Advances in Pattern Recognition

Research in Computing Science

Series Editorial Board

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 96,** junio 2015. Tiraje: 500 ejemplares. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No. : 04-2005-121611550100-102, expedido por el Instituto Nacional de Derecho de Autor. *Certificado de Licitud de Título* No. 12897, *Certificado de licitud de Contenido* No. 10470, expedidos por la Comisión Calificadora de Publicaciones y Revistas Ilustradas. El contenido de los artículos es responsabilidad exclusiva de sus respectivos autores. Impreso en la Ciudad de México, en los Talleres Gráficos del IPN – Dirección de Publicaciones, Tres Guerras 27, Centro Histórico, México, D.F. Distribuida por el Centro de Investigación en Computación, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, México, D.F. Tel. 57 29 60 00, ext. 56571.

Editor responsable: *Grigori Sidorov, RFC SIGR651028L69*

Research in Computing Science is published by the Center for Computing Research of IPN. **Volume 96**, June 2015. Printing 500. The authors are responsible for the contents of their articles. Printed in Mexico City, in the IPN Graphic Workshop – Publication Office.

Volume 96

Advances in Pattern Recognition

José Arturo Olvera-López
Jesús Ariel Carrasco-Ochoa
José Francisco Martínez-Trinidad
Juan Humberto Sossa-Azuela
Fazel Famili (eds.)

Instituto Politécnico Nacional, Centro de Investigación en Computación
México 2015

ISSN: 1870-4069

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

Indexed in LATINDEX and Periodica / Indexada en LATINDEX y Periódica

Printing: 500 / Tiraje: 500

Printed in Mexico / Impreso en México

Editorial

The 2015 Mexican Conference on Pattern Recognition (MCPR 2015, June 24-27) was the seventh event in the series. The conference was jointly organized by the Center for Computing Research of the National Polytechnic Institute (CIC-IPN) and the Computer Science Department of the National Institute for Astrophysics Optics and Electronics (INAOE), under the auspices of the Mexican Association for Computer Vision, Neurocomputing and Robotics (MACVNR), which is affiliated to the International Association for Pattern Recognition (IAPR). MCPR series of conferences aim to provide a forum for the exchange of scientific results, practice, and new knowledge, as well as, promoting co-operation among research groups in Pattern Recognition and related areas in Mexico and around the world.

This year MCPR included the third Postgraduate Students' Meeting (MCPR2015-PSM) allowed discussing their research work in order to receive feedback from experienced researchers and advices for future directions, as well as promoting their participation in conference events.

This volume contains original contributions carefully selected which are derived from both Master and PhD students' researches about Pattern Recognition and related areas. We cordially thank all authors who submitted their contributions to build this volume as well as the Reviewing committee for evaluating the submissions that were received.

We hope this volume from the MCPR2015-PSM will be useful to the reader, and hope that the meeting itself will provide a fruitful forum to enrich the collaboration between students and the broader Pattern Recognition community.

The submission, reviewing, and selection process was supported for free by the EasyChair system, www.easychair.org.

José Arturo Olvera-López
Jesús Ariel Carrasco-Ochoa
José Francisco Martínez-Trinidad
Juan Humberto Sossa-Azuela
Fazel Famili
June 2015

Table of Contents

Page

Comparing Evolutionary Strategy Algorithms for Training Spiking Neural Networks

José S. Altamirano, Manuel Ornelas, Andrés Espinal, Raúl Santiago, Héctor Puga, Martín Carpío, and Sergio Tostado

Tecnológico Nacional de México, Instituto Tecnológico León, León, Gto., México
josesaltf@gmail.com,mornelas67@yahoo.com.mx,
andres.espinal@itleon.edu.mx

Abstract. Spiking Neural Networks are considered as the third generation of Artificial Neural Networks, these neural networks naturally process spatio-temporal information. Spiking Neural Networks have been used in several fields and application areas; pattern recognition among them. For dealing with supervised pattern recognition task a gradient-descent based learning rule (Spike-prop) has been developed, however it has some problems like no convergence. To overcome these problems, metaheuristic algorithms such as Evolutionary Strategy have been used. In this work, three variants of the Evolutionary Strategy algorithm are compared for training Spiking Neural Networks. Several well-known benchmark dataset are used to test the capabilities of the algorithms.

Keywords: Spiking Neural Network, Evolutionary Strategy, Pattern Recognition.

1 Introduction

The Artificial Neural Networks (ANNs) are capable of modeling complex no linear systems, and can be used to solve a great number of day-to-day problems such as pattern recognition, optimization, prediction, function approximation, etc. [1].

In the last years, the third generation of ANN [2], Spiking Neural Networks (SNNs), have gained importance due to the inclusion of the firing time component in the neuron's process. This is obtained by coding the information in spike trains instead of spike rates as in the Second Generation of ANNs. That makes SNNs more similar to the biological neurons [3,4,5], and increases their computational power [6].

For the training process of SNNs, there has been developed a gradient-descent based learning rule, Spikeprop [7]. However it has some drawbacks such as the limitation on using negative weight values, convergence not guaranteed due to its tendency to end trapped in local minima, etc. [8].

To overcome these disadvantages, there had been some works about the use of metaheuristic algorithms for the learning process of the SNN [9,10,8,11]. In this work we compare the performance of three variants of Evolutionary

Strategy algorithms for the training process of SNN by testing three classical benchmarks data sets: Breast Cancer Wisconsin, Iris Plant and Wine, (from the UCI Repository [12]).

This document is organized as follow: Section 2 gives fundamentals for simulating SNNs. Section 3 explains the implemented methodology used for training SNNs. The experimental design and results are showed in Section 4. Finally, in Section 5 conclusions and future work are presented.

2 Spiking Neural Networks

A neural network can be defined as an interconnection of neurons, such that neuron outputs are connected to other neurons, even with themselves; both lag-free and delay connections are allowed [13]. There are several models or topologies of ANNs, which are defined around three aspects: computing nodes, communication links and message types [14].

In this work a fully-conected feed-forward SNN was used, which is defined as follows: the computing nodes are spiking neurons defined by the Spike Response Model (SRM), the communication links are formed by synaptic weights (excitatories and inhibitories) and positive delays values, and the message types are ruled by the time-to-first spike coding scheme.

2.1 Spike Response Model

The SRM has been introduced in [15], and it is an approximation of the dynamics of integrate-and-fire neurons. The neuron status is updated through a linear summation of the postsynaptic potentials resulting from the impinging spike trains at the connecting synapses. A neuron fires whenever its accumulated potential reaches the threshold (θ) from below (Fig. 1).

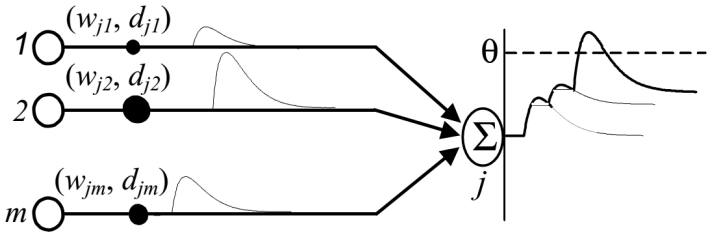


Fig. 1. Weighted input summed at the target neuron. Taken from [9]

In the SRM is consider that a neuron j has a set Γ_j of immediate predecessors called presynaptic neurons and receives a set of spikes with firing times t_i; $i \in \Gamma_j$. The internal state of a neuron is determined by Eq. (1), where w_{ji} is the synaptic

weight to modulate $y_i(t)$, which is the unweighted postsynaptic potential of a single spike coming from neuron i and impinging on neuron j.

$$x_j(t) = \sum_{i \in \Gamma_j} w_{ji} y_i(t) \tag{1}$$

The unweighted contribution $y_i(t)$ is given by Eq. (2), where $\epsilon(t)$ defines a spike response function describing a standard form of the postsynaptic potential.

$$y_i(t) = \epsilon(t - t_i - d_{ji}) \tag{2}$$

The function $\epsilon(t)$ is modeled by Eq. (3)

$$\epsilon(t) = \frac{t}{\tau} e^{1-(t/\tau)} \qquad for\ t > 0,\ else\ \epsilon(t) = 0 \tag{3}$$

where: t is the current time, t_i is the firing time of the presynaptic neuron i and d_{ji} is the associated synaptic delay. Finally the function has a τ parameter, that is the membrane potential time constant and define the decay time of the postsynaptic potential. Both θ and τ are constant and equal for all the neurons.

3 Metaheuristic Based Supervised Learning

Learning is a process by which the free parameters of a neural network are adapted through a process of stimulation from the environment in which the network is embedded. The type of learning is determined by the manner in which the parameter changes take place [16]. In this case, the learning is driven by an Evolutionary Strategy algorithm, and we refer to this learning process as Metaheuristic Based Supervised Learning.

In Metaheuristic-Based Supervised Learning, each individual contains all the free parameters of a previously structured SNN. Every individual is evaluated by means of a fitness function. To calculate the individual's fitness value the following steps are requiered: the first step makes a mapping process; this sets the individuals parameter as weights and delays in the SNN (Fig. 2). The second step uses the batch training as learning protocol, where all patterns are presented to the network before the learning takes place [17]. The third step is to calculate an error (to be minimized) according Eq. (4) (taken from [9]); where T are all training patterns, O are all output spiking neurons, $t_o^a(t)$ is the current timing output of the SNN and $t_o^t(t)$ is the desired timing output. The error calculated using Eq. (4) determines the fitness value of each individual and drives the supervised learning based on metaheuristic algorithms.

$$E = \sum_{t}^{T} \sum_{o \in O} \left(t_o^a(t) - t_o^t(t) \right)^2 \tag{4}$$

Next are presented the variants of Evolutionary Startegies used for training SNNs.

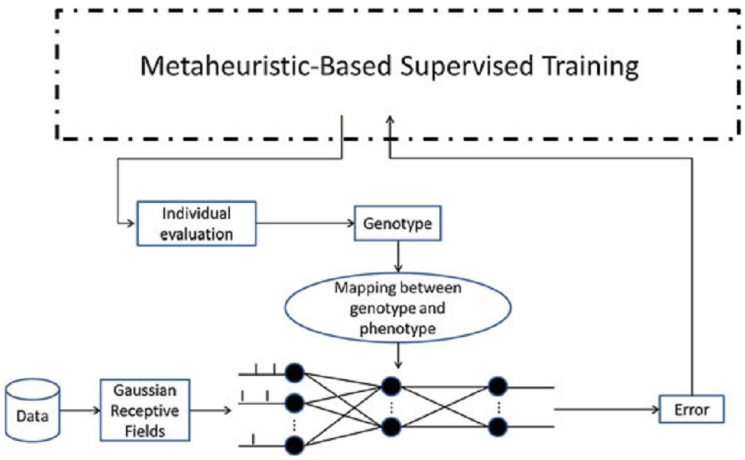


Fig. 2. Generic scheme for training SNNs with metaheuristic algorithms. Taken from [11]

3.1 Evolutionary Strategy

Evolutionary Strategy (ES), a variant of the Evolutionary Algorithms, was founded by students at the Technical University of Berlin (TUB), and although in the beginning it was not devised to compute minima or maxima of real-valued functions, it has proved to produce competitive solutions in such space ([18,19]). Next are presented the ES variants used in this work.

Evolutionary Strategies($\mu + \lambda$), (μ , λ) and Evolutionary Strategy There exists some variants for the Evolutionary Strategy, some of which depend on the selection for the new population, and others have a different mutation method.

One of the variants is the $(\mu + \lambda)$-ES, in which from a population of μ parents, there is generated λ descendants that are added to the original population, and, to keep the population size constant, the worst out of all $\mu + \lambda$ individuals are discarded [20].

Another variant is the (μ,λ)-ES, where from a population of μ parents, there is generated λ descendants and the selection takes place only among the generated offsprings, whereas their parents are forgotten no matter how good or bad their fitness was compared to that of the new generation. This strategy requires that $\lambda > \mu$ [20].

The third variant that is included in this work is a slightly modified Evolutionary Strategy (modified-ES), which is similar to the previous ones mentioned, but where the reproduction stage is not necessary, and there is the possibility of another type of operator for the mutation (Cauchy distribution mutation) [9].

The algorithm includes the parameter ρ, which is the number of parents that are going to take part in the reproduction. In the case of the first two variants, we considered a number of two parents giving the configuration $(\mu/2\overset{+}{,}\lambda)$; and $(\mu/1+\lambda)$ for the modified-ES, because only one parent was used.

The Algorithm 1 is based in [18], with some modifications to make it more general. The representation of each individual (χ) is composed of the object variables $(x_1, \ldots, x_n)$, being n the dimension of the problem, and some strategy parameters $(\eta_1, \ldots, \eta_n)$ of the mutation operator (the standard deviations). In the $(\mu/2\overset{+}{,}\lambda)$ version, there is a parent's selection, a recombination and a final modification using a random number from a Normal Distribution. On the other hand, in the $(\mu/1+\lambda)$ version the mutations are applied directly on every individual to generate the offspring. Finally, the population is replaced according to the applied version.

Algorithm 1 $(\mu \ / \ \rho \overset{+}{,} \lambda)$-ES

```
Begin
g ← 0
Initialize and evaluate pop_μ^g = {χ_i | i = 1, ..., μ}
repeat
    for l=1 to λ do
        if variant != modified-ES then
            parents = marriage (pop_μ^g, ρ) // selection through binary tournament
            χ̃_i= recombination(parents) // using intermediate recombination
            χ̃'_i= mutationNormal(X̃_i) // mutation using a Normal distribution
        if variant == modified-ES then
            r ← U[0, 1]
            if r < 0.5 then
                χ̃'_i= mutationNormal(χ̃_i) // mutation using a Normal distribution
            else
                χ̃'_i= mutationCauchy(χ̃_i) // mutation using a Cauchy distribution
    if typeSelection == (μ, λ) then
        pop_μ^{g+1} = selection(pop_λ^g) // Replace population with the created offspring
    else if typeSelection == (μ + λ) then
        pop_μ^{g+1} = selection(pop_μ^g,pop_λ^g) // Select new population from both the parents
        // and offspring populations
    g ← g + 1
until Stopping criteria
End
```

The marriage refers the way in which the ρ parents will be selected, in this work it was determined by using binary tournament selection. The intermediate recombination was made using Eq. (5) for the object variants and Eq. (6) for the standard deviations:

$$\widetilde{x}_l(j) = r\widetilde{x}_{r_1}(j) + (1-r)\widetilde{x}_{r_2}(j) \qquad \forall j \mid j = 1, \ldots, n \tag{5}$$

$$\widetilde{\eta}_l(j) = r\widetilde{\eta}_{r_1}(j) + (1-r)\widetilde{\eta}_{r_2}(j) \qquad \forall j \mid j = 1, \ldots, n \tag{6}$$

where r is a U[0,1] and r_1 and r_2 are the parents selected in the marriage. In the case of modified-ES it is not necessary to choose the parents due to the fact that each offspring is only a mutation of one parent.

The Normal mutation is made using the Eq. (7), and the Cauchy mutation for the modified-ES is made using Eq. (8). The standard deviations updates are part of each mutation and are made using Eq. (9).

$$\widetilde{x}'_l(j) = \widetilde{x}_l(j) + \widetilde{\eta}'_l(j)N(0,1) \tag{7}$$

$$\widetilde{x}'_l(j) = \widetilde{x}_l(j) + \eta(j)\delta_j \tag{8}$$

$$\widetilde{\eta}'_l(j) = \widetilde{\eta}_l(j)e^{\tau' N(0,1)+\tau N_j(0,1)} \tag{9}$$

Where:

- n represent the problem dimension
- $\tau' = 1/sqrt(2 \times (n))$ and $\tau = 1/sqrt(2 \times sqrt(n))$
- *N(0, 1)* denotes a normally distributed one dimensional random number with mean 0 and standard deviation 1. $N_j(0,1)$ indicates that the random number is generated anew for each value of j.
- δ_j is a Cauchy random variable, and it is generated anew for each value of j (Scale = 1).

The selection for the $(\mu + \lambda)$ includes elitism, to keep track of the better individuals, and is made through tournament selection.

4 Experiments and Results

Three classical benchmarks of pattern recognition from UCI[12] were used for experimentation: Brest Cancer Wisconsin (BCW), Iris Plant and Wine dataset.

4.1 Brest Cancer Wisconsin

The BCW data set consists of 683 samples belonging to two groups, namely benign and malignant cell tissues. Each data point is described with 9 attributes, represented by an integer ranging from 1 to 10 with larger numbers indicating a greater likelihood of malignancy. The data set is split into two parts, training and test data sets with 342 and 341 samples in each set respectively. The desired timing outputs where set to 6ms. for the benign class, and 10ms. for the malign class.

4.2 Iris Plant

The Iris plant dataset contains 3 classes of which 2 are not linearly separable, each class is formed by 50 patterns, where each one of them is described by 4 features. The desired timing outputs for setosa, versicolor and virginica classes are respectively 6, 10 and 14 ms.

4.3 Wine Data Set

These data are the results of a chemical analysis of wines grown in the same region in Italy but from three different cultivars. The analysis determined the quantities of 13 constituents (variables) found in each of the three types of wines. The desided timing outputs for each clase are 6ms. for class 1, 10ms. for class 2 and 14ms. for class 3.

4.4 Experimental Methodology

Due to computing times and statistical reasons, the experiments were carried out by applying 33 independently trainings for each dataset with every variant of Evolutionary Strategy. For feeding the SNN, pattern's dataset were codified by using four Gaussian Receptive Fields (GRFs) [7]. The SNN configuration for all problems was as follows: the neurons input layer depends on the GRFs, which varies by dataset features, 10 neurons into the hidden layer and 1 neuron into the output layer. All neurons from hidden and output layers had $\tau = 9$ and $\theta = 1$. The simulation time was 20ms. [11].

The Evolutionary Strategy configuration for all experiments was: $\mu = 30$, $\lambda = 30$ individuals, 15000 function calls as end criteria, and initial Standard deviation in the range $U[0, 1]$, which were choosen by empirical experimentation. The weight boundaries were [-1000, 1000] and the delay boundaries were [0.1,16] [9].

Table 1 shows the best fitness values achieved for each ES in each dataset over 33 training runs. The classification performance for both training and testing sets by the using the best results achieved by each ES are showed in Table 2.

Table 1. Results of the best fitness values for the training process of SNNs

	Fitness		
Data Set	(modified-ES)	$(\mu + \lambda)$-ES	(μ, λ)-ES
BCW	32	29	33
Iris Plant	0	16	15
Wine	6	21	14

Table 2. Comparison of the classification performance for the trained SNNs

	Training Set			Test Set		
Data Set	(modif-ES)	($\mu + \lambda$)-ES	(μ, λ)-ES	(modif-ES)	($\mu + \lambda$)-ES	(μ, λ)-ES
BCW	95.1%	94.13%	94.72%	95.91%	96.78%	95.61%
Iris Plant	100.0%	89.13%	84.0%	94.67%	73.33%	73.33%
Wine	93.18%	76.14%	90.91%	83.33%	60.0%	92.22%

The results show that the modified-ES version had a better performance in the training process for the three data sets, being able to achieve 100% in the classification for the training set of the Iris Plant dataset. On the other hand, even with lower fitness performance, the (μ, λ)-ES achieved good classification in the BCW and Wine datasets. The $(\mu+\lambda)$-ES version only had good classification performance in the BCW dataset.

5 Conclusions

This work compares three metaheuristics on the training of SNNs, and under the experiment circumstances, it was visible that even when the achieved fitness value was not too low, it is possible to obtain acceptable classification performance.

Based on the best results, the modified-ES showed better performance on both fitness value and classification.

For future work it is proposed to conduct experimentations with more metaheuristics and in more data sets, aiming for a more robust statistical analysis. Also we propone the research for some different fitness functions, and investigate the use of Grammar Evolution and Genetic Programming to evolve the neural network's structure.

Acknowledgments. Authors thank to Tecnológico Nacional de México, Instituto Tecnológico de León. The first author wants to thank to Consejo Nacional de Ciencía y Tecnología (CONACYT) for the economical support to his MS work.

