Editorial

IT is my great pleasure to present to the readers the jubilee, 50th issue of the journal Polibits. It was a long journey to go, from a local, institutional journal founded 25 years ago to an internationally recognized journal with both readers and authors from all over the world. This is a result of hard work of many people. Most importantly, this is thanks to our authors, who contributed excellent papers during the 25-year history of the journal.

My congratulations and warmest gratitude go to the editorial team, the Editorial Board, and all reviewers who contributed their knowledge, talent, and time to the hard work of reviewing, with the only reward for their unselfish labor of love being the honor to contribute to the advance of human knowledge. I want to congratulate very gratefully past Editors in Chief of the journal who laid the foundation to its current success, and most of all Prof. Grigori Sidorov, the previous Editor-in-Chief, under whose guidance and thanks to whose constant work the journal flourished as never before.

I also want to thank and to congratulate the authorities of the Centro de Innovación y Desarrollo Tecnológico en Cómputo of the Instituto Politécnico Nacional, our hosting organization that provides all the support and help to the journal. Finally yet importantly, the success of the journal is thanks to the constant and significant support from CONACYT, Mexican Ministry of Science.

This issue of Polibits includes ten papers by authors from thirteen different countries: Belgium, Brazil, Chile, Colombia, Ecuador, France, Italy, Mexico, the Netherlands, Peru, Russia, Sweden, and USA. The papers included in this issue are devoted to such topics as computer vision and video coding, scheduling, robotics, and control, neural networks, databases, and analysis of the web.

H. Brahmasury Jain and K.R. Rao from the **USA** in their paper *Fast Intra Mode Decision in High Efficiency Video Coding* propose a fast predicting algorithm for video encoding, which allows to improve the speed and resolution and to decrease complexity of encoding video for storage and transmission. They give a detailed comparison of several relevant algorithms. K. R. Rao, IEEE fellow, is a co-author of the discrete cosine transform, which has revolutionized image processing and, in particular, is the basis for the JPEG image encoding.

D. G. S. Santos et al. from **Brazil** in their paper *A Dynamic Gesture Recognition System based on CIPBR Algorithm* address a problem at the crossroad of computer vision and human-computer interaction. The authors classify video clips that represent images of human hand movements (palm and fingers) and recognize the type of hand moving gesture basing on adequate feature extraction and supervised machine learning.

A. Moran Cardenas et al. from **Peru**, **Colombia**, and **Sweden** in their paper *Design of High Accuracy Tracking Systems with H_∞ Preview Control* consider a specific type of automatic positioning and tracking control system, which is an important component of autonomous robots. In the paper the authors present a novel method for designing positioning and tracking systems based on so-called H_∞ preview control mechanism.

N. Eloe et al. from **USA** in their paper *A More Efficient Representation of Obscuration for VRCC-3D+ Relations* address yet another problem of high importance for computer vision and robotics: representation and reconstruction of three-dimensional scene basing on the observable two-dimensional image. With this paper, the authors continue their research on spatial knowledge representation, part of which has been published in a past issue of Polibits; the next part will also be published in a forthcoming issue of Polibits.

G. Sidorov et al. from **Mexico** and **Russia** in their paper *Modelo computacional del diálogo basado en reglas aplicado a un robot guía móvil* (*Computational Model of Dialog Based on Rules Applied to a Robotic Mobile Guide*) continue the topic of robotic applications and human-computer interaction. They present a formal model for a dialog between a visitor to a museum, exhibition, university, of another similar space and a mobile robot that serves as a guide for this space. At a verbal request of the visitor, the robot can show him or her places of his or her interest and can also give information about objects or places and answer relevant questions, in a way similar to how a human museum guide would do it.

A. Maccioni et al. from **Italy**, **Belgium**, the **Netherlands**, and **France** in their paper *NoXperanto: Crowdsourced Polyglot Persistence* present an approach to solve very complex database queries in an unusual way: by mining a huge amount of queries produced by human expert users and database administrators. In this way the computer not only processes the formal data to which it has access but instead orchestrates a complex interplay and collaboration between distributed human knowledge and formal computation and involves humans into a global infrastructure of data processing.

N. Rodriguez et al. from **Chile** and **Ecuador** in their paper *Haar Wavelet Neural Network for Multi-step-ahead Anchovy Catches Forecasting* develop a novel multi-step-ahead model for forecasting fish-catch. This is in important problem for the economy of countries that heavily depend on fishing industry either for internal consumption or for export. The model that the authors present is based on a specific type of neural

network, and thus is not specific for fishery; I think this model can be successfully applied to other tasks of forecasting, possibly in financial or in seismic settings.

O. Durán A. et al. from **Chile** in their paper *A Comparison between Two Metaheuristics Applied to the Cell Formation Problem with Alternative Routings* consider a manufacturing scheduling task: there are a number of "machines" that can "process" a flow of products in different order and in different configuration. The task is to optimize the workload on those "machines" to achieve the best utilization of the available processing power for fastest processing of the products. The authors provide a design of a genetic algorithm for such an optimization.

J. A. Castán et al. from **Mexico** in their paper *Control de tráfico basado en agentes inteligentes* (*Traffic control based on intelligent agents*) continue the topic of optimization of processing flow. In this case, the authors address the task of optimal regulation of traffic lights in such a way to avoid congestion and large waiting time for cars. The authors use a cognitive approach based on collaboration of artificial intelligent agents. The authors present a tool for simulating traffic situation and studying the impact of different policies of traffic light control. This work has a potential to improve the quality of life of millions of people who currently spend hours every day in traffic jams.

E. Zurek et al. from **Colombia** in their paper *Acoustic Fingerprint Recognition Using Artificial Neural Networks* apply neural network technology to recognizing marine vessels by audible signal. This is a task important both for civil navigation and, more importantly, for military operations and defense. While most of the authors of this paper are electrical engineers, one of the authors is an NCO Chief of Colombian navy. The practical importance of the task is especially high for Colombia, a country whose military and navy is at the front of heroic combat against organized crime, protecting both its own citizen and all of us from illegal drug trade—much of which passes by sea. Currently, the task of identification of marine vessels is performed by human experts. The use of artificial neural networks allows for a significant increase in its precision.

This issue of the journal will be useful to researchers, students, and practitioners working in the corresponding areas, as well as to general public interested in advances in computer science and engineering.

Alexander Gelbukh
Editor in Chief

Fast Intra Mode Decision in High Efficiency Video Coding

H. Brahmasury Jain and K.R. Rao

***Abstract*—In this paper a coding unit early termination algorithm resulting in a fast intra prediction is proposed that terminates complete full search prediction for the coding unit. This is followed by a prediction unit mode decision to find the optimal modes HEVC encoder 35 prediction modes. This includes a two-step process: firstly calculating the Sum of Absolute Differences (SAD) of all the modes by down sampling method and secondly applying a three-step search algorithm to remove unnecessary modes. This is followed by early RDOQ (Rate Distortion Optimization Quantization) termination algorithm to further reduce the encoding time. Experimental results based on several video test sequences for 30 frames from each test sequence show for HEVC a decrease of about 35%–48% in encoding, with negligible degradation in peak signal to noise ratio (PSNR). Metrics such as BD-bitrate (Bjøntegaard Delta bitrate), BD-PSNR (Bjøntegaard Delta Peak Signal to Noise Ratio) and RD plots (Rate Distortion) are also used.**

***Index Terms*—HEVC, Fast intra coding, Early CU Termination, Early RDOQ Termination, PU Splitting.**

I. Introduction

HEVC is the latest video standard introduced by the Joint Collaborative Team on Video Coding (JCT-VC) in January, 2013 which contains three profiles namely; main (8-bit), main10 (10-bit) and still frame [1]. Here only the main (8-bit) profile is considered since it is most widely used profile. The HEVC standard is designed to achieve multiple goals, including coding efficiency, ease of transport system integration, data loss resilience, and implementation using parallel processing architectures. The HEVC standard has been designed to address essentially all the existing applications of the H.264/MPEG-4 AVC standard [1] and to focus particularly on two key issues: increased video resolution and increased use of parallel processing architectures [1]. The major achievements of the HEVC standard in comparison with the H.264 [1] standard are flexible prediction modes, larger transform block sizes, better partitioning options, improved interpolation and deblocking filters, prediction, signaling of modes and motion vectors and support efficient parallel processing [1]. The HEVC syntax should be generally suited for other applications and not specifically to two applications mentioned above [1]. This is not the result of optimizing a single step in the encoding process, but a combined result of optimization of many processes together.

HEVC supports 2k and 4k video coding and hence with increase in video resolution encoder complexity of HEVC has increased. In order to reduce the encoder complexity there are many fast intra prediction algorithms [20-28] proposed for HEVC. In this paper, a fast intra coding algorithm is proposed to reduce the encoder complexity. The block diagram of HEVC encoder is shown in Figure 1.

Manuscript received on September 15, 2014, accepted for publication on October 1, 2014, published on November 15, 2014.

H. Brahmasury Jain is with the University of Texas at Arlington, Electrical Engineering Department, USA (e-mail: harshdeep.brahmasuryjain@mavs.uta.edu).

K. R. Rao (corresponding author) is with the University of Texas at Arlington, Electrical Engineering Department, USA (rao@uta.edu).

II. Proposed Algorithm

A three step method is proposed as a solution. In CU splitting, decision is made whether to split the current CU further by analyzing the CU texture characteristics. In PU partition, down sampling prediction followed by three—step search is exploited similar to one proposed in [24]. In the last step the early RDOQ termination is implemented [25, 28].

