

Advances in Pattern Recognition

Research in Computing Science

Series Editorial Board

Editors-in-Chief:

Grigori Sidorov, CIC-IPN, Mexico
Gerhard X. Ritter, University of Florida, USA
Jean Serra, Ecole des Mines de Paris, France
Ulises Cortés, UPC, Barcelona, Spain

Associate Editors:

Jesús Angulo, Ecole des Mines de Paris, France
Jihad El-Sana, Ben-Gurion Univ. of the Negev, Israel
Alexander Gelbukh, CIC-IPN, Mexico
Ioannis Kakadiaris, University of Houston, USA
Petros Maragos, Nat. Tech. Univ. of Athens, Greece
Julian Padget, University of Bath, UK
Mateo Valero, UPC, Barcelona, Spain
Rafael Guzmán, Univ. of Guanajuato, Mexico

Editorial Coordination:

Alejandra Ramos Porras
Carlos Vizcaino Sahagún

Research in Computing Science es una publicación trimestral, de circulación internacional, editada por el Centro de Investigación en Computación del IPN, para dar a conocer los avances de investigación científica y desarrollo tecnológico de la comunidad científica internacional. **Volumen 148, No. 9**, septiembre de 2019. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No.: 04-2019-082310242100-203, expedido por el Instituto Nacional de Derecho de Autor. *Certificado de Licitud de Título* No. 12897, *Certificado de licitud de Contenido* No. 10470, expedidos por la Comisión Calificadora de Publicaciones y Revistas Ilustradas. El contenido de los artículos es responsabilidad exclusiva de sus respectivos autores. Queda prohibida la reproducción total o parcial, por cualquier medio, sin el permiso expreso del editor, excepto para uso personal o de estudio haciendo cita explícita en la primera página de cada documento. Distribuida por el Centro de Investigación en Computación, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, México, D.F. Tel. 57 29 60 00, ext. 56571.

Editor responsable: *Grigori Sidorov, RFC SIGR651028L69*

Research in Computing Science is published by the Center for Computing Research of IPN. **Volume 148, No. 9**, September 2019. The authors are responsible for the contents of their articles. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of Centre for Computing Research.

Advances in Pattern Recognition

J. Arturo Olvera-López
J. Francisco Martínez-Trinidad
J. Ariel Carrasco-Ochoa
Joaquín Salas-Rodríguez (eds.)



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



Instituto Politécnico Nacional, Centro de Investigación en Computación
México 2019

ISSN: 1870-4069

Copyright © Instituto Politécnico Nacional 2019

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

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

<http://www.ipn.mx>

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

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

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

Indexed in LATINDEX, DBLP and Periodica

Electronic edition

Editorial

This volume of the *Research in Computing Science* journal (RCS) contains original selected contributions related to the field of Pattern Recognition theory, principles and applications.

Members of a Technical committee evaluated the contributions published in this volume. In particular, the topics covered by the seven contributions published in this issue are: Computer Vision, Bio-signal Analysis, Knowledge graphs, Deep Learning, Navigation Systems, Bioinformatics, and Classification. The first contribution proposes a method to integrate environmental and physiological data through Knowledge graphs in the context of thermal comfort; the second and fourth papers present a general overview of existing methods about speech separation based on Deep Learning and navigation systems for bipedal robots, respectively. The third contribution proposes an algorithm to solve the problem of hepatitis C classification through heuristic search; approaches based on Computer Vision for detecting bumps and obstacles in self-driving environments are proposed in the fifth and sixth papers. Finally, the last contribution reports results related to a detailed experimentation for determining fundamental parameters in the field of EEG signal analysis.

We cordially thank to the RCS editorial board for allow us the opportunity of publishing this volume; in addition, we thank to our reviewing committee for their valuable participation and to authors for their submitted contributions to build this volume.

Finally, we hope the contributions in this volume will be useful to the reader interested in the field and topics previously described.

J. Arturo Olvera-López
J. Francisco Martínez-Trinidad
J. Ariel Carrasco-Ochoa
Joaquín Salas-Rodríguez

Guest Editors

August 2019

Table of Contents

	Page
Knowledge Graphs for Analyzing Thermal Comfort: A Proposal <i>Jorge-Luis Jácome-Domínguez, Edgard Benítez-Guerrero, Guillermo Molero-Castillo</i>	9
Analysis of Speech Separation Methods based on Deep Learning <i>Jessica Rincón-Trujillo, Diana Margarita Córdova-Esparza</i>	21
An Algorithm to Classify DNA Sequences of Hepatitis C Virus Based on Localized Conserved Regions and Heuristic Search <i>Sarahi Zúñiga, Iván Olmos, Javier Garcés, Mario Rossainz</i>	31
Towards the Development of a Navigation System for a Bipedal Robot: Preliminary Analysis of the Literature..... <i>Miguel Angel Ortega-Palacios, Josefina Guerrero-García, Juan Manuel González-Calleros</i>	43
Obstacle Detection and Trajectory Estimation in Vehicular Displacements based on Computational Vision <i>Lauro Reyes-Cocoletzi, Ivan Olmos-Pineda, J. Arturo Olvera-López</i>	57
Speed Bump Detection on Roads using Artificial Vision..... <i>Ana L. Ballinas-Hernández, Ivan Olmos-Pineda, J. Arturo Olvera-López</i>	71
Parameter Experimentation for Epileptic Seizure Detection in EEG Signals using Short-Time Fourier Transform..... <i>Ricardo Ramos-Aguilar, J. Arturo Olvera-López, Ivan Olmos-Pineda, Susana Sánchez-Urrieta, Manuel Martín-Ortiz</i>	83

Knowledge Graphs for Analyzing Thermal Comfort: A Proposal

Jorge-Luis Jácome-Domínguez¹, Edgard Benítez-Guerrero¹,
Guillermo Molero-Castillo²

¹Universidad Veracruzana, Xalapa, Veracruz, Mexico
jorge.jacomdiguez@gmail.com, edbenitez@uv.mx

²Universidad Nacional Autónoma de México, Mexico
gmoleroca@fi-b.unam.mx

Abstract. Knowledge Graphs (KGs) can be used to provide a unified, homogeneous view of heterogeneous data, which then can be queried and analyzed. In this paper, we explore the use of KGs to analyze the Thermal Comfort (TC) of users in specific environments (e.g. classrooms, hotel rooms). This implies the integration of several data sources that provide environmental variables (i.e. temperature, humidity), but also users' physiological variables (i.e. temperature). Thus, this paper proposes a method to integrate and analyse environmental and physiological data through KGs in the context of TC.

Keywords: Knowledge Graphs, Thermal Comfort, Data Integration.

1 Introduction

A Knowledge Graph (KG) is a structure designed to represent and organize knowledge [1–3], such as the structure of an information system [1] or specific topics for intelligent services, among others [4, 5]. In recent years, KGs have attracted the interest of diverse research groups and companies (Google, Microsoft, University of Leipzig, University of Mannheim, to name a few) [2, 6], which have implemented KG-based solutions in diverse fields, such as the Internet of Things, big data, semantic web, information management, among others [4, 5, 7, 8].

KGs can be used to integrate several, potentially heterogeneous data sources [5], with the objective of organizing the content of the sources and present them as a single source of information [3, 7]. This is particularly needed in Context-Aware systems, where the system uses context information to provide data and/or services to its users. That information usually needs to be acquired from heterogeneous data sources and integrated to enable its use in decision-making processes. An application domain of this kind of systems, that is of special interest to our research project, is the analysis of Thermal Comfort (TC). It refers to the level of satisfaction of a person with respect to the thermal conditions of the space (building, room) where he/she is [9, 10]. It has been recognized that several physical (temperature, humidity, air quality) and physiological factors (heart rate, stress) influence the thermal comfort of an individual. The physical factors

refer to environmental features, while the physiological are people characteristics [9, 11, 12]. The factors described above can be found in the ANSI/ASHRAE 55-2017 standard [9, 13] and some recommendations of the World Health Organization (WHO) [10], and have been traditionally used for the design of spaces in static contexts. However, environmental and individual conditions change over time, and thus it is necessary to consider this dynamism in a computer system in order to provide a more exact and personalized solution to guarantee the comfort of users. Our work is aimed to propose such a system.

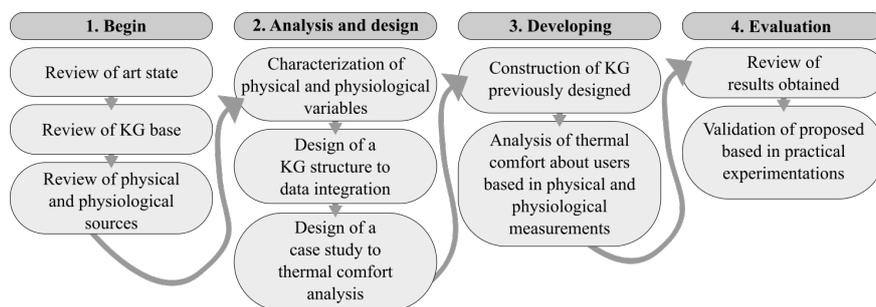


Fig. 1. Research Method.

Let us note that to achieve the aims of our work, we are following a research method consisting of four stages as shown in Figure 1. The first stage is aimed to analyze the literature related with the topics of Knowledge Graph, integration of heterogeneous data sources, and Thermal Comfort. The second stage corresponds to the characterization of physical and physiological variables needed to determine Thermal Comfort and, based on that characterization, design a KG-based solution to connect sources exposing that data (e.g. sensors). The third stage refers to the development and testing of the KG-based solution. For the development it is foreseen the use of tools for creating and querying KG. Finally, the fourth stage encompasses the evaluation of the results obtained. This will be done from the point of view of the capacity of the solution to connect heterogeneous data sources and to determine the Thermal Comfort of the users.

This paper particularly proposes a framework to integrate and analyse environmental and physiological data through KGs in the context of the study of TC. This implies a) characterization of physical and physiological variables, b) Knowledge Graph design, c) construction, and d) evaluation of solution proposal.

2 Background

2.1 Knowledge Graphs

A Knowledge Graph (KG) is a structure designed to store, organize and represent knowledge [1–3]. A KG is formed by an indefinite number of entities that are

linked to other entities through relationships; the entities and relations of a KG represent facts [14]. A KG is usually represented as a directed acyclic graph, where each node contains an entity and an arc between two nodes represent a relationship [2, 14]. Associated to a KG it is possible to find tools to query, interpret and present its content [1, 14].

Nowadays, there are no formal classification of KGs. However, it has been possible to observe three different types [1]:

1. **Prior knowledge.** A KG of this kind is constructed to work with a set of existing data and its objective is to solve some problems and inconsistencies in a well-defined context [1, 2].
2. **Existing model.** KGs of this kind model information through a highly expressive and complex semantic scheme [2, 3], that is usually an ontology [1]
3. **Defined structure.** A KG of this kind is based on a structure of data or objects, as the Resource Description Frameworks (RDF) [15], and its objective is the study of disordered data and the discovering of implicit hierarchies between its entities [1, 5].

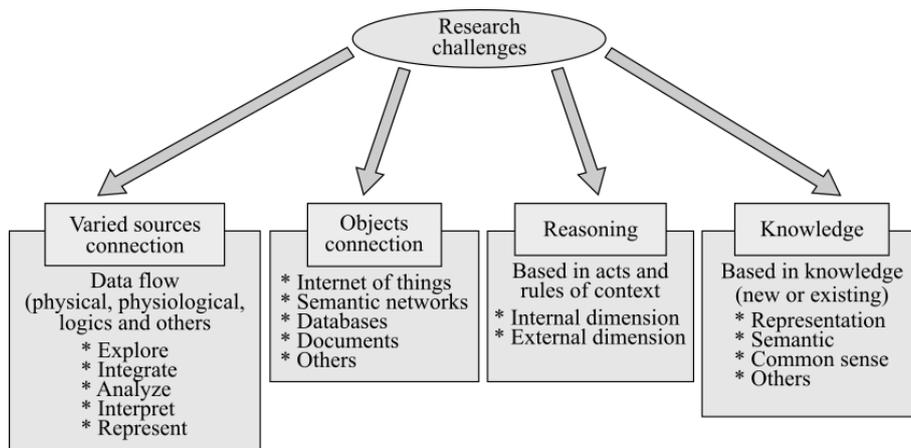


Fig. 2. KGs research challenges.

The popularity of KGs has grown in recent years, as shown by the increasing number of research groups and companies that work on the topic [2, 6]. KGs are usually used in works that focus on the representation and use of knowledge in the Semantic Web and Natural Language Processing [5, 6, 16]. However, there are still a number of research challenges related to KGs, as shown in Figure 2.

The role of KGs regarding the representation, querying and analysis of heterogeneous data, especially in dynamic contexts [5, 16], needs to be more explored.

Here, a number of problems that arises while acquiring data from highly sensitive environments need to be addressed [5, 8, 11]: errors during the measurement of context parameters, data loss, speed, size and meaning [5, 11, 13].

Other challenges associated with KGs are related to the connection of objects (e.g. IoT-devices), and the representation of knowledge (e.g. common sense [8]) and associated reasoning mechanisms. This project, as previously mentioned, focuses on the integration of heterogeneous data sources, in specific, physical and physiological data sources.

2.2 Heterogeneous Data Sources

Traditionally, data that is needed to make an informed decision does not come from a single source, but from several potentially heterogeneous data sources. This heterogeneity relates to differences among the sources with respect to formats, amount of data issued, acquisition speed, meaning, among others [5, 9, 11, 13]. In the case of the analysis of thermal comfort, it is possible to identify the need for physical and physiological data that comes naturally from heterogeneous sources [11, 17, 18].

Physical Data and its Sources. These are the result of the measurement of environmental aspects, for example temperature, humidity, air quality, airspeed, intensity of light, noise level, among others [9–11, 13]. These can be obtained through sensors such as: a) DHT11, for measuring temperature and humidity, b) MQ-135, for measuring air quality, c) KY-038, for measuring noise, and d) LDR, for measuring slight intensity.

Physiological Data and its Sources. These are the result of measuring aspects such as heart rate, or brain activity, among other parameters [9–11, 13]. These data is usually obtained using devices with integrating sensors, such as smartwatches, smartphones, headbands, among others.

2.3 Thermal Comfort

Thermal Comfort is the level of acceptance of thermal conditions (e.g. temperature or air quality [9]) of a specific space by its occupants. TC is studied because it has been shown that it favors not only the comfort, but also the wellness, the health, the productivity and the performance of the occupants of a space [13, 19]. So, the research in this domain is aimed to determine parameters and variables to help upgrading the conditions of habitability and quality of specific spaces [19].

In this context, it is important to know about the aspects that might change the conditions of a space [20]: the space itself, external environment, and its occupants. Concerning the space itself and its external environment, variables such as radial temperature and relative humidity are of importance, while regarding the occupants, variables of interest are for instance the body mass index and the

level of isolation of thermal clothing [11, 17, 20]. Based on this discussion, the approaches for the study of Thermal Comfort are two [10, 13, 17, 18]:

- **Traditional:** It is based on the measurement of different internal and external environment aspects, and statistical models are used to analyze the collected data.
- **Varied:** It involves the analysis of unusual variables, for example, characteristics of the users. Also, the revision of adaptive models is considered, such as the adaptive model of ASHRAE 55 standard.

Thermal Comfort involves then the analysis of a highly complex context, that can be easily influenced by a large number of aspects [10, 17, 19]. However, in spite of being a highly studied topic, it is commonly tackled with a traditional perspective. The study of TC considering other physiological or even psychological variables is then of relevance [17, 18].

3 Proposal

The disconnection of different heterogeneous data sources is a problem that persists today [6, 12]. So, the tasks of acquiring, integrating, and analyzing diverse data flows, are challenging [3, 21]. To solve this problem, KGs are an alternative with useful advantages, since its construction implies the use of structures with the ability to represent, explore and analyze different types of data [5, 6].

This project is aimed to develop a KG-based solution to the problem of integrating physical and physiological data sources in a dynamic context. Thus, the Thermal Comfort provide an ideal case of study in a highly dynamic environment and with distinct types of data. Moreover, in this proposal we consider to set our study in the context of a classroom. Thus, we are considering the physical and physiological variables shown in Table 1, which are considered by the ASHRAE 55 standard and the literature [17, 22–25].

Table 1. Physical and physiological variables.

Variables	Type	Units
Air temperature	Physical	°Celsius
Relative humidity	Physical	Water vapor % in 1 m ³
Airspeed	Physical	Distance/Time
Amount of energy expended by user	Physiological	MET
User heart rate	Physiological	BPM

In order to tackle the problem, we propose to follow a process divided in four stages, as seen in Figure 3: a) physical and physiological data acquisition, b) integration of physical and physiological data, c) representation and analysis

of context based on the obtained physical and physiological data, and finally d) determine the Thermal Comfort of users. Each stage is further explained next.

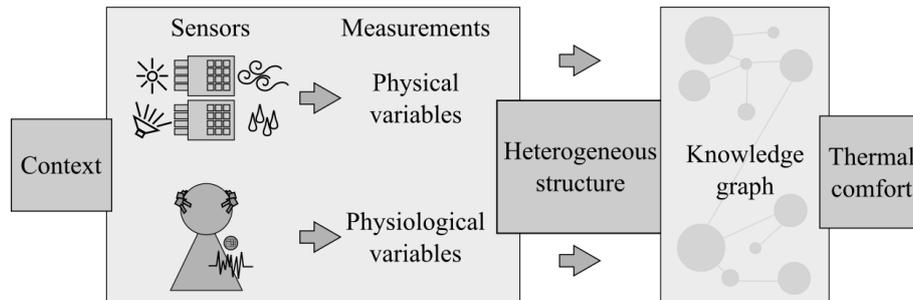


Fig. 3. Process for the acquisition, integration, and analysis of data to determine thermal comfort.

Stage one is aimed to the acquisition of physical and physiological data. In this stage, we will implement three tools, one for each of the steps shown in Figure 4: a) construction of a sensor network to measuring the variables of Table 1, b) creation of a database in the cloud, and connecting the sensor network to it, c) implementation of a wrapper to export the database as a RDF repository. For this last step, it is foreseen to use the Semantic Sensor Network (SSN) ontology, that was developed by W3C based on SSO (Stimulus-Sensor-Observation) patterns, which was especially create to represent sensor information [5, 26].

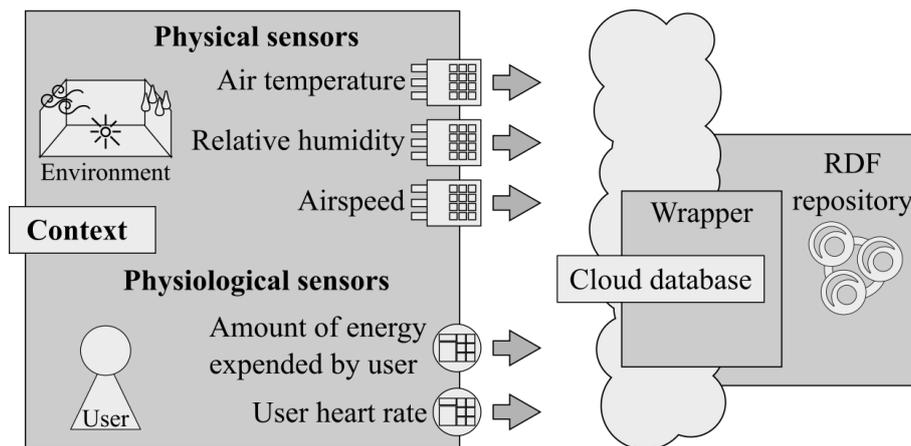


Fig. 4. Physical and physiological data acquisition.

Stage two is aimed to integrate and analyze the acquired physical and physiological data. In this stage, we will use of the RDF repository which it made in stage one, as seen in Figure 5, and involves ones steps: a) query the RDF repository by SPARQL [27], to get RDF Triples which are an integrated representation of the physical and physiological data, b) analyze the RDF Triples acquired through the ASHRAE 55 standard methods and statistical methods according to the literature, c) use the results obtained to expand the knowledge and get ontologies.

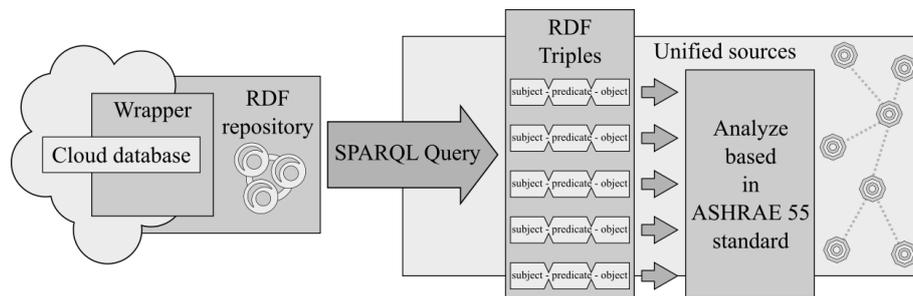


Fig. 5. Integration and analysis of physical and physiological data.

Stage three is aimed to represent and analyze the context through the RDF objects. In this stage, we will use of the RDF Triples analyzed in the stage two, to get a KG, as seen in Figure 6. To this, we will use the RDF Triples to get an ontology for each variable referred in Table 1, which we will combine with the user marks "provide a good comfort" and "provide a bad comfort" in order to distinguish the state of user comfort. Thus, each ontology will have connect with between both to make up a KG, that will have the ability to be querying by SPARQL. Also, KG obtained will have an inference engine to expand its knowledge and have more precision to know and customize the users comfort.

Finally, in stage four, based on the case study, we propose to analyze and understand the user Thermal Comfort through the application of ASHRAE 55 standard, and compare the results obtained against the knowledge that could be acquired by the KG.

4 Related Work

KGs enable the possibility of having a unified view for a set of heterogeneous data sources, which can be further queried and analyzed it. This has been investigated in several works. Particularly, in [5], the authors observed a big potential in the connection of an enormous amount of data that are generated by the billions of devices of the Internet of Things (IoT), and in [16], the authors similarly identified the need of unify data coming from energy systems.

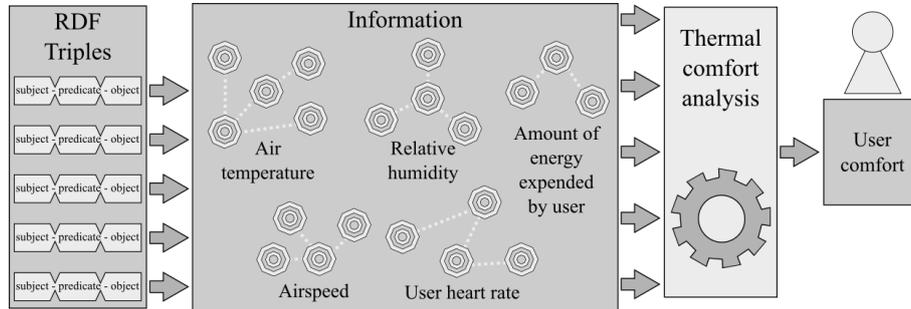


Fig. 6. Analysis and representation of context with base in the obtained physical and physiological data.

In the case of studies on Thermal Comfort, in general they follow a traditional approach, do not exploit the use of tools for handling heterogeneous data, and they are limited to following standards such as ASHRAE 55 [17]. In addition, these studies do not usually check the influence of physiological variables on the comfort of a given environment [17, 22]. For example, in [28], the authors explained that in the south of Europe exist uninhabited and obsolete homes, which need to be rehabilitated in favor of human occupation. Given this, the authors reviewed the Thermal Comfort conditions in those homes through on a model developed for the prediction of Thermal Comfort based on Artificial Intelligence. So, they provided a mechanism capable of supporting the rehabilitation tasks in these places. However, the proposed model focused on the revision of environmental variables, and was unable to analyze the conditions of the environment in real-time.

In [29], the authors mentioned that in Cuba, the approach to determine the punctual thermal load of a building used by the companies that install HVAC systems was wrong, causing high maintenance costs and high energy consumption by this type of machines. Thus, the authors studied Thermal Comfort conditions in spaces with the double-folded objective of exposing the bad implementation of HVAC systems and of determining how to improve the Thermal Comfort level in those spaces. However, parameters related to users were not considered, only taking into the occupational affluence in those spaces.

In [30], the authors presented RoomFort, an intelligent system for the management, adjustment and improvement of the comfort conditions in hotel rooms. They used ontologies and a sensor network in hotel rooms to acquire, process and reason about different aspects related to the comfort status and how to improve it. However, although they considered parameters related to the hotel guest, this was not significant on the Thermal Comfort of the hotel room, since the study was mainly aimed to study Visual Comfort.

In [31], the authors proposed E-Cabin, a framework based on IoT which take advantage of the potential of ontologies to represent knowledge and maintain a

consistent relationship between ontology knowledge and user profiles. The authors used this framework to design and construct an intelligent cabin on a cruise, equipping the cabin with the ability to provide its inhabitants with personalized interior comfort and, at the same time, better management of the energy consumed. However, in the case of Thermal Comfort, the variables related to the user were not significant in the environment thermal settings.

In [32], the author explored the capabilities of the Wireless Sensor Network (WSN) for the study of the Human Comfort problem and the creation of Habitable Environments. Thus, he revised the potential of a WSN to analyze the environment and determine the Visual, Acoustic, Indoor Air Quality and Thermal Comfort state. However, in the case of Thermal Comfort, his analysis was based on physical parameters related to the environment.

5 Conclusions

The popularity that KGs have gained in the last years, has allowed to improve the form of construction of this kind of tools, highlighting the ability to manage large volumes of heterogeneous data.

Thus, KGs have served as a solution to satisfy problems related to the use of data sources of different nature, providing a base to develop systems for presenting an integral view of those sources.

Nonetheless, KGs have still some challenges related to the management, connection and integration of data. Particularly, the challenges related with acquiring and analyzing data flows coming from sources with high levels of variability and interoperability.

In this sense, Thermal Comfort analysis through KGs can provide a dynamic panorama with heterogeneous variables of multiples sources, helping and expanding the study of KG abilities to integrate, connect, storage and represent heterogeneous data sources. Also, contribute to Thermal Comfort studies, and the use of ontologies on topics related to intelligent environments and the development of spaces for human occupation.

Thus, as part of our research, this paper presents a proposal to manage and analyze heterogeneous data based on their integration through KG by knowing the user Thermal Comfort which are the study case. This proposal consists of a process of four stages: a) physical and physiological data acquisition, b) integration and analysis of physical and physiological data in a structure, c) representation and analysis of context with base on the obtained physical and physiological data, and finally d) determine the Thermal Comfort of users. Then, the knowledge obtained by this proposal could be used to develop intelligent, healthy and productive environments, as well as a help to manage heterogeneous sources.