References

1. Jain, A., Mao, J., Mohiuddin, K.: Artificial neural networks: a tutorial. Computer 29(3), 31–44 (Mar 1996). URL http://dx.doi.org/10.1109/2.485891
2. Maass, W.: Networks of spiking neurons: The third generation of neural network models. Neural Networks 10(9), 1659–1671 (Dec 1997). URL http://dx.doi.org/10.1016/S0893-6080(97)00011-7
3. Abeles, M.: Corticonics: Neural circuits of the cerebral cortex. Cambridge: Cambriedge University Press (1991)
4. Abeles, M., Prut, Y.: Spatio-temporal firing patterns in the frontal cortex of behaving monkeys. Journal of Physiology-Paris 90(3-4), 249–250 (Jan 1996). URL http://dx.doi.org/10.1016/S0928-4257(97)81433-7

5. Hopfield, J.J.: Pattern recognition computation using action potential timing for stimulus representation. Nature 376(6535), 33–36 (Jul 1995). URL http://dx.doi.org/10.1038/376033a0
6. Maass, W.: Noisy spiking neurons with temporal coding have more computational power than sigmoidal neurons. Advances in Neural Information Processing Systems 9, 211–217 (1997)
7. Bohte, S.M., Kok, J.N., La Poutr, H.: Error-backpropagation in temporally encoded networks of spiking neurons. Neurocomputing 48(1-4), 17–37 (Oct 2002). URL http://dx.doi.org/10.1016/S0925-2312(01)00658-0
8. Belatreche, A.: Biologically Inspired Neural Networks: Models, Learning, and Applications. VDM Verlag Dr. Mller, Saarbrcken (2010)
9. Belatreche, A., Maguire, L.P., McGinnity, M., Wu, Q.: An evolutionary strategy for supervised training of biologically plausible neural networks. In: The sixth international conference on computational intelligence and natural computing (CINC), proceedings of the 7th joint conference on information sciences, pp. 1524–1527. USA (2003)
10. Belatreche, A., Maguire, L.P., McGinnity, M.: Advances in design and application of spiking neural networks. Soft Computing 11(3), 239–248 (Oct 2006). URL http://dx.doi.org/10.1007/s00500-006-0065-7
11. Espinal, A., Carpio, M., Ornelas, M., Puga, H., Melin, P., Sotelo-Figueroa, M.: Comparing metaheuristic algorithms on the training process of spiking neural networks. Studies in Computational Intelligence pp. 391–403 (2014). URL http://dx.doi.org/10.1007/978-3-319-05170-3_27
12. Lichman, M.: UCI machine learning repository (2013). URL http://archive.ics.uci.edu/ml
13. Zurada, J.: Introduction to Artificial Neural Systems. West Publishing Co., St. Paul, MN, USA (1992)
14. Judd, J.: Neural network design and the complexity of learning. Neural Network Modeling and Connectionism Series, Massachusetts Institute Technology (1990)
15. Gerstner, W., Kistler, W.M.: Spiking Neuron Models: Single Neurons, Populations, Plasticity. Cambridge University Press (2002)
16. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice Hall PTR, Upper Saddle River, NJ, USA (1998)
17. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification (2Nd Edition). Wiley-Interscience (2000)
18. Beyer, H.G., Schwefel, H.P.: Evolution strategies. a comprehensive introduction. Natural Computing 1(1), 3–52 (2002). URL http://dx.doi.org/10.1023/A:1015059928466
19. Rechenberg, I.: Evolutionsstrategie: optimierung technischer systeme nach prinzipien der biologischen evolution. Frommann-Holzboog (1973)
20. Schwefel, H.P.: Numerische optimierung von computer-modellen mittels der evolutionsstrategie (1977). URL http://dx.doi.org/10.1007/978-3-0348-5927-1

Advances in the Study of Hand Gesture Recognition Systems for Human Computer Interaction

P. Rodrigo Díaz-Monterrosas, Rubén Posada-Gómez, and Albino Martínez-Sibaja

Tecnológico de Orizaba, División de Estudios de Posgrado e Investigación, Orizaba, Veracruz, México
diaz.monterrosas@gmail.com, pgruben2@hotmail.com, albino_mx@yahoo.com

Abstract. Getting three-dimensional pose and orientation of parts of the body observed by one or more cameras is of great theoretical interest and widely applicable. Usually, computing devices interaction is accomplished by means of a mouse and a keyboard or by touching the screen, but otherwise, human beings relate to their surrounding world using hands, body, and voice in most of their daily activities, therefore, development of more natural and intuitive techniques for interacting with a variety of user interfaces is critical. In this paper, a review of recent research efforts in Human Computer Interaction (HCI), specifically in hand gesture recognition, is performed, analyzing the state-of-the-art methodology and discussing some important issues about.

Keywords: Hand Gesture Recognition, Human Computer Interaction, Image Processing, Computer Vision.

1 Introduction

Analyzing the different techniques used in literature to achieve location, tracking, description, and object recognition, has led to the development of tools that tend to improve robustness and naturalness in handling HCI devices to be fully functional in real world. Taking hands, face, body, voice, or even the eyesight as objects of study, research evolves allowing Douglas Engelbart's augmentation dream to increasingly become tangible, and disciplines such as Artificial Intelligence (AI) whose philosophy since its beginning has been considered as opposed to HCI's vision, devote part of its efforts to study and try to simulate human communication processes with the same purpose: interact in a friendlier way with machines. Among main achievements of these disciplines, Automatic -visual, speech, gesture, etc.- Recognition Systems designed to recognize and translate what they hear and/or see (voice, lips or body movements, facial gestures, hand signs or other objects movements) via microphones, video cameras or other sensors, could be mentioned.

Recently, new technologies such as RGB-D cameras, which have sensors to capture RGB images along with depth information of each pixel, have been taken

into account. These high precision cameras are capable of delivering high-quality three-dimensional information (color and depth) to understand the whole shape of an object [1]. For example, Microsoft® Kinect® technology allows players to enjoy video games simply by moving their bodies in front of a screen, the data taken as input from the device are the tracked body skeleton. The extension of this technology to individual finger 3D tracking is an active research area, having more complete information on user's hand pose would lead to grasping, pointing, and manipulation capable richer applications [2]. However, despite the growing amount of research in the area, there are still existing problems having a variety of theoretical and practical challenges.

The following section describes some actual marketed devices features, regarding their interaction drawbacks. The major current challenges in image-based gesture recognition are cited and later the resolution methods generally reported in literature are defined along with a description of some of the most representative recent projects. Finally, a future contribution and discussion to the described state-of-the-art is submitted.

2 Existing Devices, their Drawbacks and Health Impact

Nowadays, there are several devices with the adjective "smart" on them: digital still and/or video cameras with facial/smile recognition and tracking, automatic washing machines capable of weighing clothes and dosing optimum amount of water and detergent together with the estimation of washing time, computers and smartphones which light-up when passing the hand over them or with a voice command, or the new smart TVs, very trendy by additionally providing a web browser, access to several services, and likewise have the ability to perform all typical orders of their modern remote control by gestures or voices.

Even with such technological advances, these devices are far from being even slightly smart at least for now, because although they averagely meet what they promise, there are more than a few shortcomings in their performance, causing most of the time user prefers to disable smart features, and use traditional interaction ways (it might also be the case of the Leap Motion Controller, described as "rather limited" in [3]). For some smart TVs, for example, keeping alive gesture detection feature implies the stress of its continuous unwanted activation simply by raising arms or doing a hand sign to someone else in the room. Voice detection behaves similarly, activating itself even by the loudness of a movie dialog. As if this were not enough, when someone does want to use these features, orders hardly execute at first attempt, having the user to desperately repeat the gesture or key word once and again.

Clearly, at the moment of this study, there are still many challenges related to accuracy, speed, naturalness, and even usefulness of these devices, since although occasional user might be pleased or amazed, for regular user the continuous use of features such as gesture recognition, for example, could harm his health and comfort due to a phenomenon known as "gorilla arm syndrome", a problem that arises from continued use of the arms in the air, that is, without a place to

stand, causing a feeling of heaviness, fatigue, or discomfort (imagine a designer working eight hours on a gesture based software). On the other hand, it is true that this is a less significant problem if a similar disease suffered for decades due to extensive use of the mouse, known as "carpal tunnel syndrome", comes to mind (a hand and arm condition caused by the inflammation of a tendon in the wrist and triggering chronic pain), yet, it still remains one of the most used interaction device.

3 Current Challenges in Hand Gesture Recognition Area

Leaving aside negative issues on the development of image-based gesture recognition devices, there have been challenging factors since the beginning, such as the high dimensionality and speed of hand motion, while image capturing is performed through low resolution cameras, besides the diversity of existing cameras that hinder calibration or standard lighting conditions, the ambiguity of image elements identification due to its color uniformity, the similarity of fingers, the large number of degrees of freedom (DOF), or the absence of observations when parts of hands (or other object) obstruct each other. Searching for solutions, special hardware for motion capture has been used, such as magnetic tracking devices, bracelets with optical [4] or electromyographic [5] sensors, and visual markers placed on gloves [6] or the bare hands. Unfortunately, such methods require complex and expensive hardware, interfere with the observed scene, or add restrictions to user's pose, preventing their use in real world [7], not to mention their use for people with disabilities. Moreover, emerging projects addressing hand tracking interacting with objects, have increased the challenge, since although they can help to reduce the number of potential poses, there are limitations still being solved, such as the desirability of the hardiness and non-similar to skin color of the object.

4 Classification of Hand Gesture Recognition Methodology

From a historical perspective, starting with the development of articulated hands to explore issues related to grip and object manipulation at early 80s, a growing attention from a variety of disciplines to modeling, 3D simulation, tracking, and interpretation of hands and other body parts motion has been found (it is accepted that the study of hand tracking began with the Zimmerman et al "VPL Dataglove" [8]). Several applications have emerged from these studies, which have been useful to many science and technology areas, to name a few: medicine and biotechnology, robotics, computer animation, movies, e-commerce, and virtual reality.

At first systems, selection of key points (articular joints) was performed manually on the computer screen. Obviously those had serious restrictions, the most significant: selection subjectivity, process slowness, and calibration

system sensitivity. Such systems, typical of the 90s, are deprecated, and from the first decade of 21st century, finding semi-automatic methods with reflective, magnetic or infrared light sensitive optical markers is usual, allowing, when scanned, determine joint connections. Currently, research is mainly oriented to the development of processes that avoid the use of invasive elements.

Depending on the type of input, gesture translation approaches may vary. However, most techniques are based on key clues represented in a 3D coordinate system, by tracking their relative motion a gesture can be detected with high precision. Several methods have been used in the literature to estimate the hands pose, and have been classified by some researchers according to certain common properties; regarding the output completeness, Erol et al [9] describe them as "Partial hand pose estimation methods" which can be viewed as extensions of "Appearance-based methods" to provide information on continuous motion in navigation, handling or pointing; and "Full DOF hand pose estimation methods", which get all the kinematic parameters of the skeleton of the hand, such as joint angles and hand position or orientation for a full reconstruction of hand motion. The latter class is divided into: "Model-based tracking", which can be subdivided into methods using a single hypothesis and those who manage multiple hypotheses, and "Single frame pose estimation", that is, they are not committed to time coherence. Both full DOF hand pose estimation classes are addressed in [10].

4.1 Appearance-based Methods

This approach, also known as "Discriminative", uses classification or regression techniques directly into the image data. An offline training process is used to establish a nonlinear mapping (due to the different hand views) from the image feature space to a finite set of hand poses, depending on specific parts of the hand, such as palm or fingertips and their orientation. These methods process each image independently, but may be used with image sequences; they work well when recognition of a small well-known and distinct hand configuration set is required and are not recommended when there is free hand motion and high recognition accuracy is required. Velocity, offline training, computationally efficient online execution, low computational cost and hardware complexity, the requirement of a single camera, and generalization if training is suitable are some of their advantages. Their inherent disadvantages lie in the need for very large training data sets and that their accuracy and reduced number of hand recognition poses rely on those data, therefore requiring high degree of user intervention.

Recent research involving two-hand recognition introduce a new challenge if this approach is used, due to the fact that the offline training must include the combinatorial space of both hands configuration and the changes that different points of view cause in their appearance.

4.2 Model-based Methods

These methods are called "Generative" because they generate hypothetical 2D or 3D hand models and compare model projection to the observed images. This is done through an optimization problem whose objective function measures the discrepancy between the model key indicators and the observations, however, the optimization method should be able to evaluate the objective function at arbitrary points in the multidimensional space of model parameters, so the search must be carried online, causing a high computational cost, which is their major drawback, besides relying entirely on visual information available, usually provided by a multi-camera system. On the other hand, these aspects also imply their major strengths: there is no training need and they can easily be extended to any gesture recognition problem. If researcher decides to use this approach, dimensional reduction of the configuration space, efficient construction of realistic three-dimensional hand models, and development of quick and reliable estimating techniques would be interesting contributions [11].

The usual visual features to match are silhouettes, edges, shades, color, optical flow, and recently depth. Among the optimization techniques that have been proposed are, to name a few, belief propagation, particle swarm optimization, and local optimization, one of the first and still used because of its efficiency. Similarly, stochastic optimization techniques such as Kalman filter and particle filter have been used, the latter together with local optimization in [12]. In [13] and [14] linear subspaces are used to reduce the hand pose space.

In short, appearance-based methods allow fast processing with a loss of generality, whereas model-based ones give generality at a high computational cost.

Another classification is based on how partial evidence of individual rigid parts of an articulated object contributes to the final solution [15]:"Disjoint evidence methods" consider individual parts in isolation before evaluating them against observations, usually requiring less computational power but needing to handle explicitly part interactions (such as collisions and occlusions); in contrast, "Joint evidence methods", consider all parts in the context of full object hypotheses, their computational requirements are high, but part interactions do not represent much of a problem.

5 Current Research in Literature

Oikonomidis et al [11] introduce a model based multiple-view method to recover 3D position of the hand given by 27 geometric primitives that redundantly encode a 26 DOF 3D hand model. Observations are acquired from a static, pre-calibrated camera network, computing reference features for each acquired view based on skin color and edge detection. Mapping of these features is rendered and compared directly with the respective view. Discrepancy between a 3D hand pose and the actual observation is quantified by an error function minimized through particle swarm optimization. The pose for which this error

function is minimal constitutes the output of the proposed method at a given moment in time. As a temporal continuity in hand motion is assumed, initial hypotheses for current time instance are restricted in the vicinity of the previous time instant solution. Being computationally expensive, the method is implemented in a GPU, resulting in near real-time performance. Their study is improved in [15] using Kinect® and a single hypothesis, where observation is the RGB-D image segmented to locate the hand through skin color and depth of the scene; therefore, error function is different, computational requirements are lower, camera array is simplified, and resulting system works even under variable lighting conditions.

On systems that do not handle occlusions or interactions with other objects, certainty in estimating hand positions is seriously affected, so the role of context in object recognition is very significant [16]. Several researchers have tried to exploit the contextual constraints on Computer Vision problems, in [17] a brief count of researching work considering the context in the classification of human-object interaction activities can be found, differentiating between those who have focused on the human body or hands and those who provide a detailed 3D model of them and the object. This project is an extension of that presented in [11] by considering jointly the hand and the manipulated object. It is an optimization problem whose solution is the 26 DOF hand pose along with the pose and parameters of the manipulated object model using a multi-camera system. In each of the acquired observations, skin color maps and edges of the hand are extracted; depending on the point of view, the presence of an object can obstruct the presence of the hand, their incomplete observation provides evidence of the type and pose of the manipulated object and at the same time the object improves the estimation of hand pose. The process seeks the hand-object model that best explains the incompleteness of the resulting observations of the occlusions derived from their interaction and also be physically plausible (that the hand does not share the same physical space with the object) by penalty the objective function. Regarding methodology, the authors use Canny edge detection to build an edge map, compute a distance transform for each one, and use a previous own method to generate the color map. Thus, the image observations are given by the skin color maps and the transform. The authors claim that this is the first model-based work that efficiently solves the continuous full-DOF, joint hand-object tracking problem based solely on markerless multi-camera input, further demonstrating that hand-object interaction can be seen as a context that facilitates hand pose estimation, instead of being a problem factor.

Ren et al [18] propose a distance metric called "Finger-earth mover's distance" to measure the dissimilarity of the noisy hand shape provided by a Kinect® sensor, as method just matches fingers and not the whole hand shape, it can better distinguish hand gesture subtle movements. This metric sees each finger as a "cluster", penalizing unmatched fingers. The method is proposed to address the problem that, due to the low resolution of the depth map delivered by the Kinect® sensor (640x480), it is hard to detect and segment a small object

like the hand and all its joints.

Oikonomidis et al [19] extend again their work with a model-based, joint-evidence method, where a two-hand tracking is performed as an optimization problem whose objective function quantifies the discrepancy between the structure and 3D appearance of hypothetical configurations of both hands and the corresponding Kinect® observations. Optimization is performed by a variant of a particle swarm optimization method, adapted to the needs of the specific problem. The methodology combines the steps performed in their previous studies [15] and [17], especially in the latter idea to model the hand-object relations and to treat occlusions as a source of information rather than see them as a complicating factor. Furthermore, in this work the problem is more complex since it focuses on both hands with only one Kinect® sensor instead of a multi-camera system. An update of this work can be found likewise in [7].

In [20], a method to capture the articulated motion of two hands while interacting with each other and with an object is proposed. Salient points such as finger tips are scanned through a multi-camera system, however, since these points cannot be tracked continuously due to excessive occlusions and similarity in their features and color appearance, avoiding a fixed association between the salient points and the respective fingers, an approach that solves the salient point association jointly with the hand pose estimation problem is proposed. Also, a quite differentiable objective function for pose estimation is implemented, taking into account edges, optical flow, salient points, and collisions. Thus, authors may use simple local optimization instead of a sampling based one as in [19]; in fact, they say their approach achieves significantly lower pose estimation errors than the sampling optimization. In conclusion, they suggest the possible desirability of researching the combination of both optimization techniques.

In [2], a new approach for tracking 3D articulated skeletal models using an augmented rigid body simulation is presented, being able to follow a human hand from a depth sensor. The method allows robust, real-time results using only an x86 processor. The system generates constraints that limit motion orthogonal to the rigid body model's surface, these constraints, along with prior motion, collision constraints, and joint mechanics, are solved by a Gauss-Seidel solver. To improve tracking accuracy, multiple simulations are generated at each frame and fed some heuristics, constraints, and poses.

Kulshreshth et al [21] present preliminary results of a real-time, markerless finger tracking technique using a Kinect® sensor as an input device. The technique calculates feature vectors based on Fourier descriptors of equidistant points chosen on the silhouette of the detected hand and matches templates to find the best fit.

Karnan et al [22] propose a method to control the movement of a mouse pointer using simple hand gesture 2D images and a webcam. An algorithm for real-time tracking based on adaptive skin detection and motion analysis is implemented. Using the history of motion, the trajectory of movement of the hand is drawn and used to identify a gesture. The image database consists of four different gestures. In order to scale the motion when user is far away from

the point of capture, an algorithm is used to define the region of interest, motion of the mouse pointer is scaled accordingly. The system is fully automatic, real time, and does not need a uniform background.

In [23] a method for real-time continuous pose recovery of markerless complex articulated objects from a single depth image is described. In order to generate the training data, the system can use multiple depth cameras, however, only a single depth camera for real-time tracking is required. The method can be generalized to track any articulated object that (a) can be modeled as a 3D boned mesh, (b) can be fed to a binary classifier to label pixels belonging to the object, and (c) that the projection from bones pose space to a 2D depth image be approximately one to one. Four stages are distinguished:

1. a randomized decision forest classifier for image segmentation,
2. a robust method for labeled dataset generation,
3. a convolutional network for dense feature extraction, and
4. an inverse kinematics stage for stable real-time pose recovery.

6 Main Expected Contribution

Oikonomidis et al [11] suggest that there is great interest in the development of markerless, computer vision based solutions, since they are not invasive and maybe less expensive. Furthermore, by fully understand hands configuration thanks to their 3D pose estimation, systems that understand human activities and interaction with their physical and social environment could be built. The economic benefit that areas such as ludic get globally, and all the advantages that could bring the development of these new ways of communication to the daily life of every human being, encourage scientific community to further research and improve or develop new methods looking for a higher efficiency and accuracy. But above all, this study was conducted to provide background on the research area as a basis for developing a set of tools that can be applied in the handling of HCI devices by people with motor disabilities, whose condition has not been actually addressed by the current hand gesture recognition methods.

7 Discussion of the Results and their Validity

In this paper, a brief review of recent research efforts in hand gesture recognition has been performed. Table 1 is a comparative summary of the tools, features, techniques, and objectives reviewed. As shown, several areas of opportunity can be derived from these data, regarding current research in the literature. The following are of particular interest for the purpose of this study, therefore will be addressed in the development of the project.

1. The use of two or maybe more RGB-D cameras (whether Kinect® or other brands), and/or other technology such as optical flow or infrared light sensors, could mean a significant advantage mainly to avoid occlusions in scanned objects.

2. A combination of techniques concerning feature extraction and optimization methodology to check if there is an improvement (or optimization) on recognition.
3. The application of these approaches to people with motor or speech disabilities, which has not been addressed in the state-of-the-art, and besides being a relevant research topic, becomes an increasing needing for them to interact with various technological devices.