III. CU Early Termination

When the CU texture is complex, the CU is split into smaller sub units to find the best size and when the CU texture is flat, the CU is not divided further into subunits. This has already been proved [12].

In the first stage, to decrease the computational complexity, the down-sampling method is exploited by applying a 2:1 down sampling filter by a simple average operator to the current CU and other CU have the similar operation as shown in Figure 2.

After the downsampling, the complexity of the original LCU can be calculated by the following formula:

$$Ecom = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\left[p(i,j) - \frac{1}{N}\frac{1}{N}\left(\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} p(i,j)\right)\right]$$

where *Ecom* is the texture complexity, N is the size of the current CU, $p(i,j)$ is the pixel, and (i, j) is the coordinates in CU.

Depending on the texture calculation, two thresholds are set with a tradeoff on coding quality and complexity reduction as Thres1 and Thres2. The CU is split when the complexity is greater than Thres1 and when complexity is less than Thres2, the CU is not split further. If the complexity is between the Thres1 and Thres2, HEVC reference software is referred [4].

IV. PU Mode Decision

At the second stage, PU modes decision is obtained by calculating the Sum of Absolute Differences (SAD), which is

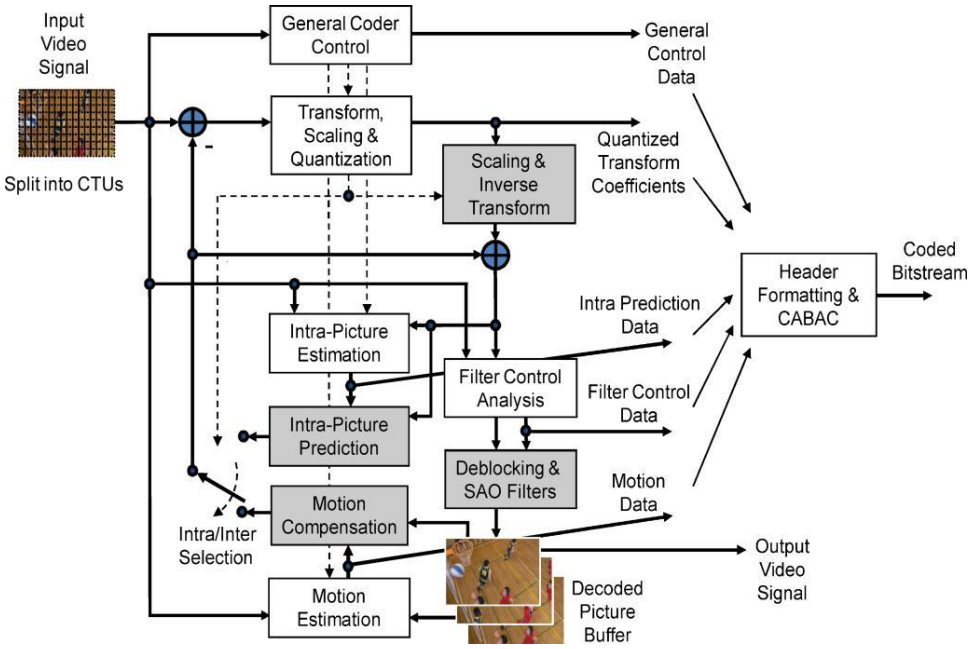


Fig. 1. HEVC encoder block diagram [1]

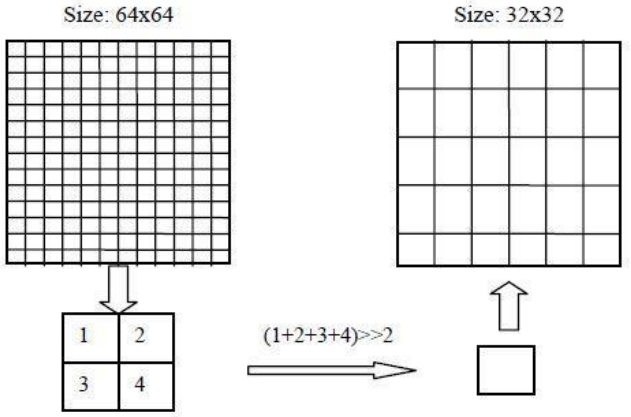


Fig. 2. Simple averaging based on down-sampling on 64 × 64 CU [27]

performed by downsampling and then by applying similar three step search algorithm. The detailed operation is as follows.

- List of candidates is created, S1 = {0, 1, 2, 6, 8, 12, 16, 20, 24, 28, 30, 32, 34} from the 35 prediction modes and then 5 optimal modes by SAD is check on S1, suppose 5 modes are S2 = {0, 3, 12, 16, 34}.
- From the three-step algorithm [27], list S2 is extended on to the 2-distance neighbors and S3 = {2, 10, 20, 32} for both the modes 0 and 2 and then S1, S2, S3 are checked for optimal modes S4 = {8, 14, 24}. Suppose modes of upper and left PUs are S5 = {1, 6}, then checking optimal modes and if the optimal two modes are S6 = {2, 6}.
- Then 1-distance neighbors of S6 are S7 = {3, 5, 9} then we choose the best M modes as the candidates for RDOQ. Tables 1 and 2 show the HEVC encoder complexity for CU and PB blocks. Figure 3 shows the luma intra prediction modes for HEVC.

TABLE 1.
LUMA INTRAPREDICTION MODES SUPPORTED BY DIFFERENT PU SIZES [14]

PU Size	Intraprediction Modes
4 × 4	0–16,34
8 × 8	0–34
16 × 16	0–34
32 × 32	0–34
64 × 64	0–2, 34

TABLE 2.
CURRENT PROBLEM-COMPLEXITY FOR HEVC [33]

Size of PB	Number of PBs in a 64x64 CU	Number of modes to be tested in each PB	Total number of modes to be tested at this level
32 × 32	4	35	140
16 × 16	16	35	560
8 × 8	64	35	2240
4 × 4	256	35	8960
		Total	11900

V. EARLY RDOQ TERMINATION

At the third stage, there are M modes selected from the result of the second step which are put into a group, Ψ, that go through the RDOQ process to get the best mode, mopt. An early RDOQ termination is proposed for further encoder time reduction. For each intra mode $m \in \Psi$, its overall cost $J(m)$ as the combination of SATD cost and associated mode index bits consumption is calculated. Within Ψ there is a mode with minimal J_{min} defined as rough best mode mopt_rough. If mopt_rough is Planar or DC

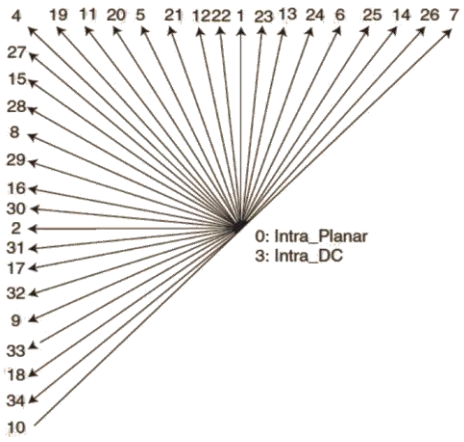


Fig. 3. Luma intra prediction modes of HEVC [14]

mode, all other modes in Ψ are skipped. If mopt_rough is not 0 or 1 and |m−mopt_rough| > 3, such mode *m* is skipped also; meanwhile, if $J(m) > \alpha J_{min}$, mode m will not be checked and α = 1.08 is considered. After such early termination procedure, all the remaining modes are checked by RDOQ. The next section outlines the experimental results of the implemented algorithm vs. the original HM 13.0 [4].

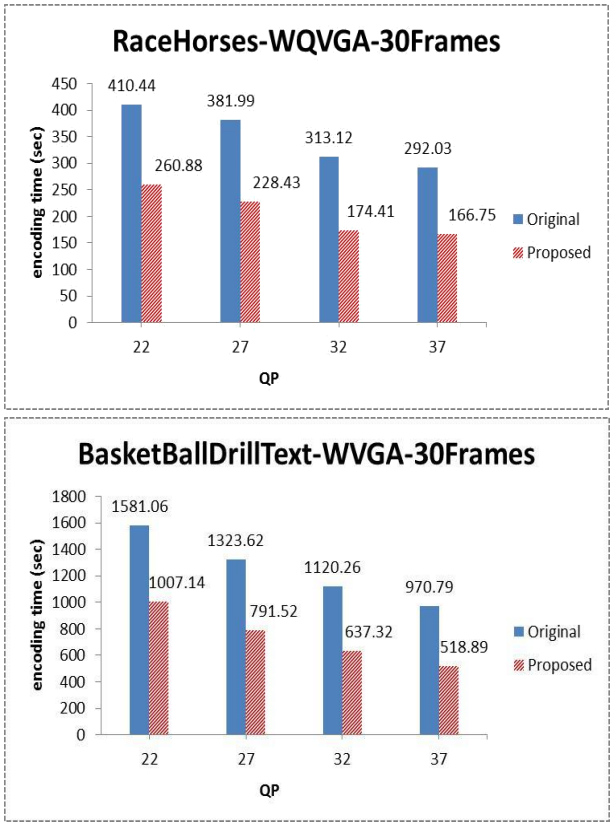


Fig. 4. Encoding time vs. quantization parameter for Racehorses and BasketBallDrillText

VI. Results

In order to evaluate the performance of the proposed intra prediction algorithm, the algorithm is implemented on the recent HEVC reference software (HM 13.0) [4]. The intra main profile is used for coding with the intra period set as 1 and frame rate set at 30 fps. The proposed algorithm is evaluated with 4 QPs of 22, 27, 32 and 37 using the test sequences recommended by JCT-VC [35].