As future work of this project, it is planned to design a network of sensors to obtain physical and physiologic sources, obtained from a specific working environment and users, respectively, to integrate them using a KG and, then analyze the users' Thermal Comfort in closed environments.

Acknowledgments. This work was partly sponsored by the Universidad Veracruzana (UV) and the National Council of Science and Technology (CONACYT) of México as part of the project "Infraestructura para agilizar el desarrollo de sistemas centrados en el usuario" (ref. 3053). Also, the first author was supported by CONACYT through the MSc. scholarship, Ref. 930925.

References

1. Ehrlinger, L., Wöß, W.: Towards a definition of knowledge graphs. In: Posters and Demos Track of 12th International Conference on Semantic Systems. Volume 1695., Leipzig, Germany (2016)
2. Paulheim, H.: Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web* 8 (2016) 489–508
3. Fathalla, S., Vahdati, S., Auer, S., Lange, C.: Towards a knowledge graph representing research findings by semantifying survey articles. *Lecture Notes in Computer Science* 10450 (2017)
4. Yu, T., Li, J., Yu, Q., Tian, Y., Shun, X., Xu, L., Zhu, L., Gao, H.: Knowledge graph for tcm health preservation: Design, construction, and applications. *Artificial Intelligence in Medicine* 77 (2017) 48–52
5. Le-Phuoc, D., Mau-Quoc, H., Ngo-Quoc, H., Tran-Nhat, T., Hauswirth, M.: The graph of things: A step towards the live knowledge graph of connected things. *Journal of Web Semantics* 37-38 (2016) 25–35
6. Wang, C., Gao, M., He, X., Zhang, R.: Challenges in chinese knowledge graph construction. In: 2015 31st IEEE International Conference on Data Engineering Workshops. (2015) 59–61
7. Ernst, P., Siu, A., Weikum, G.: Knowlife: a versatile approach for constructing a large knowledge graph for biomedical sciences. *BMC Bioinformatics* 16 (2015) 157
8. Freitas, A., Da-Silva, J., Curry, E., Buitelaar, P.: Approximate and selective reasoning on knowledge graphs: A distributional semantics approach. *Data and Knowledge Engineering* 100 (2015) 211–225 Special Issue of 19th International Conference on Applications of Natural Language Processing to Information Systems.
9. De-Vecchi, R., Sorgato, M., Pacheco, M., Cândido, C., Lamberts, R.: Ashrae 55 adaptive model application in hot and humid climates: the brazilian case. *Architectural Science Review* 58 (2015) 93–101
10. Tartarini, F., Cooper, P., Fleming, R.: Thermal perceptions, preferences and adaptive behaviours of occupants of nursing homes. *Building and Environment* 132 (2018) 57–69
11. Sharif, M., Alesheikh, A.: Context-aware movement analytics: Implications, taxonomy, and design framework. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8 (2018) 1–19
12. Aggarwal, N., Shekarpour, S., Bhatia, S., Sheth, A.: Knowledge graphs: In theory and practice. In: *Conference on Information and Knowledge Management' 17*, Singapore (2017)
13. Kim, J., De-Dear, R.: Thermal comfort expectations and adaptive behavioural characteristics of primary and secondary school students. *Building and Environment* 127 (2018) 13–22
14. Ebisu, T., Ichise, R.: Toruse: Knowledge graph embedding on a lie group. In: *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*. (2018)
15. W3C: Resource description framework (rdf). <https://www.w3.org/RDF/> (2014)

16. Chun, S., Jin, X., Seo, S., Lee, K., Shin, Y., Lee, I.: Knowledge graph modeling for semantic integration of energy services. In: 2018 IEEE International Conference on Big Data and Smart Computing (BigComp). (2018) 732–735
17. Forgiarini-Rupp, R., Kim, J., De-Dear, R., Ghisi, E.: Associations of occupant demographics, thermal history and obesity variables with their thermal comfort in air-conditioned and mixed-mode ventilation office buildings. *Building and Environment* 135 (2018) 1–9
18. Lu, C.Y., Tsai, M.C., Muo, C.H., Kuo, Y.H., Sung, F.C., Wu, C.: Personal, psychosocial and environmental factors related to sick building syndrome in official employees of taiwan. In: International journal of environmental research and public health. Volume 15. (2017) 7
19. Ghaffarianhoseini, A., AlWaer, H., Omrany, H., Ghaffarianhoseini, A., Alalouch, C., Clements-Croome, D., Tookey, J.: Sick building syndrome: are we doing enough? *Architectural Science Review* 61 (2018) 99–121
20. Corona-Patricio, C., Dueñas-Sosa, C., Castro-Enríquez, J., Avendaño-Juárez, J., Olivieri, F.: Diseño e implementación de un sistema inalámbrico para el monitoreo de las condiciones ambientales externas y de confort térmico en casetas prototipo. *Pistas Educativas* 36 (2015) 207–227
21. Molero-Castillo, G., Benítez-Guerrero, E., Mezura-Godoy, C.: Interacción humano computadora y minería de datos para la generación y representación de conocimiento útil. *Ciencias de la Información* 48 (2017) 3–5
22. Kim, J., Schiavon, S., Brager, G.: Personal comfort models – a new paradigm in thermal comfort for occupant-centric environmental control. *Building and Environment* 132 (2018) 114–124
23. ASHRAE: Thermal environmental conditions for human occupancy (2017)
24. Raish, J.: (Thermal comfort: Designing for people)
25. Pancardo-García, P.: Personalización basada en sensores para la estimación del estrés por calor en ambientes laborales. PhD doctor en ciencias de la computación, Univesidad Autónoma de Aguascalientes, Universidad Juárez Autónoma de Tabasco, Universidad Veracruzana (2016)
26. Gim, J., Lee, S., Joo, W.: A study of prescriptive analysis framework for human care services based on ckan cloud. *Journal of Sensors* (2018) 10 Art. ID 6167385.
27. Doan, A., Halevy, A., Ives, Z.: Principles of Data Integration. 1st edn. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2012)
28. Escandón, R., Ascione, F., Bianco, N., Mauro, G.M., Suárez, R., Sendra, J.J.: Thermal comfort prediction in a building category: Artificial neural network generation from calibrated models for a social housing stock in southern europe. *Applied Thermal Engineering* 150 (2019) 492–505
29. Madrigal, J., Cabello, J., Sagastume, A., Balbis, M.: Evaluación de la climatización en locales comerciales, integrando técnicas de termografía, simulación y modelado por elementos finitos. *Información tecnológica* 29 (2018) 179–188
30. Spoladore, D., Arlati, S., Carciotti, S., Nolich, M., Sacco, M.: Roomfort: An ontology-based comfort management application for hotels. *Electronics* 7 (2018)
31. Nolich, M., Spoladore, D., Carciotti, S., Buqi, R., Sacco, M.: Cabin as a home: A novel comfort optimization framework for iot equipped smart environments and applications on cruise ships. *Sensors* 19 (2019)
32. Mohamed-Rawi, M.: Sensor Network Embedded Intelligence: Human Comfort Ambient Intelligence. PhD doctor of philosophy, School of Engineering Auckland, New Zealand (2013)

Analysis of Speech Separation Methods based on Deep Learning

Jessica Rincón-Trujillo, Diana Margarita Córdova-Esparza

Universidad Autónoma de Querétaro, Facultad de Informática, Querétaro, Mexico
{jesssy_1993,diana_mce}@hotmail.com

Abstract. In this paper, we perform an analysis of speech separation methods using deep learning. We provide a review of the literature, discussing the main features of the works based on audio and audiovisual processing, as well as the deep learning method. For the analysis of the articles, we provide a description where we identify the category, characteristics, methodology, results, and application. According to the study, we observed that the acoustic-based algorithms require audio portions of voices with external interferences to improve the intelligibility and the quality of the voice signals. Audio-visual-based methods use thousands of hours of segments of noisy videos to obtain stable performances and enhancing the quality of the separation. This latter category, also known as the cocktail party problem represents an ongoing open problem in the deep learning community.

Keywords: Deep Learning, Speech Separation, Computer Vision.

1 Introduction

According to Wang & Chen [23], speech separation is the task of dividing the target voice from background interference. Speech separation from multiple sources of sound is defined by Cherry [3] as cocktail party problem. It can be said that the separation of speech isolates the source of the sound. For this, the separation corresponds to the segregation of the auditory current. Speech separation can be used for assistive technologies such as hearing aids and cochlear implants, automatic captioning, dictation systems, and solving multi-speaker simultaneous speech [5, 4].

The first systematic study on flows segregation carried out by Miller [18] establishes that listeners divide a signal with two alternating sinusoidal wave tones in two flows. He also reviewed human intelligibility scores when he interfered with a variety of tones, broadband noise, and other voices.

In the 1950s, Rosenblatt created some useful brain analogs for analytic tasks [20]. His research began the learning of machines, developing techniques that allow computers to learn and classify as human beings do. Having as main objective *the ability to recognize and process complex patterns of information related from all dimensions, such as the human brain*. Rosenblatt was the one who invented the Perceptron, an artificial neuron or basic unit of inference, on which the

multi-layered learning networks are based, which are the basis of what is known as Deep Learning [24]. In the context of artificial intelligence (AI), deep learning according to Gómez, et al. [12] refers to the automatic activity of knowledge acquisition, through the use of machines that use several levels for extraction.

Nowadays, due to massive amounts of data, deep Learning-based methods have pushed the state-of-the-art beyond what was possible with traditional techniques either using audio-only sources or audiovisual involving audio signal processing and computer vision [19].

1.1 Audio-based Methods

Below we discuss articles that address this topic in the acoustics category and the results accomplished.

The work developed by Simpson, et al. [21] proposes a statistical analysis based on the hypothesis of the separation of sources, making use of twelve models already tested on voice separation and with fifty pieces of music which were produced professionally. They used non-parametric statistics, establishing reliable evidence for significant conclusions about the performance of the various models. They conclude with the design of a basic procedure based on the hypothesis for the non-rigorous statistical analysis of the results of the source separation model. They obtained reliable evidence in the study. However, they found no evidence of any significant difference between their two best models based on Deep Neural Networks (DNN).

The proposal implemented by Gao, et al. [11] describes a unified voice improvement framework for jointly managing both background noise and interfering speech in a speaker-dependent scenario based on deep neural networks (DNN). They explore the speech improvement that depends on the speaker optimizing the performance of the system in comparison with the independent speakers; they considered the interference as a type of noise and demonstrated that their system could achieve performance comparable to specific systems where only noise or voice interference is present. They then use a framework based on joint learning to further improve system performance in low signal-to-noise (SNR) environments.

In the present year, Dadvar et al. [7] describe a robust binaural voice separation system based on a deep neural network (DNN). The system is based on three main stages of processing. The first stage deals with the spectral processing from multiresolution cochlea characteristic extracted from the signal beamforming. The second stage comprises spatial processing, based on a spatial feature of a softly-masked binaural cue (smITD + smILD) obtained by soft masking missing data from binaural signals. The third stage contains a deep neural network with combined spectral and spatial characteristics, designed for noisy and reverberant conditions. The proposed system is evaluated and compared with the two recent binaural voice separation systems as baselines in various noisy

and reverberant conditions. They show that their system exceeds current baseline systems because improving the intelligibility and quality of separate speech signals in reverberant and noisy conditions. The results confirm the efficiency of each component of the system, especially in highly reverberant scenarios.

1.2 Audiovisual-based Methods

Next, we present the articles in the audio-visual category and the results they have obtained.

Khan, et al. [15] proposes a study into whether visual speech information can be used to assist in the estimation of the mask for the separation of audio speakers to improve speech quality and intelligibility. They developed two visual methods for the separation of the speakers, which use a deep neural network that assigns the visual characteristics of the speech to a space of audio attributes from which the visually derived binary masks and the visually derived relationship masks are estimated before the application to the mix of voices. The other method entails masking in relation to the audio, forming a line focus for speaker separation. The collection of the audio mix from the speakers is made from a single microphone, and the visual characteristics of the speech are extracted from the mouth of each speaker in the combination. The audio speech characteristics are subsequently estimated for each speaker from the corresponding visual components and then used within the proposed mask estimation methods. Obtaining good results in situations in which the two speakers are equidistant from the microphone and have a similar volume. However, the accuracy of the mask reduces when the audio power of the two speakers differs.

In the same year Afouras, et al. [1] propose an audiovisual neural network that can isolate the voice of a speaker from other people, obtained from a noisy audio signal and the corresponding speaker video, producing an improved audio signal containing only the speaker's voice with the rest of the speakers and the background noise suppressed. The proposed model is evaluated for five simultaneous voices, demonstrating both qualitative and quantitative reliable performance. The performance of the model was put to the test in open environments. It consists of two modules: a magnitude sub-network and a phase sub-network. The first sub-network receives the magnitude spectrograms of the noisy signal and the video of the speaker as inputs and generates a smooth mask. Then, the input magnitudes are multiplied elementally with the mask to produce a spectrogram of filtered scale. The prediction of the scale, together with the phase spectrogram obtained from the noisy signal, is introduced into the second sub-net, which produces a residual phase. The residue is added to the noisy stage, generating the improved phase spectrograms. Finally, the enhanced magnitude and phase spectra are transformed back into the time domain, producing the improved signal.

Another work, developed by Ephrat, et al. [9] establishes an audiovisual model to isolate a single voice signal from a mix of sounds such as speakers and background noise. The input is a video with one or more people speaking, the speech of interest interfered with by other speakers and background noise.

Both auditory and visual characteristics are extracted and incorporated into the joint model for audiovisual separation. The output is a decomposition of the audio track input into clean voice tracks, one for each person detected in the video. Videos are obtained in which the speech of specific people is improved while the rest of the sound is suppressed. The model is based on a neural network that incorporates visual and auditory signals. They introduced an audiovisual dataset called AVSpeech composed of thousands of hours of video segments from the web. Demonstrating the applicability of their method to classical tasks of separation of voices, as well as to real situations. Their demo only requires the user to specify the face of the person in the video whose speech they wish to isolate. They make two main contributions: A model of audiovisual separation by voice that surpasses the audiovisual and unique models in the classic tasks of voice separation. A new set of audiovisual data on a large scale, AVSpeech, composed of segments of videos in which the audible sound belongs to a single person, visible on the video, and without audio background interference.

The purpose of this work is to provide an overview of the state-of-the-art techniques for speech separation using either audio-only or audiovisual input sources.

2 Methodology

In this section, we present the analysis of multiple deep learning-based speech separation studies using acoustic and audiovisual sources.

For our analysis, we considered the following: category, features, method, results, application, advantages, and disadvantages.

2.1 Acoustic Category

Simpson et al. [21] analyzed separation results on twelve different models, as data they used 50 pieces of recorded music. They applied statistical analysis of the results for audio-based speech separation discovering that the highest performance models were the ones that used deep learning.

Chen et al. [2] used the ideal Radio Mask (IRM) as the target of supervised learning. The IRM is estimated from the 64-channel cochleagrams of a combination of plain speech and noise. According to [22], the cochleagram is a time-frequency representation of an acoustic signal. They showed that random frequency noise perturbations on the spectrogram gives the best speech separation results in classification accuracy. They found that when training a DNN, the quality of the relationship mask improves as the classifier has been exposed to more noise interference scenarios. A disadvantage is that when a training set is created from limited speech and noise resources, a classifier will likely adapt to the training set and make wrong predictions in a test set, especially when the background noise it is highly non-stationary, it is suggested to expand the noise resources.

Gao et al. [11] used clear speech and speech interference followed by log-power spectral (LPS) features [8] of both speeches to train an ensemble of deep neural networks with dropout. Their system were able to decrease background noise and speech interference in speaker-dependent situations. They found that a speaker-dependent system is much more robust than one independent of the speaker, can unify speech improvement and speech separation and that it is possible to achieve better performance for mixed conditions of noise and voice interference. However, improvements can still be made for environments with high SRN and speech improvement architecture.

Liu and Wang [16] inspired by auditory scene analysis [22], they decompose the task into two stages, a frequency-domain concurrent grouping step, and a time-domain grouping step. The two stages are trained combined and using recurrent neural networks achieving state-of-the-art performance using permutation invariant training (PIT) and deep clustering (DC). The experiments show that the proposed system improves on the best-reported results of PIT and DC. The training of the first two stages can be done together and sequentially (1st phase of simultaneous grouping to separate two speakers in the frame level, 2nd sequential grouping to transmit spectra).The proposed approach takes advantage of both variable permutation training and deep grouping and has been shown to produce better results than both methods. The disadvantage is that the algorithm does not perform in real-time.

Dadvar and Geravanchizadeh, [7] used Multiresolution cochleagram (MRCG) features extracted from the beamformed signal and spatial features of softly-masked binaural cues (smITD + smILD) to train a supervised deep neural network with mixed spectral and spatial features to determine an Ideal Radio Mask (IRM) based on the signal-to-reverberant noise ratio.The results revealed that MRCG is the best spectral feature in the binaural speech separation task in noisy and reverberant environments. The application of their system is the binaural speech separation task in noisy and reverberant environments. They show that the proposed system exceeds baseline systems in a variety of simulated conditions, especially in low SNR and high RT scenarios. They compared their system proposed with two binaural separation systems and RT demonstrate the superiority of the proposed system in terms of the gains of STOI, ESTOI and PESQ. As a disadvantage, the system needs to be extended to be applicable in real environments.

2.2 Audiovisual Category

Khan et al. [15] used visual features taken from Active Appearance Models (AAM) from each speaker in the mixture. Audio speech features are afterward estimated for each speaker from the equal visual features and used within the suggested mask estimation methods. They trained a deep neural network to map a stack input spectral features to a ratio mask. The mask is later applied to the noisy mix to determine the target speaker spectral features. They obtained the

highest performance when merging audio and visual information to create the masks. Applications of their system entail situations where the two speakers are at the same distance from the microphone and have comparable loudnesses. They found that the audiovisual masking that combines visual and audio masks offers higher performance in all the tests carried out as well as in all the SNR. As disadvantage, the resulting voice of the resulting masks was of lower quality and sometimes intelligible compared to the result of the relationship masks. - When the audio power of the two speakers differs the accuracy of the mask is reduced.

Afouras et al. [1] used visual features extracted from images with a spatio-temporal residual network. Acoustic features are derived from the audio waveforms using Short Time Fourier Transform (STFT) with a Hann window function to generate spectrograms. They trained a Convolutional Neural Network (CNN) capable of producing clear speech from noisy audio segments recorded in real environments. They used two databases: LRS2 and VoxCeleb2. As a disadvantages, the LRS2 database contains information only in English and the method can fail in conditions where there is a lot of noise.

Ephrat et al. [9] used the AVSPEECH Dataset containing around 4700 hours of video fragments with nearly 150,000 distinct speakers to train a multi-stream architecture that takes visual data from faces and noisy audio to generate spectrogram masks for each face detected in the video. The spectrograms are then used to obtain isolated speech signals for each speaker suppressing other interfering signals. They obtained state-of-the-art results on speech separation as well as a potential application to video captioning and speech recognition. The proposed model was tested in different videos which contained various types of noise (in a bar, restaurant, debates, etc.). The model works well in situations with a lot of background noise or several people talking. They carried out several tests, that allowed them to observe the results of the proposed model in different scenarios. They used 3 Mandarin databases: TCD-TIMIT and CUAVE databases. The only current disadvantage is that the method does not work in real time without a powerful GPU due to faces movement.

Lu et al. [17] used the WSJ0 Dataset (audio-only) [14] that contains 30 hours of training data as well as the GRID Dataset [6] containing 34 speakers each with 1000 frontal video recordings. They trained an Audiovisual matching network obtaining improvements on speech separation quality over the state of the art of audio-only speech source separation. They found that When the audio is not separated correctly, the audiovisual focus remains stable and in some cases, the correct audio separation is not achieved.

Gabbay et al. [10] used the GRID Dataset [6] and the TCD-TIMIT dataset [13] consisting of 60 speakers with about 200 videos each to train a Convolutional neural network that takes the frames of silent video as input, and predicts sound features that are converted into intelligible speech. Compared to audio-only techniques, their method is not influenced by similar speech vocal components generally observed in same-gender speech separation. Their system has no problems caused by the gesticulation. As disadvantages, they did not perform testing with

multiple people talking and according to the tests carried out when working with TCD-TIMIT vid2speech generates unintelligible content.

3 Results

According to our analysis, we observe that in the case of the acoustic category the deep learning methods that use pieces of music or audios voices with external interferences (noise) improve the intelligibility and quality of voice signals.

The articles presented in this category use deep neural networks, based on non-parametric supervised recurrent neural networks characterized by having the required information to ensure adequate data sets to improve the intelligibility and the quality of signals for speech separation in reverberant situations. In the case of the Audio-Visual category, we observed that its main feature is that it requires high amounts of training data, comprising thousands of hours of noisy video segments.

Some of the advantages found in the combination of audio and visual sources are stable performances improving the quality of speech separation, therefore associating the separate voice tracks with the speakers visible in the video.

4 Conclusions

In this article, we carried out an analysis of deep learning-based speech separation methods. Compared to traditional techniques, deep learning techniques have had a breakthrough using acoustic-only sources, improving the clarity and quality of voice signals.

For audiovisual sources, there have been significant advances, obtaining stable performances, and enhancing the quality of the separation. However, there is still active work in progress to solve the so-called cocktail party problem on unconstrained environments.

Acknowledgements. The authors wish to acknowledge the financial support for this work by the Consejo Nacional de Ciencia y Tecnología (CONACYT) through financial support scholarship number (CVU): 925734. We also want to thank Universidad Autónoma de Querétaro (UAQ) through project number FIF-2018-06.

References

1. Afouras, T., Chung, J.S., Zisserman, A.: The conversation: Deep audio-visual speech enhancement. arXiv preprint arXiv:1804.04121 (2018)
2. Chen, J., Wang, Y., Wang, D.: Noise perturbation for supervised speech separation. *Speech communication* 78, 1–10 (2016)
3. Cherry, E.C.: Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America* 25(5), 975–979 (1953)

4. Chung, J.S., Senior, A., Vinyals, O., Zisserman, A.: Lip reading sentences in the wild. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3444–3453. IEEE (2017)
5. Chung, J.S., Zisserman, A.: Lip reading in the wild. In: Asian Conference on Computer Vision. pp. 87–103. Springer (2016)
6. Cooke, M., Barker, J., Cunningham, S., Shao, X.: An audio-visual corpus for speech perception and automatic speech recognition. *The Journal of the Acoustical Society of America* 120(5), 2421–2424 (2006)
7. Dadvar, P., Geravanchizadeh, M.: Robust binaural speech separation in adverse conditions based on deep neural network with modified spatial features and training target. *Speech Communication* 108, 41–52 (2019)
8. Du, J., Huo, Q.: A speech enhancement approach using piecewise linear approximation of an explicit model of environmental distortions. In: Ninth Annual Conference of the International Speech Communication Association (2008)
9. Ephrat, A., Mosseri, I., Lang, O., Dekel, T., Wilson, K., Hassidim, A., Freeman, W.T., Rubinstein, M.: Looking to listen at the cocktail party: A speaker-independent audio-visual model for speech separation. arXiv preprint arXiv:1804.03619 (2018)
10. Gabbay, A., Ephrat, A., Halperin, T., Peleg, S.: Seeing through noise: Visually driven speaker separation and enhancement. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 3051–3055. IEEE (2018)
11. Gao, T., Du, J., Dai, L.R., Lee, C.H.: A unified dnn approach to speaker-dependent simultaneous speech enhancement and speech separation in low snr environments. *Speech Communication* 95, 28–39 (2017)
12. Gómez Gil, M.: Aprendizaje profundo, el poder del aprendizaje automático unido al poder de cálculo de las computadoras actuales (2016)
13. Harte, N., Gillen, E.: Tcd-timit: An audio-visual corpus of continuous speech. *IEEE Transactions on Multimedia* 17(5), 603–615 (2015)
14. Hershey, J.R., Chen, Z., Le Roux, J., Watanabe, S.: Deep clustering: Discriminative embeddings for segmentation and separation. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 31–35. IEEE (2016)
15. Khan, F.U., Milner, B.P., Le Cornu, T.: Using visual speech information in masking methods for audio speaker separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26(10), 1742–1754 (2018)
16. Liu, Y., Wang, D.: A casa approach to deep learning based speaker-independent co-channel speech separation. In: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 5399–5403. IEEE (2018)
17. Lu, R., Duan, Z., Zhang, C.: Listen and look: Audio-visual matching assisted speech source separation. *IEEE Signal Processing Letters* 25(9), 1315–1319 (2018)
18. Miller, G.A., Heise, G.A.: The trill threshold. *The Journal of the Acoustical Society of America* 22(5), 637–638 (1950)
19. Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M.P., Shyu, M.L., Chen, S.C., Iyengar, S.: A survey on deep learning: Algorithms, techniques, and applications. *ACM Computing Surveys (CSUR)* 51(5), 92 (2018)
20. Rosenblatt, F.: The perceptron, a perceiving and recognizing automaton Project Para. Cornell Aeronautical Laboratory (1957)
21. Simpson, A.J., Roma, G., Grais, E.M., Mason, R.D., Hummersone, C., Liutkus, A., Plumbley, M.D.: Evaluation of audio source separation models using hypothesis-

- driven non-parametric statistical methods. In: 2016 24th European Signal Processing Conference (EUSIPCO). pp. 1763–1767. IEEE (2016)
22. Wang, D., Brown, G.J.: Computational auditory scene analysis: Principles, algorithms, and applications. Wiley-IEEE press (2006)
 23. Wang, D., Chen, J.: Supervised speech separation based on deep learning: An overview. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26(10), 1702–1726 (2018)
 24. Wason, R.: Deep learning: Evolution and expansion. *Cognitive Systems Research* 52, 701–708 (2018)

An Algorithm to Classify DNA Sequences of Hepatitis C Virus Based on Localized Conserved Regions and Heuristic Search

Sarahi Zúñiga-Herrera¹, Ivan Olmos-Pineda¹, Javier Garcés-Eisele²,
Mario Rossainz-López¹

¹ Benemérita Universidad Autónoma de Puebla, Facultad de ciencias de la Computación,
Puebla, Mexico

{sarahi.zuhe, ivanoprkl}@gmail.com, rossainz@cs.buap.mx

² Laboratorios Clínicos de Puebla, Puebla, Mexico

javier.garces@gmx.de

Abstract. In this work, the problem of classification in DNA sequences (STP Sequence-Typing Problem) is addressed. The main tool for molecular diagnosis is the polymerase chain reaction (PCR). However, the design of a PCR test is expensive (in time and economic resources) if the number of types of sequences to be classified is large and highly variable. The optimal design of PCR primers is essential to maximize the specificity and sensitivity of the test. This paper presents an algorithm to solve the problem of locating conserved sequences regions that help to discriminate between different types of hepatitis C virus using Shannon entropy concepts, information gain, a tree-based classifier and an evaluation function that considers specific parameters of PCR tests, which allow to identify patterns in conserved areas of DNA sequences.