Table 1. Abbr: DT=Distance Transform, PSO=Particle Swarm Optimization, AD=Adaptive detection, RDFC=Randomized Decision Forest Classifier, CN=Convolution Network, IK=Inverse Kinematics, HFC=Hough Forest Classifier, FEMD=Finger-Earth Mover's Distance, FD=Fourier Descriptors

Research	Camera	Features	Technique	Optimization	Objective
[11]	Multiple	Skin color, Edges	DT, Canny	PSO	One hand
[15]	Kinect®	Skin color, Depth	-	PSO	One hand
[22]	Webcam	Skin color	AD	-	One hand
[23]	Depth	Depth	RDFC, CN	IK	One hand
[2]	Depth (2 sensors)	Depth	-	Gauss-Seidel	One hand to two hand
[17]	Multiple	Skin color, Edges	DT, Canny	PSO	One hand-object interaction
[19]	Kinect®	Skin color, Depth	-	PSO	Two hand interaction
[20]	Multiple	Edges, Optical flow, Collisions, Salient points	HFC	Local	Two hand-object interaction
[18]	Kinect®	Skin color, Depth	FEMD	-	Fingers
[21]	Kinect®	Depth, Silhouette	FD	-	Fingers

References

1. Nakashika, T., Hori, T., Takiguchi, T.: Depth Spatial Pyramid: A pooling method for 3D-object recognition. Advances in Computer Science and Engineering 12, 15–30 (2014)
2. Melax, S., Keselman, L., Orsten, S.: Dynamics based 3D skeletal hand tracking. In: Proceedings of Graphics Interface 2013, pp. 63–70 (2013)
3. Bachmann, D., Weichert, F., Rinkenauer, G.: Evaluation of the Leap Motion Controller as a new contact-free pointing device. Sensors 15(1), 214–233 (2014)
4. Kim, D., Hilliges, O., Izadi, S., Butler, A.D., Chen, J., Oikonomidis, I., Olivier, P.: Digits: freehand 3D interactions anywhere using a wrist-worn gloveless sensor. In: proceedings of the 25th annual ACM symposium on User interface software and technology, pp. 167–176 (2012)

5. Jung, P., Lim, G., Kim, S., Kong, K.: A wearable gesture recognition device for detecting muscular activities based on air-pressure sensors. IEEE Transactions on Industrial Informatics 11, 485–494 (2015)
6. Wang, R.Y., Popović, J.: Real-time hand-tracking with a color glove. ACM Transactions on Graphics 28(9), 63:1–63:8 (2009)
7. Oikonomidis, I., Kyriazis, N., Argyros, A.A.: Tracking the articulated motion of human hands in 3D, ERCIM News 95, pp. 23–25 (2013)
8. Zimmerman, T.G., Lanier, J., Blanchard, C., Bryson, S., Harvill, Y.: A hand gesture interface device. In: Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface, pp. 189-192. ACM Press, New York, USA (1987)
9. Erol, A., Bebis, G., Nicolescu, M., Boyle, R.D., Twombly, X.: Vision-based hand pose estimation: A review. Computer Vision and Image Understanding 108(1), 52–73 (2007)
10. Salzmann, M., Urtasun, R.: Combining discriminative and generative methods for 3D deformable surface and articulated pose reconstruction.In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 647–654 (2010)
11. Oikonomidis, I., Kyriazis, N., Argyros, A.A.: Markerless and efficient 26-dof hand pose recovery. In: Kimmel, R., Klette, R. and Sugimoto A. (eds.), ACCV 2010, Part III. LNCS, vol. 6494, pp. 744–757. Springer Verlag Heidelberg (2011)
12. Bray, M., Koller-Meier, E., Van Gool, L.: Smart particle filtering for high-dimensional tracking. Computer Vision and Image Understanding 106(1), 116–129 (2007)
13. Heap, T., Hogg, D.: Towards 3D hand tracking using a deformable model. In: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, pp. 140–145 (1996)
14. Lin, J.Y., Huang, T.S.: Capturing natural hand articulation. In: Proceedings of the Eighth IEEE International Conference on Computer Vision, pp. 426–432 (2001)
15. Oikonomidis, I., Kyriazis, N., Argyros, A.: Efficient model-based 3D tracking of hand articulations using Kinect. In: procedings of British Machie Vision Conference, pp. 1–11 (2011)
16. Oliva, A., Torralba, A.: The role of context in object recognition. Trends in cognitive sciences 11(12), 520–527 (2007)
17. Oikonomidis, I., Kyriazis, N., Argyros, A. a.: Full DOF tracking of a hand interacting with an object by modeling occlusions and physical constraints. In: International Conference on Computer Vision, pp. 2088–2095 (2011)
18. Ren, Z., Yuan, J., Zhang, Z.: Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In: proceedings of the 19th ACM international conference on Multimedia, pp. 1093–1096 (2011)
19. Oikonomidis, I., Kyriazis, N., Argyros, A. A.: Tracking the articulated motion of two strongly interacting hands. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1862–1869 (2012)
20. Ballan, L., Taneja, A., Gall, J., Van Gool, L., Pollefeys, M.: Motion capture of hands in action using discriminative salient points. In: Fitzgibbon et al. (eds.), Part VI. LNCS, vol. 7577, pp. 640–653. Springer Verlag Heidelberg (2012)
21. Kulshreshth, A., Zorn, C., LaViola, J.J.: Poster: Real-time markerless kinect based finger tracking and hand gesture recognition for HCI. In: IEEE Symposium on 3D User Interfaces, pp. 187–188 (2013)
22. Karnan, J., Ramkumar, M., Sivaraman, K., Santhakumar, G., Karthik Kumar, R.: Real-time gesture based human computer interaction for office applications.

International Journal of Review in Electronics and Communication Engineering 2(1), 42–47 (2014)
23. Tompson, J., Stein, M., Lecun, Y., Perlin, K.: Real-time continuous pose recovery of human hands using convolutional networks. ACM Transactions on Graphics 33(5), 69:1–69:10 (2014)

Development of an Interpreter for LRT using the Exact Real Number Paradigm

J. Leonardo González-Ruiz, J. A. Hernández-Servín, J. Raymundo Marcial-Romero

Facultad de Ingeniería, Universidad Autónoma del Estado de México, Toluca, México
leon.g.ruiz@gmail.com, xoseahernandez@uaemex.mx, jrmarcialr@uaemex.mx

Abstract. The exact real number representation can grow arbitrarily so it does not truncate or rounds up as opposed to floating point number representation. The advantage of this representation is that not rounding errors are generated and any operation can be achieved with the desired accuracy. The LRT is a proposal that implements exact real number. This paper develops an interpreter for LRT using this paradigm whose operational semantics is based on sixteen rules so the programs based on LRT libraries are less complicated to debug. One of the main problems in implementing LRT has been memory consumption. The main contribution of this work is that LRT has its own administrator for infinite lists as well as its own lazy evaluation. The performance of programs written in LRT interpreter is shown to be superior to the libraries of functions in both execution time and use of memory.

Keywords: Exact real number, Interpreter, lazy evaluation, LRT, interval arithmetic.

1 Introduction

The paradigm of exact real numbers arises from the need to represent real numbers on a computer, compared to traditional floating point representation [6]. The real numbers may be described and represented in various forms for formal purposes in mathematics, for example, represented by the intersection of intervals [15].

For practical purposes, the common representation of real numbers is performed by finite strings of decimal digits{0, 1, 2, 3, 4, 5, 6, 7, 8, 9} and a dot {.}.This representation is commonly used and is called "floating point". The set of digits can represent only a subset of the real numbers exactly, this means that most real numbers must be represented by other real numbers that are close to, or, for a interval of rational numbers which delimit the actual number requested, causing rounding errors. Despite this, floating point representation is acceptable for a wide range of applications, however, the accumulation of these rounding errors produced by a large number of computations produces inaccurate results. For example, programming languages such as C, which uses floating point arithmetic causes rounding errors in complex operations that require precision of

significant digits, as an alternative of floating point we have used the interval analysis[15]. Interval analysis involves expressing a real as a pair of numbers which represent an interval containing the number. Interval arithmetic operations compute new upper and lower bounds on the result after the operation has been performed. One way to represent a real number using interval analysis is digit strings of potentially infinite size, this representation is known as exact real numbers paradigm[3]. There are exact real number implementations like IRRAM [12], IC-Reals [4], RealLib [7], that expect to become the standard for use in program languages.

Another option for calculating real numbers is the language called LRT (Language for Redundant Test) [11]. The language LRT is a variant of PCF [5] and an extension of PCF (Programming Computable Functions) with a ground type for real numbers and suitable primitive functions for real-number computation. LRT efficiency is given by the choice of linear functions, handling of potentially infinite lists and type of language evaluation. An implementation of this language was developed by Lucatero [10], that developed a library for handling exact real numbers in the programming language `C++`.

In this paper an interpreter for the LRT language is described, this interpreter has the ability to handle infinite lists and *lazy* evaluation. The development of this interpreter merge two languages, LRT (for the real number calculation) and `C-` [9] (for the basic grammar in a language) producing the new language LGR. The scripts written for the interpreter will be compared with developed implementations of Lucatero's library because it is considered the continuation of her research.

2 Implementations

A real number is computable if there is an algorithm that can produce all its digits. The representation of real number in a computer can be handled by different approaches ranging from libraries of functions in the C programming language to applications in a functional language [10].

2.1 LRT Language

The LRT language (Language for Redundant Test), its a variant of *Real PCF* [5] developed by Marcial et al. in 2004. LRT is a nondeterministic language, eliminates the parallel conditional of *Real PCF* that is a parallel language and add a nondeterministic operator `Rtest` [8].

A real number x can be represented in the LRT language as a tuple:

$$(y, 2^e)$$

where $y \in [-1, 1]$ y $e \in \mathbb{Z}$.

LRT uses the representation (mantissa, exponent), where the mantissa is represented by a infinite list and the exponent as a integer number. This representation allows LRT extends its rank to the entire real number line, with the restriction that the mantissa must be in the interval $[-1, 1]$.

The constructors `Cons`, `Tail` y `Rtest`, are the base of the real number construction in the LRT language.

- *Constructor* `Cons`: It is a function that accepts as input two intervals and returns an interval, is the one that performs a reduction intervals and is defined as follows:

$$\texttt{Cons}([\underline{x}, \overline{x}]) = [\underline{x}\overline{a} + \underline{b}, \overline{x}\overline{a} + \underline{a}]$$

  where$[\underline{a}, \overline{a}]$ y $[\underline{x}, \overline{x}]$ are two intervals of rational numbers in the interval$[-1, 1]$.

- *Constructor* `Tail`: It is a function that uses as input two intervals of rational numbers belonging to the interval $[-1, 1]$. It is considered the inverse of the constructor `Cons` only when the interval $[\underline{x}, \overline{x}]$ is contained in $[\underline{a}, \overline{a}]$. In this case the following equation is considered:

$$\texttt{Tail}_{[\underline{a},\overline{a}]}[\underline{x}, \overline{x}] = \frac{x - (\frac{\overline{a}+\underline{a}}{2})}{(\frac{\overline{a}-\underline{a}}{2})}$$

When $[\underline{x}, \overline{x}]$ is not within $[\underline{a}, \overline{a}]$, has to guarantee that the result is within the interval $[-1, 1]$, since this is the range of the function.

- *Constructor* `Rtest`: This function determines which side the rational number lies within the interval $[-1, 1]$, that is the `Rtest` returns $True$ if the number to be calculated lies in the positive side and $False$ otherwise. LRT language is called "redundancy check language" because the constructor `Rtest` is an operator of redundant verification, ie that there are numbers for which the operator can return both true and false [10].

2.2 LRT Implementations

Two calculators were developed based on LRT language. One developed by Villanueva as described in [16], and the second one is a modified version developed by Linares [8]. Villanueva implements the basic operation using the functional language Haskell by considering real number in the interval $[-1, 1]$, his calculator gives accurate results but it is limited to basic operations. Specifically, he implemented addition, subtraction, multiplication and division only. Linares, on the other hand takes Villanueva's implementation and manages to add improved algorithms to compute transcendental functions such as sin, cos, tan, tan^1, e, and log. fundamentally based on algorithms developed in [14] by adapting the type of data used in the LRT language.

The calculator developed by Linares is more efficient than the one by Villanueva and also the range operation is expanded to the entire line of real numbers; in spite of that neither of them improve implementations developed in the programming language `C++` such as iRRAM [10].

To improve these calculators, Lucatero [10] developed a library for handling exact real numbers in the programming language C++ [8]. The algorithms and the operational semantics of LRT language are implemented in C++ language using FC++ which allows the use of calls by need (*lazy* evaluation) and the definition of infinite lists on an imperative language as is C++, both necessary for the operation of the LRT semantics; in addition to use the floating point number of GMP library for greater accuracy in the LRT constructors.

The comparisons made by Lucatero, show significantly improved runtime operations compared with the implemented in a strictly functional language, which also holds the operational semantics of the LRT language. The resulting time is longer compared to the one developed in C++ that also uses the exact real computation, as iRRAM. Lucatero concluded that memory consumption increases as more precision digits are required in the calculations; this is due of the recursive call used in the algorithms, because it must be stored in memory of the new state of function call. The library developed by Lucatero has been compared to commercial implementations as mentioned above, so, this research is considered a continuation of her previous work in search of an interpreter implementation more efficient at runtime.

3 Methodology

To implement the interpreter a methodology based on a standard implementation [1] of a compiler is used, which consists of the following stages of development: Lexical Analyzer, Syntactic analyzer, Semantic Analyzer, generate intermediate code and optimize code.

The implementation of this methodology begins with two basic languages, LRT and C- [9], the merge of these languages forms LGR language, having the main characteristics of the two languages. The first one can compute exact real numbers, and the second one, is a reduced programming language which grammar is used as the basis for the development of interpreter in this research, its semantics can be consulted in [9]. The interpreter is developed using the programming language C, alongside with the grammar of LRT and C-.

3.1 Design the Lexical Analyzer

At this stage the source code of LGR is read, which is composed by a character string. The analyzer collects character sequences into meaningful units called *tokens* to be used for the following stages. The lexical units used in this interpreter are integers, floats, booleans, real, rational comments, reserved words and blanks. For design the lexical analyzer we use Flex[13], that is a tool for generating lexical analyzers used for the development of the interpreter. This program reads a text file which contains C code that shows rules of regular expressions, this file can be compiled, generating an executable file, which is capable of analyze an input code in order to find regular expressions and execute its corresponding code in C.

For the develop of an lexical analyzer we must define some rules called regular expressions which represent patterns of strings. A regular expression is defined by a set of strings that matches [1]. This language made by a pattern of characters, is used to define different lexical elements of the interpreter, for a example, a regular expression of a real number is shown as follow:

A real number can be represented as an interval succession of the form: $[(((digit+/digit+),(digit+/digit+))),((digit+/digit+),(digit+/digit+)))*,...]$.

An other example of regular expressions are the reserved words of LRT language:

- *cons*: `Cons` function.
- *tail:* `Tail` function.
- *rtest:* `Rtest` function.
- *succ:* `Succ` function.
- *pred:* `Pred` function.
- *iszero:* `Iszero` function.

3.2 Design the Syntactic Analyzer

The second stage of an interpreter is the syntactic analyzer. In order to develop it, we use GNU bison [2] that is a general-purpose parser generator available for almost all operating systems, normally used in conjunction with Flex. Bison converts the formal description of a language, written in BNF (a formal way to describe formal languages), in a program written in C that performs parsing.

The parser receives the LGR source code in a token form from the lexical analyzer and performs the parsing call, which determines the structure of the program.

A part of the LGR grammar in its real part, is described below:

⟨*type specifier*⟩ ::= 'INT'|'FLOAT'|'REAL'|'BOOL'|'RATIONAL'|'VOID'

⟨*statement*⟩ ::= ⟨*expressionstmt*⟩ | ⟨*compoundstmt*⟩ | ⟨*selectionstmt*⟩
| ⟨ *iterationstmt* ⟩ | ⟨*returnstmt*⟩ | ⟨*natstmt*⟩ | ⟨*realstmt*⟩

⟨*natstmt*⟩ ::= 'SUCC' '('⟨*statement*⟩')' | 'PRED' '('⟨*statement*⟩')'
| 'ISZERO' '('⟨*statement*⟩')'

⟨*realstmt*⟩ ::= 'TAIL' '[' 'RATIONAL' ',' 'RATIONAL' ']' '('⟨*statement*⟩')'
| 'RTEST' 'NUM' 'NUM' '('⟨*statement*⟩')'

⟨*factor*⟩ ::= '('⟨*expression*⟩')' | ⟨*var*⟩ | ⟨*call*⟩ | 'NUM' | 'NFLOAT' | 'FALSE'
| 'TRUE' |'RATIONAL' | 'CONS' | '['RATIONAL',RATIONAL']'
'('⟨*statement*⟩')'

3.3 Design the Semantic Analyzer

The next stage of development is the semantic analyzer, which is responsible for giving meaning to the *tokens* and the structures formed on a user entered code. LGR is a typed language and a type declaration has the form $C\ x$ where x is a variable and C is a type. Each variable must have a type declaration before is used, the types of the attributes of a function must also be declared. Types used for this interpreter are:

- int, Integer type.
- float, Float type.
- real, Real type.
- bool, Bool type.
- rational, Rational type.

When a function is invoked in the interpreter, the parameters join arguments and the expression is evaluated; the value obtained is the meaning or reason for the invocation of the function. In other hand LGR allows recursion, adding elements such as lazy evaluation that is discussed below. An structure that the interpreter uses for general purpose is a symbol table, which is a data structure that uses the translation process of a programming language, where each symbol in the source code of a program is associated with information such as location, type of data and the scope of each variable, constant or function. A common implementation of a symbol table can be a dynamic table, which will be maintained throughout all stages of the compilation process.

The lexical analyzer in the previous step forms a parse tree, which is a representation of the code structure of a string tokens, the tokens appear as leaves of the parse tree from left to right and internal nodes tree depicting the steps of a derivation [1]. However, a parse tree contains information that is not necessary for the interpreter to be able to produce an executable code. Therefore, a new tree is generated using only the information needed for compilation and interpretation.These trees represent abstractions of token sequences of source code that the user entered, and sequences of tokens can not be recovered from them. However, it contains all the information needed for making more efficient the grammatical trees translation, these are known as abstract syntax trees, or AST [9]. A parser will cover all the steps represented in a parse tree, in order to obtaining a correct syntax, then an abstract syntax tree is constructed.

An other major task of the interpreter is to keep the information integrated and updated of the data types, the use of this information ensures that each part of the program makes sense and consistency under the rules of language, in this case LGR, ie , type checking. This test is developed in the syntactic part. When the interpreter produce the grammar tree, also performs a verification type. This operation is based on finding a variable, looking for the variable in the symbol table to know its type which it was declared. Subsequent to solve the expression, either an assignment, an operation or an invocation to a function, the interpreter verifies that the data type in each of them is the same, otherwise sends an error in the output of the interpreter. In this way ensures that the program does just

what it must do, to have uniformity in the type of data, their operations and their return values.

3.4 Generate Intermediate Code

At this stage an interpreter that takes as input the AST generated by the above stage. The functions described below are responsible for interpreting the user-written code, including its resolution using lazy evaluation.

- *reduce* function*:* The purpose of the interpreter is to solve the operations described in the AST, reductions are performed in order to leave a code in its minimal form to achieve the intended operation specified by the user.
- *lazyreduction* function*:* To solve a source program in a *lazy* way, functions are implemented that allow code execution using this form of evaluation.

To achieve *lazy* evaluation, specific functions are implemented to solve the parameters of a function in a lazy way, in other hand functions that allow handling of recursion, detailed below are implemented.

- *ftocall* function*:* It is performed whenever there is a call to a function. A stack that contains two columns is created, the first stores the variable that invoked the call or on which the result will be reflected back . In the second column, the name of the function that was invoked is stored. In this way, the interpreter is able to be solving the recursive calls and go freeing up memory once the evaluation of functions is finished.
- *fparameters* function*:* The lazy evaluation can decide when to solve a parameter, this is that if you need it, it will be evaluated. For this, the *fparameters* function creates a stack, which stores in its first column a copy of the symbol table of the function being invoked, in the second column pointers to AST where the unresolved parameters are stored.

The above two functions are invoked at the same time whenever there is a call to a function in the source code. In this way joins a call to a function with its unresolved parameters. Once a recursive function is evaluated, the line in both stacks is removed and freed from memory.

- *interpret_parameters* function*:* The moment that a function requires a parameter to solve their instructions, *interpret_parameters* function is responsible for searching in the parameter table the AST code that solves it. In this way, the function makes necessary operations to solve a function, arithmetic expression, or stored variable, in order to return its resolved value and enable the function that needs the value to continue with its operations.

When solving a parameter, the parameter table is updated with the value obtained, avoiding solve each time the parameter required.

Functions for calculating real numbers are implemented, which use the functions described above for *lazy* evaluation and recursion. The interpreter has predefined functions where the LRT language is integrated.

- *evaluate* function*:* This function is predefined by the interpreter. Receives two parameters to calculate the actual number and the desired accuracy. The function *evaluate*, work as any function in the interpreter in a *lazy* way.
- *evaluate_cons* function*:* This function receives an AST pointer, the received node has the return instruction from a `Cons`. Is the main function of LRT, because performs the reduction of two `Cons` and evaluate if has come to the accuracy desired by the user. This function stores a list of results, which is the actual number determined at the end of the program.
- *fcons* function*:* The reduction of two intervals is performed by this function. Receives two intervals of type `Cons` and returns the reduction. Since operations are performed between rational separately (numerator, denominator), the integers in the numerator and denominator tend to grow quickly, so a factorization function that factorize the rational numbers to store smaller values within the INTERVAL structure is implemented.
- *frtest* function*:* This function receives an interval where are stored the delimited numbers l and r and an interval. The function returns an integer that indicates whether the input range is or not, between the delimiting content.
- *ftail* function*:* The opposite process of determining the reduction of two intervals, is made by the *ftail* function, this function receives an interval of type `Tail`and returns an interval of type `Cons`, this as a result of the operational semantics rules of LRT.
- *evaluate_intervals* function*:* Because *ftail* function returns an interval of type `Cons`, an evaluation function is implemented which together with the above, determine the reduction of two intervals using the rules of the operational semantics of LRT, recursively calling the *ftail* function . This function receives a pointer of INTERVAL type, which contains the first position of the doubly linked list of generated intervals.
- *split_head* function*:* This function receives a pointer to a node in the AST and returns an interval generated type `Cons`. The primary function of *split_head* is to obtain an interval type `Cons`. This function is called whenever it is necessary to calculate an other interval in different functions of the LRT language.
- *asignParamAst* function*:* For proper operation and code and memory optimization, a specific function is implemented to store the resulting values in the calculation of real numbers. The *asignParamAST* function receives a pointer to an AST node, and returns a pointer to a node of the same type. Its primary function is to store the resulting intervals in the parameter stack to prevent calculate again.

3.5 Optimize Code

The handling of pointers along the implementation of the interpreter is essential for fast and efficient memory usage. By not having to store each value in memory space, we opt for using pointers to reference calculated values, resolve the *lazy* evaluation and recursion over interpretation. Similarly, for optimal usage, calls and parameter stacks are built in such a way that when a recursive call is finished

memory is released. One of the main advantages of the interpreter, despite using infinite lists scheme, the list that stores the result does not grow more than two elements, unlike other implementations that store a list of intervals calculated as the final result.

4 Results

The tests of the interpreter are performed on a computer with a processor at 2.66 GHz Intel Core 2 Duo, 4GB of DDR3 RAM. For testing and validation of the interpreter ten different scripts written in LGR are executed. The results of the performance are shown below.

4.1 Calculation of Real Numbers with LGR

The *faverage* function receives two real numbers and returns the average of them. Ten different scripts are written and ran to calculate the arithmetic expressions of the first column of table 1, in these scripts *faverage* and *fractolist* functions are implemented in LGR, the last one converts a rational to a intervals list to do later basic operation of average. The comparative with the `GMP-FC++` library for LRT is shown in Table 1.

The results in the Table 1, show that the runtime is lower than Lucatero's, Likewise the memory consumption is minimal compared to her implementation; Regarding the arithmetic expressions 9 and 10, interpreter results surpass those of `FC++` due to inefficient factorization function used by the interpreter, showing that the implementation of the program takes due to the number of operations performed to factor and not the overflow of RAM memory. In cases where takes longer time is only consuming $348K$ of available memory.