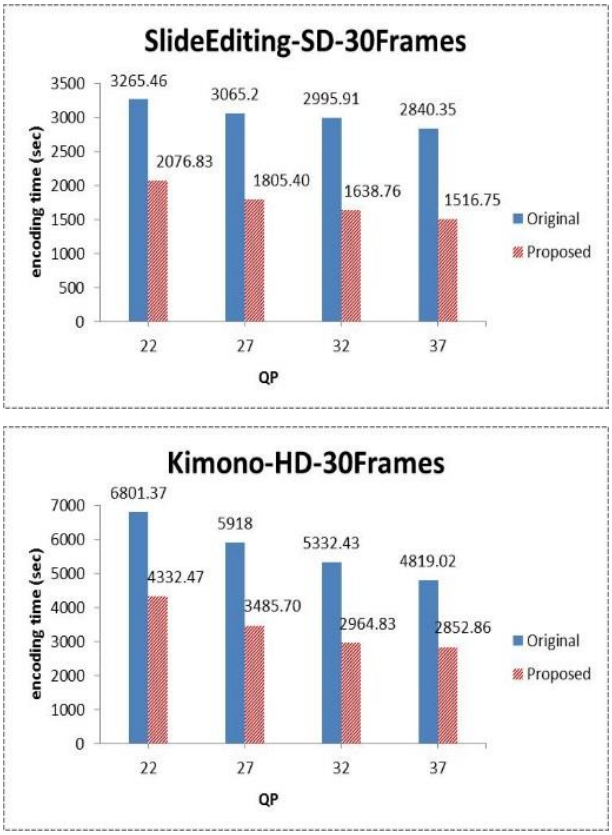


Fig. 5. Encoding time vs. quantization parameter for SlideEditing and Kimono

A. Encoder Complexity Reduction

With the proposed CU early termination algorithm, encoder complexity in terms of encoding time for the test sequences is reduced by 35-48% as compared to the unmodified encoding HM13.0 [4].

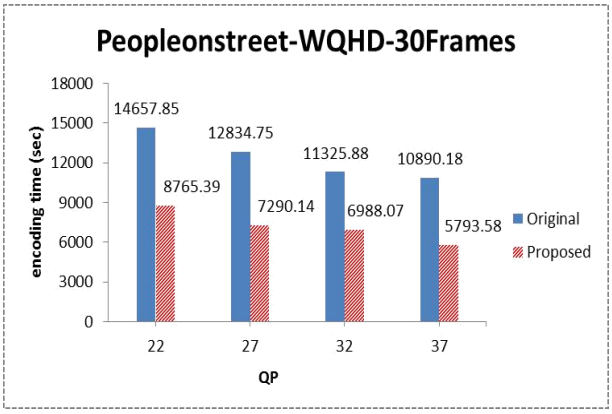
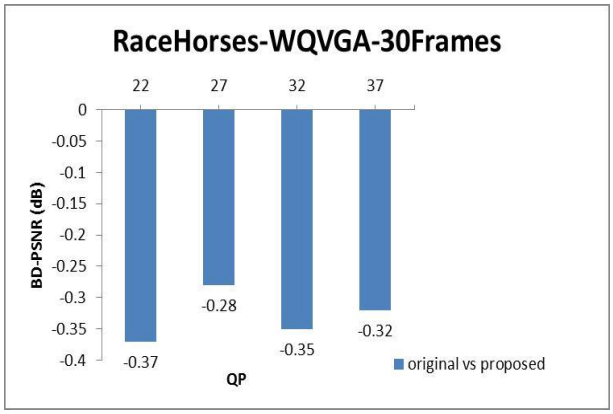


Fig. 6. Encoding time vs. quantization parameter for PeopleOnStreet



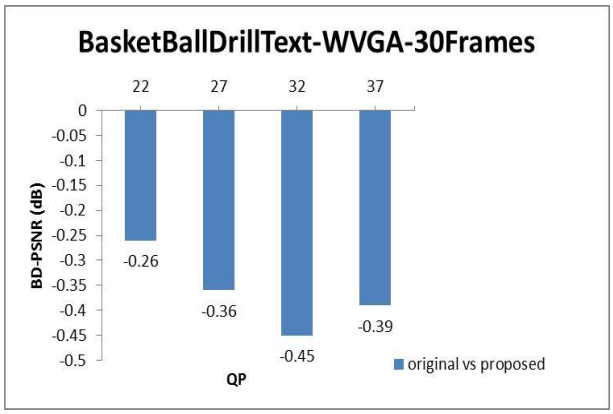
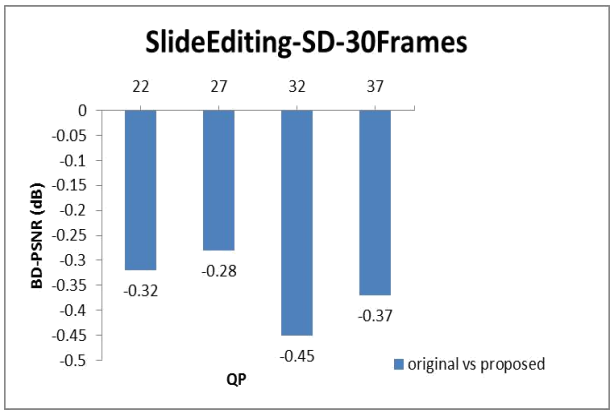


Fig. 7. BD-PSNR vs. quantization parameter for RaceHorses and BasketBallDrillText



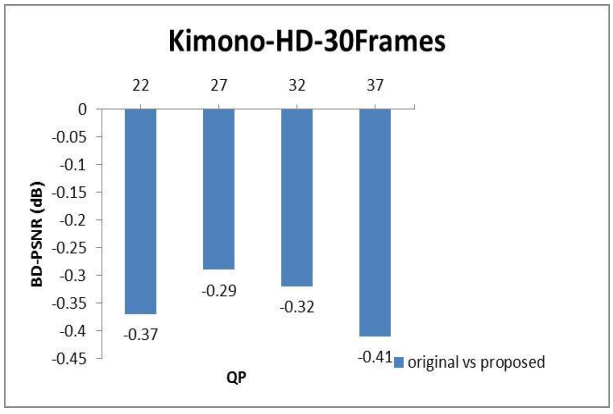


Fig. 8. BD-PSNR vs. quantization parameter for SlideEditing and Kimono

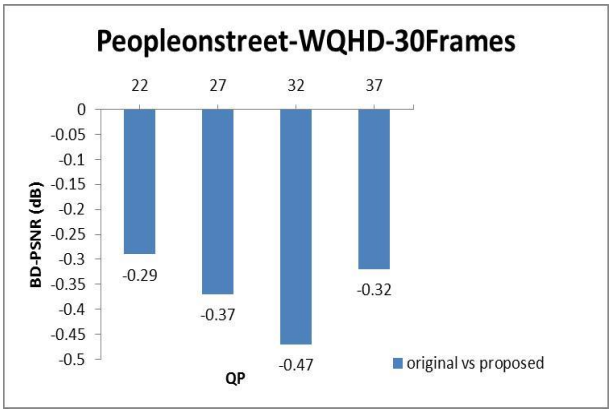
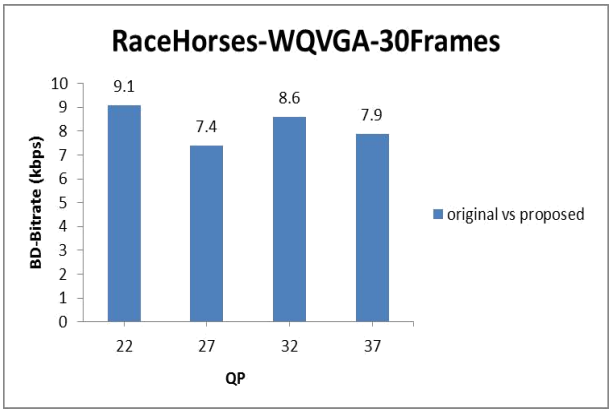


Fig. 9. BD-PSNR vs. quantization parameter for Peopleonstreet



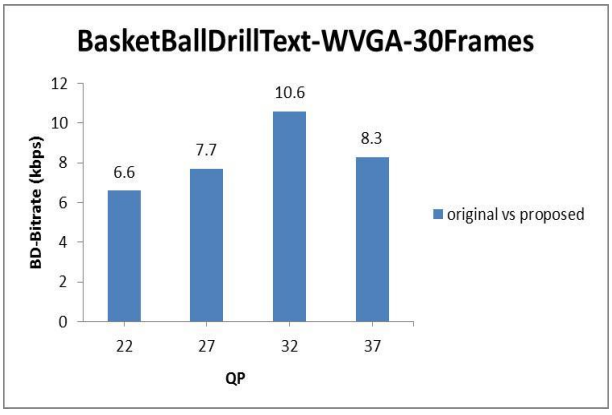
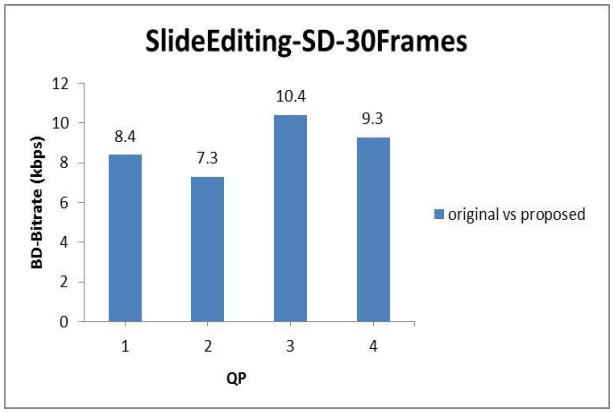


Fig. 10. BD-bitrate vs. quantization parameter RaceHorses and BasketBallDrillText

B. BD-PSNR

To evaluate objectively the coding efficiency of video codecs, Bjøntegaard Delta PSNR (BD-PSNR) was proposed [36]. Based on the rate-distortion (R-D) curve fitting, BD-PSNR is able to provide a good evaluation of the R-D performance [36]. BD-PSNR is a curve fitting metric based on rate and distortion of the video sequence. However, this does not take into account the complexity of the encoder, but the BD metric tells a lot about the quality of the video sequence [30, 31]. Ideally, BD-PSNR should increase and BD-bitrate should decrease. The following results show a plot of BD-PSNR versus the quantization parameter (QP).