Keywords: Shannon Entropy, Information Gain, DNA Sequences, Classification.

1 Introduction

DNA is made up of four molecules called nucleotides or nitrogenous bases. It contains all the genetic information of living organisms [1]. If we know a DNA sequence or part of it, we can determine to which organism it belongs. Thus, different types of sequence tests can be used in clinical analysis laboratories to identify: infectious agents present in a blood sample taken from a patient, for diagnostic tasks, and others applications.

One of the most frequent problems in diagnosis is to classify a given sequence and determine the class or subclass, which is known as sequence-typing problem (STP) [2]. There are different instances for the STP. Currently, the problem of classifying DNA sequences with a high variability rate is of great interest for the development of diagnostic tests. In this paper, the case for the hepatitis C virus is addressed. The hepatitis C virus (HCV), since its discovery in 1989, has been recognized as a serious

health worldwide problem [3]. The World Health Organization estimated that 71 million people had chronic HCV infections in 2015 [4]. At present, this virus represents the most frequent cause of chronic liver disease, cirrhosis and liver transplantation. HCV has a high genetic variability, identifying from its discovery to date seven main types associated with different behaviors of the virus in the host and responses to treatment [5]. To determine the type or subtype of HCV it is necessary to use a molecular diagnosis test which allows the doctor to be guided about the treatment for the patient.

One of the main tools in molecular diagnostics is the polymerase chain reaction (PCR) [1]. PCR revolutionized the field of molecular diagnostics, to the point that it currently represents the fastest growing segment in the clinical laboratories of the world [5].

The PCR process was originally developed to amplify short segments of a longer DNA molecule [6]. A typical amplification reaction includes target DNA, a thermostable DNA polymerase, two oligonucleotide primers, deoxynucleotide triphosphates (dNTPs), reaction buffer and magnesium. Once assembled, the reaction is placed in a thermal cycler, an instrument that subjects the reaction to a series of different temperatures for set amounts of time. This series of temperature and time adjustments is referred to as one cycle of amplification. Each PCR cycle theoretically doubles the amount of targeted sequence (amplicon) in the reaction. Ten cycles theoretically multiply the amplicon by a factor of about one thousand; 20 cycles, by a factor of more than a million in a matter of hours. Each cycle of PCR includes steps for template denaturation, primer annealing and primer extension [7]. PCR is characterized for being a technique with high sensitivity, reproducibility and efficiency, which generates reliable results in a short time and easy to analyze [8,9]. However, the design of a PCR diagnostic test can be very complex, if the number of sequences to be classified is large and, in addition, they are highly variable. The optimal design of primers is essential to maximize the specificity and efficiency of a PCR [10]. Poor design of the primers can result in small, nil or non-specific quantities of the amplification product. An appropriate design is one of the most important factors for the success of a PCR [8]. In general, the design of primers is summarized in 4 main points.

The first one is to obtain a database with the target genetic sequences, this database can be obtained in banks of international sequences such as GenBank or more selective sources such as ViRP where genetic sequences of viral pathogens are located, including HCV.

The second step is to process the database by aligning it using any of the currently available computational tools such as Jalview, Strap, ClustalX or Clustal Omega, among others [11] to locate the homologous and conserved regions. This is the process by which the sequences are compared by searching for common characters and establishing the correspondence residues between the related sequences. These regions are interesting because if there is a high degree of conservation, the probability of amplification increases.

The third step corresponds to the identification of the oligonucleotides or primers. Select those nucleotides that meet the chemical criteria that guarantee specificity and sensitivity.

The fourth corresponds to the verification and validation of the oligonucleotides proposed by PCR reactions.

The third step is described as the most expensive part of the entire design in terms of time and economic resources, especially for tests where you want to perform a classification of sequences as mentioned is the case for HCV, since two factors should be considered mainly: the first is to select regions that allow us to distinguish between one class and another; the second is that established chemical and thermodynamic conditions that guarantee the amplification of the PCR.

On the second factor, there are currently several computer programs that perform these tasks, such as oligo7 [11]. The level of development of these programs is so high that for identification tests it is enough to enter the sequence to the program and this will yield a series of the best proposals for the primers, leaving the researcher little or nothing to improve and accept the proposal made by the software. However, for the first factor and based on the study carried out in the state of the art, there are no computational tools that allow the selection of those primers that could classify the sequences of interest. Currently researchers are limited to observe the sequences and determine manually which is the right region to solve the classification problem.

2 State of the Art

The problem consists in deciding whether the unclassified DNA sequences belongs to a particular class. There are many highly efficient string-matching algorithms in the current state of the art that could be used to solve the problem of classification like Machine learning algorithms. However, in the clinical diagnosis a sample is extracted from a patient that contains a genetic material of which the sequence of nucleotides is unknown, so the algorithms available for the machine learning cannot be used since they require that the nucleotide sequence be known. Therefore, algorithms must be devised to help solve the classification problem by means of clinical diagnostic methods that currently exist, as has been mentioned one of the most used in PCR.

The researchers who design the primers for a PCR currently do the selection of the regions to solve the classification problem manually, this method is slow and complex, therefore it is necessary to design an algorithm that automates this process. Consequently, it is proposed to use the information gain criteria and look for those regions that after a mathematical analysis turn out to be better candidates to solve the classification problem.

To solve the STP problem on different pathogens, different proposals have been proposed. For example, in 2002 and 2004, two works were developed under the direction of Javier Garcés and his team, in which a solution to the STP problem for the human papillomavirus (HPV) is proposed using alignment tools of sequences, clustering processes, tools of Shannon's information theory such as entropy and information gain and decision theory [12, 13]. His proposal helped to design a molecular diagnostic test by RFLP-PCR (Restriction Fragment Length Polymorphism coupled to Polymerase Chain Reaction) that is currently used in laboratories for the diagnosis of HPV [14, 15]. On the other hand, Benish WA [16] applied the Shannon Information Theory

to clinical analysis tests calculating the Gain of Information pre- and post-test. Information theory was also successfully applied to the codon classification problem to reveal the order in the genetic code [17]. It has also been used to clarify the interrelationships between structure, function and evolution of a family of genes or gene products [18]. Ebeling and Frommel [19] applied the concept of entropy as the ability to describe the structure of information carriers such as DNA, proteins, text and musical notes. The research of Solis et al. [20] proposes a method to extract the maximum amount of information available from peptide structures in fragments of sequences, finding that the manner in which the structure is represented affects the quantity and quality of structural information that can be extracted from sequences.

Despite the work already done and the tools designed, these tools focus on solving the problem of identifying conserved regions and designing primers regardless of which regions optimize the classification process and without consider hybridization criteria for PCR tests. In the case of the work done by Garcés, the entropy analysis and information gain focuses on the design of an RFLP-PCR. Therefore, it is necessary to do applied research that allows the development of software tools to solve the problem of classification of sequences by PCR diagnostic tests.

3 Formal Description of the Problem

A nucleotide is a fundamental organic chemical compound of nucleic acids (DNA and RNA), constituted by a nitrogenous base, a sugar and a molecule of phosphoric acid [1]. In a DNA sequence there are four nucleotides: adenine (*A*), cytosine (*C*), guanine (*G*) and thymine (*T*). A DNA sequence or chain represented by G_w is a sequence of letters representing the structure of a DNA molecule, with the capacity to transport information. The alphabet of a DNA sequence is composed of letters *A, T, G* and *C* which symbolize the four nucleotides:

$$G_w = \langle v_1, v_2, v_3 \dots v_n \rangle. \quad (1)$$

A string of elements of the alphabet Σ_{ADN} where:

$$\Sigma_{ADN} = \{A, T, G, C\}, v_x \in \Sigma_{ADN}. \quad (2)$$

To be able to analyze a database with DNA sequences it is necessary to have the sequences aligned. Alignment is the process by which sequences are compared by searching for common characters and establishing the residues of correspondence between related sequences, to highlight areas of similarity, which could indicate functional, evolutionary or interesting relationships for analysis. These regions are interesting because if there is a high degree of conservation, the probability of amplification for a PCR increases. A set of aligned strings is assigned the name set *S*:

$$S = \{G_w: w = 1, 2, \dots m\}, |S| = m. \quad (3)$$

Each instance G_w belongs to a class C_y where *y* represents the *y*-th class value that is, the type of HCV. Each position or nucleotide of a set of aligned sequences *S* was

assigned the attribute name A_i where i indicates the position of the nucleotide. The domain $D(A_i)$ is equal to the set of values $v_x^i : x$ is the x -th value in the domain of the attribute A_i (See Figure 1).

Considering the concepts and the nomenclature previously described, we can formally describe the problem that we want to solve. In a set S of DNA sequences or instances G_w , we want to locate those attributes A_i that provide more information and are considered as the best attributes to solve the classification problem of the seven C_y classes of the hepatitis C virus that currently exist. These attributes must belong to a conserved region and consider the criteria that favor a molecular diagnostic test by PCR.

4 Proposed Solution

In the context of classification, the quality of an attribute A_i has to do with its capability to separate the instances G_w , between the different possible classes. If there is a direct relationship between the values of the attributes and the possible classes, it means that the attribute is very good to classify. The quality of an attribute has to do with what classes can be separated each time we instantiate that attribute A_i .

The classes are well separated when each subgroup is generated by the division of the product is homogeneous, that is: in each subgroup all the G_w belong to the same class once we instantiate the attribute with x . This is represented as S_x^i where x -th of the domain of A_i . Therefore, a homogeneity metric is necessary.

4.1 Shannon Entropy

Shannon entropy comes from information theory, which can be interpreted as the degree of error or the certification of a classification problem, which is a good way to measure homogeneity. It is defined as $H(S)$:

$$H(S) = - \sum_{y=1}^N P(C_y) \times \ln(P(C_y)). \quad (4)$$

Where $P(C_y) = \frac{|C_y|}{|S|}$ corresponds to the probability that an element belongs to the C_y class and N the total number of classes.

4.2 Information Gain

The information gain allows quantify the information provided by an attribute A_i with respect to the classification problem. The information gain is defined as the difference between the entropy of Shannon before $H(S)$ and then $H(S | A_i)$ of knowing the value of the attribute A_i :

$$I_G(A_i) = H(S) - H(S | A_i). \quad (5)$$

The attribute A_i subdivides the instances of S into z_i subgroups S_x^i ($x = 1, \dots, z_i$) where $z_i = |D(A_i)|$ that is, the number of values that the attribute can present. To calculate the entropy of $H(S|A_i)$, it is calculated as the weighted average $\frac{|S_x^i|}{|S|}$ of the Shannon entropy in each subgroup S_x^i :

$$H(S|A_i) = \sum_{x=1}^{z_i} P\left(\frac{|S_x^i|}{|S|}\right) \times H(S|S_x^i), \tag{6}$$

Where:

$$H(S|S_x^i) = - \sum_{y=1}^N P(S_y^C(S_x^i)) \times \ln(P(S_y^C(S_x^i))). \tag{7}$$

The function $P(S_y^C(S_x^i)) = \frac{|S_y^C(S_x^i)|}{|S_x^i|}$, is the probability that an element $S_y^C(S_x^i)$ belongs to the class C_y and if the element belongs to the subgroup S_x^i .

C_y	G_w	A_1	A_2	A_3	A_4	A_5
C_1	G_1	A	T	T	A	T
C_1	G_2	A	T	T	C	T
C_2	G_3	A	G	C	A	C
C_2	G_4	A	G	C	G	C
C_2	G_5	A	G	C	A	T
C_2	G_6	A	G	C	G	T
C_3	G_7	G	T	C	T	C
C_3	G_8	G	T	C	T	C
C_3	G_9	G	T	C	A	G
C_3	G_{10}	G	T	C	T	G
C_4	G_{11}	A	G	A	A	C
C_4	G_{12}	A	G	A	G	C

$S_x^1 = S_C^1$ $S_y^C(S_x^i) = S_A^3(S_x^i)$

Fig. 1. Representation of S set of instances G_w , where each instance belongs to a class C_y . Each position or nucleotide of a set of aligned sequences S was assigned the attribute name A_i where i indicates the position of the nucleotide. The attribute A_i subdivides the instances of S into z_i subgroups S_x^i ($x = 1, \dots, z_i$) where $z_i = |D(A_i)|$ $S_y^C(S_x^i)$ expresses that the subset S_x^i belongs to the C_y class.

To understand the above, in Table 1 an example associated with the values is shown. When applying the formula 5 for each of the attributes of the set S it is observed that the attribute A_3 is evaluated with the greatest information gain and divides in three subsets to the set S . The first one is S_C^3 where its instances belong to the classes C_2 and C_3 . The second subset is S_A^3 with all its instances belonging to class C_4 . The last subset is the S_T^3 where all instances belong to the class C_1 . Because the subset S_C^3 does not have instances of a single class, the information gain analysis is performed again applying formula 5. If two instances do not have the same value for each attribute and belong to different classes, the attributes are suitable to carry out

the classification, as can be seen, the subgroup S_C^3 the attributes A_1 and A_2 allow to classify the elements correctly for the classes C_2 and C_3 .

This example shows that it is possible to classify DNA sequences using the concepts of entropy and information gain. For this case the attribute A_3 with greater I_G is selected, this attribute allows to quickly discriminate the classes C_1 and C_4 . When calculating again I_G of all the attributes, it is obtained that both A_1 and A_2 allow to discriminate the classes C_2 and C_3 . The above can be displayed in a simple way in a decision tree where each vertex has a maximum of 4 possible values (Figure 2).

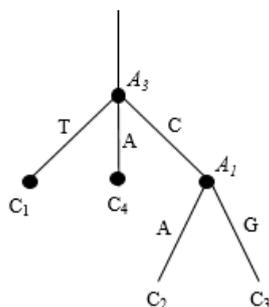


Fig. 2. Decision tree for the classification of the data presented in Fig. 1.

In addition to using the information gain criterion to select the attribute that best separates the classes, it is necessary to contemplate that the results will be used to mount a PCR diagnostic test. It is necessary to consider some criteria that can optimize the design of the test.

Lefever et al. [21] demonstrated the effect that the type of mismatch has on the alignment and position in a primer on extension efficiency during the first cycles of the PCR. Finding minimal or no extension step [7] when they introduced a mismatch in the last 3 or 4 nucleotides of the primer at the 3' end. Their hypothesis was that the low extension was caused by the reduction in the binding of the enzyme DNA polymerase to the binding site [7]. He concluded that the closer the mismatch to the extreme 3' was, the greater the impact it had during the extension of the PCR, increasing the number of cycles where the fluorescence signal was detected in the PCR.

Using the analysis carried out by Lefever, an evaluation function was designed to consider its contributions and, in addition, to consider the criterion of information gain.

The criteria used for the evaluation function $E(A_i, d)$ was, the attribute of interest A_i must be evaluated in I_G to consider it a good attribute to discriminate between classes. The attributes that surround it must be also be evaluated to establish if it is a good option for the primer design.

The attributes surrounding the A_i attribute were called window $\varphi(A_i, d)$ (see Figure 3), the window $\varphi(A_i, d)$ can be expressed as:

$$\varphi(A_i, d) = \varphi(A_i, d)^- \cup \varphi(A_i, d)^+ ; d = \text{distance}, \quad (8)$$

Such that:

$$\varphi(A_i, d)^- = \{A_j: J = i - d, \dots, i - 1\} \text{ and } \varphi(A_i, d)^+ = \{A_j: J = i, \dots, i + d\}. \quad (9)$$

C_y	G_w	A_i					
C_1	G_1	A	T	T	A	T	
C_1	G_2	A	T	T	C	T	
C_2	G_3	A	G	C	A	C	
C_2	G_4	A	G	C	G	C	
C_2	G_5	A	G	C	A	T	
C_2	G_6	A	G	C	G	T	
C_3	G_7	G	T	C	T	C	
C_3	G_8	G	T	C	T	C	
C_3	G_9	G	T	C	A	G	
C_3	G_{10}	G	T	C	T	G	
C_4	G_{11}	A	G	A	A	C	
C_4	G_{12}	A	G	A	G	C	

$\longleftarrow A_i \longrightarrow$
 $\varphi(A_i, d)^+ \qquad \varphi(A_i, d)^-$

Fig. 3. Representation of the evaluation window $\varphi(A_i, d)$ where d indicates the number of attributes that relate to the analysis of the window to the right and to the left of the attribute of interest A_i .

Therefore, the selection criterion was established by means of the evaluation function $E(A_i, d)$ is simply the product between $I_G(A_i)$ and the evaluation of the window $E\varphi(A_i, d)$:

$$E\varphi(A_i, d) = -I_G(A_i) \times E\varphi(A_i, d). \quad (10)$$

Where $E\varphi(A_i, d) = \sum_{-d}^d (W_i^{-1} \times H(A_i))$ such that $E\varphi(A_i, d)$ is the evaluation of window $\varphi(A_i, d)$ with d attributes before and after position A_i .

W_i is the weight of the attribute W_i taken from the Lefever experiments defined as $\sum \frac{dC_q}{R_i}$ the sum of the differences between C_{qMM} number of cycles for the amplification with the unadjusted alignment and the number of cycles for the amplification with the alignment perfect C_{qP} [21]. It is simplified in Table 1, based on the results obtained by Lefever.

Table 1. Weights W_i and its associated value by position based on Lefever’s results.

W_i	Value associated with W_i
0	0.687
1	0.057
2	0.031
3	0.016
4	0.012
5-20	0.014

The entropy of the attribute A_i is defined as:

$$H(A_i) = \sum_{x=1}^{z_i} P(v_x^i) \times \ln(P(v_x^i)). \quad (11)$$

Such that $P(v_x^i) = \frac{|v_x^i|}{S}$.

The previous analysis indicates that the concepts of entropy and information gain allow to classify DNA sequences and due to the study of Lefever can be considered criteria that can favor the design of primers for a PCR. Based on the above analysis we design the Algorithm 1 that receives a database with instances of DNA sequences and selects the best attributes by analyzing them using (10). Finally, it returns a decision tree where each node is an attribute that solves the classification problem.

Algorithm 1. Receive from a database with instances of DNA sequences. Returns a decision tree where each node is an attribute that solves the classification problem.

```

id3Ev (instances, target_attribute, attributes)
begin
    Create a new root node to the tree;
    If all instances have the target_attribute belonging to the
    same Cv
        Return the tree with single root node with label Cv;
    If attributes is empty, then
        Return the tree with single root node with the most common
        label of a target_attribute in instances;
    Else
        Ai := The attribute in attributes which best classifies in-
        stances;
        root decision attribute := Ai;
        For each possible value vi of Ai
            Add new ramification below root, corresponding to the test
            Ai = vi;
            Let instancesvi be the subset of instances with the value
            vi for Ai;
            If instancesvi is empty then
                Below this ramification, add a new leaf node with the
                most common value of target_attribute in instances;
            Else
                DTC (instancesvi, target_attribute, attributes);
end
    
```

5 Discussion

The in silico design [21] of the primers can allow a tremendous saving of time and money in the development of diagnostic tests, however, they can never replace the experimental verification tests since it is not possible to predict the specificity of a first in silico.

The proposed methodology seems to be an adequate solution to help researchers design primers to solve STP problems. As mentioned in the introduction, researchers are currently limited to observing the sequences and manually determining which regions are suitable for solving the classification problem. This paper presents the design of a methodology as a solution proposal to solve the classification problem through the concepts of information gain and an evaluation function that quantifies the value of the conserved area where the attributes are of interest. This evaluation function uses the W_i weight, which is a heuristic value, calculated with the results offered by Lefever in its work.

Algorithm 1 is currently in the implementation stage to perform tests later. It is expected that the algorithm will be evaluated not only by a mathematical analysis but also be analyzed by experts in the area of molecular biology and diagnostics that can determine the quantitative and qualitative quality of the method.

6 Conclusions

The example presented in section 4.2 shows that it is possible to find a way to classify all instances of the database with DNA sequences using the concepts of entropy and information gain.

The proposed algorithm in addition to using these concepts incorporates an evaluation function that establishes criteria that could favor the design of primers for a PCR. This can be useful to solve the task of classifying genetic sequences with high variability rates. Algorithm 1 finally proposes a decision tree where the nodes are the suggested attributes for researchers to design primers and a PCR diagnostic test to detect the seven types of Hepatitis C virus that currently exist.

References

1. Panduro, A.: *Biología molecular en la clínica*. McGraw-Hill Interamericana (2000)
2. Eisele, J. G., Roldán, C. Y. C., Galindo, M. O., Gil, M. P.: Usefulness of solution algorithms of the traveling salesman problem in the typing of biological sequences in a clinical laboratory setting. In: 14th IEEE International Conference on Electronics, Communications and Computers CONIELECOMP 2004, pp. 264–269 (2004)
3. Mohd Hanafiah, K., Groeger, J., Flaxman, A.D., Wiersma, S.T.: Global epidemiology of hepatitis C virus infection: new estimates of agespecific antibody to HCV seroprevalence. *Hepatology* 57(4), 1333–1342 (2013)
4. World Health Organization, <http://apps.who.int/iris/bitstream/10665/255016/1/9789241565455-eng.pdf?ua=1>
5. Messina, J. P., Humphreys, I., Flaxman, A., Brown, A., Cooke, G. S., Pybus, O. G., Barnes, E.: Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology* 61(1), 77–87 (2015)
6. Sharkey, D.: Antibodies as thermolabile switches: High temperature triggering for the polymerase chain reaction. *Biotechnology* 12, 506–509 (1994)

7. Eckert, K.A., Kunkel, T.A.: DNA polymerase fidelity and the polymerase chain reaction. *PCR Methods Appl.* 1, 17–24 (1991)
8. Mas, E., Poza, J., Ciriza, J., Zaragoza, P., Osta, R., Rodellar, C.: Fundamento de la Reacción en Cadena de la Polimerasa (PCR). *AquaTIC* 15, 70–78 (2016).
9. Tamay de Dios, L., Ibarra, C., Velasquillo, C.: Fundamentos de la reacción en cadena de la polimerasa (PCR) y de la PCR en tiempo real. *Investigación en discapacidad* 2(2), 70–78 (2013)
10. Abd-Elsalam, K. A.: Bioinformatic tools and guideline for PCR primer design. *African Journal of biotechnology* 2(5), 91–95 (2003)
11. Bioinformatics at COMAV, bioinf.comav.upv.es/courses/intro_bioinf/multiple.html
12. Lozano Yécora J.: La genotipificación del virus de papiloma humano: una familia de problemas de optimización combinatoria. Undergraduate thesis, Universidad de las Américas, Puebla(2004)
13. Fernández, R.; Desarrollo de Algoritmos para la Clasificación de Secuencias. Master's Thesis, Universidad de las Américas, Puebla (2002)
14. Rodríguez, R. S., Roldán, C. Y. C., Eisele, J. G., Gil, P. G., Galindo, M. J. O.: Algorithms for the typing of related DNA sequences. In: 15th IEEE. International Conference on Electronics, Communications and Computers CONIELECOMP'05, pp. 268–271 (2005)
15. Benish, W.A.: Relative entropy as a measure of diagnostic information. *Med Decis Making* 19(2), 202–206 (1999)
16. Jiménez, M.: On the syntactic and redundancy distribution of the genetic code. *BioSystems* 32, 11–23 (1994)
17. Gotoh, O.: Multiple sequence alignment: algorithms and applications. *Adv Biophys* 36, 159–206 (1999)
18. Ebeling, W., Frommel, C.: Entropy and predictability of information carriers. *Biosystems* 46(1-2), 47–55 (1998)
19. Solis, A.D, Rackovsky, S.: Optimized representations and maximal information in proteins. *Proteins* 38(2), 149–164 (2000)
20. Lefever, S., Pattyn, F., Hellemans, J., Vandesompele, J.: Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. *Clinical chemistry* 59(10), 1470–1480 (2013)
21. Okada, M., Tsukamoto, M., Ohwada, H., Aoki, S.: Consensus scoring to improve the predictive power of in-silico screening for drug design. In: Proceedings of the 2nd International Conference on Engineering and Meta-Engineering, pp. 27–30 (2011)

Towards the Development of a Navigation System for a Bipedal Robot: Preliminary Analysis of the Literature

Miguel Angel Ortega-Palacios, Josefina Guerrero-García,
Juan Manuel González-Calleros

Benemérita Universidad Autónoma de Puebla,
Language & Knowledge Engineering (LKE), Puebla, Mexico
miguel.ortegap@alumno.buap.mx,
{josefina.guerrero,juan.gonzalez}@correo.buap.mx

Abstract. In the literature, several works are reported about the gait of bipedal robots. Some papers are associated with stability or automatic learning and navigation, but no paper was found that includes the variables of bipedal gait, stability, navigation and automatic learning, all applied in an indoor bipedal robot. Our objective is, after the analysis of the state of the art, to propose a methodology that will allow the autonomous navigation of a bipedal robot and validate its degree of autonomy with respect to the evaluation metrics reported in the literature.

Keywords: Bipedal Robot, Bipedal Gait, Stability, Navigation.

1 Introduction

As many mobile robots begin to integrate into different areas of society, they will need to operate in a wide variety of environments. Often, these environments will be dynamic; the objects move and the structures change physically. Less dynamic environments can be characterized by physical changes that occur over days, weeks or months, while more dynamic environments involve objects in continuous motion, such as human beings or vehicles.

Mobile robots cannot assume that the world is static if we expect them to work effectively. For a mobile robot to work autonomously in a dynamic environment, it must have a way of detecting its environment. Cameras are ubiquitous among modern robots and can provide a large amount of information about the environment. Robotics applications can employ techniques such as computer vision to perform object recognition, 3D reconstruction, mapping and localization tasks [1]. Nowadays, most robots require moving and performing tasks in a variety of environments that are sometimes even unpredictable. Navigation with mobile robots is a challenging problem in the field of robotics where numerous studies have been conducted that have resulted in a variety of solutions. Four integral parts are identified in the navigation of a biped robot: perception, location, movement control and trajectory planning.

The rest of this document is structured as follows: the second section addresses the importance of bipedal robots, section three presents the state of the art analysis, the

limitations found in the literature are mentioned in the fourth section, the proposed methodology is presented in the fifth section and finally the conclusions are presented.

2 Importance of Bipedal Robots

The study of walking anthropomorphic robots has been a challenge for the scientific community until today, the fact to replicate the movements and the way of walking of human beings is not trivial and is something that has not been completely imitated.