The development of this work, follows a standard methodology for creating compilers and interpreters. This allows the LRT language perform more autonomously and not depend on functional languages like Haskell or as GMP library functions for handling infinite lists. Since, the interpreter is written in the C programming language, is possible to run on different operating systems and computer architectures. One problematic factor found in the development of this work is the handling integers of variable lenght due to buffer overflow in the *long long* data type that handles the C programming language; because of that, the interpreter fails to finish on time because rational numbers are variably growing, causing instability in the interpreter. To partly cope with this situation a factoring algorithm that reduces the rational number is implemented to store it appropriately in the type of data set; considering that an algorithm for prime factorization is located within a problem of type NP, their performance at runtime effects the output of the interpreter making it particularly slow. Because of the difficulty of this problem we propose some guidelines in future work section to remedy the situation. Making a comparison with the library developed by Lucatero shows that the memory usage for recursion and efficiency of *lazy* evaluation solves the problem encountered in the previous research because it

Table 1. Comparison between implementations

No	Arithmetic Expression	Runtime (seconds)				
		GMP-FC++	**LGR + factorization**	**LGR without factorization**	RAM used by GMP-FC++(K)	**RAM used by LGR (K)**
1	$\sum_{1}^{5} \frac{(\frac{1}{4}+\frac{1}{4})}{2}$	0.342	0.000693	0.000659	740	328
2	$\sum_{1}^{5} \frac{(\frac{1}{3}+\frac{1}{3})}{2}$	0.342	0.000923	0.000885	796	324
3	$\sum_{1}^{10} \frac{(\frac{1}{2}+\frac{1}{2})}{2}$	0.685	0.001929	0.000899	820	344
4	$\sum_{1}^{10} \frac{(\frac{2}{5}+\frac{2}{5})}{2}$	0.685	0.001234	0.000813	1,020	340
5	$\sum_{1}^{15} \frac{(\frac{1}{3}+\frac{1}{3})}{2}$	1.028	0.009086	0.008132	1,252	332
6	$\sum_{1}^{15} \frac{(\frac{1}{4}+\frac{1}{4})}{2}$	1.028	0.009270	0.001066	904	356
7	$\sum_{1}^{20} \frac{(\frac{1}{2}+\frac{1}{2})}{2}$	1.41	0.135286	0.001587	984	372
8	$\sum_{1}^{20} \frac{(\frac{2}{5}+\frac{2}{5})}{2}$	1.41	0.406574	0.005135	1,544	372
9	$\sum_{1}^{25} \frac{(\frac{1}{4}+\frac{1}{4})}{2}$	1.7625	8.527817	0.011391	1,076	384
10	$\sum_{1}^{25} \frac{(\frac{2}{5}+\frac{2}{5})}{2}$	1.7625	12.953400	0.025865	1,868	380

could not control how memory is freed in `FC++` runtime, unlike this interpreter that performs releasing memory when finished using a recursive call or parameter. Making a comparative in runtime this interpreter surpass the reported by Lucatero, due to handling infinite lists.

5 Conclusions

In this paper an interpreter using the methodology shown in Section 3 is developed, for calculating exact real numbers in the programming language C. The functions necessary to implement recursion, *lazy* evaluation and infinite lists are implemented, using Flex tools for building a lexical analyzer and Bison for the construction of a parser. The operational semantics of the programming language LRT and `C-` language are merged in order to obtain the language LGR. The results show the efficiency of the management of recursion and *lazy* evaluation. In a matter of precision and accuracy of the results of operations that can be performed with the interpreter, the results show that the implementation provides an accurate result using the specified accuracy. The development of the

interpreter shows evidence of having a standalone product and controlled for LRT language .

The problem of memory management in the library created by Lucatero has been resolved in this implementation, however, problems encountered in this development are the following:

An interval consists of two rational, which are treated separately as the numerator and denominator, performing operations using these integers. The problem arises when these integers start growing. Integers use the type *long long*, the larger type that stores integers in C. Since the numbers are growing rapidly, a factorization function is used to decrease the size of the resulting integer. When making an interval reduction by *fcons* function, function *factorize* is invoked. When trying to store these numbers in variables of type *long long*, at reach its maximum storage capacity, overflow variables, avoiding the interpreter to finish its execution.

The development of this interpreter leaves open lines for improvement in a matter of implementation, to solve the problems encountered in this development and add more grammar to the language to make a more robust LGR language. A further work may be the following:

- Development of a data structure capable of storing a variable length integer.
- Implementation of the trigonometric functions for LRT.
- Perform the implementation of three-address code.

References

1. Aho, A.V., Sethi, R., Ullman, J.D.: Compilers, Principles, Techniques. Addison wesley (1986)
2. Corbett, R., Stallman, R.: Bison: Gnu parser generator. Texinfo documentation, Free Software Foundation, Cambridge, Mass (1991)
3. Edalat, A.: Exact real number computation using linear fractional transformations. `http://www.doc.ic.ac.uk/exact-computation/exactarithmeticfinal.ps.gz`
4. Errington, L., Heckmann, R.: Using the ic reals library. `http://www.doc.ic.ac.uk/exact-computation`
5. Escardó, M.H.: PCF extended with real numbers. A domain-theoretic approach to higher-order exact real number computation. Ph.D. thesis (1997)
6. Goldberg, D.: What every computer scientist should know about floating-point arithmetic. ACM Computing Surveys (CSUR) 23(1), 5–48 (1991)
7. Lambov, B.: Reallib: An efficient implementation of exact real arithmetic. Mathematical Structures in Computer Science 17(01), 81–98 (2007)
8. Linares, D.: Diseño e implementación de una calculadora científica con el lenguaje de programación lrt. Licentiate thesis (2009)
9. Louden, K.C.: Construcción de compiladores. Principios y prácticas (2005)
10. Lucatero, A.: Desarrollo e implementación de una librería para el cómputo real exacto basado en lrt. Licentiate thesis (2013)
11. Marcial-Romero, J.R., Escardó, M.H.: Semantics of a sequential language for exact real-number computation. In: Logic in Computer Science, 2004. Proceedings of the 19th Annual IEEE Symposium on. pp. 426–435 (2004)

12. Müller, N.T.: The irram: Exact arithmetic in c++. In: Blanck et al. (eds.). LNCS, vol. 2064, pp. 222–252. Springer Verlag Heidelberg (2001)
13. Paxson, V., Estes, W., Millaway, J.: Lexical analysis with flex. `http://flex.sourceforge.net/`
14. Plume, D.: A calculator for exact real number computation. Ph.D. thesis (1998)
15. Potts, P.J., Edalat, A.: Exact real computer arithmetic. Draft report, Imperial College, London (1977)
16. Villanueva, G.: Diseño e implementación de una calculadora para cómputo con números reales exactos empleando información redundante. Licentiate thesis (2007)

Business Intelligence in Educational Institutions

Agustín León-Barranco, Susana N. Saucedo-Lozada, Iselt Y. Avendaño-Jimenez, Ricardo Martínez-Leyva, Luis A. Carcaño-Rivera

Universidad del Valle de México, Coordinación de Ingeniería, Puebla, Pue., México
agustinleonb@inaoep.mx, snsl_nicol@hotmail.com, iselt.avendanoji@my.uvm.edu.mx, ricardomtzleyva@naatbot.com, luis.carcanori@uvmnet.edu

Abstract. Thanks to the Information and Communication Technology (ICT), nowadays it is possible to access to a lot more quality information and faster too, but analyzing such an amount of information it is not a simple task for a human. However is in charge with their management suffers needs to make decisions based on their experience. Knowledge becomes a competitive advantage in the fundamental tool makers of educational management to make decisions that best promote the organization. The integration of advances in ICT, data mining and business intelligence is a field of extensive exploration to bring more intelligence to business, and in particular educational institutions.

Keywords: Business intelligence, big data, decision making, data mining, educational organizations, educational management.

1 Introduction

With the modern conditions of competence, making decisions has become a very difficult and risky task for directors these days. It is true that thanks to the technologies of information and communication technology (ICT) it is now possible to access to much more information, higher quality and faster, but it is also true that analyzing such amounts of information is no longer a simple task for a human. Management of information on business organizations is a key tool in surviving the dynamic, global market from nowadays. This document presents an educational management based on business intelligence whose purpose is helping in the "decision making". Taking better decisions in pro of education could increase the quality of educational institutions which gives them more competitiveness in this fierce market. Learning to compete with this information is crucial for decision making, growth and management of educational organizations. The most competitive countries are those that are making the best use of ICT, those that dominate and productively apply the knowledge. Attracting and retaining more students and resources, for instances, has become the main driver of the management of university governments.

2 Proposed Solution

This work is focused on solving problems in the decision making for an educational institution by using ICT and other disciplines such as data mining and business intelligence. The way to do it is by means of an online system connected to databases which helps directors to make better decisions about the school.

Today, integrating advances in ICT, data mining and business intelligence is a great opportunity for academic competitiveness, in the future it will be a necessity to bring more intelligence to educational institutions. On the one hand, data mining encompasses a range of techniques to achieve the efficient operation of the data, by extracting actionable knowledge that is implicit in the databases, knowledge with which it is possible to solve problems of prediction, classification and segmentation. On the other hand, business intelligence encompasses understanding the actual operation of the organization by anticipating future problems by means of knowledge obtained with data mining.

3 Main Contribution

The purpose of this document is to show an educational management based on business intelligence to aid decision making in a university institution.

Educational management based on business intelligence covers not only decision-making in top management, but also other levels as academic direction and marketing. There are four items on which this research on management is focused: student monitoring, teacher follow-performing, loans and market opportunity.

4 Project Justification

Educational institutions are organizations [1], since there is a structure (director, deputy director, coordinators, teachers, etc.), there are operation and communication, there are goals and objectives, etc. Note that the educational institution is an organization of difficult management [2], because the core of it is the people, and sometimes, the interests are not the same for everyone. Making a clever tracking of each entity is not an easy task but it is necessary to make decisions that best suit the organization.

This new information era demands changing the paradigm of educational management to improve competitiveness and development of education, with the development of information technologies and new algorithms for treating information directors are potentially capable of processing and analyze vast amounts of information to be taken into account when making a decision. Hence lays the importance of proposing a new educational management based on business intelligence to improve decision-making at the University Institution.

5 Theoretical Background

The educational administration requires, besides the effective and efficient management of human and material resources, information management to provide a strategic and forward-looking vision.

Large amounts of information generated in educational institutions sometimes have hidden knowledge difficult to be seen for one person, but thanks to the development of data mining and business intelligence, it is possible to exploit their progress to troubleshoot educational management, one of these advances is the ability to extract knowledge from large databases to help managers make better decisions that allow growth and strengthening of educational organization.

"It is possible to extract useful knowledge from huge amounts of existing databases of educational institution to aid decision making on issues of educational management information."

Business Intelligence, what does it means?: It refers to the development of strategies and relevant aspects that focus on the creation and management of knowledge. It is closely related to the extraction of information from databases on a large scale.

Business Intelligence encompasses understanding the actual operation of the company anticipating future problems by using information obtained from the Data Mining, it is a powerful tool for academic intervention. Use a combination of explicit knowledge and sophisticated analytical skills to uncover secret information patterns (using pattern recognition technology as well as mathematical and statistical techniques); these patterns are the basis of predictive models that allow analysts to produce new observations of existing data.

Technology of information and communication: ICTs are comprised of four key sectors: communications, information infrastructure or hardware, software (packaged and custom) and services (support and professional implementation and outsourcing).

Their purpose is to help maximize the economic potential of individuals and companies and at the same time increasing their standard of living beyond the intrinsic potential of their own resources and technological capabilities, regardless of their economic fluctuations.

The Web 2.0 gave the user the opportunity and freedom to become a creator and editor of dynamic pages open content, enabling collaboration and social appropriation of the Internet. [3] The incorporation of products, tools, services, applications, etc., are enabling the creation of groups, communities and social networks to create, manage and share information, spaces and events by growing the personal, professional, social, political, economic, etc.

The development of ICT allowed the world to become smaller and competition among nations unaware of borders. For Mexico this may be the opportunity to attract investment, or conversely, to lose most attractive nations. Leaning on ICT is a key to improving rates of productivity and competitiveness of the country condition.

There is a high dependency on the competitiveness of countries with their competitiveness in information technology, if you want to measure the competi-

tiveness of countries this should be measured with the competitiveness they have in information technology.

Competitiveness is an objective on which the Mexican government should work, mainly promoting the development and diffusion of Information Technology and Communication sector. Therefore, it is important to begin to build competitive economy in the country offering quality goods and services at accessible prices, creating favorable conditions for the development of business conditions; it is important to generate the necessary human capital to contribute to the development of service industry information technology and make it internationally competitive and ensure its growth in the long term.

New environment of education systems in Mexico: Advances in science and technology have led to profound and significant changes in the economic and productive processes, social organization and the conception of the world and life. "An organization never exists for itself. She meets social functions that correspond to expectations of the society around it".

All organizations are forced to review both their goals and missions they seek to fulfill in society, and their modes of organizing and operations to meet their goals. The school is one of the few social institutions whose importance makes it the center of analysis and questioning from within and from outside it. Those charged with its management are faced with a task of such complexity they never before had, which requires a high level of professional competence[4].

Specifically it is considered that the director of the organization must show his ability as a leader and innovator, know how to apply the changes to practice, determine the extent of change, the ability to support and encourage the skills necessary to foster an organization that assumes learning [5].

Education and ICT, the way to connect two worlds: At the school level relevant factors include: meaningful and understood by all goals, attention to daily academic performance, coordination between programs and between school and parents, faculty development and organization of the school to support learning for all.

Strategies and resources that appear decisive in resolving problems are mainly: communication and participation, knowledge of the school system and external resources to school, cognitive flexibility, etc.

We can also appreciate the influence and impact that ICT industry has on the educational competitiveness, same that is reflected in the detonation of new production and development capabilities; Likewise, it has been observed that, unfortunately, Mexico still has a big job ahead to provide its sectors the necessary ICT enabling it to raise its competitiveness in their industries.

The influence of ICT in the education sector has an enormous responsibility because they are center of knowledge. In the contemporary world, universities must play a leading role in the progress of science and technology and dilute doubts about his ability to adapt to new social contexts and offer products that meet the needs of so-called knowledge society.

The quality of education is influenced by factors such as wealth of a society or national education policy objectives, standards, and methods of teacher recruitment. Many institutions have implemented analysis to improve enrollment management. The “actionable intelligence” is Technological achievements in communications have

revolutionized the spread of information. ICT offers us the opportunity to perform multiple tasks simultaneously knocking our barriers of time and space. Thanks to advances in communications technology people are increasingly interconnected.

Organizations are facing new challenges because, although they have greater access to valuable information, they do not know how to extract value or knowledge of them to make better decisions.

As never before in history, companies are now able to store anything digitally, but as they can generate huge amounts of information, it also decreases the percentage of data that businesses can process. The term Big Data applies to information that cannot be processed or analyzed using traditional tools or processes [6].

Big Data, the big amount of data: It is the term used today to describe the set of processes, technologies and business models that are based on capturing the value contained in the data itself. This can be achieved both through improved efficiency through data analysis, and by the appearance of new business models supposing an engine of growth. There is much talk of the technological aspect, but keep in mind that it is critical to find ways to give value to the data to create new business models or help existing ones.

Big Data is no longer a promise or a trend. Big Data is here and is causing profound changes in various industries. The analysis of information in large volumes, from various sources, at high speed and with unprecedented flexibility can be a differentiating factor for those who choose to adopt it [7].

With many institutions launching a Big Data strategy, the ability of a competitor to take your best customers is a growing threat. While internal organizational data represent a clear competitive advantage, unstructured knowledge available online, via mobile channels and social networks, is equally valuable. "Innovation in technology is a huge differentiator in today's financial services market"[8].

Operations research, something really needed: Also known as management science is a scientific approach based on mathematical models for decision-making, example: mathematically representing real situations to make better decisions.

Operations Research (OR) models are designed to "optimize" objective criteria specified subject to a set of constraints; the quality of the resulting solution depends on the accuracy with which the model represents the real system. Operations research is both an art and a science; the art of describing and modeling the problem, and science to solve the model using precise mathematical algorithms.

Data Analysis: Data analysis is the process in which the raw data are sorted and organized to be used in methods that help to explain the past and predict the future [9]. Requires skills in three areas: computer science, artificial intelligence (machine learning, data mining, etc.), and finally statistics and mathematics.

Academic Analysis: This analysis focuses on improving the admission and retention of students and related operational performance through the implementation of executive dashboards that provide points of leverage to improve performance and accountability.

In many ways, the analytic action is like a heterogeneous mix of options, all aligned with corporate objectives and strategies. To really put the action in analysis, higher education institutions need leaders committed to organizational ca-

pacity to measure and improve performance and change the organizational culture and behavior[10].

Data mining: is a powerful tool for academic intervention. Use a combination of explicit knowledge and sophisticated analytical skills to uncover secret information patterns (using pattern recognition technology as well as mathematical and statistical techniques); of these patterns the basis of predictive models that allow analysts to produce new observations of existing data emerge [11].

Social Business: The pioneers in adopting social business are registering the positive benefits and opportunities promised by social technologies. Along with other key trends (portability, cloud and big data), the social business begins to offer some suggestions of interesting value.

Bet on social business requires serious rethinking business results and the social nature of people, fusing technology, culture and values. Merging social media (social media), media work, social listening, social work, social intelligence and other social technology companies improve their business models maximizing efficiency and creating value. Additional benefits also impact other building blocks of the business model, such as Value Propositions and Customer Relationship [12].

6 Results

The purpose of this research is to establish the basis of an educational management based on business intelligence to help monitor the quality of education. The integration of advances in science and technology with the knowledge and demands for quality in education will be materialized in a digital platform to support decision-making in educational institutions.

The development of ICT today allow us to process large amounts of data and extract knowledge in real time, data mining has the appropriate set of techniques for efficient use of data by extracting actionable knowledge that is implicit in databases, knowledge with which it is possible to solve problems of prediction, classification and segmentation. The educational administration requires, besides the effective and efficient management of human and material resources, information management to provide a strategic and forward-looking vision.

On the other hand, business intelligence provides an understanding of the current operation in anticipation of future problems from knowledge obtained by data mining.

Education systems are in new social contexts with permanent and faster changes where control or reduce environmental uncertainty has become an impossible task, it is time to invest in a new management methodology that exploits the richness and benefits of the information business, a methodology supported on knowledge, a methodology for smarter business.

The problems encountered are becoming more complicated day by day and require complex thinking to build strategies to address the unpredictable, the random and qualitative, is no longer enough simply thought that establishes programs to control what is certain, calculable and measurable .

Management is one of the most important human activities, has been essential to ensure coordination of individual efforts. Management is the process of designing and maintaining an environment in which, working in groups, individuals efficiently meet specific objectives.

Science produces knowledge, while technology makes use of them. The term technology refers to the sum total of knowledge that we have about how to do things.

In the future, the planning / institutional management and associated knowledge flows can be seamlessly integrated, including strategic planning, expeditionary strategy execution, budgeting, reinventing processes and continuous improvement, accreditation, academic and critic programmable planning, and individual performance management.

Goals, objectives and performance can be measured, monitored, managed and aligned in all planning processes. The ability of colleges and universities to continue to provide leadership in knowledge is at risk. Information overload, increased competition, costly investments in technologies, increasing energy costs and benefits, heavier administrative requirements, lower net revenue per student (through initiatives to broaden participation), and a number of other forces are at fault. One facet of the solution is the use of knowledge services (and supporting technologies and systems) that offer higher value of investments in every area of operations of the institution. But colleges and universities must do so in a sustainable way, providing a transition from legacy to more accessible and flexible service-oriented and component-based approaches and aggregation of knowledge systems[13].

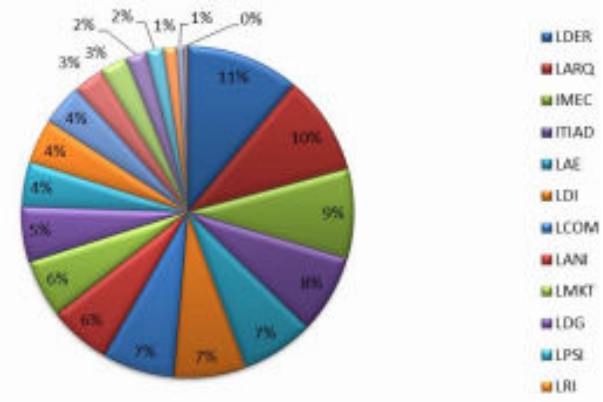


Fig. 1. Desertion per career

A study was developed within an educational institution of higher level, for reasons of confidentiality of information, can not disclose the name; in order to identify in real time the amount of students defection, providing more reliable information to managers, for better decision making.

The study was conducted in 2013, covering two semesters of study. Based on an exhaustive study of low educational institution of higher level conducted in 2013, can be seen in Figure 1 race predominates at low is LDER with 11%, following the LARQ with 10% and IMEC with 9%.

Table 1. Selected careers for this research study

NAME	CAREER
LDER	Law
LARQ	Architecture
IMEC	Mecatronics
ITIAD	Digital Animation
LAE	Administration
LDI	Industrial Design
LCOM	Communication
LANI	International Business
LMKT	Marketing
LDG	Graphic Design
LPSI	Psychology
LRI	International Relations

It was also studied and analyzed the reason for desertion, for instance, the economic reason shows 17.8%, 16.4% of the casualties were due to a change of Institution of students, while 12.8% made only one change of campus. Through the analysis we can also note that a 5.4% casualties were due to low student achievement, and 1% by dissatisfaction with academic and administrative services. Different grounds of the 298 students killed in 2013 are shown in Figure 2.

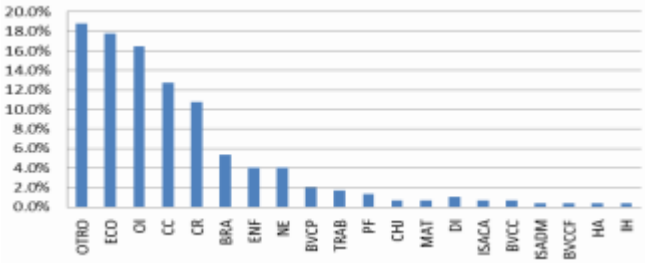


Fig. 2. Reason for desertion

Based on the analysis of the study, is also identified the month with bigger desertion: August with 22%, January with 15% and September with 14%. In months like March, May, November, December just between 2 and 3% of students desertion. In Figure 3 is shown in detail.