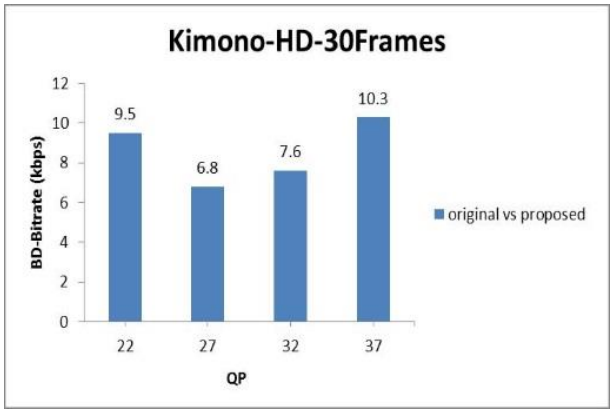


Fig. 11. BD-bitrate vs. quantization parameter SlideEditing and Kimono

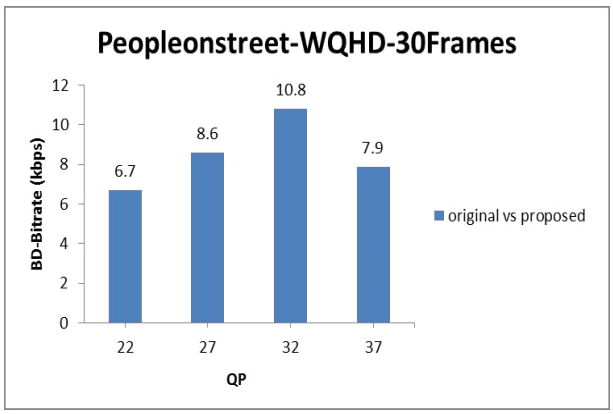


Fig. 12. BD-bitrate vs. quantization parameter Peopleonstreet

C. *BD-Bitrate*

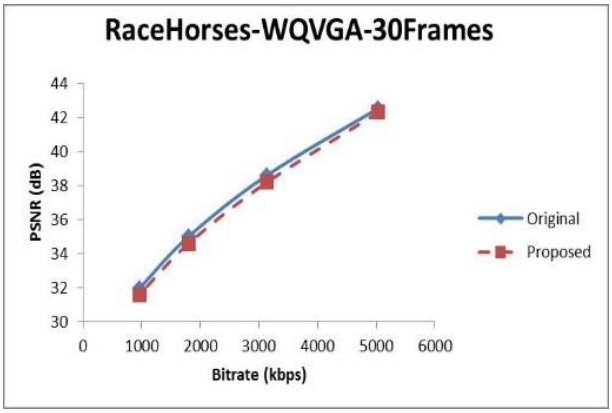
BD-bitrate is a metric similar to the BD-PSNR metric which determines the quality of encoded video sequence along with the measure of the output bitstream of encoded video sequence.

D. *Rate Distortion Plot (RD Plot)*

The results related to the Rate Distortion Plot (RD Plot) are shown in Figures 13 to 15.

E. *Percentage Decrease In Encoding Time*

The results related to the Percentage Decrease in Encoding Time is shown in Figures 16 to 18.



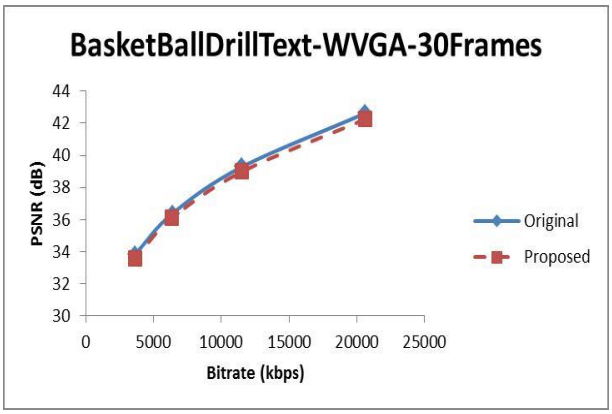
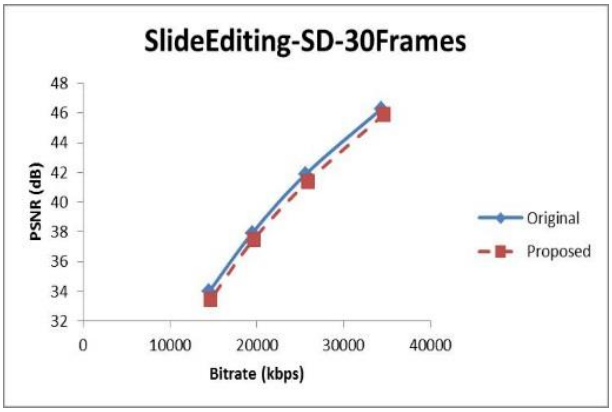


Fig. 13. PSNR vs. bitrate for RaceHorses and BasketBallDrillText



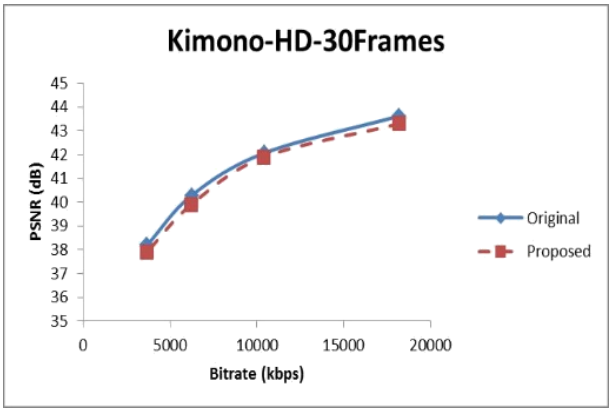


Fig. 14. PSNR vs. bitrate for SlideEditing and Kimono

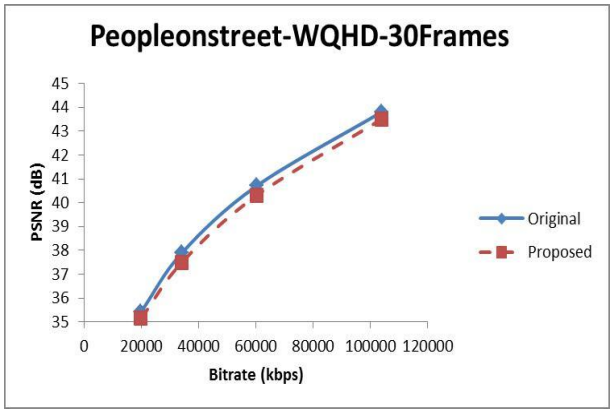


Fig. 15. PSNR vs. bitrate for Peopleonstreet

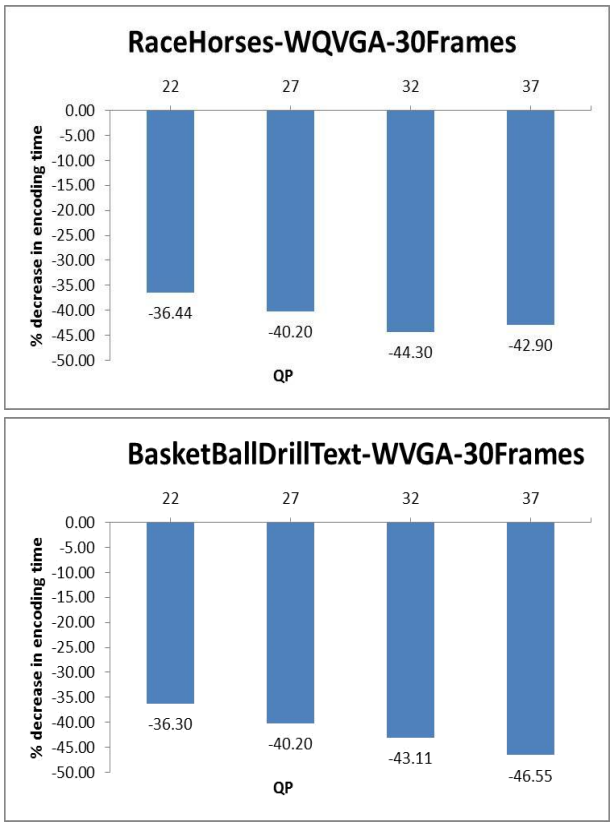


Fig. 16. Percent decrease in encoding time vs. quantization parameter for RaceHorses and BasketBallDrillText