Moreover, there is the inherent difficulty of giving artificial intelligence to a humanoid robot so that it is capable of learning to walk on its own in a dynamic environment, just as a human being does. Hence, the interest in developing an autonomous navigation system for a biped robot that is capable of navigating indoors through a horizontal plane, which can detect and evade static or mobile obstacles during its trajectory.

The advantages of this type of robots are not only limited to their ability to move on irregular terrain, but also have the advantage of having omnidirectional movement regardless of the orientation of its body. In addition, they have multiple applications that allow executing different tasks, such as surveillance, search, cooperation, transportation, navigation, and even carry out work that includes risks in the health or in the lives of people. In addition, the use of this type of robots can reduce execution times of some tasks, reducing fatigue in people, and therefore greater productivity when doing some task.

3 State of the Art

3.1 Results of the Analysis of the Literature

Among the works related to the navigation of mobile robots, most of the works are focused on robots with wheels [3-6, 9, 11-27], in second place the bipedal robots [8, 28-38] and in the last place are another type of robots which navigate on water or air [2, 7, 10, 39, 40, 41], as it can see in the graph of the Figure 1.

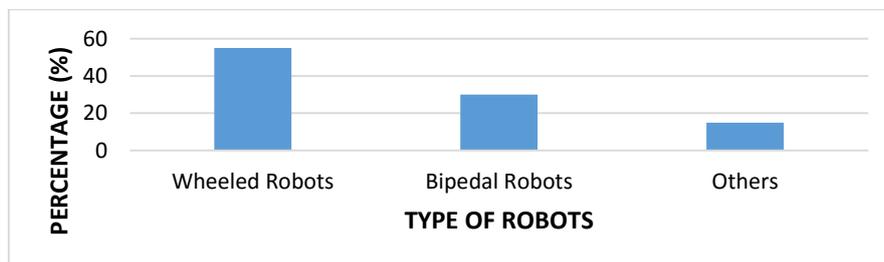


Fig. 1. Types of mobile robots studied in the literature.

As can be identified in Figure 2, most of the works reviewed so far take into account navigation elements or navigation elements in conjunction with elements of machine learning, or only the variables of running and stability separately, but not found any that integrates all these concepts at the same time to develop an autonomous navigation system for bipedal robots.

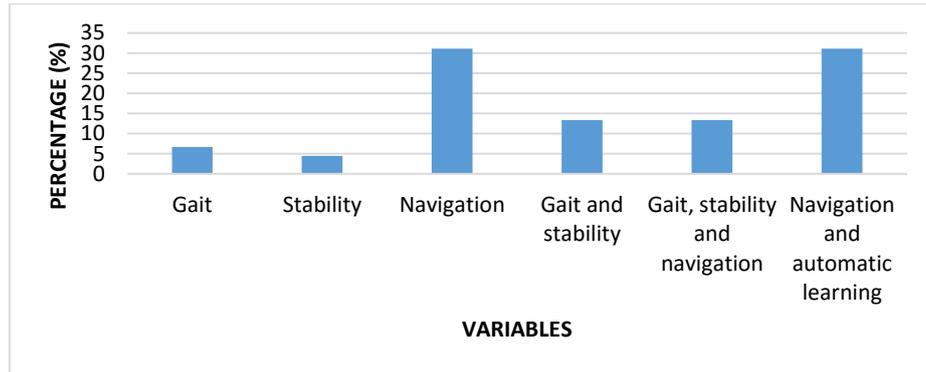


Fig. 2. Distribution of the study of the variables in the literature.

Table 1 identifies the works of the literature focused on the study of each of the variables: gait, stability, navigation and automatic learning.

As it can be observed, there are more researches reported that study stability in conjunction with navigation, so there are works that study navigation with the automatic learning, but it was not possible to find one that takes into account all the variables of interest.

Table 1. Variables in reviewed literature papers.

Paper	Gait	Stability	Navigation	Automatic learning
[42], [43], [50]				
[65], [78]				
[44], [51], [52], [53], [54], [55], [62], [64], [66], [69], [70], [71], [68], [75]				
[47], [49], [59], [60], [61], [77]				
[45], [56], [57], [58], [63], [67]				
[72], [73], [79], [80], [81], [82], [83], [84], [85], [86], [87], [88] [89], [90]				

There is a great variety of robots that are used to obtain the experimental results in the reviewed works of the literature, the most used reported in the investigations are the robots: Nao, Lola and HRP-2, as shown in Table 2.

Table 2. Robots used in literature papers.

Robot	Paper	Percentage (%)
HRP-2	[45], [63], [66], [80]	9.52 %
Amber	[42]	2.38 %
Nao	[42], [44], [52], [62], [64], [69], [70], [75], [81]	21.44 %
iCub	[47]	2.38 %
Lola	[53], [56], [57], [58], [59], [60], [61], [65], [67]	21.44 %
Lego mindstorm	[54]	2.38 %
Huro Evolution	[68]	2.38 %
Tibi y Dabo	[71]	2.38 %
SCITOS G5	[73]	2.38 %
Bioloid	[74], [77], [78]	7.14 %
E-puck	[79]	2.38 %
PR2	[80]	2.38 %
Kephera	[82], [83]	4.76 %
Robot móvil	[84], [86], [88]	7.14 %
Robotino	[85], [89]	4.76 %
Robot Raving	[90]	2.38 %
Daryl	[87]	2.38 %

As it can be observed, the Lola and Nao robot are the most used in the investigations, with respect to the Lola robot, its great use is due to the fact that there are multiple works belonging to the same authors, and they have followed it up to improve it through the years. While the use of the Nao robot is due to the fact that it is a robot with last generation technology, it has a high degree of interactivity for any type of user; it is fully programmable, and currently more than 400 universities are working on this platform. It makes this robot the platform with the most advanced technology in entertainment robotics for research and development.

Most of the works reviewed so far are focused on the detection and avoidance of obstacles in static or dynamic environments, that is, the obstacles can be fixed or can be moved during the navigation of the robot. In addition, the robots are regularly navigated on 1-level flat terrain [42, 44, 45, 47, 51, 54, 56-62, 64, 66, 68-73, 75, 77-90] or multi levels [52-67], even in irregular terrain [63]. The distribution of the type of land used in the experimentation of the reviewed papers is shown in Figure 3.

In the reviewed works, the application given to each robot is not specifically specified, only in [62] it is mentioned that the robot is intended to send a message to a staff member in an office environment, or for example in [68] the robot is required to be able to meet the technical challenge to avoid obstacles of the RoboCup Humanoid

League. While in the other works there is the context of the need to navigate and avoid obstacles to reach a goal. Therefore, the context of the works is treated from a very general approach, as shown by the graph of Figure 4.

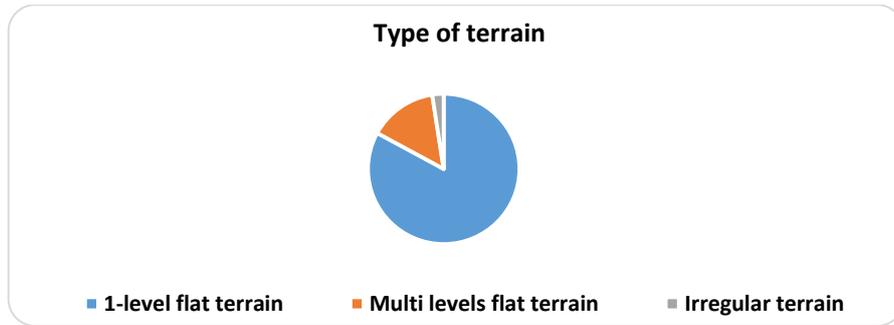


Fig. 3. Type of terrain used in the experimentation of the works reviewed in the literature.

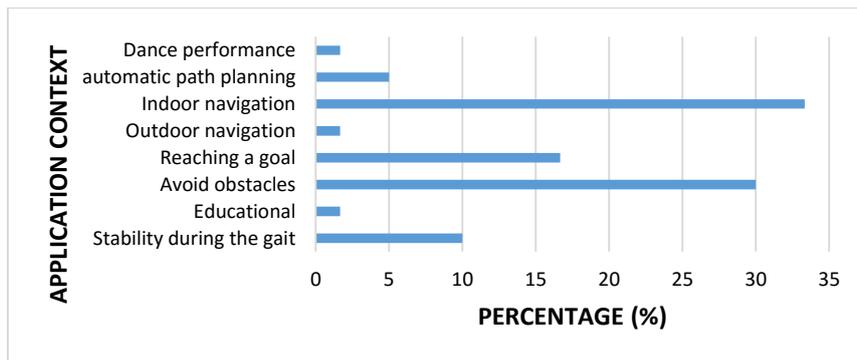


Fig. 4. Application context of the papers reviewed in the literature.

4 Limitations Found in the Literature

The main limitations found in the review of the literature are indicated below:

Physical restrictions of the robot. Due to the height of the robot, it is not always possible to capture good images of the reference points in the environment, because they could be placed at a height where the robot does not have access [62].

Restrictions on the movement of the robot. The planner only considers the quadrant directions (up, down, left, right) to prevent the robot from spinning while walking, which would increase the imprecision of the odometer data [62].

Computationally expensive. The computational resources required for pose estimation based on characteristics are quite expensive, and the method cannot be executed during the whole trip of the robot [62, 75].

Use of a 3D map for the location. The location system may depend on the availability of a 3D map to locate, which must be acquired by the robot itself or provided by the user in advance [52, 57].

Fixed marks are used to locate the robot. The problem of localization is solved by using reference points placed by the user previously on the walls of the environment where the robot travels. It also remains a challenge to build more computationally efficient algorithms that can be used continuously to detect landmarks [62, 64].

Use of an external computer. The construction of the terrain map and the planning of the steps are carried out on a computer that is not mounted on the robot [63].

Wire data transmission. The sensor data and the step planning result are sent to / from the robot via Ethernet [63].

Use of the robot's predetermined functions. The functions that are already predefined in the programming of the robot are used, for example, a function for the recognition of objects or for the running of the robot [64].

The resolution of the images. The resolution of the images acquired by the camera is an important factor in defining the processing time of the camera, which can affect the efficiency of the algorithm if a real-time response is required [66, 68, 69, 81, 84].

Maximum speed of the robot. The time required for the robot to reach its goal will not only depend on the efficiency of the algorithm, but also on the maximum speed at which the robot can move [58, 67, 70, 74].

The size, shape, and texture of the obstacles. This is important, because the efficiency for the detection of obstacles depends very much on the size, texture or shape of the objects to be detected and also on the way in which the robot acquires the information of the environment, either through sensors or from a camera [52, 66].

Restricted field of view. The vision system, like most built-in vision systems, may fail to recognize reference points due to too large viewing angles (which significantly reduces the quality of the image) and variation in lighting conditions. In the case of sensors, depending on the type and its characteristics, it has a maximum and minimum distance for the detection of objects [52, 53, 58, 62, 79].

Limitation on the size of the test space. Experimental tests are carried out in user-controlled environments, with either models or the use of harnesses so that the robot can move in a limited space [57, 83].

Static environment. The detection and avoidance of obstacles during the navigation of the robot only contemplates static obstacles and some are added in real time [69, 82, 83].

5 Proposed Methodology

After performing this research in the literature; it is proposed to use a Bioloid Premium Robot with 18 degrees of freedom, since there is no access to a complete list of the samples that make up the population, a non-probabilistic sampling technique was used. In this case, convenience sampling was used, since a sample of the population was selected due to the fact that it was accessible, but not because it was selected by a statistical criterion.

The study variables, presented graphically in Figure 5 are the following:

-Gait kinematics (independent), which corresponds to the analysis and modeling of the biped robot cycle.

-Stability (independent), which is related to the balance point of the robot to stand up during the march.

-Navigation (independent) will be defined so that the robot can move from one point to another in a horizontal plane.

-Level of autonomy (dependent), which will depend on the integration of the three previous variables so that the robot is able to navigate autonomously indoors, using an automatic learning algorithm.



Fig. 5. System variables.

The proposed methodology aims to reduce the gap in the limitations related to wire data transmission, use of the robot's predefined functions and the detection of objects limited by their size, shape, and texture. In addition, this methodology will provide autonomy to the entire system, so that the control of each of the variables will be modified to obtain the best result at all times. The stages of the proposed methodology are:

-Study and analyze the operation of the robot. There will be an in-depth study of the Bioloid robot mechanism, analyzing its kinematics, that is, the limitations of the movements of each of its joints. Therefore, it is necessary to analyze the operation of the robot servomotors.

-Implement the human walking dynamics in the robot, using the dynamic model of the robot. As you can see in Figure 6.

Following the base of the scheme of Figure 7, the same process is applied to evaluate the methods of automatic learning, stability and navigation, continuing with the following steps of the methodology:

-Select and apply a stability method for the robot. The concept of zero moment point will be taken into account; it is useful in determining if a bipedal robot is in a stable configuration or not. It has been used extensively in bipedal robots.

-Select and apply a navigation method for the robot. Different navigation methods will be evaluated, such as fuzzy control or reinforcement learning to develop automatic path planning, selecting the most satisfactory method.

-Implement all variables in the robot: gait, stability, navigation and automatic learning indoors.

-Validate the degree of autonomy in the navigation of the robot indoors. Evaluate the performance of the robot's automatic navigation, measuring the degree of autonomy of the robot, which is, making a comparison between the expected result and the real

result, evaluating if the robot arrives satisfactorily at its goal in the expected time, acting as autonomous form. The criteria to determine the degree of autonomy can vary and will be attached to standardized metrics in the literature. If the result is not satisfactory, the entire process will have to be repeated.

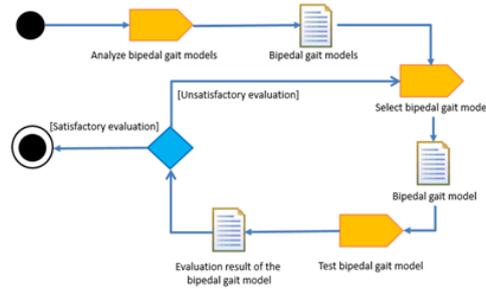


Fig. 6. Selection process of suitable gait model for the bipedal robot.

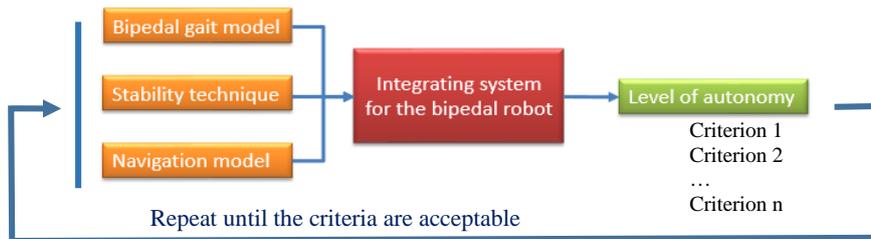


Fig. 7. Evaluation of the proposed solution based on criteria and parameters reported in the literature.

6 Conclusions

During the review of the literature, a work including all the study variables at the same time, i.e. bipedal gait, stability, navigation and automatic learning was not found. It is very important to design a methodology that integrates all the variables in a biped robot to be able to navigate autonomously indoors without using pre-established routines in order to generate a panorama of multiple applications for future researchers.

Regarding the study variables, most of the works are focused on the study of navigation and in conjunction with machine learning, but in this last approach only robots with wheels are used. It was also found that most of the robots are tested on flat terrain either one or more levels, probably because they do not include the study of the variable of stability in the case of bipedal robots, in the case of robots with wheels this does not apply since its function is to move on flat terrains. It is also simpler to avoid the use of stability in the autonomous navigation of bipedal robots when they have to walk on flat terrain.

As mentioned above, the most used robots in the literature is the Nao and Lola robot. Another important point is that the specific application of each of the works reviewed in the literature is almost never mentioned, so it will be very important to find the specific context for which the proposed autonomous navigation system will be used and which will serve to justify your investigation.

Finally, it is important to point out that although there has been a lot of work in the area of navigation and bipedal robots. There are still many limitations that cause a high computational cost in the application of the algorithms, problems in acquiring environmental information in a correct manner, restrictions in the visual field of sensors, cameras or even restrictions on the physical characteristics of the robot itself.

References

1. Russell, S. J., Norvig, P.: *Artificial Intelligence: a modern approach*. Pearson Education Limited (2004)
2. Guan, X., Cai, C.: A New Integrated Navigation System for the Indoor Unmanned Aerial Vehicles (Uavs) Based on the Neural Network Predictive Compensation. In: 33th Youth Academic Annual Conference of Chinese Association of Automation (YAC), pp. 575–580 (2018)
3. Grewal, H. S., Jayaprakash, N. T., Matthews, A., Shrivastav, C., George, K.: Autonomous Wheelchair Navigation in Unmapped Indoor Environments. In: IEEE International Instrumentation and Measurement Technology Conference (I2MTC), pp. 1–6 (2018)
4. Catane, A., Feinberg, Y., Dor, O.: Indoor Navigation System. U. S. Patent Application Publication, Pub. No.: US 2013 / 0262223 A1 (2013)
5. Schtz, M., Dobrev, Y., Carlowitz, C., Vossiek, M.: Wireless Local Positioning with SILO-Based Backscatter Transponders for Autonomous Robot Navigation. In: IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM), pp. 1–4 (2018)
6. Luo, R. C., Shih W.: Autonomous Mobile Robot Intrinsic Navigation Based on Visual Topological Map. In: IEEE 27th International Symposium on Industrial Electronics (ISIE), pp. 541–546 (2018)
7. Beak, D., Hwang, M., Kim, H., Kwon, D.: Path Planning for Automation of Surgery Robot Based on Probabilistic Roadmap and Reinforcement Learning. In: 15th International Conference, Ubiquitous Robot, pp. 342–347 (2018)
8. Mrudul, K., Mandava, R. K., Vundavilli, P. R.: An Efficient Path Planning Algorithm for Biped Robot Using Fast Marching Method. *Procedia Comput. Sci.* 133, 116–123 (2018)
9. Pattanayak, S., Sahoo, S. C., Choudhury, B. B.: An Effective Path Planning of a Mobile Robot. In: *Soft Computing in Data Analytics*, pp. 175–182 (2019)
10. Follestad M. H.: *Autonomous Path-Planning and Following for a Marine Surface Robot*. Master's thesis, Norwegian University of Science and Technology (2017)
11. Hsu, H. Y., Fu W., Lin Y.: Robot Surveillance System, U. S. Patent Application Publication, Pub. No.: US 2018 / 0181141 A1, (2018)
12. Fink, W., Brooks, A. J. W., Tarbell, M. A.: An Adaptive Hierarchical Approach to Lidar-Based Autonomous Robotic Navigation. *Micro-Nanotechnology Sensors, Syst. Appl.* X(69), 1–14 (2018)
13. Graven, O. H., Srisuphab, A., Silapachote, P., Sirilertworakul, V., Ampornwathanakun, W., Anekwirotj, P., Maitrichit, N.: An Autonomous Indoor Exploration Robot Rover and 3D Modeling with Photogrammetry. In: *International ECTI Northern Section Conference on*

- Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON), pp. 44–47 (2018)
14. Metka, B., Franzius, M., Bauer-Wersing, U.: Efficient Navigation Using Slow Feature Gradients. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1311–1316 (2017)
 15. Kim, J. T., Choi, Y. H., Lee, J., Hong, S. H.: Floor-to-Floor Navigation for a Mobile Robot. In: 10th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 362–363 (2013)
 16. Yun, C. H., Moon, Y. S., Ko, N. Y.: Vision Based Navigation for Golf Ball Collecting Mobile Robot. In: 13th International Conference on Control, Automation and Systems (ICCAS 2013), pp. 201–203 (2013)
 17. Zhou, X., Weber, C., Wermter, S.: A Self-Organizing Method for Robot Navigation Based on Learned Place and Head-Direction Cells. In: 2018 International Joint Conference on Neural Networks (IJCNN), pp. 1–8 (2018)
 18. Obo, T.: Intelligent Fuzzy Controller for Human-Aware Robot Navigation. In: 12th France-Japan and 10th Europe-Asia Congress on Mechatronics, pp. 392–397. IEEE (2018)
 19. Lim, J., Lee, S., Tewolde, G., Kwon, J.: Indoor Localization and Navigation for A Mobile Robot Equipped with Rotating Ultrasonic Sensors Using A Smartphone as The Robot's Brain. In: 2015 IEEE International Conference on Electro/Information Technology (EIT), pp. 621–625 (2015)
 20. Chinnaiah, M. C., Priyanka, G., Vani, G. D., Jyoti, M. A., Vennela, K.: A New Approach: An FPGA Based Robot Navigation for Patrolling in Service Environment. In: International Conference on Research Advances in Integrated Navigation Systems (RAINS), pp. 1–4. (2016)
 21. Alves, P., Costelha, H., Neves, C.: Localization and Navigation of a Mobile Robot in an Office-Like Environment. In: 13th International Conference on Autonomous Robot Systems. pp. 1–6 (2013)
 22. Le, C., Zhenghua, L.: Design of Two-Stage Fuzzy Controller for Mobile Robot Using Vision Navigation. In: Guidance, Navigation and Control Conference (CGNCC), 2016 IEEE Chinese. pp. 872–877 (2016)
 23. Kim, T., Kim, J.: Beaconless Navigation for Mobile Robots Using Grid Line Pattern. In: Ubiquitous Robots and Ambient Intelligence (URAI), 2013 10th International Conference on, pp. 200–205 (2013)
 24. Ko, B., Choi, H. J., Hong, C., Kim, J. H., Kwon, O. C., Yoo, C. D.: Neural Network-Based Autonomous Navigation for a Homecare Mobile Robot. In: Big Data and Smart Computing (Bigcomp), 2017 IEEE International Conference on Big Data, pp. 403–406 (2017)
 25. Ravankar, A. A., Ravankar, A., Emaru, T., Kobayashi, Y.: A Hybrid Topological Mapping and Navigation Method for Large Area Robot Mapping. In: 56th Annual Conference of Society of Instrument and Control Engineers of Japan (SICE), pp. 1104–1107. (2017)
 26. Novitsky, A., Yukhimets, D.: The Navigation Method of Wheeled Mobile Robot Based on Data Fusion Obtained from Onboard Sensors and Camera. In: 15th International Conference on Control, Automation and Systems (ICCAS), pp. 574–579 (2015)
 27. Crépon, P. A., Panchea, A., Chapoutot, A.: Reliable Navigation Planning Implementation on a Two-Wheeled Mobile Robot. In: Second IEEE International Conference on Robotic Computing (IRC), pp. 173–174 (2018)
 28. Mufti, S., Alam, M. N.: Real-Time Obstacle Avoidance and Path Planning of a Self-Navigated Autonomous Biped Robot Using Rugbat Sonar Sensors and Modular Digital Image Processing. In: 2012 International Conference on Robotics and Artificial Intelligence (ICRAI), pp. 75–80 (2012)

29. Elmogy, M., Habel, C., Zhang, J.: Online Motion Planning for Hoap-2 Humanoid Robot Navigation. In: International Conference on Intelligent Robots and Systems, pp. 3531–3536 (2009)
30. Gutmann, J. S., Fukuchi, M., Fujita, M.: A Floor and Obstacle Height Map for 3D Navigation of a Humanoid Robot. In: Proceedings of the 2005 IEEE International Conference on Robotics and Automation, pp. 1066–1071 (2005)
31. Wahrmann, D., Hildebrandt, A. C., Wittmann, R., Sygulla, F., Rixen, D., Buschmann, T.: Fast Object Approximation for Real-Time 3D Obstacle Avoidance with Biped Robots. In: 2016 IEEE International Conference on Advanced Intelligent Mechatronics (AIM), pp. 38–45 (2016)
32. Brooks, G., Krishnamurthy, P., Khorrami, F.: A Multi-Gait Approach for Humanoid Navigation in Cluttered Environments. In: The 26th Chinese Control and Decision Conference (2014 CCDC), pp. 2708–2713 (2014)
33. Arias, L. E., Olvera, L. I., Pámanes, J. A., Núñez, J. V.: Patrón De Marcha 3D de Tipo Cicloidal para Humanoides y su Aplicación al Robot Bioloid. *Revista Iberoamericana de Ingeniería Mecánica*, 18(1), 3–22 (2014)
34. Fierro, J., Pámanes, J. A., Santibanez, V., Ruiz, G., Ollervides, J.: Condiciones para una Marcha Elemental del Robot NAO. *Amrob J. Rob. Theor. Appl.*, 4(1), 13–18 (2014)
35. De Robótica, A. M., Industria, A. C.: Movimientos de un Robot Humanoid para el Pateado de un Balón de Fútbol. *AMRob Journal, Robotics: Theory and Applications* 3(2) 105–112 (2015)
36. González, J. J. T., Magé, H. F. N., Rodas, C. F. R.: Simulación del Ciclo de Marcha del Robot Bioloid en el Entorno Virtual V-Rep. In: *Memorias del XVII Congreso Mexicano de Robótica 2015*, pp. 1–8 (2015)
37. Brenes, J. C., Blanes, F.: Interfaz de Programación de Aplicaciones para Locomoción de Robot Humanoid Bioloid. Tesis de maestría, Universidad Politécnica de Valencia (2016)
38. Cuperman, D., Verner, I. M.: Fostering Analogical Reasoning Through Creating Robotic Models of Biological Systems. *Journal of Science Education and Technology* 28(2), 90–103 (2018)
39. Qi, X., Chen, L., An, K., Wang, J., Zhang, B., Xu, K., Ye, D., Li C., Ran, L.: Bio-Inspired in-Grid Navigation and Positioning Based on an Artificially Established Magnetic Gradient. *IEEE Transactions on Vehicular Technology* 67(11), 10583–10589 (2018)
40. Somov, Y. I., Butyrin, S. A., Somov, S. Y.: Navigation, Guidance and Control of a Free-Flying Robot During its Rendezvous with a Passive Space Vehicle. In: 25th Saint Petersburg International Conference on Integrated Navigation Systems (ICINS), pp. 1–5. (2018)
41. Li, H., Savkin, A. V.: A Method for Collision Free Sensor Network Based Navigation of Flying Robots Among Moving and Steady Obstacles. In: 36th Chinese Control Conference (CCC), pp. 6079–6083 (2017)
42. Ames, A. D.: Human-Inspired Control of Bipedal Walking Robots. *IEEE Transactions on Automatic Control* 59(5), 1115–1130 (2014)
43. Vasudevan, R., Ames, A., Bajcsy, R.: Persistent Homology for Automatic Determination of Human-Data Based Cost of Bipedal Walking. *Nonlinear Analysis: Hybrid Systems* 7(1), 101–115 (2013)
44. Wen, S., Chen, X., Ma, C., Lam, H. K., Hua, S.: The Q-Learning Obstacle Avoidance Algorithm Based on EKF-SLAM for NAO Autonomous Walking under Unknown Environments. *Robotics and Autonomous Systems* 72, 29–36 (2015)
45. Nishiwaki, K., Chestnutt, J., Kagami, S.: Planning and Control of a Humanoid Robot for Navigation on Uneven Multi-Scale Terrain. In *Experimental Robotics*, pp. 401–415. Springer, Berlin, Heidelberg (2014)