In Figure 4, we can also observe that a large percentage found that 18% of students, do not have the date they entered the institution therefore you cannot get actual data.

Thanks to these studies and analysis we can see the importance of real information within an educational institution, these analytical models give us key tools to get a clearer view of the problem to be solved, having real knowledge bases, giving us to attack the main variables.

Finally, the duration at the institution before defection was analyzed. In Figure 4, the percentages where we can observe that 44.3% of the students dropped out

before the first year. 20.5% of them defect in the first year of study, 8.1% in its second year, and only 1% of students dropped out after six years.

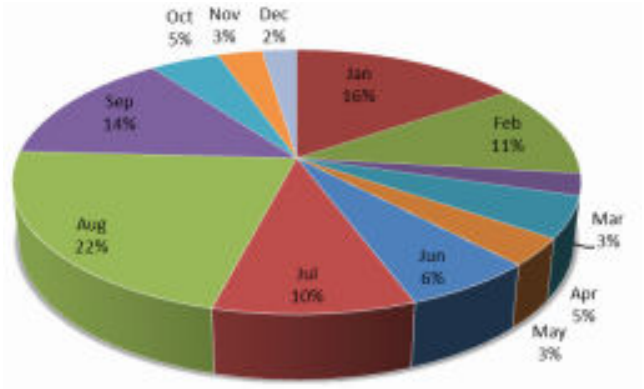


Fig. 3. Desertion per month

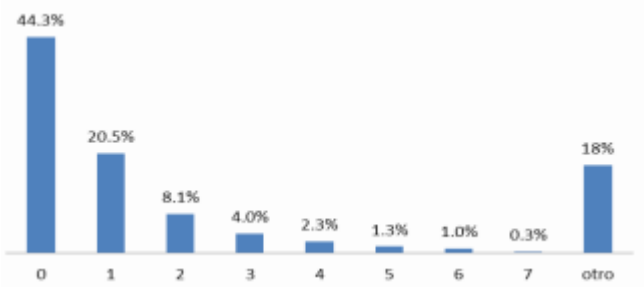


Fig. 4. Time within institution before desertion

As shown in the different graphics, the institution does not have any real information and there are significant percentages that do not have the correct or specified data.

7 Conclusions

The Analytical models for analyzing people can provide better decision-making and more efficient ones. That is why this work is focused on developing a methodology for educational management based on business intelligence to aid decision making.

In this context, all human organizations are affected because they depend largely on its adaptation to the environment they operate in. Education systems are no exception to this general situation, like other organizations, the acceleration of change

in all spheres of social life requires rethinking both their goals and in their modes of organization and conduction.

Make better decisions in favor of education certainly impact the quality of educational institutions which will give them greater competitiveness in an increasingly vicious market. The most competitive countries are those that are making the best use of ICT, which dominate and productively apply knowledge. In places like Puebla with a high number of schools of higher level, knowledge becomes a competitive advantage in the fundamental tool of the makers of educational management to make decisions that best promote the organization.

IT leaders may soon be academic and student affairs important collaborators. They can help answer accountability. through academic analysis, which is emerging as a new tool for a new era. "In the information age, one of the most influential institutions is education".

With the enterprise-wide systems that generate huge amounts of data, data warehouses aggregates different types of data and processing capacity, classifies and surfaces, academic analysis is becoming a new tool that can tackle what appears insoluble.

Richer data sets, new ways to extract and organize data, more sophisticated predictive models, and further research will drive the evolution of analytics. As the practice of analytical gets refined, colleges and universities can place more and better information in the hands of more people, enabling better decision-making. As a result, it is necessary that staff have over traditional IT skills.

The universities respond to the demand for greater accountability in higher education, the emerging practice of academic analytics likely to become a new, highly useful tool for a new, highly demanding era [13].

A key ingredient of analytic action is incorporating labor requirements in educational programs. Intervention strategies are based on powerful analytical steps to make a difference in college readiness and success for underserved students. Preparedness, awareness, financial and institutional responsibility for student success.

The world of information and knowledge is so diverse and abundant that it is vital to mastering the fundamentals associated with a skilled management methods to collect, find, interpret, analyze and recreate potentially useful knowledge required. Managing information in business organizations is, today, a key tool for survival in a changing and dynamic global market.

References

1. Carregal, M.: Planificación Estratégica. http://www.youtube.com/watch?v=fsnh5o0g_cy&safe=active
2. Cassasus, J.: Problemas de la gestión educativa en América Latina (la tensión entre los paradigmas de tipo A y el tipo B). UNESCO (2000)
3. García, L.: ¿Web 2.0 vs Web 1.0?. BENED (2007)
4. Pozner, P.: Diez módulos destinados a los responsables de los procesos de transformación educativa. IIPE - UNESCO Sede Regional Buenos Aires-Argentina (2000)

5. Louis, K.S. & Kruse, S.D.: Professionalism and community: Perspectives on reforming urban schools. Thousand Oaks, CA: Corwin Press (1995)
6. Zikopoulos, P., Eaton, C., de Roos, D., Deutsch, T. & Lapis, G.: Understanding Big Data: Analytics for Enterprice Class Hadoop and Streaming Data. IBM Corporation. Mc-Graw Hill (2012)
7. Benefitting from Big Data Leveraging Unstructured Data Capabilities for Competitive Advantage, http://www.insightdata.com.au/?p=1738
8. Cuesta, H.: Practical Data Analysis: Transform, model, and visualize your data through hands-on projects, developed in open source tools. Packt Publishing. Birmingham, UK (2013)
9. Norris, D., Baer, L., Leonard, J., Pugliese, L., Lefrere, P.: Framing Action Analytics and Putting Them to Work. EDUCAUSE Review 43(1), 1–10 (2008)
10. Siraj, F. and Ali, A. M.: Mining Enrollment Data Using Descriptive and Predictive Approaches. In: Knowledge-Oriented Applications in Data Mining, pp. 53–71 (2011)
11. Kiron, D., Palmer, D., Phillips, A.N., Kruschwitz, N.: Social Business: What Are Companies Really Doing?. MIT Sloan management review 53(4), 1–32 (2012)
12. Norris, D. M. and Baer, L. L.: Building Organizational Capacity for Analytics. EDUCAUSE 1–58 (2013)
13. Campbell, J. P., DeBlois, P. B. and Oblinger, D. G.: Academic Analytics a new tool for a new era, http://www. educause. edu/library/resources/building-organizationalcapacity-analytics

Edge detection for Very High Resolution Satellite Imagery based on Cellular Neural Network

Juan Manuel Núñez

Centro de Investigación en Geografía y Geomática "Ing. Jorge L. Tamayo", D.F., México
jnunez@centrogeo.org.mx

Abstract. In the context of Very High Resolution (VHR) satellite imagery, edge detection is one of the most important and difficult steps in image processing and pattern recognition. This paper presents the use of Cellular Neural Network (CNN) in edge detection and shows its capacity to locate and identify discontinuities in the gray levels of the objects that appear in VHR image. The results show that the characteristics of the CNN, in terms of its local connectivity, are those that allow greater extraction of continuous edges. The multi-layer processing structure design of the CNN allows to identify a definitive edge in urban environment. To evaluate the results, a metric of peak signal to noise ratio (PSNR) has been introduced as a manner to rank the accuracy of the resultant edge determined by the assessed methods. The extracted VHR features with the CNN edge detector include accuracy of edge location and better linking of edge segments.

Keywords: Cellular Neural Network, Very High Resolution satellite imagery, urban environment, edge detection, local connectivity.

1 Introduction

Today an important number of earth observation platforms are equipped with Very High Resolution (VHR) optical imagers with the purpose of increasing the visibility of terrestrial features, urban objects in particular, by reducing spectral heterogeneity per-pixel and thereby improving object segmentation [1, 2].

Successful delineation and segmentation of fine details in VHR satellite imagery, such as objects related the urban environment (buildings, cars, planters, roads crosswalks and individual trees), can be associated with the information content available in spatial, spectral and radiometric resolution in the imagery. The main idea behind edge detection is to find where abrupt changes in the intensity of an image have occurred. Therefore, edge is the connected boundary between two different regions that define different objects. Some of the usual alternatives for edge-detection algorithms in digital images are based on filtering the image through different masks such as Sobel, Prewitt, Roberts and LoG operators [3]. Although these filters work suitably and fast, they have some problems such as discontinuity in the edges, because they are very sensitive to noise which affects the images, by registering unwanted and

isolated edges. Other sophisticated methods have been developed in order to avoid this problem and find better solutions, e.g. the Canny edge detector [4, 5].

Due of the above, there is a need to explore and develop other methods, in order to improve the existing results In this paper two edge-detection methods are applied in VHR satellite imagery using cellular neural networks (CNN), to evaluate the property of the local connectivity likely associated with the high conservation of edges in the image pattern recognition. CNN is a new proposal for processing information based exclusively on local interactions. Edge-detection problem using CNN in the context of satellite imagery has presented good results in locating the edge in low spatial resolution images [6]. In a dense urban environment sensed by VHR optical sensors, the edge detection problem is very similar to the task of edge-detection problem in images affected by noise, due to buildings and trees that hide some objects of the scene [7]. The use of automatic edge detectors with a high degree of connectivity and reconstruction of edge segments lines facilitates the processes of lines's vectorization, which are used in managing geospatial databases.

The rest of this paper is organized as follows: first, a brief review the CNN model and the necessary architecture for edge detection in image processing is introduced; second, edge-detection simulations with very high-resolution satellite imagery are proposed. For the first case, we worked with a single-layer CNN application and for the second we used a multilayer CNN application. In both cases, a satellite image WorldView-3 (Mexico City) has been used. The result of choosing templates for edge detection using CNN is evaluated in comparison with other edge-detection algorithms. Finally, there is a discussion about the edge detection using CNN application.

2 Brief Description of CNN Model in Image Processing

Cellular neural networks (CNNs) introduced by [8] are locally connected analog arrays, that process large amounts of information in real time. This architecture is capable to perform time-consuming tasks such as pattern recognition and static or moving image processing [9]. In a CNN model, the basic processing unit is called cell, and its local connectivity is the most important characteristic because each cell is connected exclusively to its neighbors, therefore the adjacent cells that are defined in a neighborhood are the only ones that can interact directly with each other [10].

For a two-dimensional CNN with *MxN* array, the dynamics of each cell can be described by the following state equations:

$$\begin{cases} \dot{X}_{ij}(t) = -X_{ij}(t) + \sum_{C(k,l)\in N_r(i,j)} A(i,j;k,l)y_{kl}(t) + \sum_{C(j,l)\in N_r(i,j)} B(i,j;k,l)u_{kl} + I_{ij} \\ Y_{ij}(t) = f(X_{ij}(t)) = \frac{1}{2}\left(\left|X_{ij}(t)+1\right| - \left|X_{ij}(t)-1\right|\right) \end{cases} \quad (1)$$

where $X_{ij}(t)$, u_{ij} and $Y_{ij}(t)$ are the state, the input and the output of a cell $C(i, j)$ in the grid. The input is static, time independent, while the state and output vary with

time. Matrices $A(i,j;k,l)$ (feedback coefficients) and $B(i,j;k,l)$ (control coefficients) denote the connection templates from cell $C(k,l)$ to cell $C(i,j)$. I_{ij} represent the bias in the grid. The control template represents the coupling coefficients of the cells and it defines entirely the behavior of the CNN model with a given static input $|u_{ij}| \leq 1$ and initial condition $X_{ij}(0) = 0$. The output $Y_{ij}(t)$ is described as piecewise linear equation. A space-invariant standard CNN with 3x3 and neighborhood radius of $r = 1$ is defined uniformly by a string of 19 real numbers, called the cloning template which together with initial condition and static input can determine entirely the dynamical properties of the CNN [11]. Considering the invariant space, the interaction between the feedback operator A and the control operator B with the output and the initial state respectively can be written as follows [12]:

$$\dot{X} = X + A \otimes Y + B \otimes U + I \tag{2}$$

where $\otimes$ refers to the point-by-point multiplication, that in discrete mathematics corresponds to spatial convolution [13]. From this state equation the interconnection pattern is defined by the A and B cloning templates, the bias term I and the initial condition. They all together determine the evolution of a final state of CNN. Since image processing's point of view, these parameters determine the type of operation performed in an input image U, to obtain an output image Y, assuming an initial state of the network $X(0)$. Therefore, in an image processing context, the CNN can be understood as a sort of information processing system with interaction dynamic spatial-temporal, which can transform two-dimensional input images in a two-dimensional output image as well.

The cloning template elements constitute the CNN analog program and its determination is one of the most important problems in image processing studies [14, 15]. Regardless of what method was used to obtain the CNN parameters, it should be clear that each set of cloning template coefficients, along with the initial conditions, determines the type of processing over the input. In this paper we took advantage of existing cloning templates in the literature to solve edge-detection problem in the context of noisy images, in order to develop a CNN application related to VHR satellite imagery edge detection.

3 CNN Simulations for Edge Detection Applications

Two practical cases have been developed. The first is a single-layer CNN application whereas the second is a multilayer CNN application. In both cases a satellite image WorldView-3 (WV-3) (September 2014) for Mexico City with a GSD: 0.314 m (PAN) and 16 bits per pixel (unsigned) radiometric resolution has been used. The proposed method followed next steps: Step 1, for practical simulations MatCNN was used, a CNN toolbox for Matlab [16]. To begin implementing MatCNN in Matlab, a script file was written to define the CNN environment and to initialize multiple variables that determined the configuration of the CNN simulations. Feedback, cloning

template and a threshold value were introduced; and the time step=0.2 and number of iterations=25 were optimized for the simulation. Step 2, after CNN network configuration, the edge detection can be performed. For CNN results, the largest region was detected and the rest of the detected regions were sent to the background using a median filter with a neighborhood of 3-by-3, recovering pertaining information to the edge only. Step 3, in order to validate the results, the CNN edge-detection proposal was compared with other classics edge-detection operators also implemented in the Matlab environment: Prewitt, Roberts, LoG and Canny. To illustrate the performance of the single-layer and multilayer CNN, it was assessed local connectivity's property associated with the high preservation of boundaries in the edge-detection task through several parameters related with the pixel connection neighboring. Finally, the aspect of peak signal to noise ratio (PSNR) was calculated through computer simulations to compare the success of the edge detection.

3.1 Case 1. Single-layer CNN for Linear and Curved Edge Detection

In Chua's review [17], we find information about how the CNN can be chosen to perform many basic image processing operations. The author developed many CNN cloning templates. One of them is the template called Edgegray CNN, which can detect the edges in gray-scale images. The template has the form:

$$A=\begin{bmatrix}0 & 0 & 0\\ 0 & a & 0\\ 0 & 0 & 0\end{bmatrix}, B=\begin{bmatrix}c & c & c\\ c & b & c\\ c & c & c\end{bmatrix}, I=z \tag{3}$$

where a, b, c and z are real numbers. These parameters are optimized in order to satisfy the performance of the CNN. This is described through its global task and local rules [18]. Some parameters groups have been suggested from the literature, therefore in order to evaluate this first practical case, the chosen template is:

$$\{a,b,c,z\}=\{2{,}8,-1,0.5\} \tag{4}$$

This selection was made to evaluate the next example of edge detection from a VHR satellite image. In Figures 1 and 2, different visual results from the proposed method can be observed, compared with other edge detectors. Figure 1(a) shows an image cut of 100 x 100 pixel-wide image produced by the WV-3 satellite. Figure 1(b) shows the linear truth shape that defines the edge to extract. The program output for linear edge detection using the proposed CNN template is shown in Figure 1(c). The use of a median filter with a 3-by-3 neighborhood, to recover only information related to the edge, is shown in Figure 1(d). Figures 1(e, f, g and h) are the result of using Canny, LoG, Sobel and Prewitt operators applied to the same image, respectively. It notes that it is not possible to obtain a continuous edge, solely several separate segments.

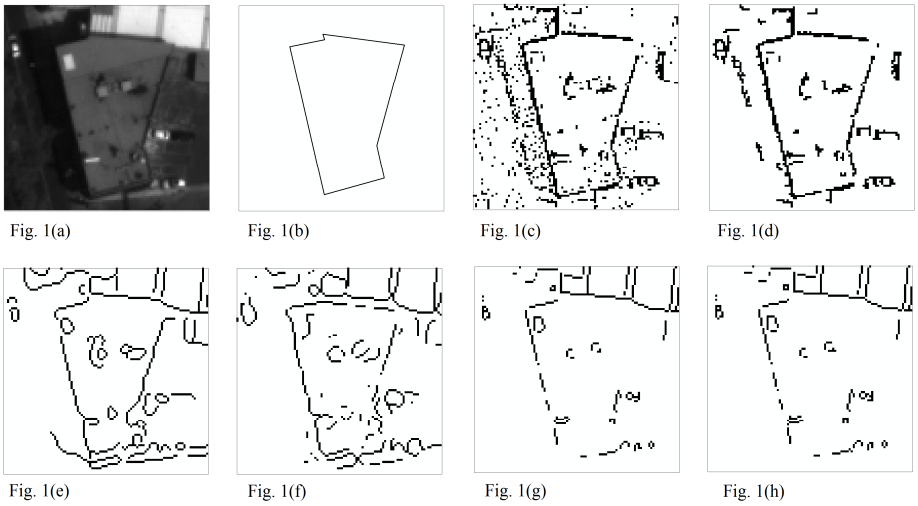


Fig. 1. Simulation result for linear edge detection

A second image cut of 100 x 100 pixel-wide image also produced by the same satellite is shown in Figure 2(a). Figure 2(b) shows the curved truth shape that defines the edge to extract. The program output for curved edge detection after using the same CNN template is shown in Figure 2(c). The use of a median filter with a 3-by-3 neighborhood, to recover only information related to the edge, is shown in Figure 2(d). Figures 2(e, f, g and h) give the same outputs for the filters proposed in Figure 1.

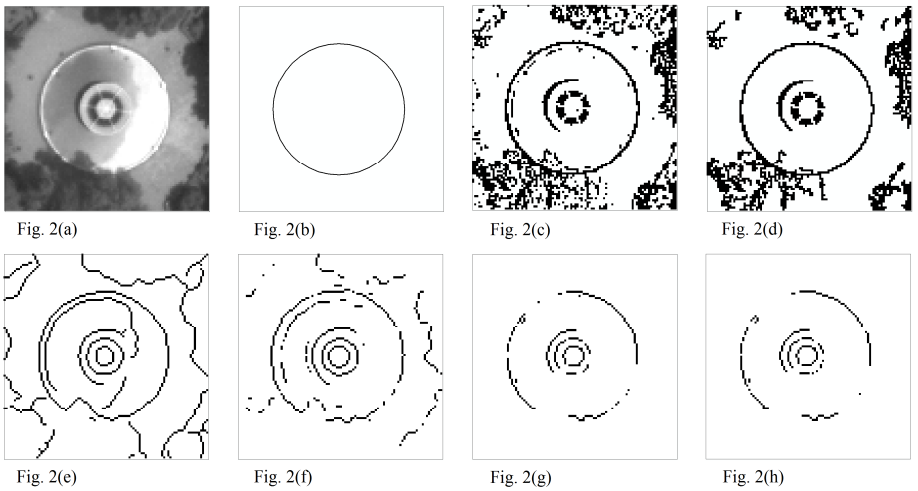


Fig. 2. Simulation result for curved edge detection

3.2 Case 2. Multi-layer CNN for Edge Detection

In the preceding exercise, a single-layer CNN was presented. This concept can be extended to multi-layer architecture, the CNN that includes a very broad multi-layer analog array which also has local interconnections [19]. The organization of this CNN is very similar to that of a multi-layer feedforward perceptron [20], although in this case, the state in each layer has continuous-variable dynamics rather than a bistable state. Multi-layer CNN array is an example of the process which is feasible through a building block approach. The example shown is a 3-layers CNN information processing system, that combines diffusion layers, threshold and edge detection. The CNN cloning templates are given in Table 1.

Table 1. Diffusion, threshold and edge detection templates [21]

Layer	A	B	I
Diffusion	$\begin{bmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & -2 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	0
Threshold	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	0
Edge detection	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0.25 & 0.25 & 0.25 \\ 0.25 & -2 & 0.25 \\ 0.25 & 0.25 & 0.25 \end{bmatrix}$	-1.5

The WV-3 image in Figure 3 is a planter located in a public square. The planter's edge building material is very similar to the square's background. A diffusion layer controlled with the values of the selected template was applied to Figure 3(a). The resulting image is displayed in Figure 3(b). The second process is thresholding and binarization of the previous image; the result is shown in Figure 3(c). Finally, an edge-detection layer with selected values of a template for a binary image was applied on the last image; the result is shown in Figure 3(d).

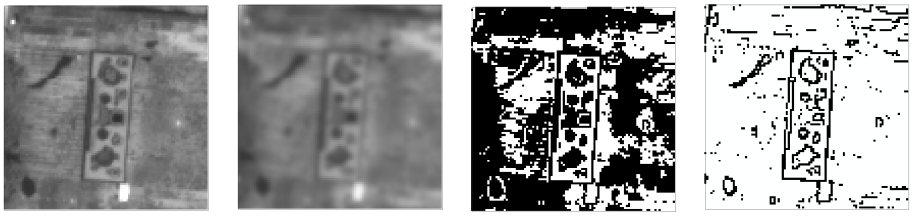


Fig. 3(a) Fig. 3(b) Fig. 3(c) Fig. 3(d)

Fig. 3. Results of applying linear templates in a multilayer CNN

The simulation result is shown in Figure 4. Edge detection results using CNN approach after using a median filter with a 3-by-3 neighborhood are evident from Fig. 4(a) and 4(b). Classic Canny and LoG operators are shown in Fig. 4(c) and 4(d) respectively.

