S. Denke. (2002). Comparison of classifier methods: a case study in handwritten digit recognition. *IEEE*. DOI: 10.1109/ICPR.1994.576879.
- [8] E. Kussul, T. Baidyk. (2004). Improved method of handwritten digit recognition tested on MNIST database. *Image and Vision Computing*, 22(12), pp. 971-981.
- [9] J. Schmidhuber, (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, pp. 85 - 117.
- [10] CENPARMI, (2010). About CENPARMI. *Centre for Pattern Recognition and Machine Intelligence*. Recuperado de: <http://www.concordia.ca/research/cenparmi.html>.
- [11] CEDAR, (2008). About CEDAR. *Handwriting Recognition*. Recuperado de: <https://www.cedar.buffalo.edu/Databases/JOCR/>.
- [12] D. Ciregan, U. Meier, y J. Schmidhuber. (2012). Multi-column deep neural networks for image classification. *IEEE*, DOI: 10.1109/CVPR.2012.6248110.
- [13] D. Ciregan, U. Meier, L. M. Gambardella y J. Schmidhuber. (2012). Deep Big Simple Neural Nets Excel on Handwritten Digit Recognition. *IEEE*, DOI: 10.1162/NECO\_a\_00052.
- [14] Ruiz-Vivanco, Omar. (2016). *Análisis predictivo para clasificar dígitos escritos a mano utilizando la base de datos MNIST* (tesis de maestría). Universidad del País Vasco / Euskal Herriko Unibertsitatea (UPV/EHU), España.
- [15] Google Brain Team. (2016). *TensorFlow is an Open Source Software Library for Machine Intelligence*. Recuperado de: <https://www.tensorflow.org/>.
- [16] L. Jianzhuang, L. Wenqing y T. Yupeng. (2002). The Automatic Thresholding of Gray-Level Pictures via Two-Dimensional Otsu Method. *IEEE*, DOI: 10.1109/CICCAS.1991.184351 .
- [17] L. Hui, C. Shi, A. Min-si y W. Yi-qi. (2008). Application of an Improved Genetic Algorithm in Image Segmentation. *IEEE*, DOI: 10.1109/CSSE.2008.794.
- [18] J. Rodríguez. (2010). *Método otsu*, Recuperado de: <https://es.scribd.com/doc/45875911/Metodo.Otsu/>.
- [19] Universidad Nacional de Quilmes. (2005). *Segmentación por umbralización - método de otsu*. Recuperado de: [www.iaci.unq.edu.ar](http://www.iaci.unq.edu.ar).
- [20] G. Jin-Sheng, y J. Wei-Sun. (2007). The Haar wavelets operational matrix of integration. *International Journal of Systems Science*, 27, pp. 623-628.
- [21] J. Yang, D. Zhang y A.F. Frangi. (2004). Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE*, DOI: 10.1109/TPAMI.2004.1261097.
- [22] M. Nunes. (2014). *Package binhf*, Recuperado de: <https://cran.rproject.org/web/packages/binhf/binhf.pdf>.
- [23] J. Johnson y A. Karpathy. (2016). *Cs231n: Convolutional neural networks for visual recognition*. Recuperado de: <http://cs231n.github.io>.
- [24] H2O Community. (2016). H2O Open Source Software Documentation. *H2O and Sparkling Water Documentation*. Recuperado de: <http://docs.h2o.ai/h2o/latest-stable/index.html>.
- [25] KDnuggets, (2016). *Top 10 Deep Learning Projects on Github*. Recuperado de: <http://www.kdnuggets.com/2016/01/top-10-deeplearninggithub.html>.
- [26] MXNet Community. (2016). MXNet Architecture. *Flexible and Efficient Library for Deep Learning*. Recuperado de: <https://mxnet.readthedocs.io/en/latest/>.



**Omar Alexander Ruiz-Vivanco** Ingeniero en Sistemas por la Universidad Nacional de Loja (2008), Máster Universitario en Ingeniería Computacional y Sistemas Inteligentes por la Universidad del País Vasco / Euskal Herriko Unibertsitatea (UPV/EHU) (2016). Actualmente es Docente - Investigador en la Universidad Técnica Particular de Loja en la Sección de Inteligencia Artificial del Departamento de Ciencias de la Computación y Electrónica.





Published by:

Escuela Politécnica Nacional  
Facultad de Ingeniería de Sistemas  
Departamento de Informática y Ciencias de la Computación  
Ecuador

<http://lajc.epn.edu.ec/>  
lajc@epn.edu.ec

November 2017