46. Nolfi, S., Bongard, J., Husbands, P., Floreano, D.: Evolutionary Robotics. In: Springer Handbook of Robotics, pp. 2035–2068. Springer (2016)
47. Ibanez, A., Bidaud, P., Padois, V.: Automatic Optimal Biped Walking as a Mixed-Integer Quadratic Program. In: Advances in Robot Kinematics, pp. 505–516. Springer (2014)
48. Todd, D. J.: Walking Machines: An Introduction to Legged Robots. Springer Science & Business Media (2013)
49. Choi, Y. J., Kim, D. I., Oh, Y. H., Kim, C. H., You, B. J., San Cho, J.: U.S. Patent No. 8,738,178. Washington, DC: U.S. Patent And Trademark Office (2014)
50. Brasseur, C., Sherikov, A., Collette, C., Dimitrov, D., Wieber, P. B.: A Robust Linear MPC Approach to Online Generation of 3D Biped Walking Motion. In: 15th International Conference on Humanoid Robots (Humanoids), pp. 595–601 (2015)
51. Ha, E. T., Kim, T. K., Ahn, D. K., Jeong, S. H., Yoon, I. R., Han, S. H.: A Stable Control of Legged Robot Based on Ultrasonic Sensor. In: 15th International Conference on Control, Automation and Systems (ICCAS), pp. 1256–1258 (2015)
52. Hornung, A., Oßwald, S., Maier, D., Bennewitz, M.: Monte Carlo Localization for Humanoid Robot Navigation in Complex Indoor Environments. International Journal of Humanoid Robotics 11(2), 1–27 (2014)
53. Hildebrandt, A. C., Klischat, M., Wahrmann, D., Wittmann, R., Sygulla, F., Seiwald, P., Buschmann, T.: Real-Time Path Planning in Unknown Environments for Bipedal Robots. IEEE Robotics and Automation Letters 2(4), 1856–1863 (2017)
54. Mac Thi, T., Copot, C., Ionescu, C. M.: Design and Implementation of a Real-Time Autonomous Navigation System Applied to Lego Robots. In: 3rd IFAC Conference on Advances in Proportional-Integral-Derivative Control, pp. 340–345 (2018)
55. Tai, L., Zhang, J., Liu, M., Boedecker, J., Burgard, W.: A Survey of Deep Network Solutions for Learning Control in Robotics: From Reinforcement to Imitation. Arxiv Preprint Arxiv:1612.07139 (2019)
56. Hildebrandt, A. C., Demmeler, M., Wittmann, R., Wahrmann, D., Sygulla, F., Rixen, D., Buschmann, T.: Real-Time Predictive Kinematic Evaluation and Optimization for Biped Robots. In: International Conference on Intelligent Robots and Systems (IROS), pp. 5789–5796 (2016)
57. Hildebrandt, A. C., Wittmann, R., Wahrmann, D., Ewald, A., Buschmann, T.: Real-Time 3D Collision Avoidance for Biped Robots. In: International Conference on Intelligent Robots and Systems, pp. 4184–4190 (2014)
58. Wahrmann, D., Hildebrandt, A. C., Wittmann, R., Sygulla, F., Rixen, D., Buschmann, T.: Fast Object Approximation for Real-Time 3D Obstacle Avoidance with Biped Robots. In: IEEE International Conference on Advanced Intelligent Mechatronics (AIM), pp. 38–45 (2016)
59. Wittmann, R., Hildebrandt, A. C., Wahrmann, D., Sygulla, F., Rixen, D., Buschmann, T.: Model-Based Predictive Bipedal Walking Stabilization. In: Humanoids, pp. 718–724 (2016)
60. Wittmann, R., Hildebrandt, A. C., Ewald, A., Buschmann, T.: An Estimation Model for Footstep Modifications of Biped Robots. In: International Conference on Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ, pp. 2572–2578 (2014)
61. Wittmann, R., Hildebrandt, A. C., Wahrmann, D., Rixen, D., Buschmann, T.: State Estimation for Biped Robots Using Multibody Dynamics. In: International Conference on Intelligent Robots And Systems (IROS), 2015 IEEE/RSJ, pp. 2166–2172 (2015)
62. Wei, C., Xu, J., Wang, C., Wiggers, P., Hindriks, K.: An Approach to Navigation for the Humanoid Robot Nao in Domestic Environments. In: Conference Towards Autonomous Robotic Systems, pp. 298–310 (2013)

63. Nishiwaki, K., Chestnutt, J., Kagami, S.: Autonomous Navigation of a Humanoid Robot over Unknown Rough Terrain. In: *Robotics Research*, pp. 619–634 (2017)
64. Wen, S., Othman, K. M., Rad, A. B., Zhang, Y., Zhao, Y.: Indoor SLAM Using Laser and Camera with Closed-Loop Controller for NAO Humanoid Robot. In: *Abstract and Applied Analysis*, pp. 1–9 (2014)
65. Wahrmann, D., Knopp, T., Wittmann, R., Hildebrandt, A. C., Sygulla, F., Seiwald, P., Buschmann, T.: Modifying the Estimated Ground Height to Mitigate Error Effects on Bipedal Robot Walking. In: *International Conference on Advanced Intelligent Mechatronics (AIM)*, pp. 1471–1476 (2017)
66. Kita, N., Kanehiro, F., Morisawa, M., Kaneko, K.: Obstacle Detection for a Bipedal Walking Robot by a Fisheye Stereo. In: *International Symposium on System Integration (SII)*, pp. 119–125 (2013)
67. Wahrmann, D., Hildebrandt, A. C., Bates, T., Wittmann, R., Sygulla, F., Seiwald, P., Rixen, D.: Vision-Based 3D Modeling of Unknown Dynamic Environments for Real-Time Humanoid Navigation. *International Journal of Humanoid Robotics* 16(1), 1–34 (2019)
68. Kuo, C. H., Chou, H. C., Chi, S. W., Lien, Y. D.: Vision-Based Obstacle Avoidance Navigation with Autonomous Humanoid Robots for Structured Competition Problems. *International Journal of Humanoid Robotics* 10(3), 1–36 (2013)
69. Fakoor, M., Kosari, A., Jafarzadeh, M.: Revision on Fuzzy Artificial Potential Field for Humanoid Robot Path Planning in Unknown Environment. *International Journal of Advanced Mechatronic Systems* 6(4), 174–183 (2015)
70. Fakoor, M., Kosari, A., Jafarzadeh, M.: Humanoid Robot Path Planning with Fuzzy Markov Decision Processes. *Journal of Applied Research and Technology* 14(5), 300–310 (2016)
71. Goldhoorn, A., Garrell, A., Alquézar, R., Sanfeliu, A.: Continuous Real Time POMCP to Find-and-Follow People by a Humanoid Service Robot. In: *International Conference on Humanoid Robots (Humanoids)*, pp. 741–747 (2014)
72. Motlagh, O., Nakhaeinia, D., Tang, S. H., Karasfi, B., Khaksar, W.: Automatic Navigation of Mobile Robots in Unknown Environments. *Neural Computing and Applications* 24(7), 1569–1581 (2014)
73. Dash, T.: Automatic Navigation of Wall Following Mobile Robot Using Adaptive Resonance Theory of Type-1. *Biologically Inspired Cognitive Architectures* 12, 1–8 (2015)
74. Han, J. K., Ha, I. Y., Kim, B. S.: Educational Robotic Construction Kit: Bioloid. In: *17th World Congress of The International Federation of Automatic Control (IFAC 2008)*, pp. 6–10 (2008)
75. George, L., Mazel, A.: Humanoid Robot Indoor Navigation Based on 2D Bar Codes: Application to the NAO Robot. In: *13th International Conference on Humanoid Robots (Humanoids)*, pp. 329–335 (2013)
76. Pandey, A., Pandey, S., Parhi, D. R.: Mobile Robot Navigation and Obstacle Avoidance Techniques: a Review. *International Robotics & Automation Journal* 2(3), 1–12 (2017)
77. Meggiolaro, M. A., Neto, M. S., Figueroa, A. L.: Modeling and Optimization with Genetic Algorithms of Quasi-Static Gait Patterns in Planar Biped Robots. In: *Congreso Internacional de Ingeniería Mecatrónica y Automatización (CIIMA 2016)*, pp. 1–10 (2016)
78. Segura, C. G., García, M. G., Díaz, L. N., Pech, R. B.: Robot Bioloid Premium Jaranero Controlado Remotamente por Voz. *Pistas Educativas* 39(126), 99–127 (2017)
79. Fathinezhad, F., Derhami, V., Rezaeian, M.: Supervised Fuzzy Reinforcement Learning for Robot Navigation. *Applied Soft Computing* 40, 33–41 (2016)
80. Furuta, Y., Wada, K., Murooka, M., Nozawa, S., Kakiuchi, Y., Okada, K., Inaba, M.: Transformable Semantic Map Based Navigation Using Autonomous Deep Learning Object

- Segmentation. In: 16th International Conference On Humanoid Robots (Humanoids), pp. 614–620. (2016)
81. Lobos-Tsunekawa, K., Leiva, F., Ruiz-Del-Solar, J.: Visual Navigation For Biped Humanoid Robots Using Deep Reinforcement Learning. *IEEE Robotics and Automation Letters* 3(4), 3247–3254 (2018)
 82. Konar, A., Chakraborty, I. G., Singh, S. J., Jain, L. C., Nagar, A. K.: A Deterministic Improved Q-Learning for Path Planning of a Mobile Robot. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 43(5), 1141–1153 (2013)
 83. Algabri, M., Mathkour, H., Ramdane, H., & Alsulaiman, M.: Comparative Study of Soft Computing Techniques for Mobile Robot Navigation in an Unknown Environment. *Computers in Human Behavior* 50, 42–56 (2015)
 84. Ran, L., Zhang, Y., Zhang, Q., Yang, T.: Convolutional Neural Network-Based Robot Navigation Using Uncalibrated Spherical Images. *Sensors* 17(6), 1–18 (2017)
 85. Bischoff, B., Nguyen-Tuong, D., Lee, I. H., Streichert, F., Knoll, A.: Hierarchical Reinforcement Learning For Robot Navigation. In: *Proceedings of The European Symposium on Artificial Neural Networks, Computational Intelligence And Machine Learning (ESANN 2013)*, pp. 227–232 (2013)
 86. Bacciu, D., Gallicchio, C., Micheli, A., Di Rocco, M., Saffiotti, A.: Learning Context-Aware Mobile Robot Navigation in Home Environments. In: *5th International Conference on Information, Intelligence, Systems and Applications*, pp. 57–62 (2014)
 87. Okal, B., Arras, K. O.: Learning Socially Normative Robot Navigation Behaviors with Bayesian Inverse Reinforcement Learning. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2889–2895 (2016)
 88. Pandey, A., Parhi, D. R.: Multiple Mobile Robots Navigation and Obstacle Avoidance Using Minimum Rule Based ANFIS Network Controller in the Cluttered Environment. *International Journal of Advanced Robotics and Automation* 1(1), 1–11 (2016)
 89. Xia, C., El Kamel, A.: Neural Inverse Reinforcement Learning In Autonomous Navigation. *Robotics and Autonomous Systems* 84, 1–14 (2016)
 90. Villacorta-Atienza, J. A., Makarov, V. A.: Neural Network Architecture for Cognitive Navigation in Dynamic Environments. *IEEE Transactions on Neural Networks and Learning Systems* 24(12), 2075–2087 (2013)

Obstacle Detection and Trajectory Estimation in Vehicular Displacements based on Computational Vision

Lauro Reyes-Cocoletzi, Ivan Olmos-Pineda, J. Arturo Olvera-López

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación, Puebla, Mexico
lauro.reyesc@alumno.buap.mx, {iolmos,aolvera}@cs.buap.mx

Abstract. Obstacle detection and trajectory estimation in vehicular environments is an open problem in autonomous vehicles development. The automotive industry has made significant progress in research and development of tools; however, there are still challenges to overcome and opportunity areas to be exploited in order to achieve full autonomy in vehicles. This paper presents an analysis of different methods proposed for obstacle detection and trajectory estimation, leading into a proposal of solution for solving the problem of trajectory estimation based on computer vision techniques. This proposal covers the context of traffic environments in Latin America, where basic signage (such as the dividing lines of the road) is absent or not easily visible, among other typical characteristics of countries (like Mexico and others) where the infrastructure for maintaining safe road conditions (speedbumps, potholes) is limited.

Keywords: Obstacle Detection, Trajectory Estimation, Autonomous Vehicles, Computer Vision.

1 Introduction

Over the years, the automotive industry has made efforts to provide vehicles with autonomous displacement (autonomous taxis, autonomous private cars). Recently, cars have been equipped with Driver Assistance Systems (ADAS), but their applications are limited for specific conditions and controlled environments; therefore, research in this area is incomplete. A vision-based DAS consists of modules for obstacle detection and recognition, obstacle tracking and displacement prediction [13].

It is important to clarify that the meaning of an obstacle for the purposes of this research is: an above-ground object that presents a collision risk when the vehicle is traveling along a road. Note that the basic functions for autonomous movement are obstacle detection and tracking in traffic environments.

The information provided by the automotive industry is limited [2] due to the commercial implications. However, it should be noted that research work

has been carried out not only in private business, but also in universities and academic research centers.

Some limitations of research into autonomous displacement are: low performance in terms of detection rate [3], a limited number of objects perceived as obstacles (pedestrians and vehicles only) [6], limited information acquired from the environment (sensors) [13], the high cost of sensors, in some cases (laser radar), and high processing runtimes [1], among other issues.

Video cameras commonly provide information about the environment, these devices are low cost, do not emit signals into the environment (electromagnetic or sonic radiation), and these do not suffer from frequency interference. Video cameras are used in this research due to the aforementioned advantages over other sensors (sonic or laser radar systems), additionally, using a stereoscopic computational system provides spatial information from the environment (depth and position of objects).

This paper is organized as follows: Section 2 describes some methods performed in academic research centers, Section 3 shows the results obtained in related research works. Section 4 discusses an initial stage of the proposed solution. Section 5 presents some initial experiments with respect to the proposal, Section 6 briefly analyzes the main problem of tracking objects in realistic scenarios. Section 7 offers a brief conclusion about the state of the research and the work to be developed.

2 Detection and Estimation of Obstacle Trajectory

In general, methods that address the problem of obstacle trajectory detection and estimation are diverse and depend on the type of sensor used and information obtained from the environment. The most common sensors are ultrasonic, radars, laser radars and video cameras (stereoscopic or monocular).

Due to the amount of information obtained from a digital image (color, texture, resolution, spatial position, etc.), computational vision techniques are similar to environmental human perception [13].

For these reasons, it was decided to work with the information captured on video in order to overcome the detection rates of recent works compared against the use of other types of sensors (laser radars).

There are several strategies for detecting and monitoring obstacles via computational vision; some of these methods are mentioned briefly in following paragraphs.

Algorithms based on convolutional neural networks [8] for recognition of cars and pedestrians classifies these into three collision categories (high risk, low risk, medium risk) by means of a deep learning system that considers multiple sources of local pattern information and depth in the scene.

Techniques based on stereo vision produce a dense disparity map, which is used as the input for adaptive disparity algorithms U-V [11] when mapping the regions of interest (ROI) and the consequent in-scene spacial obstacle localization.

Evolutionary algorithms (EA) aim to detect multiple obstacles in a video through the use of a genetic regression algorithm (GA) and prevent false detection through the perception of movement [15]. This basis method is based on the operation of the fitness function with the heuristic parameters, which means choosing the optimal fitness parameters using EA and the least squares method.

Table 1 presents some reference works in this area: the methods implemented to solve the problem are reported, as well as the test sets, the hardware, the software used, and the results obtained, such as detection rate (DR), mean average precision (mAP), frames per second (fps), or multi-objective tracking accuracy (MOTA).

Table 1. Quantitative results and tools used in some related works about obstacle detection.

Reference	Method	Technique (detection / tracking)	Quantitative results	Test set	Software hardware
[17]	Stereo map	-Disparity maps -Stereo comparison -Local binary patterns (LBP) -Gradient histogram	-DR = 82.5 % -mAP = 79.5 -fps = 14	-KITTI Benchmark [5] -HCI EISAT	-CPU xeon 2.67 GHz -Tesla K40 GPU
[20]		- Conditional random field -Trajectory model	-MOTA = 58.9 % -Recall = 69.2 % -Accuracy = 88 %	-Training: 490 frames 1,578 pedestrians labeled resolution 640 x 480 pixels	-nVidia GeForce 8800
[23]	CNN	-LBP -K-medias -Classifiers -Local trust patterns	(Optimal lighting conditions) -DR = 77 % -Fault alarm rate (FA) FA = 1.25 % (variable lighting conditions) -DR = 50.8 %, FA = 1 % -mPA = 63.4 % , fps = 45	-KITTI Benchmark -Daimler dataset: 15,560 pedestrian samples -21,790 images with 5692 pedestrians labeled	-CPU Intel-xeon -Tesla K40 GPU
[6]		-Fast RCNN -Polynomial prediction	-Regions of trust with 95 % -Risk of estimated collision higher than 50 % - Estimated Travel to 8.6 m at a speed 50 km / hr -50 fps	-150 pedestrian paths -They are divided into three different classes: Begin: consists of 50 scenes of approx. 1 sec Stopped: consists of 50 scenes, 1 sec Walking: consists of 50 movements	2 Monochromatic cameras resolution 1920x1080 pixels
[10]	Deep learning	-Progressive convolutional layers -Decoding stage	-DR =81.5 to 90.6 % depending on the class -mAP = 78.6 % -Runtime 110 ms average	-PASCAL VOC 2007 -40 k iterations -COCO benchmark	-GPU TitanX -Caffe
[18]		-Deconvolutional detectors	-MOTA = 11.8 % -fps = 87.9	-DETRAC 60 sequences of training. -40 test sequences.	-AlexNet -Caffe -Titan Black GPU -iiRAV
[15]	Evolutionary algorithms	-Genetic algorithm -Function fitness -Hypothesis of generation (HOG)	-Value fitness = 0.1216 to 0.4315 -Processing time = 48 ms -DR = 96.08 %	-3025 images captured in a controlled set and another real environment vel. from 20 to 70 km/hr	-FPGA xilinx

Table 1 also shows important research parameters for each technique; it can be seen that the detection rate is close to 100% in some cases, however, the total number of obstacles detected must be taken into consideration, as well as the class or type (pedestrian, vehicle, animals, etc.), which in many cases only detect limited classes. To track the obstacles, even the MOTA parameter is below 100 % as the error with respect to the next position or movement in

the video increases due to the complexity and quantity of data to be processed, the algorithm processing speed, the object's movement speed, and the number of objects present on the scene among others.

In order to test the approaches proposed in the literature, some well-known datasets are used, such as the KITTI Vision Benchmark suite from the Karlsruhe Institute of Technology, datasets are captured in urban areas on highways, and up to 15 cars and 30 pedestrians are visible per image [5].

The PASCAL Visual Object Classes Challenge also has data blocks available to researchers [4], while the MOT challenge benchmark provides ready to use information [12]. COCO is a large-scale object detection, segmentation, and labelling dataset [9], the UA-DETRAC Benchmark Suite is a multi-object tracking benchmark with a dataset consisting of 10 hours of surveillance cameras video at 24 different locations in Beijing, China [7].

Depending on the method to be used, there are some limitations; a summary is provided in the following section.

3 Analysis of Related Works

The method proposed in the literature has some positive aspects, but also some limiting factors. Information about the deficiencies found in related works is covered below, including detection, tracking, and estimation issues.

With respect to the mapping ROI models: the problem of trajectory prediction has areas of opportunity such as detection failures, objects with similar appearances, occlusions, and variations in illumination and points of view [11, 20]. Only two kinds of obstacles (vehicles and pedestrians) are taken into consideration [3, 11, 17, 20, 19]. High processing runtime rate and low obstacle detection rate [3, 14]. Monocular mapping (ROI), in some cases acquires data through static surveillance cameras [11].

For the Probability models case: trajectory estimation methods with a high uncertainty have problems with complexity (with many samples) or low accuracy (with an insufficient number of samples) [1, 19, 22]. Kalman filter algorithms and particle filters used for the purpose of tracking obstacles suffer from drift problems caused by changes in appearance [19].

Deep learning models have the following characteristics: they only consider characteristics such as color or depth, which implies a limitation to obtaining a higher level of abstraction of the representative characteristics of the obstacles [23, 18, 21]. Tracking methods based on online-classifiers learning suffer the problem of error accumulation during the self-learning process [21, 16]. Some algorithms based on the convolutional neural networks approach only capture translational invariances and not the rotational invariance or out-of plane rotation, which makes them susceptible to error when classifying and identifying obstacles [6, 23, 16].

Evolutionary algorithm models tend to fail in object classification, as they confuse similar forms, which implies an erroneous obstacle detection, for example, detecting a car in the scene when it is not in fact in the image [15].

The general stages that an algorithm must follow in order to solve the problems of obstacle detection, classification, and tracking as well as of trajectory prediction, obtained from related works, are presented in Table 2 below.

Table 2. Problem stages addressed by different authors.

#	Stage	Paper reference															
		[3]	[5]	[7]	[10]	[11]	[12]	[14]	[15]	[16]	[17]	[18]	[19]	[20]	[21]	[22]	[23]
I	Selection ROI	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
II	Spatial position estimation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
III	Object detection	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
IV	Feature extraction	Training							✓			✓	✓				✓
		Neighborhood Mapping		✓		✓	✓			✓	✓			✓			
V	Classification	Specific class										✓		✓		✓	
		No obstacle	✓	✓	✓		✓	✓	✓	✓	✓		✓				
VI	Rank (danger level)			✓		✓	✓	✓	✓				✓			✓	
VII	Tracking	Multiple obstacle			✓									✓			
		Unique obstacle	✓			✓				✓		✓	✓				
VIII	Trajectory estimation			✓								✓				✓	

From these stages and the associated review of the works, we can obtain an approximation of the strengths of the work done until now and the work to be solved. In the table 2 we can see the distribution of related works and points (stages) that have less development progress. Based on this information, it can be affirmed that the pending tasks are tracking and trajectory estimation.

Obstacle classification has been developed and has reported rates greater than 90%; however, it still requires a class for obstacles other than pedestrians and vehicles, this is due to the fact that peculiar obstacles can be found on the roads. Figure 1 shows obstacles that are common to find on Latin American roadways, but difficult to identify with current obstacle recognition/detection models.



Fig. 1. Obstacles present in road, a) road sign in sewer, b) light pole in the middle of the street, c) fallen tree blocking the street, d) and tires in sewer.

In order to observe the performance of obstacle detection in Latin American environments, a test was performed with a deep learning algorithm (designed by Rendom and Divvala [16]).

Figure 2 shows that the algorithm cannot detect the tire or the road barrier as obstacles blocking the road. To date, the different models and computational techniques have failed to detect these types of obstacles in the images.

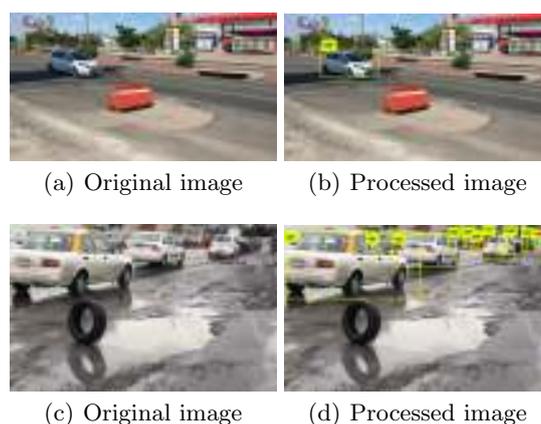


Fig. 2. Original and processed images for obstacle detection.

This is due to the fact that in some cases the algorithm focuses on detecting the objects (automobile) in the scene through previously known and defined characteristics. However, objects dis-similar to the known detection characteristics are not detected or identified.

In the following paragraphs the contributions that can be made to the research problem are discussed; this is a challenge due to the amount of information to be processed and the complexity of the tasks to be performed:

Multiple-obstacle tracking. The inference method approach for trajectory and the on-scene prediction of multiple obstacles based on a Gaussian probability distribution. This is possible due to roads being visualized as a structured environment, where the driver's infinite number of possible movements can be approximated by a limited number of maneuvers.

Block trajectory estimation. Trajectory detection and estimation in unstructured vehicular environments, that is, cities with problems such as potholes, a lack of signs, speedbumps without height allowed, abrupt lane changes by drivers, jaywalking pedestrians, cyclists or motorcyclists cutting lanes, or animals crossing the road, among others.

Feature extraction. Occlusion when tracking real scenarios in traffic environments. Training of the learning model should include some sets with partially hidden objects that consider the disparity map to complete the missing information.

Classification. Characterize cars, pedestrians, cyclists, animals or (foreign) objects that obstruct the road into-pattern classes. On Latin American roadways it is common to find animals crossing or traveling next to vehicles, cars cutting

lanes, cyclists ridding very close to cars, or even tires or trash cans to signal a damaged manhole.

4 Obstacle Detection and Trajectory Estimation Proposal

The initial proposed system is structured into stages; the flowchart of the stages is provided below (Figure 3) and a description of each is provided in the following paragraphs.

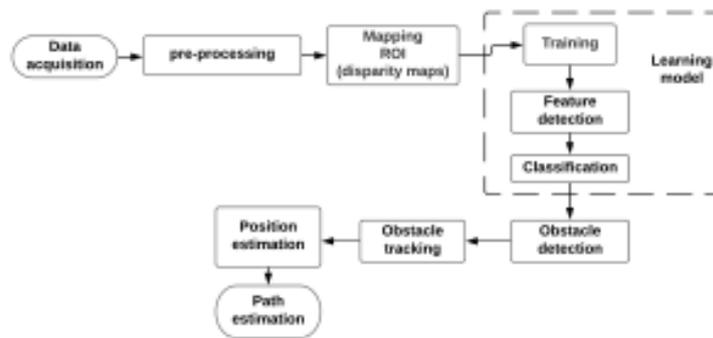


Fig. 3. Flow diagram method for obstacle detection and trajectory estimation.