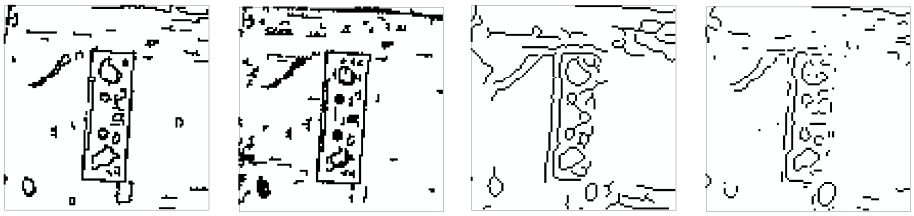


Fig. 3(a) Fig. 3(b) Fig. 3(c) Fig. 3(d)

Fig. 4. Simulation result for edge detection using multi-layer CNN, single-layer CNN, Canny and LoG

4 Analysis of Simulation Results

In order to evaluate the accuracy of set performance of the CNN under different shapes of edge, we will introduce the Peak Signal to Noise Ratio (PSNR) as a quality metric in order to estimate the number of connected pixels, denoted by p_{ij}, involved in edge definition.

$$PSNR = 20 * \log_{10} \frac{MAX_p}{\sqrt{MSE}} \tag{5}$$

where MAX_p is the maximum possible pixel value of the image represented by bits per pixel. Mean Square Error (MSE) is the cumulative squared error between the extracted edge and the truth shape that defines the edge. The MSE is given by:

$$MSE = \frac{1}{MxN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\tilde{p}_{ij} - \hat{p}_{ij} \right)^2 \tag{6}$$

where $\tilde{p}_{ij}$ are the total pixels that define the edge in an ideal image, $\hat{p}_{ij}$ are the connected pixels which belong to the detected edge at the output by using the edge-detection method selected. The result will show the difference between the methods, which allows to know the edge's shape for each example. This kind of metric is a benchmark to evaluate the grade of connectivity of the resultant edge determined by the assessed methods [6]. Results of PSNR calculation for both the single-layer CNN for linear and curved edge detection (Figures 1 and 2) and multi-layer CNN for edge detection (Figure 4) are shown in Table 2.

Table 2. Edge-detection results: Comparisons CNN with other methods

Method	Percent of connected pixels	8-connected segments	Average pixels per edge segment	MSE	PSNR
Case 1. An example of linear edge detection					
Single-layer CNN	37.3	4	25	2.220	44.7
Canny	28.3	5	17.2	3.534	42.6
LoG	27.2	8	8.6	4.162	41.9
Sobel	21.9	10	5.9	4.368	41.7
Prewitt	21.1	20	3.1	4.410	41.7
Case 1. An example of curved edge detection					
Single-layer CNN	72.5	5	37.4	0.504	51.5
Canny	43.8	6	18.2	2.103	44.9
LoG	43.4	7	16	2.132	44.8
Sobel	40.7	9	11	2.340	44.4
Prewitt	38.0	10	9.6	2.560	44
Case 2. An example of Multi-layer CNN					
Multi-layer CNN	59.9	6	7.5	0.578	50.5
Single-layer CNN	29.4	9	6.4	0.828	48.6
Canny	13.4	13	2.2	1.988	45.1
LoG	10.0	19	2.1	2.592	43.6

For each of the evaluation methods, the table shows the connectivity analysis, which includes the percent of pixels that belongs to the detected edge in reference to the truth edge, the number of connected segments that defines the edge for 8-connected neighborhood as well as the pixels average per edge segment. A higher percent of connected pixels that belongs to the detected edge means a lower MSE, and as can be noticed, this generates a high value of PSNR. A higher value of PSNR generally indicates that the reconstruction is of higher quality. Here, the "signal" is the truth shape that defines the edge to extract in the original image, and the "noise" is the error represented as the non-identified pixels of the edge by the segmentation process.

It can be observed first, that the percent of pixels which belong to the detected edge obtained with the proposed CNN approach is composed of a large number of pixels, in comparison with the other edge-detection methods for all cases. Furthermore, the number of connected segments is always the smallest and the pixels average per segment has the largest number of pixels. In terms of PSNR, the single-layer CNN proposal showed the best result; however, for a curved edge the results are better than those from the linear edge. This can be explained in terms of the object's height to be detected and the context of its shadow that describe the edge-detection task. The multi-layer CNN proposal detected the edges as well as the ideal and noiseless edge detection. In fact, it is remarkable that the multi-layer CNN proposal is more accurate than the other evaluated methods. The method based on multilayer CNN determines the ideal edges of the image in detail and with high precision. It can be seen that the

single-layer CNN performance is satisfactory but the output image has too much noise. Results from Canny and LoG operators respectively are just unsatisfactory.

5 Conclusions

The edge in a very high spatial resolution image gives an indication of the physical extent of an object within the image, and its accurate segmentation is very important because the performance of the tasks after the edge detection, such as image segmentation and image registration, depend on the information of the edge. The edge is the connected boundary between two different regions that define different objects. The algorithms dedicated to obtain the edge must meet two principles: detection of all abrupt changes in the intensity of an image that defines an object and the largest number of connected regions that describes these changes. So that, local connectivity's property in an edge detection task is essential and this property is intrinsic to the CNN model.

In each of the simulations, single or multi-layer CNNs for edge-detection simulations in VHR satellite imagery, similar to the edge detection problem in images affected by noise, have been considered because of the objects that hide some other objects of the scene. Satisfactory results could be achieved by choosing standard templates among a few experiments available in the literature. Experimental results have shown that the Single-layer CNN for linear and curved edge detection is an excellent alternative for connected edges detection in the VHR context. In case of the multi-layer CNN, composed of cells that have several single-layer arrays stacked one above the other, it is possible a better edge-detection task than when using the other evaluated methods.

Acknowledgment. The author wishes to thank GttImagIng, authorized dealer and certified DigitalGlobe, for the WV-3 image used for this work.

References

1. Poli, D., Remondino, F., Angiuli, E., Agugiaro, G.: Radiometric and geometric evaluation of GeoEye-1, WorldView-2 and Pléiades-1A stereo images for 3D information extraction. ISPRS Journal of Photogrammetry and Remote Sensing 100, 35–47 (2015)
2. Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q.: Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote sensing of environment 115(5), 1145–1161 (2011)
3. Shrivakshan, G.T., Chandrasekar, C.: A comparison of various edge detection techniques used in image processing. International Journal of Computer Science Issues 9(5), 272–276 (2012)
4. Maini, R., Aggarwal, H.: Study and comparison of various image edge detection techniques. International journal of image processing 3(1), 1–11 (2009)
5. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence (6), 679–698 (1986)

6. Gazi, O.B., Belal, M., Abdel-Galil, H.: Edge Detection in Satellite Image Using Cellular Neural Network. International Journal of Advanced Computer Science & Applications, 5(10), 61–70 (2014)
7. Bedawi, S.M., Kamel, M.S.: Segmentation of very high resolution remote sensing imagery of urban areas using particle swarm optimization algorithm. In: Campilho A., Kamel M (eds.) ICIAR 2010, Part I. LNCS, vol. 6111, pp. 81–88. Springer Heidelberg (2010)
8. Chua, L.O., Yang, L.: Cellular neural networks: theory. IEEE Transactions on Circuits and Systems 35(10), 1257–1272 (1988)
9. Chua, L. O., & Yang, L.: Cellular neural networks: Applications. IEEE Transactions on Circuits and Systems 35(10), 1273–1290 (1988)
10. Li, H., Liao, X., Li, C., Huang, H., & Li, C.: Edge detection of noisy images based on cellular neural networks. Communications in Nonlinear Science and Numerical Simulation 16(9), 3746–3759 (2011)
11. Matsumoto, T., Chua, L.O., Suzuki, H.: CNN cloning template: connected component detector. IEEE Transactions on Circuit and System 37 (5), 633–635 (1990)
12. Chua, L. O., Roska, T.: Cellular neural networks and visual computing: foundations and applications. Cambridge University Press (2002)
13. Aydogan, D.: CNNEDGEPOT: CNN based edge detection of 2D near surface potential field data. Computers & Geosciences 46, 1–8 (2012)
14. Wang, W., Yang, L. J., Xie, Y. T., & An, Y. W.: Edge detection of infrared image with CNN_DGA algorithm. Optik-International Journal for Light and Electron Optics 125(1), 464–467 (2014)
15. Zarándy, Á.: The art of CNN template design. International Journal of Circuit Theory and Applications 27(1), 5–23 (1999)
16. Karacs, K., Cserey, G. Y., Zarandy, A., Szolgay, P., Rekeczky, C. S., Kek, L., Szabó, V., Pazienza, Roska, T.: Software Library for Cellular Wave Computing Engines in an era of kilo-processor chips, Version 3.1. Cellular Sensory and Wave Computing Laboratory of the Computer and Automation Research Inst., Hungarian Academy of Sciences and the Jedlik Laboratories of the Pazmany P. Catholic University, Tech. Rep (2010)
17. Chua, L. O.: CNN: a version of complexity. Int J Bifurcat Chaos 7 (10), 2219–2425 (1997)
18. Zhang, M., Min, L., Zhang, X.: Automatic Robust Designs of Template Parameters for a Type of Uncoupled Cellular Neural Networks. In: Foundations of Intelligent Systems. Springer Berlin Heidelberg. pp. 577–590 (2014)
19. Chua, L.O., Yang, L., Krieg, K.R.: Signal processing using cellular neural networks. Journal of VLSI signal processing systems for signal, image and video technology 3(1-2), 25–51 (1991)
20. Mrugalski, M.: Advanced neural network-based computational schemes for robust fault diagnosis. Springer. (2014)
21. Parmaksızoğlu, S., & Alçı, M.: A novel cloning template designing method by using an artificial bee colony algorithm for edge detection of cnn based imaging sensors. Sensors 11(5), 5337–5359 (2011)

Towards the Automatic Identification of Spanish Verbal Phraseological Units

Belém Priego Sánchez[1,2], David Pinto[2] and Salah Mejri[1]

[1]LDI, Université Paris 13, Sorbonne Paris Cité
Paris, France
belemps@gmail.com, smejri@ldi.univ-paris13.fr

[2]Benemérita Universidad Autónoma de Puebla, FCC
Puebla, Pue., México
dpinto@cs.buap.mx

Abstract. Verbal Phraseological Units are expressions made up of two or more words in which at least one of these words is a verb that plays the role of the predicate. Their main attribute is that this form of expression has taken on a more specific meaning than the expression itself. The automatic recognition of this type of linguistic structures is a very important task, since they are the standard way of expressing a concept or idea. This paper describes the outgoing advances of a PhD research work in which it is attempted to construe a methodology which allows to automatically identify these linguistic structures for the Mexican Spanish language. It is presented a set of hypotheses which will allow to produce novel proposals in the way of automatically identifying a verbal phraseological unit in a raw text. Additionally, we have presented experiments carried out in this sense, for example, by employing machine learning methods. Finally, we show a lexical resource which is product of the current advances in this PhD thesis.

Keywords: Verbal phraseological units, Automatic identification, Corpus linguistics.

1 Introduction

Some concepts are expressed in language through set of words or phrases, which intuitively are employed native speakers, thus characterizing different cultural communities. Phraseology, considered a cultural heritage of the linguistic community [8], aims to study these blocks of words, which are usually refered as phraseological units.

The study of phraseological units has a growing importance in recent years, in part because the linguistic and computational linguistic community has understood that this phenomenon covers all the sentence components [11], a fact that involves different dimensions of the natural language: linguistics, pragmatics, culturals, among others [12]. A phraseological unit is basically one type of multiword expression, and under this denomination one assumes a wide range

of linguistic constructions such as idioms (*storm in a teacup*, *sweep under the rug*), fixed phrases (*in vitro*, *by and large*, *rock'n roll*), noun compounds (*olive oil*, *laser printer*), compound verbs (*take a nap*, *bring about*), etc.

In this research work, we are particularly interested in studying mexican spanish phraseological units containing one verb as the grammar centre, i.e., Verbal Phraseological Units (VPU) which present a challenged degree of fixation in comparison with other phraseological units [18], for example, "*Leer entre líneas (To read between the lines)*". Actually, this paper aims to present an overview of the doctoral research that mainly focuses on identifying whether or not a verbal phraseological unit is present in a given text, a process that implies to analyze raw text and the features in the context of the phraseological unit for creating computational algorithms that allow to fullfil the mentioned task in environments highly scalable.

The remaining of this paper is structured as follows. Section 2 describes the motivation and shows a number of hypotheses we are proposing which we consider useful for the development of this research work. Section 3 presents some relevant works found in literature. Section 4 shows the results obtained in the experiments carried out in this research work, presenting first the lexical resources (lexicon and corpus), and later the results obtained up to now. Finally, in Section 5 the conclusions and perspectives of this research work are given.

2 Motivation and Advances

Phraseological Units (PU) are multiword lexical units that are characterized for presenting certain degree of fixation[1] or idiomaticity in its components. In other words, phraseological units are a combination of words whose meaning are not necessarily deduced from the meaning of its components, i.e., the words together can mean more than their sum of parts [10, 6].

These linguistic structures are also known in literature as phrasemes, fixed expressions, and multiword expressions[2]. While easily mastered by native speakers, their interpretation poses a major challenge for computational systems, due to their flexible and heterogeneous nature. Furthermore, phraseological units are not nearly as frequent in lexical resources as they are in real-world text, and this problem of coverage may impact the performance of many natural language processing tasks.

Phraseological units are widely employed by human beings. In [7], it is said that the number of phraseological units (there expressed as multiwords expressions, in the terminolgy used by the author), is in the same order of magnitude as the number of simple or isolated words.

A Verbal Phraseological Unit (VPU) is a PU that contains one verb as the grammar centre. For example, the PU *to come to one's sense* means *to change*

[1] Fixation is a inherent property of natural language that occupies a central role in the description of phraseological units.

[2] Throughout this paper we will employ the term *phraseological unit*, assuming that the terms aforementioned have a similar meaning.

one's mind, or *to fall into a rage* means *to get angry*. Verbal phraseological units perfectly illustrate the overall saturation as indicated in [13]. Taking this characteristic into account and also the fact that verbal phrases have a paradigmatic rupture, make us focus our attention in this type of phraseological units, a task that implies a very high challenge research line in terms of semantic identification and classification of phraseological units.

In this way, is it very important to study the nature of such linguistic structures, so that we can be able to construe automatic methods for dealing with those units. In particular, we have centered our research in a set of hypotheses associated to the behaviour of verbal phraseological units. These hypotheses will determine the reseach path of this PhD thesis.

- *Fixation hypothesis.* The fixer a verbal phrase is, the higher it is its likelihood of being a verbal phraseological unit. We will substitute each component of a target verbal phrase (candidate VPU) with their near synonyms in order to verify if the new verbal phrase loses the meaning. In such case, it will be considered that a fixation phenomenon exist, thus verifying that the candidate VPU is a real VPU. In order to verify the meaning of the new verbal phrase, we are considering to use a reference corpus in which we may search the evidence of such phrase.
- *Translation hypothesis.* The more literal the translation of a verbal phrase is, the lower it is its likelihood of being a verbal phraseological unit. We will translate the verbal phrase from one language to another one. Thereafter, we will look for evidence of such translation in a reference corpus written in the target language.
- *Internal attraction and contextual co-reference hypothesis.* The greater the internal attraction and the lower the contextual co-reference in a verbal phrase are, the higher it is the likelihood of the verbal phrase of being a verbal phraseological unit. We will employ statistical methods for determining the level of internal attraction and contextual co-reference between the terms of the verbal phrase and those of their context.
- *Terminology domain hypothesis.* The greater the number of vocabulary terms out of the current domain for the verbal phrase is, the higher it is its likelihood of being a verbal phraseological unit. Is quite common to employ terms out of the current domain in a real VPU, therefore, we will identify terminology out of the current domain in order to determine whether or not, such verbal phrase is a real VPU.

We are currently working in the different methods for validating or rejecting the aforementioned hypotheses. In the following sections we will show the current advances with respect to this research work.

3 Related Work

There exist at least three different methods for recognizing phraseological units in a raw text: the application of constructed "local" grammars, the employment

of dictionaries, and the use of statistical processes [1]. Dictionaries are the common method for localizing phraseological units, but it is insufficient when we are considering to discover new fixed expressions, i.e., those that were not stored in the dictionary. Constructing local grammars is a knowledge based method that provides a broad reach because it may find new PU's which have similar linguistic structures than those considered when the grammars were constructed. Statistical methods usually employ document term frequencies for searching evidence of linguistic phenomena. For example, in [16] it is presented a statistical-based method for detecting verbal phraseological units; this paper includes a particular procedure for constructing lexical resources which may be later employed for machine learning methods.

Eventhough we are using the term phraseological unit in this paper, we are aware that there are a number of research works in which the term multiword expression (MWE) is employed. The following references are examples of such works [2, 4, 9, 20, 14].

We should emphasize that many other works associated with MWE exist in literature, mainly because of the different forums that are encouraged by the computational linguistic community[3], in which we can be able to find many interesting papers. However, in literature there are other papers employing other terminology which refer to phraseological units, therefore, we mention some of them as follows.

In [19] the authors propose a statistical measure for calculating the degree of acceptability of light verb constructions, based on their linguistic properties. This measure shows good correlations with human ratings on unseen test data. Moreover, they find that their measure correlates more strongly when the potential complements of the construction are separated into semantically similar classes. Their analysis demonstrates the systematic nature of the semi-productivity of these constructions.

Paul Cook et al. presents the VNC-Tokens dataset, a resource of almost 3000 English verb-noun combination usages annotated as to whether they are literal or idiomatic [3]. This authors began with the dataset used by Fazly and Stevenson [5], which includes a list of idiomatic verb-noun combinations (VNCs), and they found that approximately half of these expressions are attested fairly frequently in their literal sense in the British National Corpus (BNC)[4]. Their study is based on the observation that the idiomatic meaning of a VNC tends to be expressed in a small number of preferred lexico-syntactic patterns, referred to as canonical forms [17].

In [5], the authors investigate the lexical and syntactic flexibility of a class of idiomatic expressions. They develop measures that draw on such linguistic properties, and demonstrate that these statistical corpus-based measures can be successfully used for distinguishing idiomatic combinations from non-idiomatic ones. They also propose a process for automatically determining which syntactic

[3] http://multiword.sourceforge.net
[4] http://www.natcorp.ox.ac.uk/

forms a particular idiom can appear in, and hence should be included in its lexical representation.

We consider that other works associated to the identification of phraseological units exist in literature, for example, in [15] a set of experiment towards the identification of polarity has been presented, however, the exhaustive discussion of the state of the art is out of the scope of this paper.

The following section presents the experiments carried out attempting to detect whether or not a VPU exist in a raw text.

4 Results

Regarding our current advances in the task of automatic identification of Spanish verbal phraseological units, we have considered the Mexican newspaper domain and a number of Mexican verbal phraseological units, thus, firstly we describe the lexical resources constructed for the proposed task. The VPU identification approach employed is based in supervised machine learning techniques, a branch of artificial intelligence that concerns the construction and study of computational systems that can learn from supervised data.

The supervised machine learning techniques are able to learn the human process of identifying verbal phraseological units based on features fed in the classifier by means of the manually annotated corpus. In order to have a perspective of the type of classifier that can best deal with the problem of automatic detection of VPUs, we have selected one learning algorithm from four different types of classifiers: Bayes, Lazy, Functions and Trees. The obtained results are discussed in Section 4.2, and the lexical resource obtained up to now is presented in Section 4.3.

4.1 Dataset

Supervised machine learning methods assume that we have supervised data from which they can learn knowledge. In this case, we need corpora manually annotated by experts indicating whether or not a certain text contains a verbal phraseological unit. Thus, we constructed a dataset for the experiments proposed in this paper by selecting a number of news stories (from a mexican newspaper) having and not having verbal phraseological units. In order to do so, firstly, we extracted all the verbal phraseological units from a dictionary named "Dictionary of Mexicanisms"[5]. In particular, we have collected 1,219 verbal phraseological units from this dictionary which have been stored in a database, considering they to be further employed for identifying their regular use in the Mexican newspaper domain. For the purpose of the experiments carried out in this paper, we have selected only the most representative ones, which in this case resulted to be 56 VPUs. In order to select those VPUs we have taken into account their frequency of occurrence in the corpus, selecting at the end the most frequent ones.

[5] http://www.academia.org.mx/

By using information retrieval techniques we have found 3,164 news stories containing at least one occurrence of some of the verbal phraseological units selected. This process considers the occurrence of original VPU any of it morphological variants; for this purpose, we have lemmatized both, the VPU and the text in the news story, so that we can be able to find variations of the VPU in the target text. The news stories have been gathered from Mexican newspapers belonging to the Mexican Editorial Organization[6]. All the texts compiled are written in Mexican Spanish and contain news stories that occurred between the years 2007 and 2013.

As a consequence of counting the occurrence of Mexican verbal phraseological units in the corpus gathered, we were able to construct a labeled corpus which may be further used as a training corpus for supervised machine learning methods with the aim of identifying whether or not a news story contains a VPU. The context gathered have been manually annotated by 5 human annotators with an agreement inter-annotators greater than 80%. Each human annotator was asked to manually clasify when a given raw text contained a VPU (Class 1), or when that text did not contained a VPU (Class 2). The description of the corpus employed is shown in Table 1.

Table 1. Description of the manually annotated corpus

Feature	**Class 1 (VPU)**	**Class 2 (¬VPU)**	**Total**
Instances	1,959	1,205	3,164
Tokens	117,715	63,600	181,315
Vocabulary	16,359	10,817	20,953
Minimum length	3	3	3
Maximum length	2,291	302	2,291
Average length	60.09	52.78	57.31

In the experiments carried out, all the texts were represented by means of a vector of n-grams frequencies, with $n = 1, 2$ and 3. Frequencies greater than two for the n-grams were only considered for the vector features. The corpus was used as both, training and test corpus by means of a v-fold cross validation process (v=10). The results obtained in the experiments are shown in Section 4.2.