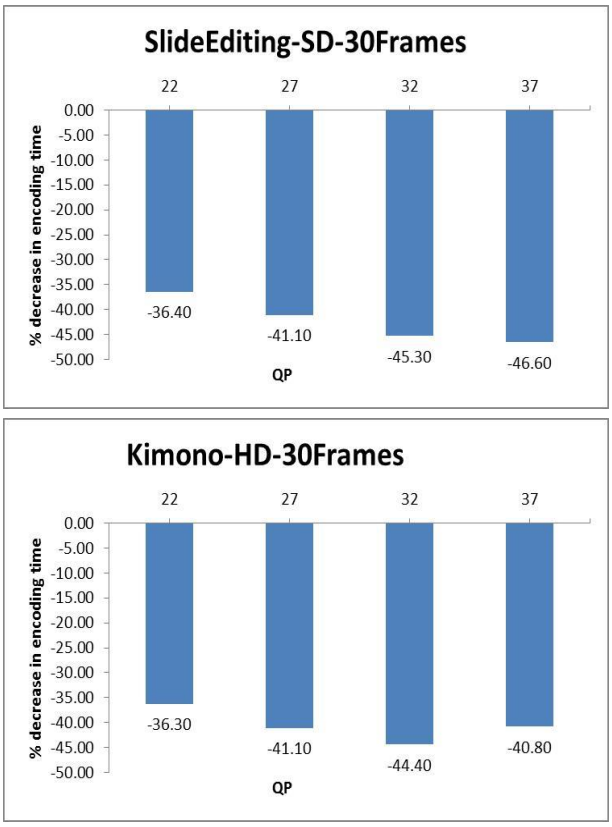


Fig. 17. Percent decrease in encoding time vs. quantization parameter for SlideEditing and Kimono

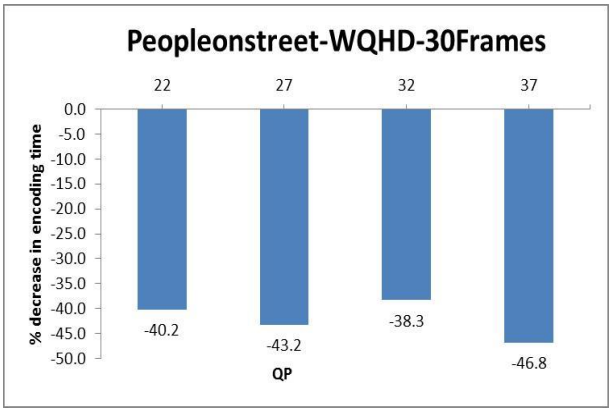


Fig. 18. Percent decrease in encoding time vs. quantization parameter for Peopleonstreet

VII. Conclusions and Future Work

In this paper, a CU early termination algorithm and fast intra mode decision algorithm are proposed to reduce the computational complexity of the HEVC encoder, which includes three strategies, i.e., CU early termination, PU mode decision and early RDOQ termination. The results of comparative experiments demonstrate that the proposed algorithm can effectively reduce the computational complexity (encoding time) by 35–48% (see Figures 4–6) on average as compared to the HM 13.0 encoder [4], while only incurring acceptable drop in the PSNR (see Figures 7–9) and a negligible increase in the bitrate (see Figures 10–12) and encoding bit stream size for different values of the quantization parameter based on various standard test sequences [29]. The results of simulation also demonstrate negligible decrease in BD-PSNR [30] i.e. 0.25 dB to 0.48 dB as compared to the original HM13.0 software [4]. RD distortion plots and percentage decrease in encoding time are shown in Figures 13–15 and Figures 16–18, respectively.

There are many other ways to explore in the CU early termination and fast intra prediction in the intra prediction area as suggested by research [25],[33]. Many of these methods can be combined with this method, or if needed, one method may

be replaced by a new method and encoding time gains can be explored.

Similar algorithms can be developed for fast inter-prediction in which the RD cost of the different modes in inter-prediction are explored, and depending upon the adaptive threshold [34], mode decision can be terminated resulting in less encoding time and reduced complexity combining with the above proposed algorithm.

Tan et al. [37] proposed a fast RQT algorithm for both intra and inter mode coding in order to reduce the encoder complexity. In [37], for all intra case, 13% encoding time can be saved, However, BD-Rate just increases by 0.1%. For random access and low delay constraints it reduces by up to 9% encoding time with 0.3% BD-Rate performance degradation. This method can be integrated with the proposed algorithm to increase the encoding time.

Tian et al [38] proposed a PU size decision algorithm to speed up the intra coding. In this method, two-stage is applied. In the pre-stage, filtering the unnecessary PU by analyzing the texture complexity of the LCU and its four sub-blocks secondly, skipping the small PU candidates by referring the neighboring PU. The simulation results show that proposed method can speed up by average of 44.91%, with only PSNR degradation less than 0.04dB. This method can be combined with the proposed algorithm.

The Bayesian decision [39] rule can be applied to calculate the CU size, and then this information can be combined with the proposed method to achieve further encoding time gains.

Complexity reduction can also be achieved through hardware implementation of a specific algorithm, which requires much computation. The FPGA implementation can be useful to evaluate the performance of the system on hardware in terms of power consumption and encoding time.

ACKNOWLEDGMENTS

This research paper is based on the research conducted by H. Brahmasury Jain towards his M.S. from UTA.

REFERENCES

[1] G. Sullivan et al., "Overview of the high efficiency video coding (HEVC) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, December 2012.

[2] I. E. Richardson, *The H.264 advanced video compression standard*, 2nd Edition, Wiley 2010.

[3] G. Escribano et al., "Video encoding and transcoding using machine learning," *MDM/KDD'08*, Las Vegas, NV, USA, August 24, 2008.

[4] *HM13.0, HEVC code*: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-13.0/

[5] *Blu-ray discs*, http://www.blu-ray.com/info.

[6] *Coding of moving pictures and audio*, http://mpeg.chiariglione.org/standards/mpeg-2/mpeg-2.htm

[7] R. Schafer and T. Sikora, "Digital video coding standards and their role in video communications," *Proceedings of the IEEE*, Vol 83, pp. 907–923, January 1995.

[8] *Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios*, http://www.itu.int/rec/R-REC-BT.601/

[9] *HEVC software manual*: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/branches/HM-13.0-dev/doc/software-manual.pdf.

[10] K. Iguchi et al., "HEVC encoder for super hi-vision," *IEEE ICCE*, pp. 61–62, Las Vegas, NV, January 2014.

[11] MPL Website: http://www.uta.edu/faculty/krrao/dip.

[12] G J. Sullivan et al., "Standardized Extensions of HEVC," *IEEE Journal of Selected topics in Signal Processing*, Vol. 7, No. 6, pp. 1001–1016, December 2013.

[13] C. Fogg, *Suggested figures for the HEVC specification*, ITU-T/ISO/IEC Joint Collaborative Team on Video Coding (JCT-VC) document JCTVC- J0292r1, July 2012.

[14] M.T. Pourazad et al., "HEVC: The new gold standard for video compression," *IEEE Consumer Electronics Magazine*, vol. 1, no. 7, pp. 36–46, July 2012.

[15] F. Bossen et al., "HEVC Complexity and Implementation Analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1685–1696, December 2012.

[16] Basics of video: http://lea.hamradio.si/~s51kq/V-BAS.HTM

[17] B. Bross et al., *HM9: High Efficiency Video Coding (HEVC) Test Model 9 Encoder Description*, JCTVC-K1002v2, October 2012; http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=6807

[18] B. Bross et al., *High Efficiency Video Coding (HEVC) text specification draft 9 (SoDIS)*, JCTVC-K1003v13, October 2012; http://phenix.it-sudparis.eu/jct/doc_end_user/current_document.php?id=6803

[19] W. Jiang, H. Ma, and Y. Chen, "Gradient based fast mode decision algorithm for intra prediction in HEVC," *2nd International Conference on Consumer Electronics, Communications and Networks*, pp. 1836–1840, April 2012.

[20] S. Lui et al., "Video Prediction Block Structure and the Emerging High Efficiency Video Coding Standard," *IEEE Proceedings on Signal & Information Processing Association Annual Summit and Conference* (APSIPA ASC), 2012 Asia-Pacific, pp. 1–4, 2012.

[21] G. Sullivan et al., "Efficient quadtree coding of images and video," *IEEE Transactions on Image Processing*, vol. 3, pp. 327–331, May 1994.

[22] P. Helle et al., "Block merging for quadtree-based partitioning in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1720–1731, May 2012.

[23] K. Choi et al., "Fast coding unit decision method based on coding tree pruning for high efficiency video coding," Proc. SPIE Optical Engineering, vol. 51, 030502, March 2012.

[24] R. Li et al., "A new three-step search algorithm for block motion estimation," IEEE Transactions on Circuits and Systems for Video Technology, vol. 4, pp. 438–442, August 1994.

[25] H. Zhang and Z. Ma, "Fast intra prediction for high efficiency video coding," *Advances in Multimedia Information Processing*, 13th Pacific-Rim Conference on Multimedia, Proceedings, Lecture Notes in Computer Science, vol. 7674, pp. 568–577, December 2012.

[26] G. Tian et al., "Content adaptive prediction unit size decision algorithm for HEVC intra coding," *Picture Coding Symposium (PCS)*, pp. 405–408, May 2012.