The initial step is environment perception by means of a pair of cameras (stereoscopic vision) together with the preprocessing.

Among the typical preprocessing stages used to improve image quality are the application of noise filters, contrast enhancement, histogram equalization, and image scaling.

The fundamental step of the detection stage is to perform disparity mapping to simplify the information to be processed from the RGB channel intensity levels in the image. Disparity mapping [17] is used to obtain basic characteristics for the representation of objects in a scene in complex vehicular environments, it also has the advantage of providing spatial depth information in the scene.

The ROI obtained in the disparity map is filtered by means of edge detection (Shi-Thomasi algorithm), horizontal or vertical pattern detection (Haar algorithm) and even LBP [13] to obtain better defined regions and disparity maps.

The subsequent stages include processing based on learning models, by means of which it is possible to classify obstacles perceived in the scene, and identify obstacle classes (it is proposed to first identify cars, pedestrians, animals, and consider additional kinds of obstacles in a future work).

The learning system includes a training block, by means of a learning database. In training, abstract features must be learned in the manner least

sensitive to rotational invariance, out-of-plane rotation, and reduced sensitivity to noise, that is, a deep learning system [8].

The training block aims to provide a sample set of instances identified to learn and classify the characteristics in specific categories such as: cars, pedestrians or dogs. Prior to this, an exploration process is carried out in order to determine if the analyzed object is an obstacle.

The proposal model implies representation of the elements with n -levels of abstraction to determine the n -representative characteristics of these elements. These parameters are used to define the number of instances (database). For example, in the case of cars body color, shape, or texture or some other characteristics might be considered. In the case of pedestrians, there is skin color, posture, body shape (2 upper extremities, 2 lower extremities), etc. as well as other yet to be identified obstacles.

The obstacles must be followed from the n -frame to the $n+1$ frame and on the consecutive frames of the video. The optical flow algorithm [3] is proposed in order to solve the tracking block, using the U-V disparity information. It is worth mentioning that Kalman filter and particle filter algorithms have been successfully used to track vehicles, although they present performance drawbacks (as noted in section 3).

Finally, current position estimation (in the image) and trajectory estimation are connected; these can be denoted as the algorithms whose task is to track multiple objects. The novel proposal to solving the trajectory estimation problem is through tracking by detection in order to calculate the estimation of the next position of the obstacles in vehicular traffic. This was first addressed as a problem of estimation probability; the object trajectories can be considered as a group of points moving simultaneously (in space and time). Based on the previous position and route information followed by the tracking stage, a trajectory hypothesis was established. The trajectory hypothesis was parameterized with a Bayesian network and a proposed mathematical model based on statistical probability distribution (a trajectory estimation model is in development).

5 Experimental Results

Experiments are performed to obtain initial results with respect to the proposal and determine its feasibility.

The stages carried out in the experiments include the object detection stage (disparity maps experiment) and the object tracking stage (optical flow experiment).

The disparity method is a first approximation of the objects in the scene; however, it is necessary to first detail the obstacle detection results. Figure 4a shows the original image, while Figure 4b shows the disparity map and the values calculated with the stereo correspondence. Blue tones indicate objects at the back of the scene and yellow tones those closer to the camera. The matrix of values obtained are from 0 to 3.5 units. Value 3.5 represents the volume ratio of the objects with respect to the background (zero). Figure 4c shows a view of the

disparity map that serves to observe the regions in 3d; this volume is constructed using the data obtained, which allows segmenting the background and objects. It is possible to observe scene-depth levels of the objects present with respect to the background. In addition, some undesirably high values can be observed, which must be eliminated (maximum peaks).

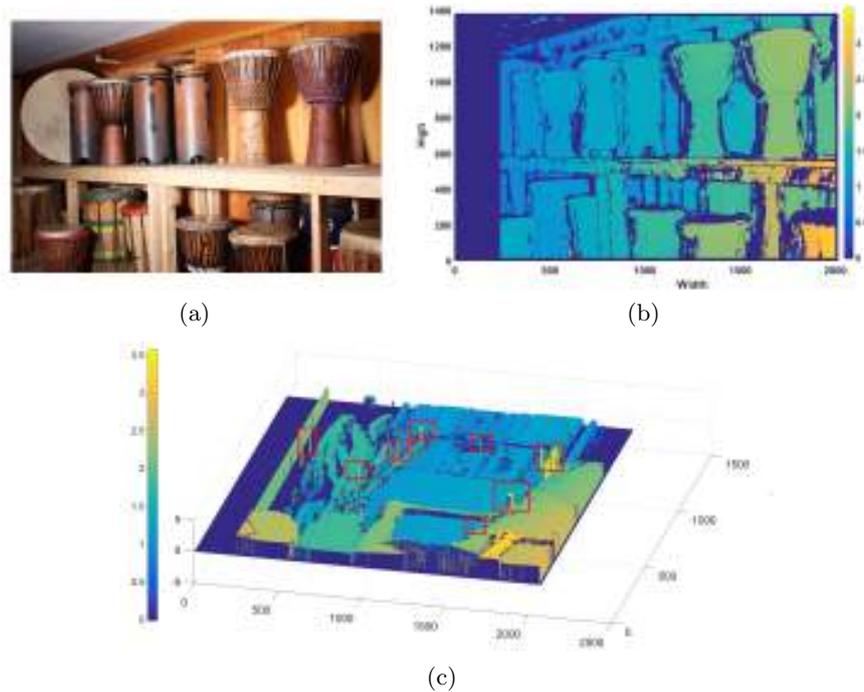


Fig. 4. a) Original image, b) dense disparity map, and c) side view disparity map.

Maximum peaks values are presented in figure 5, based on the number of peaks to the number of pixels, error estimation can be calculated and the disparity maps obtained. With the number of maximum peaks found against the total number of values obtained, a percentage of error in obtaining the disparity maps can be found:

$$Error_rate(\%) = \frac{\# \text{ maximum peaks}}{\# \text{ values obtained}} \times 100. \quad (1)$$

Table 3 shows some data regarding this error, these errors (maximum peaks) can be eliminated by means a cutting filter (from atypical data). Otherwise, the selectable regions of interest must be composed of a block of pixels sufficiently large to discard this type of outlier.

At this stage, the actual distance of the objects in relation to the data obtained from the disparity map still needs to be parameterized.

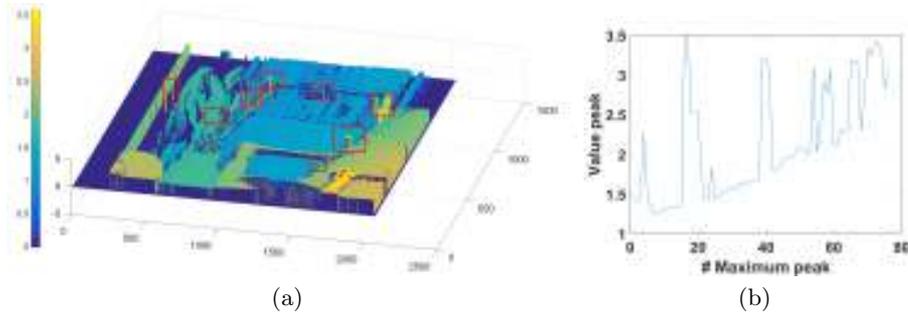


Fig. 5. a) Disparity map (peaks in red rectangle) and b) maximum unwanted peaks in the calculations.

Table 3. Analysis of the data obtained disparity maps.

# Disparity map	1	2	3	4	5	6	7	8	9	10	11
# Values obtained	971	1013	1589	1650	1037	860	600	1546	1358	1900	901
# Maximum peaks	61	76	210	150	89	45	89	75	156	341	100
Error_rate (%)	6.28	7.50	13.22	9.09	8.58	5.23	14.83	4.85	11.49	17.97	11.10
Average error (%)											10.01
Standard deviation error (%)											4.15

The next experiment consists of tracking the groups of pixels that move together (representing objects) and plotting points from the previous positions for a standard representation of the dense motion (figure 6a). Figure 6b-c shows a traffic scene captured in the real environment of a typical street in Puebla City, Mexico, where public transport, cyclists and vehicles can be observed. Every pixel is labelled by a motion vector, indicating the change in image data from time t to time $t+1$, in such a way that shows the trajectory taken by the detected pixels of interest.



Fig. 6. a) Theoretical trajectory [1], b), and c) tracing approximate trajectory points using an optical flow algorithm in a real vehicular environment.

It is necessary to define obstacles regions to be followed, since the pixel displacement information (analyzed in this experiment) provides only a preliminary approximation of local regions in the image where the obstacles must

be detected. This involves reference corner detection (Shi–Tomasi algorithm) and tracking of same (Lucas–Kanade optical flow algorithm) [12].

The characteristics to take into consideration in the process are the following. Shi–Tomasi corner detection parameters:

- Maximum number of corners.
- Block size: size of the neighborhood considered for corner detection.
- Minimal distance: distance between detected corners.

Lucas–Kanade optical flow parameters:

- Max level: 0-based maximum pyramid level number.
- Window size: size of the search window at each pyramid level.
- Count: maximum number of iterations or elements to compute.

The tracking rate is below 50% given that regions of interest for object detection are not yet limited.

Table 4. Object tracking parameters.

Shi-Thomasi corner detector	# Max corners	Block size	Minimum distance (# pixels)
	50 to 100	7 to 18	7 to 50
Lucas-Kanade optical flow	# Max level	Window size	Count
	3	15x15	12

The work to be done is to implement a form descriptor that adapts to any type of irregular object present on the road that represents an obstacle to displacement, so that whatever the obstacle while having volume will be detected. The features to extract (object descriptors) must be rotation invariant i.e., to the non-linear deformation transformation (perspective). In addition, they must be discriminant in order to distinguish objects of different classes and robustness levels so that they are insensitive to variations in noise and illumination.

Figure 7 shows the common descriptors used for the basic representation of objects.



Fig. 7. Representation of shape descriptors for the detection of objects a) cloud points, b) polygonal approximations, c) bounding box, and d) skeleton.

The typical databases used in the related works are being considered for verification and validation of initial results [5, 4, 12, 9, 7]; however, related works consider controlled environments, so the contribution focuses on uncontrolled environments, such as those found in Latin America.

Finally, these experiments serve as a basis for determining the requirements in terms of the amount of data to be processed, the scenes to be taken into consideration, the infrastructure of the video capture equipment, the traffic schedules, routes, and safety conditions when gathering the data. Similarly, it should be noted that the lighting conditions or the number of vehicles or objects on the route of travel are not controlled.

6 Discussion

In this section, the main problems that arise when tracking objects in realistic scenarios are briefly analyzed. These include the location of objects as they undergo occlusion and the constant tracking of unique object trajectories as they are viewed through multiple cameras.

Occlusion. The partial occlusion of an object by a structure in the image is difficult to detect, since it is not always possible to differentiate between the object changing shape and the becoming occluded. A common approach to handling complete occlusion during tracking is to model the object's motion by either linear or nonlinear models and, in the case of occlusion, to continue predicting the object's location until it reappears [11, 20]. The chance of occlusion can be reduced by means of an appropriate selection of camera positions. For instance, if the cameras are mounted on vehicles, that is, when a birds-eye view of the scene is available, occlusions between objects on the ground occur less frequently.

Stereo camera tracking. The need for multiple cameras for tracking arises for two reasons. The first is the use of depth information for tracking and occlusion resolution [3]. The second an increase to the area under view, since it is not possible for a single camera to observe large areas because of a finite sensor field of view [11]. Performance depends greatly on how closely the objects follow the established paths and the expected time intervals across cameras [17]. For scenarios in which spatio-temporal constraints cannot be used, for example, objects moving arbitrarily in the non-overlapping region, only the tracking by recognition approach can be employed, which uses the appearance and the shape of the object to recognize it when it reappears in the camera's view.

Significant progress has been made in object tracking over the last few years, and several robust algorithms have been developed which can track objects in real time in simple scenarios. However, it is clear that the assumptions used to make the tracking problem manageable, for example, smoothness of motion, minimal occlusion, illumination constancy, high contrast with respect to background, etc., are violated in many realistic scenarios and therefore limit a trackers usefulness in vehicle navigation applications. Thus, tracking and associated feature selection, object representation, dynamic shape, and

motion estimation problems are very active research areas and new solutions are continuously being proposed.

7 Conclusions

This paper briefly describes the work related to obstacle detection in real environments, presents a general description, analyzes the existing techniques and how they have proposed to solve the problem.

The proposed model for estimating trajectories through probability methods is focused on determining the future displacement of the detected obstacles within a confidence range. The proposed solution establishes a theoretical overview of the solution; the work to be done is extensive and the solution model is in development. Future work involves implementing the model stages and obtaining performance parameters to compare with the existing research.

From this work, it can be concluded that the problem of obstacle detection is an area of significant interest and that there is still room for improvement and opportunities for innovation. There is much research to be done and it is worth continuing with the development of techniques and tools to solve this problem. A large amount of data must usually be processed, so the computing aspects (memory, processing speed) must be robust to meet the needs of operating close to real time.

Acknowledgments. The first author thanks the support provided by the CONACYT scholarship number 708553.

References

1. Barth, A., Siedemund, J., Meißner, A., Franke, U., Förstner, W.: Probabilistic multi-class scene flow segmentation for traffic scenes. In: Joint Pattern Recognition Symposium. pp. 503–512. Springer (2010)
2. Dikmen, M., Burns, C.M.: Autonomous driving in the real world: Experiences with tesla autopilot and summon. In: Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications. pp. 225–228. ACM (2016)
3. Dur, E.: Optical flow-based obstacle detection and avoidance behaviors for mobile robots used in unmanned planetary exploration. In: 4th international conference on recent advances in space technologies. pp. 638–647. IEEE (2009)
4. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. *International journal of computer vision* 88(2), 303–338 (2010)
5. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research* 32(11), 1231–1237 (2013)
6. Gidaris, S., Komodakis, N.: Object detection via a multi-region and semantic segmentation-aware cnn model. In: Proceedings of the IEEE international conference on computer vision. pp. 1134–1142 (2015)

7. Lee, B., Erdenee, E., Jin, S., Nam, M.Y., Jung, Y.G., Rhee, P.K.: Multi-class multi-object tracking using changing point detection. In: European Conference on Computer Vision. pp. 68–83. Springer (2016)
8. Li, P., Mi, Y., He, C., Li, Y.: Detection and discrimination of obstacles to vehicle environment under convolutional neural networks. In: Chinese Control And Decision Conference (CCDC). pp. 337–341. IEEE (2018)
9. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollar, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
10. Lu, K., An, X., Li, J., He, H.: Efficient deep network for vision-based object detection in robotic applications. *Neurocomputing* 245, 31–45 (2017)
11. Mane, S.B., Vhanale, S.: Real time obstacle detection for mobile robot navigation using stereo vision. In: International Conference on Computing, Analytics and Security Trends (CAST). pp. 637–642. IEEE (2016)
12. Milan, A., Leal-Taixé, L., Reid, I., Roth, S., Schindler, K.: Mot16: A benchmark for multi-object tracking. arXiv preprint arXiv:1603.00831 (2016)
13. Mukhtar, A., Xia, L., Tang, T.B.: Vehicle detection techniques for collision avoidance systems: A review. *IEEE Transactions on Intelligent Transportation Systems* 16(5), 2318–2338 (2015)
14. Nguyen, V.D., Nguyen, D.D., Lee, S.J., Jeon, J.W.: Local density encoding for robust stereo matching. *IEEE Transactions on Circuits and Systems for Video Technology* 24(12), 2049–2062 (2014)
15. Nguyen, V.D., Nguyen, T.T., Nguyen, D.D., Lee, S.J., Jeon, J.W.: A fast evolutionary algorithm for real-time vehicle detection. *IEEE Transactions on Vehicular Technology* 62(6), 2453–2468 (2013)
16. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779–788 (2016)
17. Roggeman, H., Marzat, J., Derome, M., Sanfourche, M., Eudes, A., Le Besnerais, G.: Detection, estimation and avoidance of mobile objects using stereo-vision and model predictive control. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2090–2099 (2017)
18. Sangineto, E., Nabi, M., Culibrk, D., Sebe, N.: Self paced deep learning for weakly supervised object detection. *IEEE Transactions on pattern analysis and machine intelligence* 41(3), 712–725 (2018)
19. Schulz, J., Hubmann, C., Löchner, J., Burschka, D.: Multiple model unscented kalman filtering in dynamic bayesian networks for intention estimation and trajectory prediction. In: 21st International Conference on Intelligent Transportation Systems (ITSC). pp. 1467–1474. IEEE (2018)
20. Song, W., Yang, Y., Fu, M., Qiu, F., Wang, M.: Real-time obstacles detection and status classification for collision warning in a vehicle active safety system. *IEEE Transactions on intelligent transportation systems* 19(3), 758–773 (2017)
21. Wang, J.G., Zhou, L., Pan, Y., Lee, S., Song, Z., Han, B.S., Saputra, V.B.: Appearance-based brake-lights recognition using deep learning and vehicle detection. In: Intelligent Vehicles Symposium (IV). pp. 815–820. IEEE (2016)
22. Yoon, J.H., Yang, M.H., Lim, J., Yoon, K.J.: Bayesian multi-object tracking using motion context from multiple objects. In: Winter Conference on Applications of Computer Vision. pp. 33–40. IEEE (2015)
23. Yu, H., Hong, R., Huang, X., Wang, Z.: Obstacle detection with deep convolutional neural network. In: Sixth International Symposium on Computational Intelligence and Design. vol. 1, pp. 265–268. IEEE (2013)

Speed Bump Detection on Roads using Artificial Vision

Ana L. Ballinas-Hernández, Ivan Olmos-Pineda, J. Arturo Olvera-López

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación, Puebla, Mexico
analuisa.ballinas@correo.buap.mx, {iolmos, aolvera}@cs.buap.mx

Abstract. In recent decades, self-driving has been a topic of wide interest for Artificial Intelligence and the Automotive Industry. The irregularities detection on road surfaces is a task with great challenges. In developing countries, it is very common to find un-marked speed bumps on road surfaces which reduce the security and stability of self-driving cars. The existing techniques have not completely solved the speed bump detection without a well-marked signaling. The main contribution of this work is the design of a methodology that use a pre-trained convolutional neural network and supervised automatic classification, by using the analysis of elevations on surfaces through stereo vision, for detect well-marked and no well-marked speed bumps to improve existing techniques.

Keywords: Self-Driving Cars, Speed Bump Detection, Artificial Vision, Stereo Vision, Machine Learning.

1 Introduction

In recent years, the trend of the automotive industry has been focused on the development of self-driving cars, and several companies have been incorporating technologies aimed at achieving self-driving. One example of this is the cars that park without the need of a human conductor.

Self-driving car competitions have been taking place, such as the DARPA Grand Challenge and the DARPA Urban Challenge, organized by the Defense Advanced Research Projects Agency. In these competitions, cars start from a designated origin with the aim of reaching a specific destination, then crossing diverse extreme environments without a human driver [3]. Derived from these, several research topics in this area have emerged.

In countries like the United States, Japan, Germany, and others, self-driving cars have begun to circulate in trial versions. Self-driving cars have worked properly in environments with acceptable conditions where, through their sensors, they clearly detect road signals and different elements of road traffic. Nevertheless, the circulation of these cars in developing countries has not been possible due to lack of an adequate infrastructure and road signals.

In developing countries, such as Mexico and India, different sorts of irregularities are common on road surfaces. One of such irregularities being the presence

of high relief speed bumps, which are placed on roads with the purpose of forcing cars to reduce their speed in specific areas, which leads to potential accidents as well as loss of stability. Despite the existence of traffic rules, it is very common to find un-marked speed bumps. The techniques of digital image processing have been focused mainly on the speed bump detection when their traffic signals are well-marked. However, it is difficult to apply these techniques when there is no clear road marking, and therefore some works are being developed to generate algorithms that can address this issue.

This paper presents a methodology for speed bump detection on road surfaces without the need of well-marked signals. The proposal includes a combination of artificial vision techniques and the use of digital cameras. Specifically, digital image processing, stereo vision, machine learning and a convolutional neural network are used to infer whether or not speed bumps are present in the images. Inputs are stereo images of vehicular transit scenarios acquired through digital cameras and the output is the classification of images where it is identified whether or not there are speed bumps.

2 Related Work

There are some aspects that have been studied from vehicular transit scenarios such as: identification of streets, lanes, traffic signs and pedestrian crossings, and irregularities on road surfaces such as: garnishes, urban objects, potholes and marked and un-marked speed bumps. This work focuses on the speed bump detection for which some techniques have been applied: digital image processing, monitoring with accelerometers, LIDAR sensors, stereo vision, deep learning and hybrid techniques.

Some methods of digital image processing have been used for the speed bump detection [5, 6, 8, 32, 30]. Analysis of color patterns and segmentation of interest areas (ROI) for the detection of speed bumps have been used. The main advantage of segmentation algorithms is that they use inexpensive cameras, however, the accuracy in the obstacles detection on surfaces is low, because it is subject to the visualization of obstacle patterns and if the speed bump does not have a well-marked pattern it is not easily detectable. In addition, it is not uncommon not find that there is not a standard regarding the signals correspondent to speed bumps, which complicates the application of the same vision technique to all types of patterns.

Correspondingly, digital image processing has been used in the detection of different vertical traffic signals [11]. Specifically, segmentation algorithms and automatic learning methods are applied for the recognition of preventive signs for speed bumps [19, 16]. In these works, different aspects that affect the detection are considered as: the environment weather conditions, vision angle, sizes and shapes of the traffic signs. The downside of using these detectors is that the signals do not indicate a precise location of the speed bumps, which causes confusion in the self-driving.

There are prototypes of mobile devices to monitor road surfaces [29, 17, 18, 13]. These devices capture vibrations by finding irregularities on road surfaces through accelerometers and microcontrollers. They use ultrasonic sensors to detect speed bumps and capture coordinates of these obstacles through GPS (global positioning system). A pattern recognition system has been applied to the accelerometer readings as well as an automatic classifier using support vector machines to recognize anomalies on the surfaces [26]. The advantage of these devices is that they have a very simple design and are inexpensive. However, there are several disadvantages: precision errors are frequent, due to the lack of exactitude of the GPS coordinates, the delay when the system is tested in real time as well as the low accuracy of the sensors.

Another technique studied to detect obstacles on road surfaces is the analysis of data by LIDAR sensors [14, 28]. Surfaces have been analyzed to detect the pavement contour and the location of the road where the cars cruise [14]. For this, a parallel algorithm is designed that applies the data filtering of the floor and the objects as well as the segmentation of data using morphological operators. An algorithm for segmenting classical urban objects using digital elevation maps has also been designed [28]. Main disadvantages of these sensors is the high computational cost since large amounts of point clouds that are processed, in addition to that its cost is very high with respect to other sensors.

Some stereo vision techniques for analysing road surfaces have been studied [23, 24, 2, 1]. By means of these techniques, a three-dimensional structure of the scene can be obtained from the estimation of 2D images from different points of view. An algorithm has been designed to detect roads, traffic islands and obstacles by transforming the perceived data into a digital elevation maps, thus making it possible to perceive elevations on the road surface [23]. The main disadvantage is that it is only possible to detect surfaces with a maximum depth of 3 meters and with high heights in the elevations.

Other works can be found, which combine some of the techniques mentioned above with the aim of improving the individual performance of each one. The research carried out by Fernández presents an algorithm that combines the use of medium-cost LIDAR sensors and digital image processing to detect speed bumps as a type of obstacles represented by zebra crosses, typical of Spain and some other Spanish speaking countries [10]. Results from this work reveal that the detection rate of speed bumps is very high. However, when there are many cars with LIDAR sensors, their signals could overlap and become confused.

Figure 1 shows a plot of the numerical results reported in the literature on speed bump detection. As can be seen, most works detect speed bumps with different techniques when their road signals are clearly visible. They present good results but they do not consider the same conditions as this work. There are other works that apply the technique of surface monitoring with accelerometers and other sensors which is a non-predictive technique and it does not correspond to the conditions considered in this work.

There are very few methods that detect speed bumps when their signals are not well-marked. One of the works related to this research is carried out by

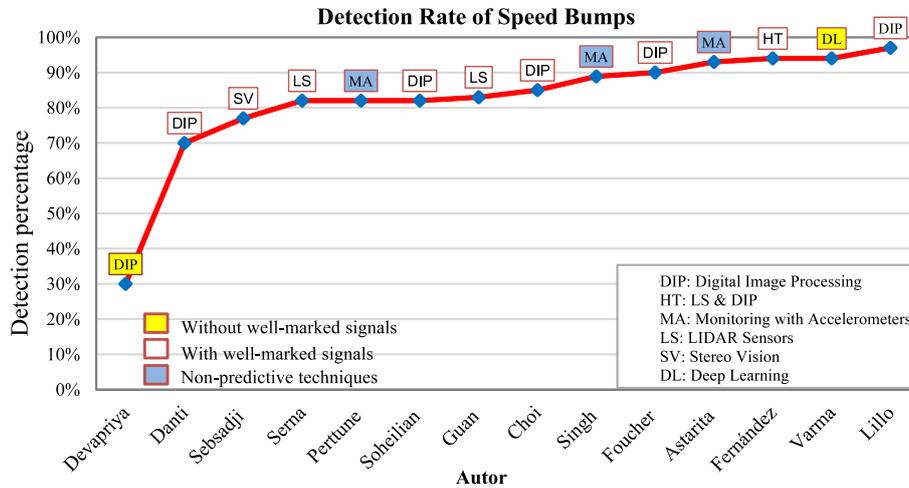


Fig. 1. Speed bump detection rate of different techniques.

Devapriya [8]. Results obtained by Devapriya have a detection rate of 30% that is below the average as seen in Figure 1. Another work related to the conditions considered in this research is that carried out by Varma [31] in which a method is proposed to detect and inform drivers in real time about the presence of speed bumps with or without signaling. For this purpose, Deep Learning techniques are applied using a pre-trained convolutional neuronal network. The speed bump detection rate is approximately 94%, which is above the average as shown in Figure 1. Still, despite showing good results, no details of the speed bumps can be obtained.

In this work, speed bumps with and without marks are detected and a convolutional neural network is also applied for detection. However, the difference with the work of Varma [31] is that when the convolutional neural network fails, stereo vision is applied to detect speed bumps by analyzing the 3D surface elevation because it is not restricted only to its signaling pattern. In addition, the methodology proposed in this work, unlike Varma [31] and Devapriya [8], would make it possible to identify the kind of speed bumps as well as some other details regarding their shape and size.