4.2 Obtained Results

In this section we are presenting the accuracy obtained by each classifier when attempting to identify whether or not a VPU exist in a given raw text.

In Table 2 we show the percentage of instances classified correctly and incorrectly. Basically, this table summarizes the weighted average results of the previously shown result tables. As it can be seen, the results obtained are highly enough to be seriously considered in the process of automatic detection of verbal

[6] http://www.oem.com.mx/

phraseological units in raw texts. All the classifiers have obtained a percentage above 71%. It is the J48 implementation of C4.5 that has generated a decision tree able to classify correctly the 76.74% of instances.

Table 2. Percentage of correctly vs. incorrectly instances classified

Classifier	Type	Correct (%)	Incorrect (%)
Naïve Bayes	Bayes	74.05	25.95
K-Star	Lazy	71.14	28.86
SMO	Functions	75.32	24.68
J48	Trees	76.74	23.26

4.3 A Lexicon of VPUs with Probabilities

The news stories were collected from the web by means of an information retrieval system, employing the candidate VPUs as input query. Thus, we obtained texts from Internet which may contain or not a real VPU inside (see Table 1). In other words, the distribution of occurrence of a given VPU can be approximated by counting the number of times the candidate phrase is really a VPU, and the number of times this sequence of words is not a real VPU. By doing so, it is possible to estimate the probability of a given sequence of words (candidate VPU) of being a real VPU in real texts. This lexical resource may be of high benefit for the computational linguistic community since, up to our knowledge, they have not been constructed for restricted domain corpora, or at least they have not been considered with that amount of data. We then, provide public access to this lexical resource to the community, by requesting it to any of the authors of this paper. Up to now, this lexicon contains only 56 entries, because we have selected only the most frequent VPUs from the total we have collected from the above mentioned dictionary of mexicanisms; however, as further work we are planning to apply exactly the same methodology for introducing more entries to this lexicon. A sample of the entries of this lexicon is shown in Table 3.

5 Conclusions

In this paper we have presented advances towards the automatic identification of the presence of verbal phraseological units in raw texts. We consider particularly important, the set of hypotheses proposed, because they will lead the current research of this PhD thesis. We are very interested in obtaining feedback about these hypotheses, and thus this is the reason of presenting this paper in this forum.

Additionally, as a manner of example, we have presented an experiment in which we compared four different supervised classifiers with the aim of determining whether or not exist significant differences among the results obtained

Table 3. Lexicon of VPUs with probabilities of being vs not being VPU in the news domain context

Verbal phraseological unit	**Probability of being a real VPU** P**(VPU)**	**Probability of not being a real VPU** $P(\neg$**VPU)**
darse por vencido (*to give up*)	0.49	0.51
salir a flote (*to keep one's head above water*)	0.83	0.17
comer el mandado (*to take advantage of*)	0.94	0.06
pegar su chicle (*to catch somebody's eye*)	0.95	0.05
ponerse la camiseta (*to put one's back into it*)	0.57	0.43
valer madre (*to be worthless*)	0.98	0.02
echar porras (*to encourage someone*)	0.52	0.48

by applying each supervised classifier in the process of automatic identification of VPU's in raw texts. The revision 8 of the C4.5 decision tree learner obtained the best results for the task executed in this paper, obtaining 76.74% of accuracy. We still interested in improving the performance obtained by analyzing other features which can be used in the classification process, this issue will be considered as future work.

An additional interesting contribution was the construction of a lexicon of 56 VPUs, each one containing an estimate of its probability of being a real VPU in a news stories domain[7]. As future work, we are planning to increase the number of entries to this interesting lexicon.

Acknowledgments. This paper has been partially supported by the CONACyT grant with reference #218862/314461 and CONACyT Project #225784.

References

1. Buvet, P.A.: Vers l'elaboration d'un dictionnaire unique des prédicats du français : DEESSE. Dictionnaire Electronique Syntactico-Sémantique. In: Description linguistique pour le traitement automatique du français. pp. 23–42 (2008)
2. Church, K.: How many multiword expressions do people know? TSLP 10(2), 4:1–4:13 (2013)

[7] The lexicon has been provided freely available for research purposes to any people that request it to any of the authors of this paper, considering this paper as the corresponding reference for every one that use the lexical resource.

3. Cook, P., Fazly, A., Stevenson, S.: The VNC-tokens dataset. In: Proceedings of the MWE workshop ACL. pp. 19–22 (2008)
4. Davis, A.R., Barrett, L.: Lexical semantic factors in the acceptability of english support-verb-nominalization constructions. ACM Trans. Speech Lang. Process. 10(2), 5:1–5:15 (2013)
5. Fazly, A., Stevenson, S.: Automatically constructing a lexicon of verb phrase idiomatic combinations. In: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL). pp. 337–344 (2006)
6. Huerta, P.M.: Estudio contrastivo lingüístico y semántico de las construcciones verbales fijas diatópicas mexicanas/españolas. In: Las construcciones verbo-nominales libres y fijas. pp. 179–198 (2010)
7. Jackendoff, R.: The Architecture of the Language Faculty, vol. 28. MIT Press (1997)
8. Lamiroy, B.: Les expressions figées: á la recherche d'une définition. In: Blumental et Mejri 2008. pp. 85–98 (2008)
9. Levin, B.: English verb classes and alternations : a preliminary investigation. Chicago Press, University (1993)
10. Martinez-Blasco, I.: Verbos soporte y fijación lexica. In: Las construcciones verbo-nominales libres y fijas. pp. 47–59 (2008)
11. Mejri, S.: Le figement lexical. Descriptions linguistiques et structuration sémantique. In: Publications de la faculté des lettres de Manouba, Tunis (1997)
12. Mejri, S.: Catégories linguistiques et étiquetage de corpus. In: L′information grammaticale, Peeters, Paris (2007)
13. Mejri, S.: Constructions à verbes supports, collocations et locutions verbales. In: La traduction des Mejri Salah (2008)
14. Nissim, M., Zaninello, A.: Modeling the internal variability of multiword expressions through a pattern-based method. ACM Trans. Speech Lang. Process. 10(2), 7:1–7:26 (2013)
15. Priego Sánchez, B., Pinto, D., Mejri, S.: Evaluating polarity for verbal phraseological units. In: Human-Inspired Computing and Its Applications - 13th Mexican International Conference on Artificial Intelligence, MICAI 2014, Tuxtla Gutiérrez, Mexico, November 16-22, 2014. Proceedings, Part I. pp. 191–200 (2014)
16. Priego Sánchez, B., Pinto, D., Mejri, S.: Metodología para la identificación de secuencias verbales fijas. Research in Computer Science 85(1), 45–56 (2014)
17. Riehemann, Z., Wasow, T., Copestake, A.A., Clark, E.V., Zwicky, A.M.: A constructional approach to idioms and word formation. Tech. rep., Stanford University. Dept. of Linguistics (2001)
18. Sfar, I.: Polylexicalite et continuite prédicative: le cas des locutions verbales figées. In: Las construcciones verbo-nominales libres y fijas. Aproximación contrastiva y traductológica. pp. 213–221 (2008)
19. Stevenson, S., Fazly, A., North, R.: Statistical measures of the semi-productivity of light verb constructions. In: Proceedings of the Workshop on Multiword Expressions: Integrating Processing. pp. 1–8. MWE '04, Association for Computational Linguistics, Stroudsburg, PA, USA (2004)
20. Vincze, V., T., I.N., Zsibrita, J.: Learning to detect english and hungarian light verb constructions. ACM Transactions on Speech and Language Processing 10(2), 6:1–6:25 (2013)

Performance Evaluation for a Multimodal Interface of a Smart Wheelchair with a Simulation Software

Amberlay Ruíz Serrano, Ruben Posada Gómez, Albino Martínez Sibaja, Alberto Alfonso Aguilar Lasserre, Giner Alor Hernández and Guillermo Cortes Robles

Instituto Tecnológico de Orizaba, Orizaba, Veracruz, México
amberlay_21@outlook.com, pgruben@yahoo.com, albino3_mx@yahoo.com, albertoaal@yahoo.com, gineralor@hotmail.com, gc_robles@hotmail.com

Abstract. Handling systems for power wheelchairs are very useful for users who have difficulty moving by themselves; however, most of the current models are designed exclusively for diplegic people and quadriplegics users with a specific disability. But it doesn't exist yet a wheelchair that assist it properly and efficiently the different disabilities that may present a quadriplegic due to the level of the injury. Therefore, the main contribution for this paper was a performance evaluation for a multimodal interface system of a smart wheelchair with different control methods for patients with severe spinal cord injuries (SCI) through a software simulation to make more secure the movement of the wheelchair and thereby standardize the basics of wheelchairs for quadriplegics in a more natural way according to their disabilities.

Keywords: Power Wheelchair, Tongue Control, Speech Recognition, Motor Disability, Inter-faces, Simulation Software.

1 Introduction

Recently studies shows that children and adults with physical disabilities such as loss of muscle control quadriplegia or paralysis, are essentially benefited because they regain their independent mobility through manual wheelchairs, electric wheel-chairs or scooters [1,2]. However, the current wheelchairs, such as electric wheel-chairs or manual wheelchairs represent an alternative to partially recover the ability to move by themselves. If the wheelchair user has a very limited upper body, electric wheelchair is the perfect choice for this person. However, this is not an option for quadriplegic users, who suffer paralysis of all their limbs. The proposal of this article is to focus only on quadriplegic users, because there is a problem, which often is not taken into account when a wheelchair is designed for them and the reason is: "Not all quadriplegics are equal." On one side, an individual may need assistance all the time and on

the other, a per-son can easily be able to live independently with appropriate assistance technology. Nowadays there is a tendency to create Human Computer Interactions (HCI) taking advantage of the skills that users with disabilities still preserve, for example, for patients who cannot use any of its members, but which can still use the movement of the eyes [3], face [4], hands [5], eyebrows [6] or voice [7]. This allows to create different control methods and apply them to a wheelchair, however each of these methods have efficiency problems and it doesn't exist smart wheelchairs developed and commercially available under these technologies in our country. The objective is to make a smart wheelchair which can enabling users with quadriplegia, retrieving a way to move themselves at the same time that this adapts to the user, regardless of whether the person has one or more disabilities. The proposed system is looking for an HCI with a multimodal interface by applying it to a smart wheelchair, to goal that different users with quadriplegia can use it, independently of the degree of SCI. To achieve this, it was also performed a study with different patients. These used the proposed system through a simulation soft-ware to learn how to use the smart wheelchair with an obstacle course in the software. After the users complete the course in a real smart wheelchair and this was compared against the results of similar systems by other researchers, to establish standards aimed at any quadriplegic people and providing a natural language for controlling the wheelchair.

2 Multimodal Interface System

The need to create interfaces that allow a more natural interaction with a wheelchair has motivated this system so the proposed multimodal interface integrate four control methods: a magnetic control system using a magnet, a voice control interface using a microphone, a control pad and a joystick that it can perform a simple command control with the basic moves to drive the wheelchair.

2.1 Tongue Movement Interface

Quadriplegics people have very limited options to care for themselves, so it is necessary to use specific skills of patients through signal pattern recognition as the movement of the tongue [8]. It is known that the tongue training with a simple protrusion task induces neural plasticity [9]. Among the different proposals to use the tongue as a method of controlling devices for assistance, most systems have direct contact with the patient [10,11]. For proposed assistance system to the wheelchair, the main goal was to use simple commands to move seamlessly through an obstacle course with precision using an interface reliable, inexpensive, discreet, minimally invasive, and easy to use. So the magnet is not placed on the tongue, but is in a dental retainer with a small rail behind the teeth, where the magnet can be moved in 5 positions by the tongue as seen in Figure 1.

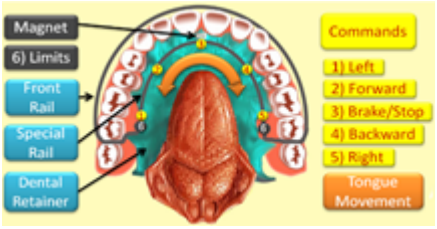


Fig. 1. Dental retainer with a magnet.

2.2 Voice Interface

In particular, the speech recognition system (SCS) consists of two parts: the first part is a section with a small vocabulary training that builds a model and the second is a speech recognition section which uses this model. The magneto-resistive sensor modules and microphone are placed in a headset, as seen in Figure 2, through a pair of flexible tubes which allow placing the sensors in different positions and after receiving the command signals, they are sent to the microcontroller system.

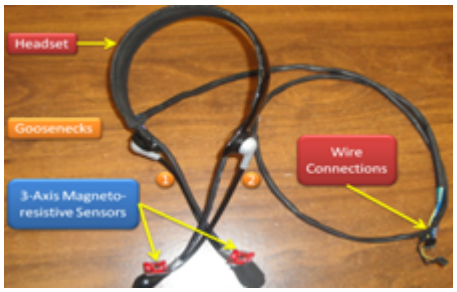


Fig. 2. Headset for Speech and Magnetic Control System

The voice commands that were used to control the system voice which can be seen in the following Table 1.

Table 1. Voice Commands for the Speech Control System

Commands	Action	Commands	Action
One	Power on the system	Right	Turn to the right
Forward	Move forward	Down	Brake motors
Backward	Move back	Up	Shut down (Stop also engines)
Left	Turn to the left		

2.3 Control Pad Interface and Joystick Interface

It was used a button pad with 9 buttons, these has been programmed with directions and speeds for the wheelchair. Of course it was included a joystick because of being the most typical interface for power wheelchairs. There were included typical inter-faces that some quadriplegics can use according to their injuries. On the other hand, pad button and joystick's circuits were added to have different ways for wheel-chair's driving.

3 Simulation Software

Before testing the multimodal interface in the wheelchair, a computer simulation software was developed using the same features as weight and speed of the actual chair to create a small game and teach users how to properly use it intuitively. There-fore a graphical user interface (software) was performed using the computer for multimodal interface (hardware) and thereby simulate the movement of the wheelchair. The programming environment that was used for the graphical interface was "Processing", which is an open programming language resource (open source) for those who want to create animations and interactions. Moreover multimodal interface has been developed in Arduino, which was created based on Processing, so the programming is similar and the interaction between both environments is very stable. In Figure 3 is seen the GUI looks like, while in Figure 4, there is a flow diagram of the internal workings.

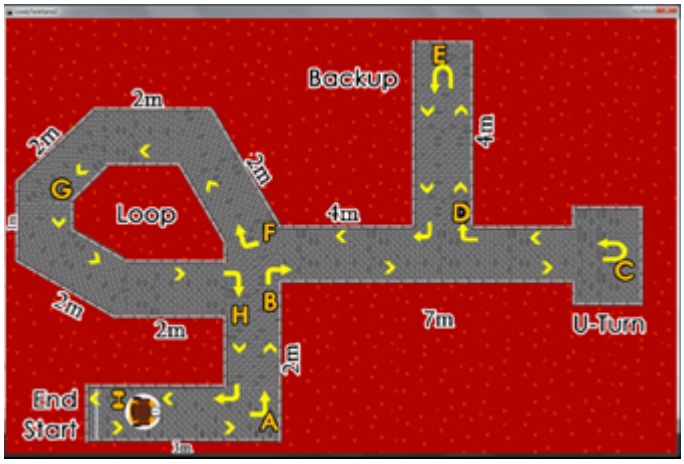


Fig. 3. GUI Software Simulation for Multimodal Interface. Track adapted for testing wheelchair performance [12,13]

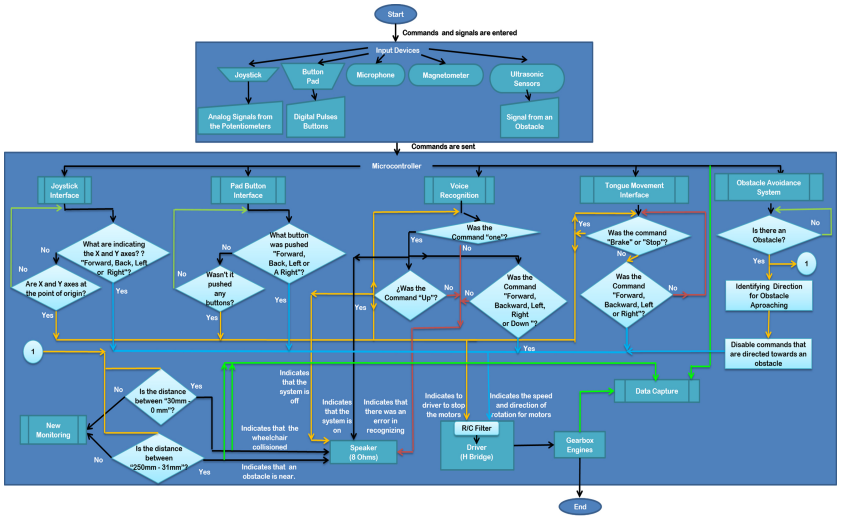


Fig. 4. Software Simulation Flow Diagram for Multimodal Interface

Flowchart shows how simulation software works. Among its main components, the input devices provide data for controlling the direction of the virtual small wheelchair. This was important, because this software is focused to achieve maximum accuracy possible to receive orders to move the wheelchair.

3.1 Performance Tests for Simulation Software

As can be seen, a small wheelchair and an obstacle course are shown with different signs to take appropriate route. The goal is to travel around the circuit and back to the starting point. Similarly through the usb port on the computer and serial port functions, control methods were adapted to move the wheelchair as shown in Figure 5.

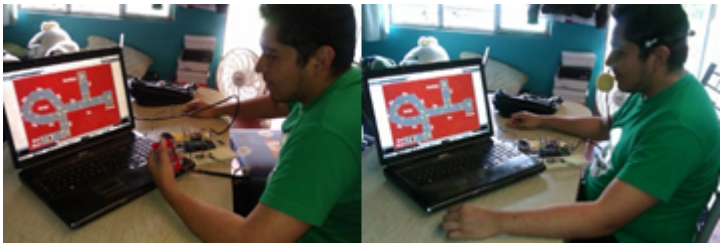


Fig. 5. GUI Software Simulation using Control Methods

Once it worked the GUI along with the multimodal interface, it proceeded to testing with people with a physical disability in the Integral Rehabilitation Center of Orizaba (CRIO). The tests consisted of 2 people assigned to each of the control methods (joystick, button pad, voice recognition and tongue movement). In total 8 different people used the multimodal interface through the GUI. Each user had to complete the obstacle course of the GUI according to the control method which they were assigned, in addition the data for each user was stored, taking particular importance in the type of injury or disability suffered. According to international standards established by the American Spinal Injury Association (ASIA), can be classified in Full Spinal Cord Injury (FSCI) and Incomplete Spinal Cord Injury (ISCI) and in 5 types from A to E. Were also obtained data from different tests that can be seen in Table 2 as the time in which they completed the obstacle course or the times they repeated the circuit. Voice and Tongue Interface just work with a slow speed to keep safer the user.

Table 2. Performance Tests using GUI software simulation

User	Injury Class	Control Method	Speed	Avg t/Lap	Avg Col/Lap
1	ISCI. Type E.	Pad button	Slow/Fast	69 s / 58 s	6.66 / 8.56
2	ISCI. Type E.	Pad button	Slow/Fast	74 s / 67 s	5.1 / 7.53
3	ISCI. Type C.	Joystick	Slow/Fast	80 s/ 65 s	3.38 / 6.73
4	ISCI. Type C.	Joystick	Slow/Fast	72 s / 63 s	4.08 / 7.15
5	ISCI. Type D.	Voice interface	Slow	83 s	32.21
6	ISCI. Type C.	Voice interface	Slow	90 s	28.37
7	ISCI. Type B.	Tongue interface	Slow	53 s	23.19
8	ISCI. Type B.	Tongue interface	Slow	58 s	24.53

3.2 Performance Tests in a Real Environment

To evaluate the different ways to handle the wheelchair, the same obstacle course was used in a real environment. This time when the users handled the wheelchair seen in Figure 6. They were more cautious, so they took longer to complete the laps. However, the number of collisions was significantly reduced among drivers as shown in Table 3. Other features of this performance test were the time response taken, the average time for a full stop after selecting slow / stop and the minimum step of the wheelchair in Table 4. It should be mentioned that were used ultrasonic sensors to prevent accidents and stop the wheelchair if this was directed against an obstacle by mistake. Also whenever the sensors are activated, these count as a collision, moreover were also used a Speaker which beeps with a certain proximity to avoid a collision. Despite this sound if the wheelchair is directed against an obstacle, a different sound is activated to indicate that there was a collision.

Fig. 6. Smart Wheelchair using the Multimodal Interface

Table 3. Performance Tests using Multimodal Interface apply to the Smart Wheelchair

User	Injury Class	Control Method	Speed	Avg t/lap	Avg Col/lap
1	ISCI. Type E.	Pad button	Slow/Fast	78 s / 67 s	0.97 / 1.53
2	ISCI. Type E.	Pad button	Slow/Fast	92 s / 84 s	0.86 / 1.29
3	ISCI. Type C.	Joystick	Slow/Fast	98 s / 76 s	0.42 / 0.88
4	ISCI. Type C.	Joystick	Slow/Fast	89 / 69 s	0.51 / 1.07
5	ISCI. Type D.	Voice interface	Slow	95 s	7.37
6	ISCI. Type C.	Voice interface	Slow	106 s	9.74
7	ISCI. Type B.	Tongue interface	Slow	72 s	5.22
8	ISCI. Type B.	Tongue interface	Slow	76 s	3.28

Table 4. General Characteristics of Speed and Distance of the Smart Wheelchair

Control Method	Speed	Resp Time	Avs t to slow	Avg d/s
Pad Button	Slow/fast	0.82s / 0.80s	0.75s / 0.93s	0.274m/s / 0.456m/s
Joystick	Slow/fast	0.83s / 0.82s	0.73 / 0.95s	0.305m/s / 0.540m/s
Voice Interface	Slow	0.81s	0.74s	0.203 m/s
Tongue Interface	Slow	0.82s	0.76s	0.236 m/s

4 Results

The main contribution for this paper were the results of making a wheelchair with a multimodal interface. Specifically magnetic control and voice recognition was develop at low cost. These results will be discussed below.