[27] H. Lei and Z. Yang, "Fast Intra Prediction Mode Decision for High Efficiency Video Coding," *2nd International Symposium on*

Computer, Communication, Control and Automation, November 2013.
[28] H. Zhang and Z. Ma, "Fast intra prediction for high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 24, No. 4, pp. 660–668, April 2014.
[29] HEVC test sequences: ftp.tnt.uni-hannover.de/testsequences.
[30] G. Bjontegaard, *Calculation of average PSNR differences between RD-curves*, Q6/SG16, Video Coding Experts Group (VCEG), 2-4, April 2001.
[31] *BD metrics code*, http://www.mathworks.com/matlabcentral/fileexchange/27798- bjontegaardmetric/content/bjontegaard.m
[32] K.R. Rao, D.N. Kim, and J.J. Hwang, *Video coding standards: AVS China, H.264/MPEG-4 Part10, HEVC, VP6, DIRAC and VC-1*, Springer 2014.
[33] S. Vasudevan and K.R. Rao, "Combination method of fast HEVC encoding" *IEEE ECTICON 2014*, Korat, Thailand, May 2014.
[34] K. Shah, *Reducing the complexity of inter prediction mode decision for HEVC*, PhD Thesis, University of Texas at Arlington, UMI Dissertion Publishing, April 2014; http://www-ee.uta.edu/Dip/Courses/ EE5359/KushalShah_Thesis.pdf.
[35] *JCT-VC documents*: http://www.itu.int/en/ITU-T/studygroups/2013-2016/16/Pages/video/jctvc.aspx
[36] X. Li et al., "Rate-Complexity-Distortion evaluation for hybrid video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, pp. 957–970, July 2011.
[37] Y. H. Tan et al., "On residual quad-tree coding in HEVC," *2011 IEEE 18th International Conference on Image Processing (ICIP)*, pp. 3469–3472, September 2011.
[38] G. Tian et al, "Content adaptive prediction unit size decision algorithm for HEVC intra coding," *Picture Coding Symposium (PCS)*, pp. 405–408, May 2012.
[39] X. Shen et al., "Fast coding unit size selection for HEVC based on Bayesian decision rule," *IEEE Picture Coding Symposium (PCS)*, pp. 453–456, May 2012.
[40] Special issue on emerging research and standards in next generation video coding, *IEEE Transactions on Circuits and Systems for Video Technology (CSVT)*, vol. 22, pp. 1646–1909, December 2012.
[41] R. H. Gweon et al., *Early termination of CU encoding to reduce HEVC complexity*, document JCTVC-F045, Torino, Italy, July 2011.
[42] A.S. Motra et al., "Fast intra mode decision for HEVC video encoder," *2012 20th International Conference on Software, Telecommunications and Computer Networks, SoftCOM*, September 2012.
[43] L. Zhao et al., "Fast mode decision algorithm for intra prediction in HEVC," Conference Article no. 6115979, *IEEE Visual Communications and Image Processing*, November 2011.
[44] T.L. Da Silva et al., "HEVC Intra Coding Acceleration based on Tree-Inter level Mode Correlation," *IEEE and SPA*, 26–28 September 2013.
[45] S.W. Teng et al., "Fast mode decision algorithm for residual quadtree coding in HEVC," Conference Article no. 6116062, *IEEE Visual Communications and Image Processing*, November 2011.

A Dynamic Gesture Recognition System based on CIPBR Algorithm

Diego G.S. Santos, Rodrigo C. Neto, Bruno J.T. Fernandes, and Byron L.D. Bezerra

***Abstract*—Dynamic gesture recognition has been studied actually for it big application in several areas such as virtual reality, games and sign language. But some problems have to be solved in computer applications, such as response time and classification rate, which directly affect the real-time usage. This paper proposes a novel algorithm called Convex Invariant Position Based on Ransac which improved the good results in dynamic gesture recognition problem. The proposed method is combined with a adapted PSO variation to reduce features and a HMM and three DTW variations as classifiers.**

***Index Terms*—Gesture recognition, computer vision, CIPBR, dynamic time wrapping, hidden Markov model.**

I. Introduction

HAND recognition gestures systems brought to several areas a more natural way for Human-Computer Interaction (HCI). Applications in sign language recognition [1], virtual reality [2], and computer games [3] have increasingly used hand gesture recognition approaches to facilitate the learning based on intuition to remember and performer a gesture.

There are three different categories for systems based in human gesture recognition: systems based on the hand gesture captured by gloves or external sensors [4], system which make the device tracking to generate a gesture path [5]. The last category use a camera to capture the gesture in images and extract features from it using computer vision techniques to interpret the gesture [6], [7].

The first class, hand gesture captured by gloves or external sensors, uses some devices connected in the user hand what turn difficult the natural articulation of hand gesture. This category has the advantage to being invariant to light conditions and complex background, achieving a better result than others applications but is more expensive, due do the high price of the devices to develop a system based on it.

The second category, is the more limited class, by the limited number of gestures which can be recognized. Most of this systems are presented in portable devices where predefined gesture execute some action. This gestures complexity is very low in compared with others categories.

The last category, system based on computer vision techniques, uses a camera to capture the hand gesture in images or videos and extract attributes and features, such as position, velocity, color, among others to identify the gesture. This class has been grown by the present facility to obtain a camera, usually coupled in a smart-phone. Another reason for this growth, is the non-invasive techniques which do not require any devices connected and nothing which hinders the natural hand gesture movement.

Systems based on vision computer techniques usually involve two steps: feature extraction and pattern classification. This paper presents a novel technique called Convexity Invariant Position Based on Ransac (CIPBR) for hand gesture image feature extraction as first module in a dynamic hand gesture system. In addition we used two classifiers to evaluate the proposed method: Dynamic Time Wrapper (DTW) [8] and Hidden Markov Model (HMM) [9].

This paper is structured as follows: Section 2 presents some related works. Section 3 describes the CIPBR algorithm. Section 4 presents the Classifier module based on HMM. Section 5 presents the Classifier module based on DTW. Section 6 presents the Experiments Results. Finally, in Section 7, the conclusions and some future work are given.

II. Related Works

Several techniques have been used in hand gesture recognition systems. Meena [10] and El-Salwah [11] use the Local Contour Sequence (LCS) algorithm to reduce the hand posture into a distances vector with the better contour points.

Calinon et al. [12] use probabilistic techniques as Principal Components Analysis (PCA) and Independent Components (ICA) to recognize and reproduce gestures.

The Speed Up Robust Features (SURF) used for Bao et al. [13] extract points of interest sets over pyramidal images using a Gaussian Laplacian [14] and recognize dynamic gestures making a track path with SURF features.

Some of those algorithms previously cited lose their accuracy when a new test image has a different rotation from the ones observed in training step. To solve this problem Wysoski et al. [15] use boundary histograms and neural networks to correct the hand posture angle. Grzeszcuk et al. [16] use an approach based on stereo vision to realize in 3D the angle correction.

Keogh et al. [17] propose a simple method to correct the angle variation based on object form, where do not matter the angle distortion the same contour pixel is extracted of the image as the first one and the rest of the contour pixels were always organized in the same order. Keogh et al. achieve good results with this technique in skulls classification, extracting

Manuscript received on July 28, 2014, accepted for publication on September 22, 2014, published on November 15, 2014.

The authors are with Universidade de Pernambuco, Brazil. Corresponding author: Diego G.S. Santos (e-mail: dgs2@ecomp.poli.br).

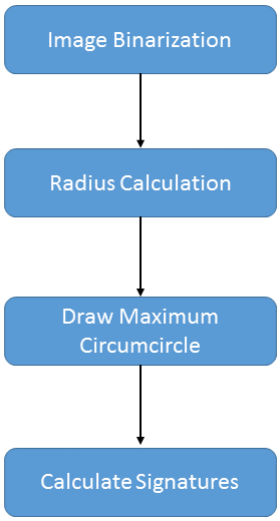


Fig. 1. CIPBR's flow

two signatures of the images, distances and angles, and turning them into a single features vector.

Barros et al. [18] propose a method to solve another problem in computer vision approaches: the feature vector size. The traditional approaches return a big vector which complicate the real-time classification task. The Convexity Approach solve this reducing the hand posture into a polygonal shape and extracting only the more external and internal contour points. For last, it is calculated the distance between each pair of points. Then, a set of points is selected in order to obtain the minimal contour representation and generate a smaller feature vector. Barros et. al. evaluates their work using three pattern classification: Elman Recurrent Neural Network (Elman RNN) [19], HMM and DTW achieving better results with the last two.

Based on Keogh et al. and Barros et al. concepts, CIPBR algorithm is proposed and evaluated in this paper using two classifiers: HMM and DTW. The last one is used in three version, the traditional one, an adaptation to CIPBR's two signatures and a variation presented by Salvador [20].

III. CIPBR Algorithm

The CIPBR algorithm is composed by simple tasks to reduce the hand posture images into two signature sets. There is four modules in cascade.

Figure 1 presents CIPBR's flow starting with the input image and finishing with the feature vector. The first module, "Image Binarization", receives a hand posture image as input and outputs two images: the first one in RGB, same as the input, and the second a binary image (based on Otsu Threshold Algorithm [21]).