3 Proposed Methodology

In order to solve the problem of detecting speed bumps on road surfaces, the methodology shown on Figure 2 is proposed. This allows to classify images with or without speed bumps. A convolutional neuronal network is used to detect speed bumps and when these are not easily detected, stereo vision is applied to identify speed bumps by elevation on 3D surfaces.

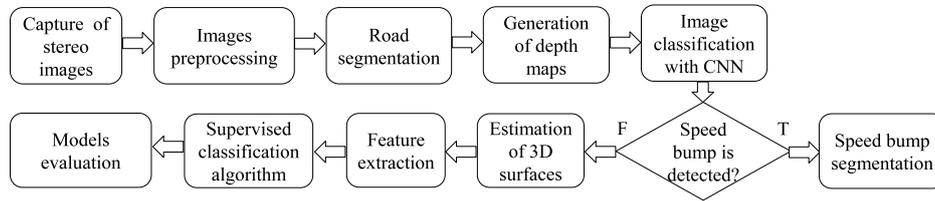


Fig. 2. Proposed methodology for automatic speed bump detection.

3.1 Capture of Stereo Images

In this phase a set of stereo images of streets with and without speed bumps is built. The speed bumps types of interest are trapezoidal shaped and pedestrian crossing because they are the most common in developing countries. In addition, the images have marked and un-marked speed bumps road signaling. It is consider environments with controlled conditions like good lighting, preferably without obstacles such as pedestrians and other cars. For capturing the images, a stereo vision system consistent of two cameras with the same characteristics and aligned in parallel with an intersection area between both is used. Streets images of the City of Puebla, Mexico, are captured with and without speed bumps.

3.2 Image Preprocessing

This phase consists on the application of digital image processing techniques to improve the quality of the images, thus preparing them for the automatic classification process. The most significant processing techniques required are: lightening the image to increase their clarity and avoiding dark images; decreasing the contrast to reduce the total color range of the image; applying a Gaussian filter to eliminate noise and details of the texture of the street in order to obtain uniform colors or gray intensities [4].

3.3 Road Segmentation

The region of interest (ROI) are the streets that will be processed and filtered in order to focus in the area to be assessed. The remaining area is irrelevant and by means of binary masks it is painted in a single color to be ignored in the following stages. A segmentation of color and textures is performed. For color segmentation, the first-order statistical descriptors are calculated considering each pixel: mean, standard deviation, median and entropy. For texture segmentation, second-order statistical descriptors are calculated by using a matrix of co-occurrences considering regions of the image: homogeneity, entropy, dissimilarity, energy and correlation [7]. The image is binarized to generate a mask with the ROI. The ranges of each statistical descriptor are adjusted to generate the mask with the ROI when the descriptor values are within the defined range.

3.4 Generation of Depth Maps

Speed bumps with a well-marked pattern are identified and located. In the event that the image contains not well-marked bumps, the stereo vision is applied to generate depth maps and estimate the 3D surface where the surface elevations can be analyzed in order to find speed bumps. The steps for applying stereo vision to the captured images of the stereo image construction phase are: camera calibration, and rectification of images. The calibration of each camera is performed to find the intrinsic parameters: focal distance, radial and tangential distortion, used to eliminate the distortions of the camera and a stereo calibration is performed to find the extrinsic parameters of both cameras: rotation and translation matrices, used to find the relationship between the coordinates of the cameras with the real world coordinates [12]. Zhan's method is one of the most used to calibrate cameras and is based on the detection of corners of chess pattern templates [9]. The aim of the rectification is to make the stereo images exactly parallel to each other horizontally. For this, correspondences between the points of both images are found through rotation and translation transformations until an exact alignment between images is achieved. Subsequently, a disparity map is generated which is a grayscale image where the differences between the pixels of the stereo pair are calculated. A depth map is also generated and the distances to the observed points are estimated.

3.5 Image Classification Convolutional Neural Network (CNN)

This allows to classify images with or without speed bumps. A pre-trained model with flat images of a CNN is used to detect speed bumps [25]. If a speed bump is detected with high precision, this is segmented using the depth map and the distance to it is calculated. Otherwise, when speed bumps are not easily detected (when the precision exceeds a threshold) an estimate of 3D surfaces is made to extract characteristics of 3D meshes.

3.6 Estimation of 3D Surface

If the CNN cannot distinguish speed bumps, an estimate of the 3D surface is made in order to detect them due to 3D surface elevation. This phase consists on reconstructing a 3D scene by obtaining spatial coordinates and the representation of point clouds relevant to the scene [27]. To generate the disparity maps, the segmented image is used and the stereo vision is applied over the road segmentation. An estimate of 3D surfaces is made by the reconstruction of the three-dimensional scene.

3.7 Feature Extraction

An image feature vector that indicates the existence of speed bumps is generated. Features are extracted from both flat images and stereo images. From flat images characteristics of color and textures of speed bumps using low-level and

shape matching [22]. From stereo images, characteristics are extracted on the 3D meshes estimated as the slope of the elevation, width, height, depth, among others calculated by interpolation functions.

3.8 Classification Models of Speed Bumps

For the automatic classification of images, a classifier is trained with a set of flat and stereo image pairs that are labelled in two classes SB and NSB (with and without speed bumps present, respectively). The automatic feature extraction is performed and supervised learning algorithms are applied for the detection of speed bumps such as: decision trees, neural networks, support vector machines, and others, and a predictive model is generated as output [15]. Once trained, the classifier is tested with another set of image pair by applying the generated predictive model to infer the class to which it belongs.

3.9 Evaluation Models

The performance of the classifier is measured to determine the combination of features and the classification algorithm that presents the best results. This phase consists on applying evaluation metrics such as: precision, recall, f-measure, and others [15]. To do this, several experiments are carried out with the classifier using different images with and without speed bumps not detected by the CNN. The results of each method are validated to determine which ones show better results. After choosing the model with the best results, when the speed bump is detected by stereo vision it is segmented using the depth map and the its distance is calculated.

When the speed bump is detected, it is segmented and the size, shape, and distance to the speed bump are calculated, since these measures are very important for the decision making of self-driving cars.

This methodology allows to identify a speed bump by two different ways. CNNs represent a good solution for speed bump detection. Visually they work properly when what they find in the images is similar to the instances with which the model was trained. However, the CNN have problems of identification of speed bumps when in their training dataset there are not enough samples similar to the cases that can occur in environments such as the streets of Puebla, Mexico. Thus, stereo vision can promote detection by analyzing the elevations of surfaces when the detection with CNN fails. This proposal can strengthen existing speed bump classifiers.

4 Preliminary Results

The expected results of this research work are: the construction of a set of stereo images of roads with and without speed bumps, the implementation of an algorithm for speed bump detection on road surfaces following the methodology proposed, as well as the elaboration of comparative plots of the algorithm results

to other algorithms. One of the most relevant preliminary results presented in this work is the implementation of the road segmentation.

A set of stereo images with and without speed bumps from the City of Puebla, Mexico, was captured using a stereo vision system built with two equal cameras. It is considered controlled conditions of the environment good climate, lighting, without shadows, and without obstacles on the streets such as pedestrians, cars, among others. These images are preprocessed to improve their quality. A bilateral filter is applied to eliminate noise and to soften the image in order to make the colors more uniform, preserving the edges [4].

A street segmentation algorithm was implemented applying segmentation by color and textures. To achieve this, a dynamic sampling area was defined to calculate a segmentation range. Samples of first-order statistical descriptors are obtained on the HSV and HLS color spaces: mean, median, min, max and the entropy of each channel of the spaces. Simultaneously, second order descriptors are obtained on the RGB space: dissimilarity, homogeneity, energy, correlation, ASM, and entropy, in order to delimit the texture characteristics of speed bumps. From the statistical samples obtained in the images, dynamic intervals are defined to construct a mask that ignores regions outside the range, combining information from each channel. This mask is processed by applying opening-closed filters to reduce the noise of the mask and avoid isolated regions. Results are presented in Figure 3. As can be seen, the region of no interest was removed from the scene and the street was reflected as a region of interest.

Disparity map of stereo images was generated and an example of this is shown in Figure 4a) where the lighter tones indicate that the point is closer and the darker ones are farther away. This map allows to perceive the depth of the scenes, to calculate the distances to different points in it and to reconstruct 3D scenes. An initial reconstruction of the three-dimensional scenes was made using point clouds as shown Figure 4b). This Figure shows some key points more representative of the real scene and the depth of them is observed. From the segmented images (Figure 3), a 3D reconstruction of the segmented street surface can be done. In addition, from the key points obtained and applying mathematical interpolation functions (such as B-splines [21]), it is possible to estimate 3D mesh surfaces. These 3D meshes are inputs to the supervised classifier to detect speed reducers by the analysis of surface elevations when the CNN failure in detection.

Some phases of the methodology have already been implemented. The applied street segmentation techniques have presented visually acceptable results for different street images. In the future, it is expected to calculate the vanishing point of the street to obtain a more precise segmentation of the street, as well as, to implement all phases of methodology: image classification with CNN, estimation of 3D surfaces with B-splines functions, feature extraction of 3D surface meshes, as well as the generation and evaluation of supervised classification models. In addition, it is expected to apply the statistical method of ROC (receiver operating characteristic curve) curves to make a sensitivity and specificity analysis of classification models.

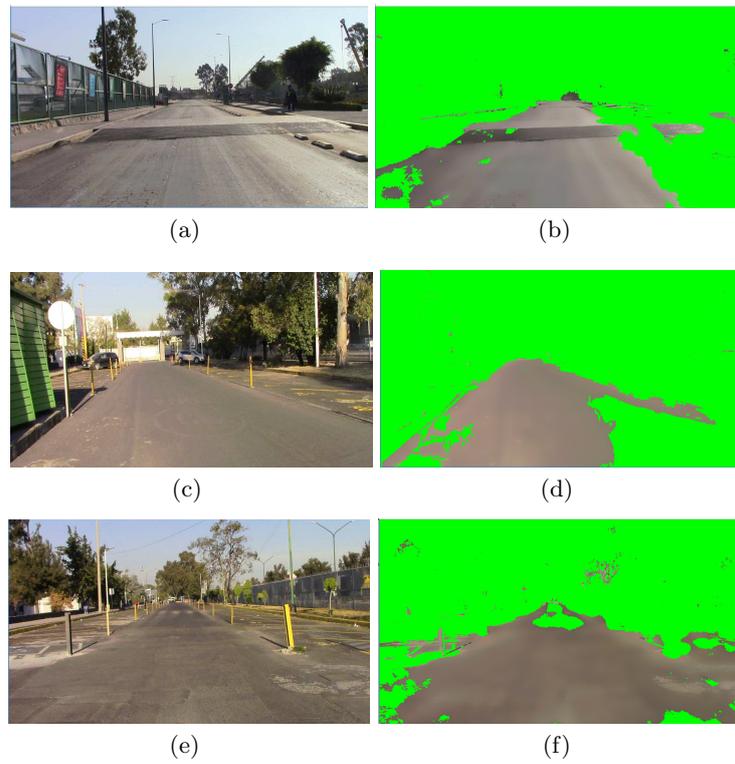


Fig. 3. Results obtained from street segmentation by color and texture analysis.

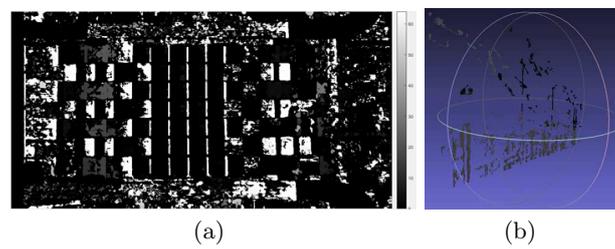


Fig. 4. a) Disparity map, b) Reconstruction of 3D scene using stereo vision.

5 Conclusions

The detection of speed bumps on roadways is key for a safe and comfortable self-driving, therefore it has been widely studied in the literature, mainly by means of digital image processing techniques. However, speed bump detection without well-marked signals has been little addressed, despite being a very frequent situation in developing countries, and the existing techniques show high errors under this restriction. The main contribution of this work is a methodological proposal for speed bump automatic detection using a pre-trained CNN and supervised automatic classification using a 3D surface estimation obtained by stereo vision without the requirement of well-marked signals, situation that has not been completely solved with existing techniques.

References

1. Badino, H., Franke, U., Pfeiffer, D.: The stixel world-a compact medium level representation of the 3d-world. In: Joint Pattern Recognition Symposium. pp. 51–60. Springer (2009)
2. Broggi, A., Cardarelli, E., Cattani, S., Sabbatelli, M.: Terrain mapping for off-road autonomous ground vehicles using rational b-spline surfaces and stereo vision. In: Intelligent Vehicles Symposium (IV), 2013 IEEE. pp. 648–653. IEEE (2013)
3. Buehler, M., Iagnemma, K., Singh, S.: The DARPA urban challenge: autonomous vehicles in city traffic, vol. 56. springer (2009)
4. Burger, W., Burge, M.J., Burge, M.J., Burge, M.J.: Principles of digital image processing. Springer (2009)
5. Choi, J., Lee, J., Kim, D., Soprani, G., Cerri, P., Broggi, A., Yi, K.: Environment-detection-and-mapping algorithm for autonomous driving in rural or off-road environment. IEEE Transactions on Intelligent Transportation Systems 13 (2), 974–982 (2012)
6. Danti, A., Kulkarni, J., Hiremath, P.: A technique for bump detection in indian road images using color segmentation and knowledge base object detection. International Journal of Scientific & Engineering Research 4 (8), 2229–5518 (2013)
7. De Siqueira, F.R., Schwartz, W.R., Pedrini, H.: Multi-scale gray level co-occurrence matrices for texture description. Neurocomputing 120 pp. 336–345 (2013)
8. Devapriya, W., Babu, C.N.K., Srihari, T.: Real time speed bump detection using gaussian filtering and connected component approach. In: Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), World Conference on. pp. 1–5. IEEE (2016)
9. de la Escalera, A., Armingol, J.M., Pech, J.L., Gómez, J.J.: Detección automática de un patrón para la calibración de cámaras. Revista Iberoamericana de Automática e Informática Industrial RIAI 7 (4), 83–94 (2010)
10. Fernández, C., Gavilán, M., Llorca, D.F., Parra, I., Quintero, R., Lorente, A.G., Vlacic, L., Sotelo, M.: Free space and speed humps detection using lidar and vision for urban autonomous navigation. In: Intelligent Vehicles Symposium (IV), 2012 IEEE. pp. 698–703. IEEE (2012)
11. Foucher, P., Sebsadji, Y., Tarel, J.P., Charbonnier, P., Nicolle, P.: Detection and recognition of urban road markings using images. In: Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on. pp. 1747–1752. IEEE (2011)

12. Hamzah, R.A., Ibrahim, H.: Literature survey on stereo vision disparity map algorithms. *Journal of Sensors* 2016 (2016)
13. Harikrishnan, P., Gopi, V.P.: Vehicle vibration signal processing for road surface monitoring. *IEEE Sensors Journal* 17 (16), 5192–5197 (2017)
14. Hernández, J., Marcotegui, B.: Filtering of artifacts and pavement segmentation from mobile lidar data. In: *ISPRS Workshop Laserscanning 2009* (2009)
15. Hernández Orallo, J., Ferri Ramirez, C., Ramirez Quintana, M.J.: *Introducción a la Minería de Datos*. Pearson Prentice Hall, (2004)
16. Jung, S., Lee, U., Jung, J., Shim, D.H.: Real-time traffic sign recognition system with deep convolutional neural network. In: *Ubiquitous Robots and Ambient Intelligence (URAI), 2016 13th International Conference on*. pp. 31–34. IEEE (2016)
17. Madhumathy P, Saurabh Singh, S.S.U.K.: Detection of humps and potholes on roads and notifying the same to the drivers. *International Journal of Management and Applied Science* 3 (1), 130–133 (2017)
18. Madli, R., Hebbar, S., Pattar, P., Golla, V.: Automatic detection and notification of potholes and humps on roads to aid drivers. *IEEE sensors journal* 15 (8), 4313–4318 (2015)
19. Maldonado-Bascón, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gómez-Moreno, H., López-Ferreras, F.: Road-sign detection and recognition based on support vector machines. *IEEE transactions on intelligent transportation systems* 8 (2), 264–278 (2007)
20. Marrón, B.S.: Co-occurrence matrix and fractal dimension for image segmentation. *Revista de Matemática Teoría y Aplicaciones* 19 (1), 49–63 (2012)
21. Masud, A., Kannan, R.: B-splines and nurbs based finite element methods for kohn–sham equations. *Computer Methods in Applied Mechanics and Engineering* 241 pp. 112–127 (2012)
22. Nixon, M., Aguado, A.S.: *Feature extraction and image processing for computer vision*. Academic Press (2012)
23. Oniga, F., Nedevschi, S.: Processing dense stereo data using elevation maps: Road surface, traffic isle, and obstacle detection. *IEEE Transactions on Vehicular Technology* 59 (3), 1172–1182 (2010)
24. Oniga, F., Nedevschi, S.: Curb detection for driving assistance systems: A cubic spline-based approach. In: *Intelligent Vehicles Symposium (IV), 2011 IEEE*. pp. 945–950. IEEE (2011)
25. Patterson, J., Gibson, A.: *Deep learning: A practitioner’s approach*. ” O’Reilly Media, Inc.” (2017)
26. Perttunen, M., Mazhelis, O., Cong, F., Kauppila, M., Leppänen, T., Kantola, J., Collin, J., Pirttikangas, S., Haverinen, J., Ristaniemi, T., et al.: Distributed road surface condition monitoring using mobile phones. In: *International Conference on Ubiquitous Intelligence and Computing*. pp. 64–78. Springer (2011)
27. Revelo, D.A., Usama, F.D., Florez, J.F.: Reconstrucción 3d de escenas mediante un sistema de visión estéreo basado en extracción de características y desarrollado en opencv. *Ingeniería y Universidad* 16 (2), 485–485 (2012)
28. Serna, A., Marcotegui, B.: Detection, segmentation and classification of 3d urban objects using mathematical morphology and supervised learning. *ISPRS Journal of Photogrammetry and Remote Sensing* 93 pp. 243–255 (2014)
29. Sujitha, M.S., Ramesh, N., Kotamraju, S.K.: Automatic speed controlling of vehicle and detection and notification of potholes and humps. *Journal of Engineering and Applied Sciences* 100 (9), 1921–1924 (2016)

30. Tedeschi, A., Benedetto, F.: A real-time automatic pavement crack and pothole recognition system for mobile android-based devices. *Advanced Engineering Informatics* 32 pp. 11–25 (2017)
31. Varma, V., Adarsh, S., Ramachandran, K., Nair, B.B.: Real time detection of speed hump/bump and distance estimation with deep learning using gpu and zed stereo camera. *Procedia computer science* 143 pp. 988–997 (2018)
32. Wang, P., Hu, Y., Dai, Y., Tian, M.: Asphalt pavement pothole detection and segmentation based on wavelet energy field. *Mathematical Problems in Engineering* 2017 (2017)

Parameter Experimentation for Epileptic Seizure Detection in EEG Signals using Short-Time Fourier Transform

Ricardo Ramos-Aguilar¹, J. Arturo Olvera-López¹, Ivan Olmos-Pineda¹,
Susana Snchez-Urrieta², Manuel Martín-Ortiz³

¹ Autonomous University of Puebla, Faculty of Computer Science, Puebla, Mexico
ricramosa1@gmail.com, {aolvera,iolmos}@cs.buap.mx

² Autonomous University of Puebla, Faculty of Electronics Science, Puebla, Mexico
susana.sanchezu@correo.buap.mx

³ Autonomous University of Puebla,
Information Technology and Communications Department(DCyTIC),
National Laboratory of Supercomputing (LNS-BUAP-CONACYT), Puebla, Mexico
manuel.martin@correo.buap.mx

Abstract. Short-time Fourier transform is a time-frequency method commonly used to analyze signals, particularly in EEGs. It has shown acceptable results for the identification of different actions, such as sleep disorders, epilepsy, and others, and in applications as brain-computer interfaces. However, the selection of short time Fourier transform parameters is not a trivial task, as the variability of these directly affects the resolution spectrogram, from which features are extracted to determine the constructed models in the classification stage. In this paper, experiments for determining STFT parameters such as window type and length, and overlapping are explored. As a case study, an EEG epilepsy database is used to identify healthy people versus patients suffering epileptic seizures, finding that the parameters modify the spectrogram visualization in terms of time/frequency and classification. Based on these experiments, it was concluded that the proposed strategy supports the correct selection of parameters that positively impact the accuracy of the results obtained.

Keywords: EEG Signals, Short-time Fourier Transform, Spectrograms, Seizure Classification.

1 Introduction

Electroencephalography (EEG) is a brain monitoring method based on measurements of electrical activity generated by the brain. An EEG shows evidence of how the brain performs the bodily functions over time, such as the pumping of the heart, gland secretion, breathing, and internal temperature regulation, among others [10]. Nowadays, EEGs are used by scientists and physicians for

analyzing brain functions and diagnosing neurological disorders, such as brain tumors, head injuries, sleep disorders, dementia, epilepsy, Alzheimer's, seizure disorder, deficit attention disorder, anxiety disorder, fetal alcohol syndrome, autism, as well as monitoring the effects of anesthesia during surgery [3, 5, 6, 8, 16]. The EEG is an appropriate tool to aid in the diagnosis of diseases, epilepsy in particular. This inexpensive tool is useful for showing the underlying manifestation of epilepsy. EEG signals in people with epilepsy show two kinds of abnormal activity: ictal (during an epileptic seizure) and interictal (between seizures) [8, 16]. EEG analysis for diagnosing epilepsy started in 1970, and since then it has been an area of interest for researchers due to its non-stationary features; at present, most problems in seizure detection are related to finding events (ictal and interictal) during epileptic seizures [8].

Different methodologies have been proposed to identify epileptic seizures in EEG signals based on frequency, time, wavelet transforms, and Gabor filters [7, 8, 14]. However, the nature of the EEG signals (non-stationary) involves specific aspects when employing techniques based on frequency or time [8], as the features obtained from these methods do not provide enough information from EEG signals alone [1]. Techniques based on Time-frequency analysis are considered more complete than some others because they decompose the signal in frequencies over signal time, to analyze non-stationary signals such as EEGs. Time-frequency representations can be applied to EEG signals, as the resulting transform can be treated as an image in order to extract features, which has shown acceptable results in terms of accuracy for different applications [13]. Short-Time Fourier Transform (STFT) is a time-frequency representation commonly used in signal analysis, as well as in digital image processing, voice processing, biology, medicine [9], and in other applications, such as Electroencephalographic (EEG) signals analysis (STFT allows a representation in time-frequency of the EEGs, which shows a different visualization for analysis).

From EEG spectrograms different features can be extracted to identify several tasks (imagined writing, motor images, epilepsy, alcoholism, etc.) [12]. In most of these, the STFT parameters are pre-defined by the programming language toolkits when beginning the analysis. Parameter selection is important, as it helps to minimize two problems: poor spectrogram resolution and a dearth of relevant features, which determine the results in the classification; in addition, the process of assigning specific parameter values for any application is a difficult task. Thus, different STFT parameters are analyzed in this work in order to identify relevant features that support the classification of epilepsy stages (seizure and normal) and their spectrograms. Initially, EEG signals are obtained from a dataset and STFT is applied with different parameters; after that, an image generated from the spectrogram is converted to grayscale to extract features; total energy and spectral peaks are obtained for training three different classifiers and evaluating STFT parameters; the results are shown in terms of the accuracy in binary classification and the visualization of some spectrograms. The STFT parameters of the experimentation were selected from main and side lobes of the windows; first, the width of the main lobe was analysed and the windows with

the thinnest were chosen, after the side lobes were observed and from its height with respect to the main lobe were selected the windows with side lobes lowest. The overlapping is proposed from window properties, where are selected in the case of lobes are added or attenuated when windows are overlapped. This paper is organized as follows: section 2 introduces the methodology and the related theory; section 3 provides a description of EEG dataset and the experimental analysis; Finally, section 4 presents our conclusions and future work.

2 Methodology

2.1 Short-time Fourier Transform

In practical applications, STFT is implemented as a sliding window adjusted to a signal. Given an input signal x^T of arbitrary duration, segments are extracted at regular intervals using a time-limited window w_n ; The segments (frames) of the signal can be expressed as [4]:

$$x_l[n] = w_n * x[n + lL], \quad 0 \leq n \leq N - 1, \quad (1)$$

where N is the length of the window, l is the frame index, and L is the hop size, that is, the sample spacing between consecutive applications of the sliding extraction window; the index n is a local time index relative to the beginning of the window displacement. The expression $x[n + lL]$ represents a position over signal, n is representing the phase to shift the window, $*$ represents a operation likewise modulation between two terms. Finally, the discrete Fourier Transform (DFT) [4] is applied to each frame of the signal, as follows:

$$X[k, l] = \sum_{n=0}^{N-1} x_l[n] e^{-i2\pi nk/K} = \sum_{n=0}^{N-1} w_n * x[n + lL] e^{-i2\pi nk/K}, \quad (2)$$

where K is the size of the DFT and k is a frequency index or bin index. The STFT $X[k, l]$ then characterizes the local time-frequency behavior of the signal around time lL and bin k ; for a sampling rate of F_s , these discrete indices correspond to the continuous time lL/F_s and frequency kF_s/K [4]. To simplify the notation, a radial frequency of:

$$\omega_k = 2\pi k/K, \quad (3)$$

is frequently included; then, the expression STFT becomes:

$$X[k, l] = \sum_{n=0}^{N-1} w_n * x[n + lL] e^{-i\omega_k n}. \quad (4)$$

Applying the window over time, the STFT returns a spectral representation of a time segment of the input signal; interpreting $X[k, l]$ as a function of the frequency k for each value of the time index l , the STFT corresponds to some series of spectra located in time. Alternatively, the STFT can be seen as a time

function for each frequency; interpreting $X[k, l]$ as a time serie that is a function of l for each bin k (points to calculate the DFT), the STFT corresponds to a filter bank that decomposes the input signal over frequency channels or sub-bands [4]. These two interpretations of the STFT are represented with respect to the time-frequency plane. These changes to time and frequency have become common in the modern literature on signal processing as a way of representing the time and frequency resolutions of signals.

One drawback of STFT is the resolution required to obtain spectrograms, caused mainly by the length of the window. A narrow window offers a better resolution in time, but not in frequency, while a wide window provides good resolution in frequency, but does not perform as well in time resolution. This means that it is difficult to achieve a good location in both time and frequency domains simultaneously, because the STFT depends on only one window.

The STFT process is a method that begins by windowing a signal into shorter segments, where phase windows can be overlapped; Fast Fourier transform (FFT) is applied to each segment separately [15] and each result is joined to form the spectrogram, which can be manipulated, and then the inverse FFT is applied to return each segment to the time domain.