4.1 Control using Tongue Movements

The control of this interface proved to be able to react to commands programmed according to the position of the magnet moving it with the tongue, also managed

to perform the scheduled tasks for each command. During the tests, one experiment was driving the wheelchair through an obstacle course proposed by [12] this experiment was repeated 100 times with 8 different people to find significant information to compare their work with both the software simulation, as the multimodal interface. (Table 5). These results show a significant advance in tongue interface without include others.

Table 5. Comparison of Results with other Research

	Speeds	Sim Software	Multimodal I.	Ghovanloo T. I.
Avg Col	Low	28.36 s	4.25 s	1.77 s
Avg t/lap	Low	55.5 s	74 s	65.5 s
Commands	Low	5 s	5 s	5 s
Time resp	Low	0.81 s	0.81s	1 s

4.2 Speech Recognition Control

The use of this control showed that it is able to recognize the commands programmed and then the system performs scheduled tasks for each command, also activate the sounds that were saved in the memory of speech recognition module to verify that the command was correct. Table 6 shows the results by repeating 100 times each command, on the other hand the runtime was taken using an oscilloscope to know how long it takes the system to recognize it, and finally the percentages of assertiveness are also shown with respect the number of times the system correctly recognized each command.

5 Discussion

Implementing this system, was a success because circuits are small. It was possible to develop a speech control that had a successful recognition rate of 95.71%, which works with any type of voice; this percentage is pretty good because the speech recognition systems usually fail to be too assertive comparing with others as Coy results [14]. However, if it is required to customize a single voice is also possible to train the module to recognize only certain tones of voice. The memory limit that owns the microcontroller and the speech recognition module, restrict and limit training commands and for this reason are planned for future works, use an external memory or a computer with enough capacity to reduce the time response between the recognition of the commands and execution of assigned tasks. Finally, the magnetic control system, proposes using the magnet in the dental retainer because other systems place the magnet on the tongue with tisular adhesive which only lasts a few hours [15], or even newer systems use magnetic piercings in patients with severe disabilities [16] making

them suffer unnecessary pain, besides the use of piercings gradually generates chipping of the dental enamel, periodontal lesions and infection tongue numbness [17]. In the other hand with the current wheelchair, the users who used it in the experiment with proper training anyone who can use the tongue will be able to drive it, although some took longer to complete the course than others.

Table 6. Percentage of assertiveness for each command of speech recognition

Commands	Runtime	Assertiveness in speech recognition
One	10 ms	100%
Forward	21 ms	95 %
Backward	20 ms	100 %
Left	12 ms	90 %
Right	16 ms	92 %
Down	13 ms	97 %
Up	11 ms	96 %

Tests conducted with the control method of the tongue, resulted in an average of 4.25 collisions, each time the track was completed, resulting satisfactory comparing it with the inductive control system proposed in [13]. Future work will include a wireless system to communicate with smart phones and android operating system. Also it will be add other interfaces as used techniques of electrooculography, electromyography, or use the computer with the help of these interfaces and those already created.

References

1. Öztürk A., Dokuztug, U.F.: Effectiveness of a Wheelchair Skills Training Programme for Community-Living Users of Manual Wheelchairs in Turkey: a Randomized Controlled Trial. Clinical Rehabilitation 25, 416–424 (2011)
2. Fliess D.O., Vanlandewijck Y.C., Manor G.L., Van Der Woude U.H.V.: A systematic review of wheelchair skills tests for manual wheelchair users with a spinal cord injury: towards a standardized outcome measure. Clinical Rehabilitation 24, 867–886 (2010)
3. Barea R., Boquete L., Rodríguez-Ascariz J.M., Ortega S., López E.: Sensory system for implementing a human-computer interface based on Electrooculography. Sensors Magazine 11, 310–328 (2011)
4. Perez C.: Real-time template based face and iris detection on rotated faces. International Journal of Opto-mechatronics 3, 54–67 (2009)
5. Mean-Foong O.: Hand gesture recognition: Sign to voice system (S2V). International Journal of Electrical and Electronics Engineering 3, 4–10 (2009)
6. Ville R., Pekka-Henrik N., Jarmo V., Jukka L.: Capacitive facial movement detection for human-computer interaction to click by frowning and lifting eyebrows. Med. Biol. Eng. Comput. 48, 39–47 (2010)

7. Po-Yi S., Po-Chuan L., Jhing-Fa W., Yuan-Ning L.: Robust several-speaker speech recognition with highly dependable online speaker adaptation and identification. Elsevier Journal of Network and Computer Applications. 34, 1459–1467 (2011)
8. Salem C., Zhai S.: An isometric tongue pointing device. In: Proceedings of the ACM SIGCHI Conference on Human factors in computing systems, pp. 22–27 (1997)
9. Lau C., O'Leary S.: Comparison of computer interface devices for persons with severe physical disabilities. American Journal of Occupational Therapy 47, 102–103 (1993)
10. Huo X., Wang J., and Ghovanloo M.: A magneto-inductive sensor based wireless tongue-computer interface. IEEE Trans Neural Syst. Rehabil. Eng. 6(5), 497–504 (2008)
11. Johnson A.N., Huo X., Ghovanloo M. and Shinohara M.: Dual-task motor performance with a tongue-operated assistive technology compared with hand operations. Journal of NeuroEngineering and Rehabilitation 9(1), 1–16 (2012)
12. Huo X., Ghovanloo M.: Evaluation of a wireless wearable tongue computer interface by individuals with high-level spinal cord injuries. Journal of Neural Engineering 7(2), 1–12 (2010)
13. Lund M. E., Christiensen H.V., Caltenco H.A., Lontis E.R., Bentsen B., Andreasen Struijk L.N.: Inductive Tongue Control of Powered Wheelchairs. In: 32nd Annual International Conference of the IEEE EMBS, pp. 3361–3364 (2010)
14. Coy A., Barker J.: An automatic speech recognition system based on the scene analysis account of auditory perception. Elsevier Speech Communication 49, 384–401 (2007)
15. Kim J., Huo X., Ghovanloo M.: Wireless Control of Smartphones with Tongue Motion Using Tongue Drive Assistive Technology. In: Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 5250–5253 (2010)
16. Huo X, Ghovanloo M.: Using unconstrained tongue motion as an alternative control surface for wheeled mobility. IEEE Trans on Biomed Eng 56(6), 1719–1726 (2009)
17. Giuca MR, Pasini M, Nastasio S, D' Ercole S., Tripodi D.: Dental and periodontal complications of labial and tongue piercing. Journal of biological regulators and homeostatic agents 26(3), 553–600 (2012)

Using Gestures to Interact with a Service Robot using Kinect 2

Harold Andres Vasquez[1], Hector Simon Vargas[1], and L. Enrique Sucar[2]

[1] Popular Autonomous University of Puebla, Puebla, Pue., Mexico
{haroldandres.vasquez,hectorsimon.vargas}@upaep.edu.mx
[2] National Institute of Astrophysics, Optics and Electronics, Puebla, Pue., Mexico
esucar@inaoep.mx

Abstract. In this paper is presented a proposal for multimodal interaction between the robot and the person, using voice and gestures, in order to make friendlier human-robot relationship. This functionality will be integrated into a service robot called Donaxi, from Robotics Lab of UPAEP. Due to this research is recent, only is showed some preliminary results what has been achieved so far. Some data from a dataset of gestures being built for service robots is shown.

Keywords: HRI, gestures,service robot, Kinect 2.

1 Introduction

Many countries, in special the European countries, are concerned about the care of the old people. This affect the social and economic aspects of any country. Given the aging of the population, in the future there will be a lack of caregivers for old people; service robots provide an alternative to assist senior persons. Due this, there are many support programs to help to develop this kind of robots, called services robots [1].

Because these robots are going to interact with a person, it is necessary a comfortable communication way so that person feels that the robot is able to understand very well the given instructions. The most natural way to achieve this is with voice, due it is the most common communication way between us. However, use only the voice to communicate is not enough, due some problems and also because not all people can use the voice.

Therefore, is necessary use other communication way. It is clear that gestures is one very useful alternative, still is not very common between persons. Gestures have been used in many projects with service robots and they have proven to be very efficient in noisy ambient, where the voice is not enough.

According to the above, we can exploit the benefits given by both of them: voice and gestures. Each one can solve the problem of another. When it's used two or more communication way with a machine, this is called multimodal interaction. Therefore, we propose a multimodal interaction with a service robot using voice and gestures.

On the first part of this document is presented the problem and the methodology proposed to solved it. Then the main expected contributions are described, followed by some results obtained at this moment and closed with the conclusions.

2 Problem

Gestures have proved be very useful to interact with a robot [2],[3],[4]. Many kind of devices have been used to achieved this, but the most commons are the video cameras and the Kinect. Among these, the Kinect is the most used in services robots, due the programmers don't have to deal with both segmentation and following persons. These functionalities are provided by the Kinect SDK, using a combination of hardware and software. With this, the problem reduced to identify when the gesture is being performing and what gesture performed the user. The "when" problem is called segmentation and the "what" problem is called recognition. Each problem can be solved by separated, but the aim is that both operate simultaneously, thus a more natural interaction is achieved [5].

Actually, this problem seems to have been solved with the Kinect version 2 (Figure 1), a device created by Microsoft Company, initially to be used with the game machine called XBox. This because the Microsoft team (the owners of Kinect) developed a IDE where is possible makes new systems with own gestures. This IDE is called Visual Gesture Builder (VGB). On this, after dispose of a good number of videos examples, we can obtain a database with the new gestures trained. With this Database of Gestures (DBG), it's possible develop a system where it's used these gestures.

Fig. 1. Kinect version 2 used for multimodal interaction with the robot using voice and gestures

In some tests conducted (showed in Section 5), this technique proved be very good solving the two problems described before. Nevertheless, all the previous work was very tedious and long. We have to capture many videos from different persons to achieve a good training of the DBG. After, we have to tag each video with the gestures where are performed.

The other hand, voice is the most natural way to interact with machines. The most common problem with this is the machine can't understand to the person. This because each person have different voice and pronunciation. Another issue that affect this is noisy ambients. This can be solved with: software only, hardware only or both. In those cases, is necessary complement the voice with another communication ways. Obviously, gestures is the best option for doing

this. Therefore, we can obtain a multimodal interaction with the service robot using voice and gestures.

The different situations in which the voice and gestures together can be used are:

1. Gesture is voice reinforcement, i.e. when the robot is not capable to understand the voice command, the person performs a gesture with the same significance of the voice command. For example, if the user commands with voice to robot to pay attention to him, and because they are in a Shopping Center, the robot can't hear its owner; then the user can wave his hand to call the attention of the robot.
2. Gesture is voice complement, i.e. at home environment, the user wants the robot gives him a new medicine unknowing by it, and therefore, the robot don't know where is. Then, with voice, the user asks for the medicine to robot and simultaneous, with hand, he shows the place where is the medicine.

In summary, this work objective is to contribute to solve these challenges of the multimodal interaction with the service robot.

3 Main contribution

There are many works about Multimodal Human Robot Interaction (MHRI) [2],[6]. Most of them use different devices by each input: voice and gesture, or only use the first combination of them described in Section 2.

We propose use only the Kinect 2 to treat both signals. This because, while fewer devices have connected the robot, it consumes less power and less weight will be loaded. Also, on software side, less communication with different devices is required and this reduces processing costs. This last is very difficult to achieve, due the problems with voice described before, so only can be solved on software side.

Another relevant aspect of our proposal is use the multimodal interaction in different ways, like was described in Section 2.

As result of this work, be going to create data sets for both signals, so these can be used by any research in robotics or associated fields.

4 Methodology

It is important mention that this work is a continuation of a previous work regarding simultaneous segmentation and recognition of gestures, which apparently is already solved with the VGB. Nonetheless, the experience obtained with that research has allowed a more rapid development of this proposal.

At this moment, this research is in the beginning phase. A general idea of a work plan to be followed, based on the model shown in Figure 2, is describe below.

As result from the model showed, the work plan to follow is:

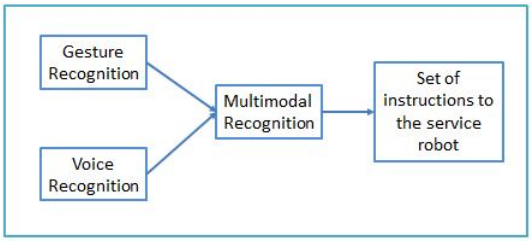


Fig. 2. General model to work

1. Know how the Kinect version 2 works with gestures and voice. Although we worked with the Kinect version 1 before, this new version has many different technologies and software that make it more capable to recognize gestures and voice separately. Therefore, it's necessary understand in depth how this works to exploit its benefits and uses them to achieve this research. One idea for this, is using the examples coming with the Microsoft Kinect SDK, designed for gestures and voice recognition.
2. Test the recognition quality from the Kinect 2 for both gestures and voice. It has to design formal experiments to measure the quality of both recognitions way and probe if these are enough in order to the service robot can interact with any person in not controlled ambients.
3. Identify the causes of any shortcomings found in existing recognition techniques from kinect 2. This will allow us determine our research hypothesis and then build proposal to solve them.
4. Once each recognition works well separately, be must to design a model to combine both results, as it showed in Figure 2. This is the multimodal recognition model, that is expected to be able to perform the service robot.
5. Finally, we must apply the experiments for evaluations, designed to prove that our robot can perform a natural interaction with any person in any ambient.

As will be shown in the next section, there are already some progress in the plan designed, allowing to demonstrate the feasibility of this work.

5 Results

This research is conducted in the UPAEP robotics laboratory , where there is a service robot called Donaxi. This robot is completely built from scratch by students from different college careers like electronics, mechatronics, bionics and computing. Donaxi will have:

1. Omnidirectional navigation system, with four wheels, each one with a DC motor with encoder. This allow to Donaxi moves in almost any direction.
2. Laser system to build navigation map. It consists of two laser, front and rear.

3. Vertical movement System, based on a rail and a motor, which moves a platform up or down.
4. One arm with five freedom degrees and a parallel gripper on the end.
5. One Kinect version 1 that moves with the vertical movement system, and it used for object recognition.
6. One Kinect version 2 for people recognition, whether the whole body, face or voice. This is fixed in the top of the robot. This it will use to multimodal recognition with gestures and voice.
7. Two laptops. One with Ubuntu and ROS, used for some of the functionality of the robot, and the second with Windows, used to recognize people, faces, gestures and voice. The laptops communicating with each other through TCP/IP messages.

Some of these devices are already available on the robot, but others are in the adaptation process. In Figure 3 is showed the preliminary version of Donaxi.

Fig. 3. The Service Robot Donaxi from the UPAEP Robotics Laboratory

Because in april this year was the Mexican Robotics Tournament (TMR2015)[3] and Donaxi team participated on it, it became necessary to have some work of both gestures and voice recognition separately. For this reason we have some progress on these two features, that was proven in a almost realistic house ambient in this tournament. In Figure 4 are showed the services robots competed at TMR2015.

For voice recognition, a Creative 3D Sense camera was used, which also was used to face recognition. The software used for this was the Intel RealSense SDK, created for these devices. Although not conducted rigorous testing of this feature, a very good preliminary results was obtained, even allowed Donaxi to win the competition in this category. In Figure 5 it showed this device.

[3] http://www.tmr2015.mx/

Fig. 4. The three services robots present in the Mexican Robotics Tournament 2015

Fig. 5. Creative Sense 3D used for voice and face recognition

For gesture recognition, the Microsoft Kinect version 2 was used. To achieve this, it was necessary complete the steps showed in Figure 6.

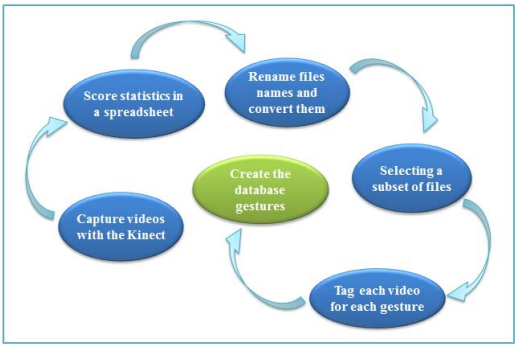


Fig. 6. Diagram where show the steps followed to obtain the DataBase of Gestures

First, we have to capture many videos from many different people in different places. This because for more different examples provided to the training algorithm, more overfitting is avoided and will provide the capacity to detect any person in any environment. Second, to keep track of the statistics of the video taken by the Kinect, it has been written relevant data into a spreadsheet. Third, to know the content of each video file without opening it, it should be rename them using a specific nomenclature, to then make a conversion using the tool KSConvert from the Kinect SDK. Fourth, because several kinds of catches with various kind of people was used, it was decided to use only part of the full set, to verify if this subset was enough for both training and testing of the recognizer. Fifth, to train each gesture, it should shown in each video where are the positive

examples and negative examples, which is called tagging and it was used the Microsoft Visual Gesture Builder. Finally, using the VGB also, can be created the DataBase of gestures, for build the application for the service robot.

In table 1 it shows the total numbers of gestures captured until now for this research.

Table 1. Total number of gestures in dataset

Gesture	Number
Stop	194
Come	177
Left	395
Right	407
Attention	417
Indication	363
Turn	207

Further details on this work will be shown in a publication about this dataset and thus put this material available to the scientific community.

Once this database of gestures is available, it can build an application for Donaxi, that can detect gestures being made by the user, though the Kinect 2. For this purpose, a sample in C# available in the Microsoft Kinect SDK was used, called "DiscreteGestureBasics". For the TMR competition, only were used three gestures from the seven contemplated. This because has not been tested the accuracy of the recognizer with the seven gestures together in the database. In Figure 7 is showed the three gestures used: attention, to make Donaxi know where is your master; right, to make Donaxi revolves on his right and stop, so Donaxi stop when approaching. These three gestures were used in TMR2015.

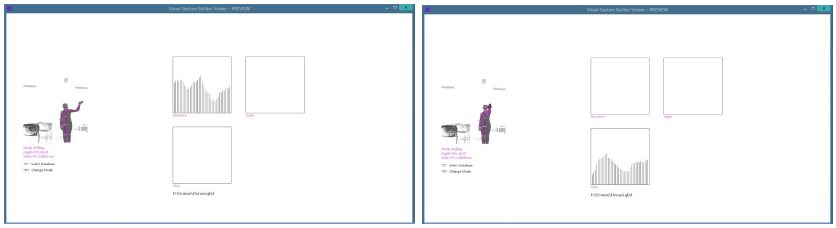
(a) Attention (b) Stop (c) Right

Fig. 7. The three gestures used at TMR2015 for Donaxi

In attention gesture, the user moves either hand from side to side, near to the head. In stop gesture, the user put the open hand, with arm extended, in front to him. In right gesture, the user extend his arm to right side. It can be appreciated in Figure 7 the arrow showing the movement at each case.

Nevertheless, the application developed with three gestures showed good results, making Donaxi speak only when the user completed any of the gestures. Only the gesture "stop" had some not detection problems (false negatives). For this reason, it is necessary to prepare a more rigorous test plan while is being adding more gestures and to evaluate the results.

In figure 8 can be appreciated the gesture recognition running in Microsoft VGB LivePreview tool, where the user perform each gesture at once.



(a) Attention (b) Stop

(c) Right

Fig. 8. ScreenShots from the Microsoft VGB LivePreview where the user perform each gesture at once

At this moment, is preparing Donaxi for her next challenge, the Rockin 2015 (http://rockinrobotchallenge.eu). This is an international competition of services robots, which this year will be held in Lisbon, Portugal in November. For this competition it will expected to have ready the speech and gesture recognizers fully functioning separately.

6 Conclusions

It is clear the need for service robots to interact with people in the most natural way possible. However, achieving this type of interaction is not so simple. Many challenges are looming to achieve this kind of robot-human relationship.

For now, most research on multimodal human robot interaction are using only voice and gestures, but obviously this set can grow to reach the expected goal.

This research aims to achieve this multimodal HRI using only the Kinect version 2 for detecting gestures and voice and allow to Donaxi can understand better its owner.

References

1. Aracil, R., Balaguer, C., and Armada, M.: Robots de servicio. Revista Iberoamericana de Automatica e Informatica Industrial 5, 6–13 (2008)
2. Goodrich, M. A. and Schultz, A. C.: Human-Robot Interaction: A Survey. Foundations and trends in human-computer interaction 1(3), 203–275 (2007)
3. Droeschel, D., Stckler, J., Holz, D and Behnke, S.: Towards joint attention for a domestic service robot person awareness and gesture recognition using time-of-flight cameras, In: Proc. of the IEEE Int. Conf. on Robotics and Automation (ICRA), pp. 1205–1210 (2011)
4. Yan, R., Tee, K. P., Chua, Y., Li, H., Tang, H.: Gesture Recognition Based on Localist Attractor Networks with Application to Robot Control. Computational Intelligence Magazine 7(1), 64–74 (2012)
5. Vasquez Chavarria, H., Escalante, H. J., and Sucar Succar, L. E.: Simultaneous segmentation and recognition of hand gestures for human-robot interaction. In: 16th International Conference on Advanced Robotics, pp. 1–6 (2013)
6. Pavlakos, G., Theodorakis, S., Pitsikalis, V., Katsamanis, A., Maragos, P.: Kinect-based Multimodal Gesture Recognition using a two-pass fusion scheme. In: Proc. International Conference on Image Processing, pp. 1495–1499 (2014)

Reviewing Committee

Vicente Alarcón
Maya Carrillo
Sergio Coria
Anilú Franco
René García
Félix F. González
J. Ángel González
Juan M. González
Josefina Guerrero
Raudel Hernández
Manuel Lazo
Agustín León
Manuel Martín
Ivan Olmos
Airel Pérez
David Pinto
Ansel Rodríguez
Guillermo Sánchez
Roberto Rosas
Daniel Valdés

Impreso en los Talleres Gráficos
de la Dirección de Publicaciones
del Instituto Politécnico Nacional
Tresguerras 27, Centro Histórico, México, D.F.
junio de 2015
Printing 500 / Edición 500 ejemplares