The second module, "Radius Calculation", has three simple tasks to complete. The first one is find the hand pulse line in RGB image using a simple linear regression. The second one is extract the contour where each pixel is marked as border, if it has one of its eight neighbors black in the binary image and calculate the mass center point by Hu Invariant moments. This module use same principle as Keogh et al. [17] to guarantees the rotation invariance starting the contour organization with the center pulse line pixel. Lastly the distance between the center of mass of the contour point and pulse center point is calculate as show Figure 2 (a).

The third module, "Draw Maximum Circumcircle", uses the distance previously calculated as radius to draw a circle inside the hand contour. In order to solve the problem of the circle exceeding the hand contour, a triangle is calculated using the three more distant contour points from mass center point and the biggest circle inside it is used.

Then, in "Calculate Signatures" module, convex Hull is calculate from hand posture contour using Andrew's monotone chain convex hull algorithm [22], thus reducing the number of contour points substantially.

As result a set $P = \{p_1, p_2, \ldots, p_n\}$ from Convex hull points are used to generate CIPBR signature sets. For each point $p_k = \rho$ is traced a line PC starting with ρ and ending with the central point C of the maximum circumcicle hand shape exported from Module 3. Then, it is measured the Euclidean distance from ρ to the point Q_k given by the intersection between PC and the maximum circumcicle. The set of all distances computed as this procedure is the descriptors for the first signature set. The distance ρQ_k can be calculate with the follow equation:

$$Distance_{\rho Q_k} = \sqrt{(\rho_x - C_x)^2 + (\rho_y - C_y)^2} - radius,$$

where C is the hand posture contour mass center point, ρ_x, C_x are the x coordinates from points ρ and C respective, ρ_y, C_y are the y coordinates from points ρ and C respective and $radius$ the radius calculated in the second module.

The second signature set consists of angles (A) obtained by calculating the angle (A) between a line composed for each point P of the convex hull hand shape point and radius (PC). All two signature sets are obtained in a clockwise direction always starting with point P as Figure 2 shows.

Finally, to create the feature vector the signature sets are normalized, distances by the radius calculated in the second module as follows:

$$Distance_i = \frac{Distance_i}{radius}$$

and angles set by 360°. The final vector is created concatenating distances and angles in a single vector as follows:

$$Angle_i = \frac{Angle_i}{360^\circ}.$$

IV. Classifier Module based on HMM

For some classification techniques, such as Hidden Markov Model and Artificial Neural Networks, among others require a fixed length feature vector as input, but most of techniques

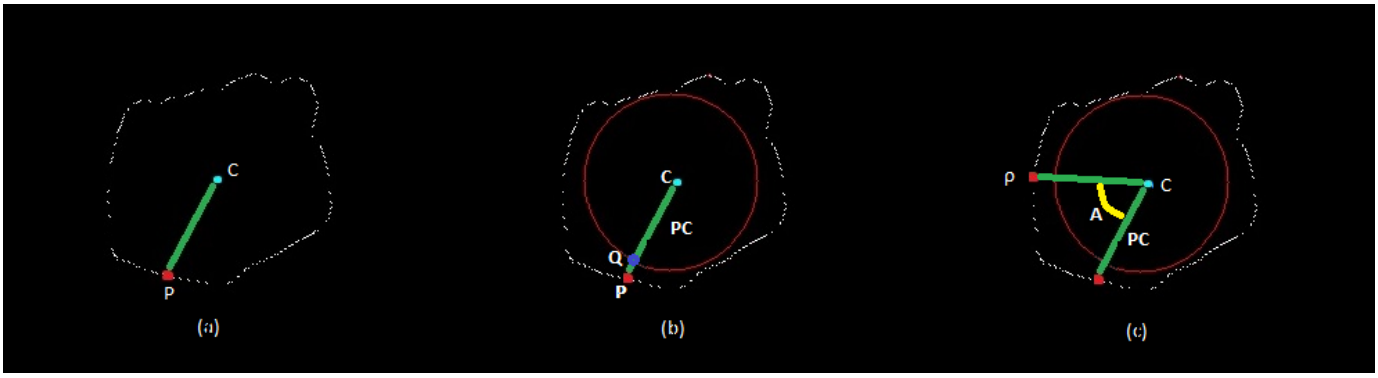


Fig. 2. (a) CIPBR second step output. The hand shape contours and central point (cyan point). The hand base middle point (red point), the radius (green line) and angle A (yellow arc). (b) and (c) CIPBR fourth step signature calculate.

return as output vectors with different sizes. CIPBR algorithm returns a set of vectors, each one containing around 200 features. To use this vectors as input in a HMM, and achieve a convergence state, it is necessary a heavy reduction of features in each vector. To solve this, two approaches is adapted in this work. At first one, Particle Swarm Optimization [23] chose the better feature vectors size and the selector algorithm proposed by Barros et al. [6] receives PSO size output and choose the candidate features for this vectors.

A. Particle Swarm Optimization

When a PSO is used to solve an optimization problem, a swarm of simple computational elements, called particles, is used to explore a solution space to find an optimum solution. The position from each particle represent a candidate solution in dimensional search space (D) represented by $X = \{x_1, x_2, x_3, \ldots, x_n\}$, where each x_n is a position in the n-th dimension. The same way, the velocity is represented by $V = \{v_1, v_2, v_3, \ldots, v_n\}$.

The fitness function evaluates how better each particle present itself in each iteration. Better position where the particle was, is saved in a variable called p_{Best} and has the value exchanged whenever particle finds a better value. To guide the swarm to the best solution, the position, where a single particle found the best solution until the current execution, is stored in a variable called g_{Best}. This way, to update the particle velocity and position the following equations are used:

$$v_i(t+1) = \omega * v_i(t) + c_1 r_1[p_i Best - x_i(t)] + c_2 r_2[g_{Best} - x_i(t)]$$
$$x_i(t+1) = x_i(t) + v_i(t+1),$$

where $i = (1, 2, 3, \ldots, N)$ and N is the size of the swarm. c_1 which represent the private experience or "cognitive experience" and c_2 represents the "social experience" interaction usually used with value 2.05. r_1 and r_2 a random number between 0 and 1 which represents how much p_{Best} and g_{Best} will influence the particle movement. The inertia factor ω, is a variable used to control the balance of the search algorithm between exploration and exploitation. x_i represent particle position in $i-th$ dimension. The recursive algorithm will go on until a break condition happen, maximum number of iteration in this paper case.

B. Binary Particle Swarm Optimization

A binary PSO is a variation from traditional version where the particle position imitates the human genetic code. In the binary version, the position particle is represented only by $\{0\}$s and $\{1\}$s. The other change in binary PSO algorithm is how position and velocity are calculated, by the following equation:

$$If\ rand\ <\ \frac{1}{1+e^{-v_i(t+1)}}\ then\ X_i(t+1) = 1;$$
$$else\ X_i(t+1) = 0.$$

For binarization process is chosen the threshold which delimit what vector cell will represent a feature. For our experiments is chosen 1.2 as threshold to be the average of all the values presented.

C. Fitness Function

Several fitness function are used to features selection in image recognition systems [24] based on binary PSO approach. After some study a fitness function based on euclidean distance is used in this paper.

First of all, the algorithm receives a binarized vector with n dimensions. Each vector reflect a image gesture and each dimension a feature from it. In this case, the fitness function is calculated with the Euclidean distance from the vector particle position to each vector of the gesture set correspondent and makes the sum of all the distances. In execution, the particles become more able as it fitness value is the smallest possible compared with the fitness obtained by other particles. The particle fitness function is:

$$fitness_i = \sum_{j=1}^{m} \sqrt{\sum_{k=1}^{n} (x_{ik} - F_{jk})^2}.$$

D. Selector Algorithm

After better feature vector size has been chose for Binary PSO, another normalization is used to choice the features which will compose the vectors. To this task the Barros et al. Selector Algorithm [18] is applied.

First, choose S, the PSO chosen size. Then, if any vector has fewer points than S, are added in the feature vector until matches the desired length (S). The feature vectors with more points than S are redefined using a selection algorithm. This algorithm consists in calculation of a window W through the division of the current vector length by S (desired length). The current vector is parsed, and each value in W position is included to the new feature vector. If the new output vector is even smaller than the desired length, the remaining positions are randomly visited and used to compose the new output vector until the desired length is achieved.

E. Binary PSO Configuration

After some tests, the following values is used to configure Binary PSO in our experiments and find the better feature vector size:

- 15 particles;
- 20 dimensions (*features*);
- 30 simulations;
- 200 iterations;
- $c_1 = c_2 = 2.05$;
- Inertia factor $\omega = 0,9 \rightarrow 0,4$;
- $r_1 = r_2 = rand[0;1]$.

For each gesture the PSO returns a number which better describe each gesture set, but HMM only accepts vectors of the same size. To decide the final size is made a tournament where the number with more incidence wins.

F. HMM Training Configuration

The Hidden Markov Model technique uses a K-Means Clustering [25] to find the best initial approximation. The Baum-Welch algorithm [26] is used to train the HMM, resulting in a fast training process.

V. Classifier Module Based on DTW

The Dynamic Time Warping (DTW) was introduced to overcome the limitation in measure the distance between two time series in specific case: when there is distortion in one of then shifting some slice. To solve this a simple approach based on Euclidean distance is proposed as follows:

Given two time series X, and Y, of lengths $|X|$ and $|Y|$, construct a warp path $W = \{1, w_2, w_3, \ldots, w_k\}$ where $max(|X|, |Y|) \leq K < |X| + |Y|$ and the k^{th} element is $w_k = (i, j)$ where i is an index from time series X and j an index from time series Y.