Therefore, to apply the STFT to signals, different parameters should be considered; for example: window type and length, main lobe length (corresponding to window type), FFT points and overlapping. These parameters that can affect the resolution spectrogram are briefly described in the following subsections.

2.2 Windowing

Windowing is a phase that involves applying a function to a signal or segment over time, which is then used in the Fourier transform or the STFT. The Fourier transform is commonly implemented for the analysis of non-periodic and stationary signals; for non-stationary signals, one option is STFT, which analyzes the signal by segments, making use of the windows, since this sub-signal is supposed to be a stationary signal. A good analysis depends essentially on the type of window and the parameters that can be modified, which will be discussed later.

First, the window type is selected; Figure 1 (a) shows different windows (Bartlett, Blackman, Hamming, Gaussian, Kaiser, rectangular, Hanning) with a length of 64 points and their spectra (b); these windows are the most common in the signal analysis literature. Initially, when a window is applied to the signal, it is modulated as the window way along time segment; the rectangular window is the ideal in time, since signals are not modified when it is applied and with this, it would not adjust to the window type due to the shape as a box, which only cut the signal and Fourier transform is applied. This window has a thin main lobe; although, its secondary lobes are very high. The main idea is to have something balanced, and the Gaussian window is optimal due to its very similar wave form spectrum; however, it cannot be obtained computationally, so approximations are implemented that show a main lobe that is average in terms of width, although the secondary lobes are not the smallest. The Blackman window shows the smallest secondary lobes, followed by Hanning, which has a small

main lobe in its window; this window is also a variant of the raised cosine, which is very similar in both spectrum and window. Taking into account the windows in Figure 1, it can be seen that the main lobes are in ascending order by window length: Blackman, Gaussian, Hamming, Hanning, Kaiser and Rectangular; the secondary lobes are listed in relation to height (from highest to lowest energy): Rectangular, Kaiser, Hamming, Gaussian, Hanning, and Blackman.

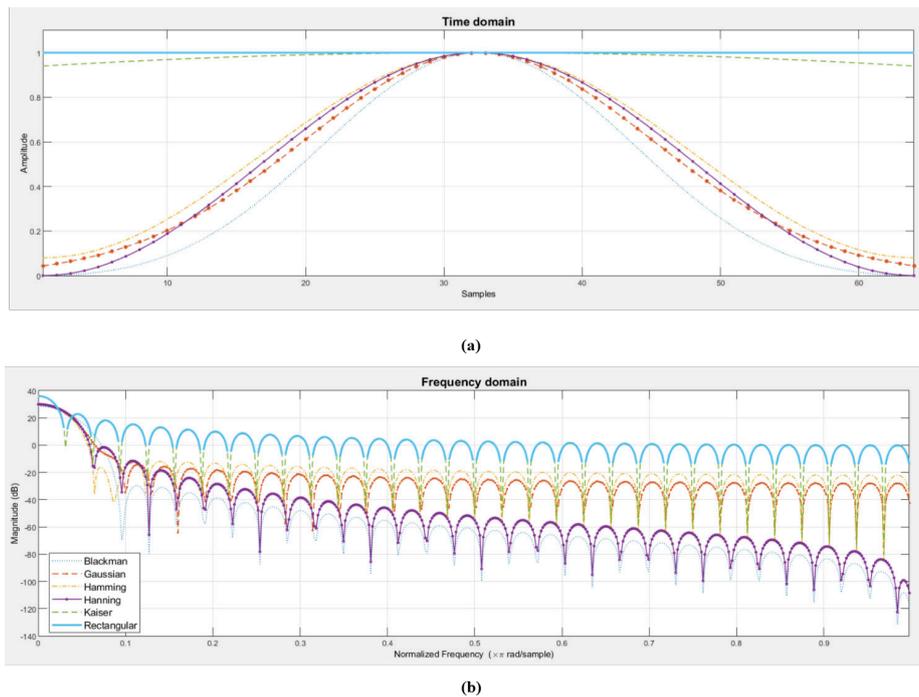


Fig. 1. Windows (a): (....)-Blackman, (-.-.-)-Gaussian, (-.-)-Hamming, (-.-.-)-Hanning, (- - -)-Kaiser, and (___)-Rectangular, 64-point length and their spectra(b).

On the other hand, temporal and frequency resolution can be obtained from the STFT parameters; Figure 1 shows the relationship between these parameters, where it can be seen that the spectrum is narrow for a wide rectangular window, so its use is not advisable, while Hamming and Blackman maintain a less strict relationship, since proportionally, it is not noticeable as in a rectangular window; this is due to both windows being variants of a raised cosine and to the lobes for both being smaller.

Considering a visual analysis from Figure 1 with main lobes and secondary side lobes as main features, any window can be selected as an acceptable option, because a thinner lobe shows frequencies close to that signal. In addition, the secondary lobes show a lower energy dispersion, as most are concentrated in

the main lobe; however, these are not the only parameters to be considered in window analysis [11], although, a method based on spectral peaks might suffice; therefore, one parameter that defines spectral peaks is the main lobe of the windows, which represents peaks, and another is the secondary lobe height, because energy is dispersed among these. Thus, considering these two parameters, the selected windows were: rectangular, which has a representation similar to the Fourier Transform, as it does not have modification in time and its spectrum has the thinnest main lobe; Gaussian, with a window similar to the spectrum; Blackman, because it has the lowest secondary lobes; and Hanning, because it has one thin main lobe and low secondary lobes.

2.3 Feature Extraction

In this work, features are obtained from EEG signal spectrograms. Spectral peaks are found first, since these are related to relevant energy over some frequency/time; after that, different features are extracted from the peaks. Finally, volume features are calculated from all intensities of the spectrogram.

Spectral peaks. These peaks are calculated using the local maxima method, which finds local minima within a region and above a threshold from energy levels, these are represented on the spectrogram. From a region or mask the peaks are found with the method mentioned. Up to this point, other parameters have not been analyzed, since these peaks depend on the window and the signal type, which will be analyzed later.

Volume. Feature extraction is based on an approach that considers the spectrogram in a three-dimensional representation as a volume. All intensity points and time and frequency coordinates were used to obtain the spectrogram volume. Likewise, all intensities were added together to obtain the approximate energy. The spectral peaks were used to calculate the surface area and the intensities from the highest peaks were added; the area formed by peaks and volume from all spectrogram points were calculated through a convex hull. Figure 2 shows these five features, which are concatenated to obtain the feature vector.

3 Results

3.1 Experimental Approach

This section presents an evaluation of STFT parameters such as overlapping and window type and length. These experiments are concerned with the identification of people experiencing epileptic seizures or healthy people. First, signals were obtained from an EEG dataset (epileptic seizure); STFT with different windows and overlapping (0% and 75%) was applied to each signal to obtain spectrograms, which were converted to grayscale images; after that, peaks were found from

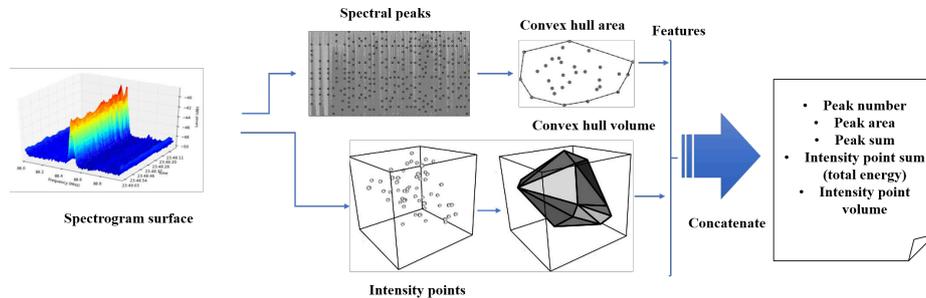


Fig. 2. Features extracted form EEG signal: peak number, peak area, peak sum, intensity point sum and intensity point volume.

regions, these are shaped by $2 \times \text{minimal distance} + 1$, where *minimal distance* between peaks is the separation between peaks; to find the maximum number of peaks the value of 1 was used as a separation distance; a threshold was selected from the minimum intensity of the spectrogram in grayscale. The features describe below were extracted from the peaks.

For the experiments, four windows were taken into account, with Gaussian being the closest to the ideal, the rectangular being the ideal in time, Blackman having the smallest lobes, and Hanning maintaining the best relationship between the minimum width of the main lobe and the height of the secondary lobes. The number of FFT points are the same for the window length, which is not relevant for resolution; a 75% overlapping was proposed, based on the properties of the windows and their spectral responses.

This methodology was tested using an open access epilepsy EEG dataset from Bonn University [2] with five subsets (Z,O,N,F, and S), each subset containing 100 EEG signals with 4097 samples, recorded at a sampling rate of 173.61 Hz using a 128-channel amplifier system with an average common reference. Sets Z and O were collected from five healthy volunteers. Sets N, F, and S were recorded from five epileptic patients for each set. Records from set S were collected during seizure activity, while sets N and F were gathered during seizure-free intervals.

From the epileptic dataset, two subsets (subset A and E) were used to classify the signals, which correspond to healthy people and patients suffering an attack, respectively. Firstly, the signals from healthy people (Figure 3) were analyzed using different STFT parameters.

STFT was applied to EEG signals from subset A, as mentioned above, and four windows with length (64) and without overlapping were used. The spectrograms generated are shown in Figure 4, which describes scattered energy, because there are different events and frequencies over time; also, the spectrogram maintains its relationship with the corresponding spectrum. The energy is more scattered along the vertical axis (frequency), while it is more noticeable in (b), due to the width of the main lobe of Gaussian spectrum and the energy in (a) not being as noticeable, since it has the narrowest main lobe (Rectangular).

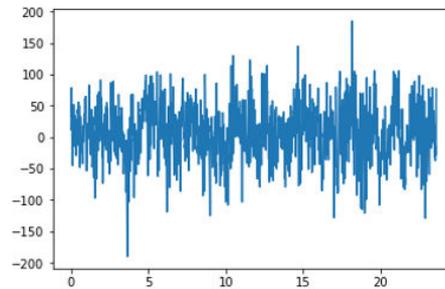


Fig. 3. EEG signal from an open eyes person in a relaxed state.

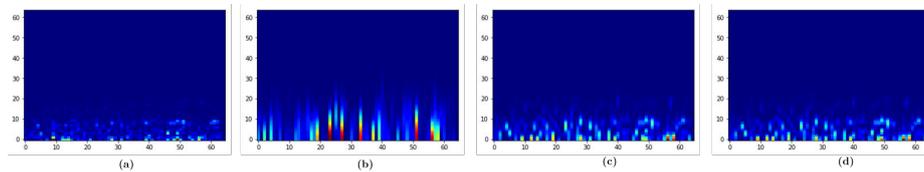


Fig. 4. Spectrogram from EEG signals of a healthy person in a relaxed state with eyes open. Window length= 64, 0% overlapping: (a) Rectangular, (b) Gaussian , (c) Blackman and (d) Hanning.

Comparing the spectra from Figures 4 and 5, only an elongation of the energy on the horizontal axis (time) can be noted, although the windows are actually focusing more on the frequencies shown along the vertical axis. Something similar occurs when a single window is placed over the entire time from a signal: it shows only one frequency; otherwise, when a narrow window is applied to short segments from a signal, it shows changes over time, i.e., frequencies for each window.

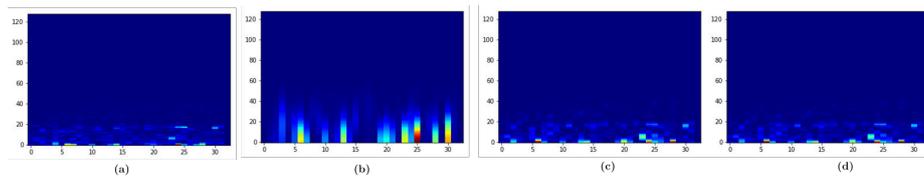


Fig. 5. Spectrogram from EEG signals from a healthy person in a relaxed state with eyes open. Windows length= 128, 0% overlapping: (a) Rectangular, (b) Gaussian , (c) Blackman and (d) Hanning.

Another interesting parameter is overlapping, which implies a more general approach to the signal. When a high percentage of overlapping is applied, window

n overlaps most of window $n + 1$, which could generate windows without side lobes (as they cancel each other out); only windows with overlapping main lobes cause the entire signal to resemble the application of a single window to the entire signal. Of the eight spectra, Figures 6 and 7 are most similar to those shown in Figures 4 and 5, but the overlap thins the energy shown over time.

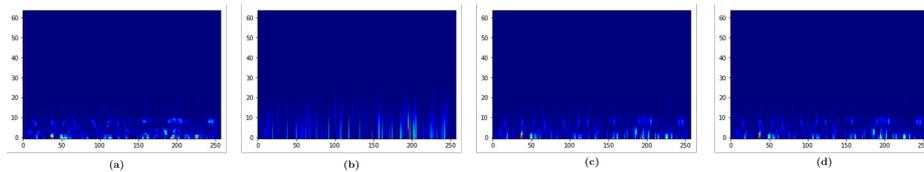


Fig. 6. Spectrogram from EEG signals from a healthy person in a relaxed state with eyes open. Window length= 64, 75% overlapping: (a) Rectangular, (b) Gaussian , (c) Blackman, and (d) Hanning.

When the overlap is greater, the energy seems to be reduced along the x-axis (time), and in terms of time-frequency resolution, a large window focused for the frequency would have a better resolution with a higher percentage of overlap.

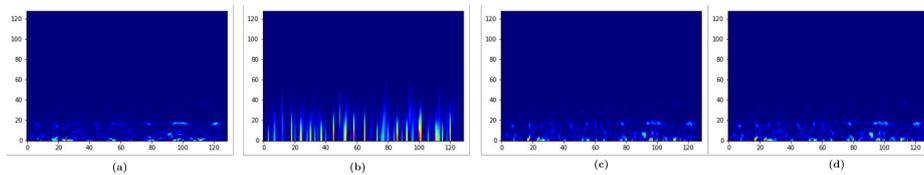


Fig. 7. Spectrogram from EEG signals from a healthy person in a relaxed state with eyes open. Window length= 128, 75% overlapping: (a) Rectangular, (b) Gaussian , (c) Blackman, and (d) Hanning.

Signals from subset E were also analyzed, which correspond to patients suffering an epileptic seizure. The signal in Figure 8 represents an EEG of a person suffering an attack; as in previous experiments, the same windows were used.

The higher-frequency spectrograms in Figures 9-12 show higher energy for some sections in comparison to the spectrograms of a healthy person. This representation is similar to both kinds of EEG signals, due to the parameters used in these experiments. From the spectrograms of healthy people, energy activity appears along time in terms of energy, possibly due to some event occurring in different time intervals; this could be noted using a 64 Gaussian window. In the opposite case, for epileptic patients, the energy is concentrated along time, showing the energy as frequency by applying rectangular, Blackman, and Hanning windows. The Gaussian window describes activity in time, and can be appre-

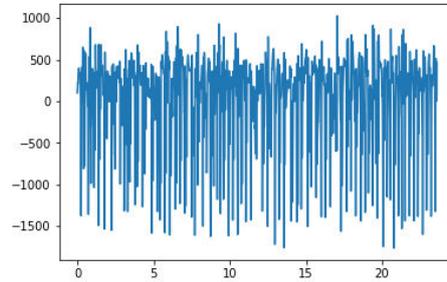


Fig. 8. EEG signal from a person suffering an epileptic seizure.

ciated when using long windows (128), for example, the window used (128) in Figure 12 with a 75% overlap.

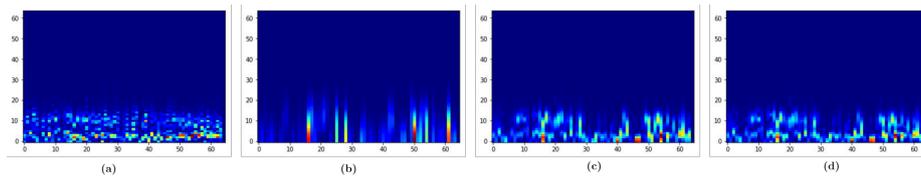


Fig. 9. Spectrograms of EEG signals from a person suffering an epileptic seizure. Window length=64 and 0% overlapping (a) Rectangular, (b) Gaussian, (c) Blackman and (d) Hanning.

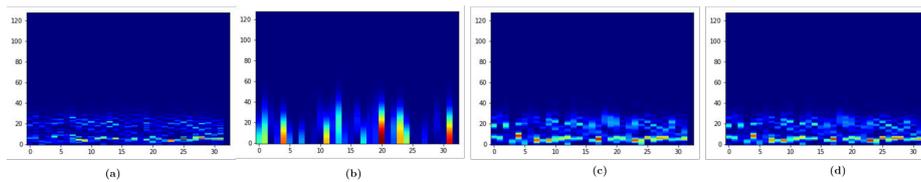


Fig. 10. Spectrograms of EEG signals from a person suffering an epileptic seizure. Window length=128 and 0% overlapping (a) Rectangular, (b) Gaussian, (c) Blackman and (d) Hanning.

From the previous spectrograms, it can be seen that the data is better displayed in narrow or long windows. In these cases, narrow windows are ideal for the EEG signals from healthy people and longer windows for the signals from patients experiencing epileptic seizures.

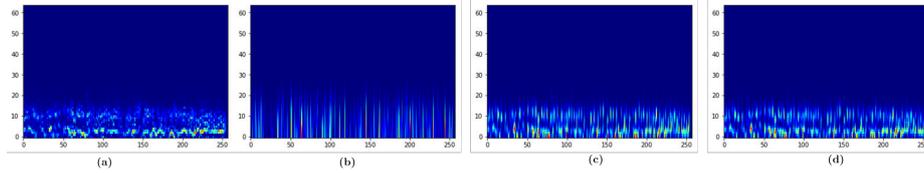


Fig. 11. Spectrograms of EEG signals from a person suffering an epileptic seizure. Window length=64 and 75% overlapping(a) Rectangular, (b) Gaussian , (c) Blackman and (d) Hanning.

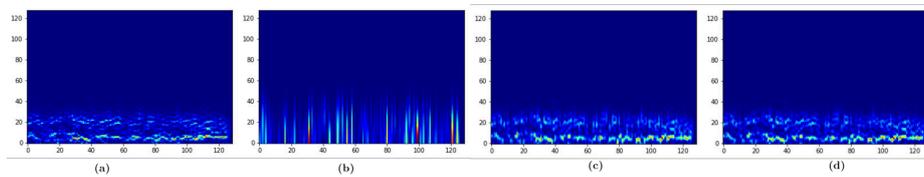


Fig. 12. Spectrograms of EEG signals from a person suffering an epileptic seizure. Window length=128 and 75% overlapping(a) Rectangular, (b) Gaussian , (c) Blackman and (d) Hanning.

3.2 Classification

EEG signals were used to generate the spectrograms. Note that the energy is distributed over time for EEG signals from healthy people, while for a person suffering an epileptic attack, the energy expands along the frequency. Based on this observation, windows could be generated for each phenomenon, as mentioned above. As part of the results, the proposed volume-based methodology was used to extract features which were then evaluated by means of classification algorithms, where the algorithm parameters were selected experimentally; the results are shown in Table 1, where the first column shows the windows used and the second describes the parameters, the first parameter is the size of the window separated by a comma from the overlapping.

Different algorithms were used in the evaluation, such as multilayer perceptron (MLP), linear and polynomial support vector machines (SVM), and 3-nearest neighbors (KNN); this are shown in columns three through six; for this, a 5-fold cross-validation was implemented, and the results are shown in terms of accuracy percentage, i.e., the average from both correctly classified classes. For MLP it had two layers with 100 units each one and 200 epochs, this classifier does not reach the highest results, although its performance reach the most with greater than or equal to 95%. Linear SVM seems to have the best results for all the windows, obtaining 99% with a narrow Gaussian window and overlapping. KNN had a performance similar to linear SVM, with an accuracy of 97.7 % with the same 99% parameter. Finally, the performance of the polynomial SVM was not as high in relation to all windows; however, it achieved a higher accuracy than KNN, with 98.3 % for one case with a Gaussian window.

Table 1 shows the highest accuracy for the Gaussian window; the characteristics of this window make it the most balanced in terms of time-frequency resolution due to the similarity of its spectrum to the window; the results were lower for Blackman and Hanning windows, due to their characteristics, although they are variants of the raised cosine and have a spectrum similar to their windows; as in a Gaussian window, this result could be caused by the overlap. Finally, the rectangular window has the lowest results; this does not mean that it is the worst, but rather that it was not useful for this application. This experiment was an initial phase of a complete analysis, which will be carried out in more depth with other parameters in order to explore their impact on the classification process, adjust them to find the best classifications, and show how they influence the feature extraction phase in a formal training process.

Table 1. Accuracy results from healthy people versus people suffering epileptic seizures, using four different length windows and two overlapping percentages, with four classifiers.

Window	Parameters (size length, overlapping)	MLP (%)	SVM Linear (%)	SVM Polynomial (%)	3-NN (%)
Rectangular	64, 0	84.5	89.7	83.7	86.7
Rectangular	128, 0	88	89.3	88.3	88.3
Rectangular	64, 48	86	87.7	84.3	88.7
Rectangular	128, 96	86.5	79.6	56.6	78.3
Gaussian	64, 0	96.5	96.3	95.7	96.7
Gaussian	128, 0	94.5	94	91.3	91
Gaussian	64, 48	97	99	98.3	97.7
Gaussian	128, 96	96.6	96.6	88.3	96.3
Blackman	92.5, 0	88.3	88.3	90.7	92.3
Blackman	128, 0	90.5	91.6	86.6	88.3
Blackman	64, 48	93.5	96.7	86.6	95.7
Blackman	128, 96	95.5	88.7	75.3	90
Hanning	64, 0	93	90.3	90.7	93
Hanning	128, 0	87	91.6	89	88.7
Hanning	64, 48	95	96.6	91	95.3
Hanning	128, 96	95.5	89.3	77.3	89.3

The reported results were obtained from a general experimentation using some STFT parameters, however, other parameters can be analysed as the energy percentage of the main lobe in relation with side lobes, and the overlapping percentage; whether the parameters proposed are applied to solve another problem the performance will depend of the feature extraction, due to the experimentation show the best results beginning with upper and lower bounds and also the overlapping percentages that best fit the windows.

4 Conclusions

In several works about EEG analysis, different parameters values are fixed in the STFT but they are neither tested nor justified. From a brief theoretical analysis and some experiments, it can be seen that these parameters are important, as they affect the spectrogram. In particular, the width of the window and the overlap of the signal must be taken into account. A good choice of window could be Blackman, Hanning, Hamming, or Gaussian, due to their properties and spectral behaviors. Gaussian window showed the best performance, and the next were Blackman and Hanning; rectangular window was the lowest in the classification, our experiments are based on peaks, thus, probably secondary lobes is affecting the main lobe, which peak is obtained.

Due to the energy shown in the spectrograms, a window should probably be proposed for each type of class, since the windows require different lengths if energy is shown over time or frequency. In this paper, we noted that different parameter affect spectrogram and performance, hence, these have to be selected with a previous analysis, an experimental analysis could be not enough.

As future work, other data sets will be analyzed, since only two signals from two classes from the epilepsy data database were analyzed in these experiments. The experimentation in this paper for different parameters of the STFT was very general and in future experimentation we consider to analyze the windows with respect to spectrum, energy percentage, lobes, and other factors. The energy percentage of the main lobe with respect to the side lobes could be a good parameter. In addition window lengths will be proposed based on the signal frequencies.

Acknowledgments. The first author thanks the support from CONACyT-Mexico through the scholarship number 457637.

References

1. Alçın, Ö.F., Siuly, S., Bajaj, V., Guo, Y., Şengür, A., Zhang, Y.: Multi-category eeg signal classification developing time-frequency texture features based fisher vector encoding method. *Neurocomputing* 218, 251 – 258 (2016)
2. Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., Elger, C.E.: Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E* 64, 1–8 (2001)
3. Gandhi, V.: Chapter 2 - interfacing brain and machine. In: Gandhi, V. (ed.) *Brain-Computer Interfacing for Assistive Robotics*, pp. 7 – 63. Academic Press, San Diego (2015)
4. Goodwin, M.M.: *The STFT, Sinusoidal Models, and Speech Modification*, pp. 229–258. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
5. Grossi, E., Olivieri, C., Buscema, M.: Diagnosis of autism through eeg processed by advanced computational algorithms: A pilot study. *Computer Methods and Programs in Biomedicine* 142, 73 – 79 (2017)

6. Ilyas, M.Z., Saad, P., Ahmad, M.I.: A survey of analysis and classification of eeg signals for brain-computer interfaces. In: 2015 2nd International Conference on Biomedical Engineering (ICoBE). pp. 1–6 (2015)
7. Kovacs, P., Samiee, K., Gabbouj, M.: On application of rational discrete short time fourier transform in epileptic seizure classification. In: 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 5839–5843 (2014)
8. Kumar, J.S., Bhuvaneshwari, P.: Analysis of electroencephalography (eeg) signals and its categorization study. *Procedia Engineering* 38, 2525 – 2536 (2012), international Conference on Modeling Optimization and Computing
9. Mateo, C., Talavera, J.A.: Short-time fourier transform with the window size fixed in the frequency domain. *Digital Signal Processing* 77, 13 – 21 (2018), digital Signal Processing and SoftwareX - Joint Special Issue on Reproducible Research in Signal Processing
10. Mustafa, M., Taib, M.N., Murat, Z.H., Hamid, N.H.A.: Glcm texture classification for eeg spectrogram image. In: 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES). pp. 373–376 (2010)
11. Prabhu, K.: *Window Functions and Their Applications in Signal Processing*. CRC Press (2013)
12. Ramos, R., Olvera-López, J.A., Olmos, I.: Analysis of EEG signal processing techniques based on spectrograms. *Research in Computing Science* 145, 151–162 (2017)
13. Ridouh, A., Boutana, D., Bourennane, S.: Eeg signals classification based on time frequency analysis. *Journal of Circuits, Systems and Computers* 26(12), 1–26 (2017)
14. Şengür, A., Guo, Y., Akbulut, Y.: Time–frequency texture descriptors of eeg signals for efficient detection of epileptic seizure. *Brain Informatics* 3(2), 101–108 (2016)
15. Sethares, W.A.: *Rhythm and Transforms*. Springer London, London (2007)
16. Suly, S., Li, Y., Zhang, Y.: *Electroencephalogram (EEG) and Its Background*, pp. 3–21. Springer International Publishing, Cham (2016)

Electronic edition
Available online: <http://www.rcs.cic.ipn.mx>