The warp path must start at $w_1 = (1, 1)$ and finish at $w_k = (i, j)$ in order find the cost matrix

To find the minimum-distance warp path, every cell of the cost matrix must be filled. The value of a cell in the cost matrix is:

$$Dist(i,j) = Dist(i,j) + min[D(i-1,j), D(i,j-1), D(i-1,j-1)],$$

The warp path to $D(i, j)$ must pass through one of those three grid cells, and since the minimum possible warp path distance is already known for them, all that is needed is to simply add the distance of the current two points to the smallest one. Since this equation determines the value of a cell in the cost matrix by using the values in other cells, the order that they are evaluated in is very important.

A. DTW for CIPBR

The DTW algorithm has a complexity problem of $O(N^2)$ level. This has a direct effect on time rating increasing it exponentially higher order time serie which compromises a real time system. To solve this problem two approaches are used in DTW. The first one in the way that DTW calculates the cost matrix and second on in the presentation series to classifier.

The first change is made in the classifier in the follow way:

$$Dist(i,j) = abs[(D_i - D_j) + (A_i - A_j)] + min[D(i-1,j), D(i,j-1), D(i-1,j-1)].$$

Where D_i, D_j are distances and A_i, A_j angles, from CIPBR feature vectors. All the rest of the DTW algorithm follows the same traditional way.

B. Coarsening

Another way to reduce the time rating in DTW execution is reduce the gesture feature vector size. *Coarsening* [20] is a technique to reduce the size (or resolution) of a time series by averaging adjacent number of points decided by a *radius* which say how many neighbors will be used. The resulting time series is a factor smaller than the original time series. In our experiments the gesture can be reduced several times what reduce the time rate for several times. If a $radius = 2$ is given to reduce a time-series $X = x_1, x_2, \ldots, x_n$, each element of the time-series X' will consist of:

$$x'_n = \frac{x_{i-2} + x_{i-1} + x_i + x_{i+1} + x_{i+2}}{5}.$$

The resulting time series is a factor of two smaller than the original time serie. *Coarsening* is run several times to produce many different resolutions of the time series

VI. Experiments

The effectiveness of the proposed method is shown in comparison with the results presented by Barros et al. [18], using the same dynamic gesture dataset, RPPDI Dynamic Gestures Dataset[1].

[1] Available at http://rppdi.ecomp.poli.br/gesture/database

RPPDI Gesture dataset contains 7 different dynamic gestures each one composed by 14 frames with an uniform background and a size of 640×480 pixels in RGB format. In order to make the hand detection easier, the user wears a black belt to define the pulse in each hand posture. The Figure 3 shows an example of each gesture in the dataset.

In order to evaluate our approach, a classification system composed of two modules connected in cascade is used. The first module contains the CIPRB algorithm to extract features from the hand gesture images, while the second module classify the gesture using HMM or DTW.

To evaluate the proposed system as presented by Figure 4, four scenarios are executed. For all scenarios a system based on computer vision is tested with CIPBR algorithm as features extractor module. The scenarios are presented as follows:

1) The first one scenario consists in use Classifier Module based on HMM as system classifier, normalizing the feature vectors with PSO and Barros et al Selector Algorithm;
2) The second scenario consists in traditional DTW as classifier;
3) Third scenario uses the modified DTW for CIPBR (accord section 5.1);
4) The last scenario reduce feature vectors using *coarsening* method with $radius = 4$ and uses the traditional DTW to classify.

All scenarios are repeated 30 times in a dataset randomly chosen containing 2/3 of the sequences in each gesture class for training and 1/3 for test. The results present the average among all repetitions.

A. Results and discussion

Table I presents the resume results in classification for this work. The table shows that all classifiers have achieved rates nearing complete resolution of the RPPDI dataset.

HMM recognize the lower number of gesture, achieving only 94% of classification rate, making more mistakes than all DTW versions. But the average time rated is several times smaller than the faster version of DTW. Due to the low number of features generated by the combination of PSO with the selector algorithm, generating vectors with size 10.

TABLE I
CLASSIFICATION RATE RESUME FOR ALL EXPERIMENTS.

Scenario	Classification rate	Standard Deviation	Time Average
1.	94.11%	6.72%	1.78ms
2.	99.79%	0.54%	4516.74 ms
3.	98.85%	1.41%	16337.94ms
4.	96.82%	1.85%	956.74ms

Table II presents a comparison between the classification rate obtained by Barros et al. [6] using the Convexity Approach applied to Local Contour Sequence (LCS) and Speed Up Robust Features (SURF). The combination of the techniques LSC and SURF with Convexity Approach are called CLCS and CSURF, respectively.

TABLE II
CLASSIFICATION RATE RESUME COMPARING BARROS ET AL. [18] RESULTS WITH OURS.

Method	Classification rate (%)
CLCS + HMM	91.00
CSURF + HMM	91.00
CIPBR + HMM	94.11
CLCS + RNN	90.00
CSURF + RNN	92.00
CSURF + DTW	93.00
CLCS + DTW	97.00
CIPBR + Coaserning + DTW	96.82
CIPBR + DTW for CIPBR	98.85
CIPBR + DTW	99.79

CIPBR and DTW combination increases the classification rate in 2.79 percentage points in the better Barros' result. For several executions the proposed system do not commit any classification mistake, solving the gesture recognition problem in the RPPDI database.

The results comparative have a smaller difference among them, been necessary apply a Shapiro-Wilk statistic test between the result using the CLCS and CIPBR for HMM and DTW as classifiers. The test with HMM presents a p-value equals to 3.89336E-26 what discard the possibility of equality between scenarios for 95% significance. To DTW versions the p-values are smaller than 0.5 for scenario 2 and 3. For scenario 4, p-value has 0.508361 as value demonstrating the equality among the CLCS and CIPBR scenarios. This demonstrates the effectiveness of CIPBR approach achieving better results than CLCS in RPPDI dataset, where in the worse case our proposed system has an effectiveness equal an system based on CA.

VII. CONCLUSION

This paper presents a new approach for feature extraction and classification of hand gestures called CIPBR. CIPBR + DTW combination improve the previously results for RPPDI dataset.

Also is presented a new method to use the binary version of PSO to find the optimal number of features for CIPBR vector being selected by a simple selector algorithm and classified using the Hidden Markov Model as classifier. This selector algorithm is necessary because PSO fitness function generated inconclusive results for classification, being necessary another method to choice the feature to final vectors.

The results showed promise in that first moment of finding the best number necessary to describe every gesture extracted by CIPBR. Using dynamic gestures RPPDI dataset, our approach achieved 100% of classification rate in almost all executions solving the classification problem of RPPPDI dataset on both assertiveness as at runtime with both classifiers. As can be seen in the results session, DTW is much more assertive than HMM but the last got a lot shorter rating

Fig. 3. RPPDI Dynamic Gesture Dataset [18]

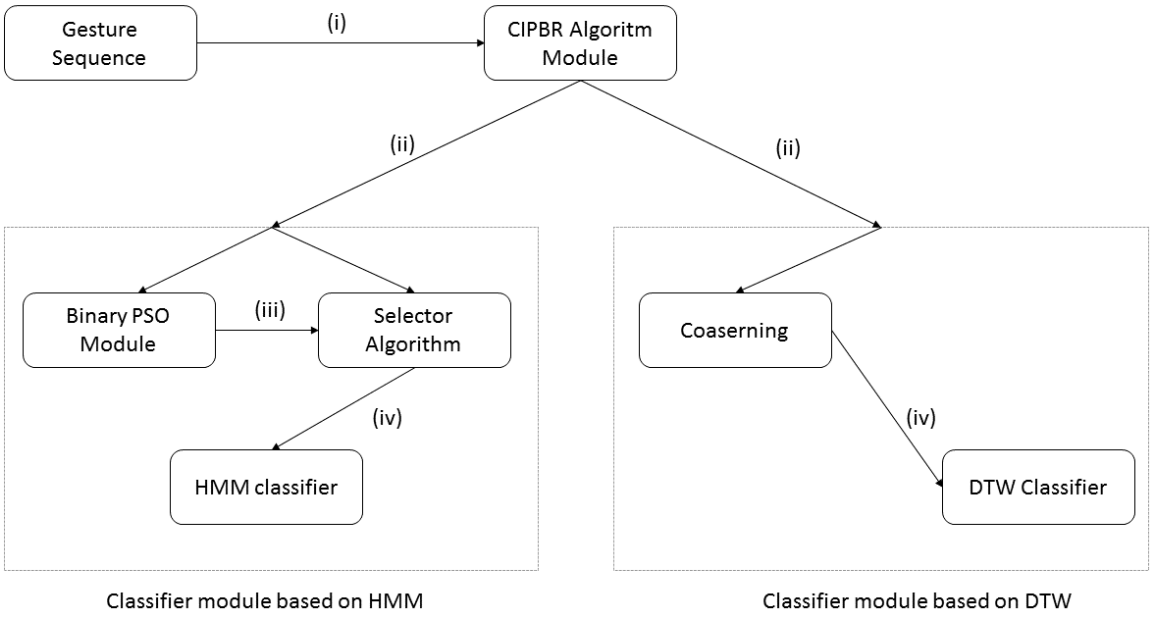


Fig. 4. General architecture of the system. (i) Gesture images sequence as input for CIPBR module. (ii) Gesture features vector. (iii) The best size for gesture features vector. (iv) Reduced vector for respective classifier.

time, being the best for a system to classify gestures in real time.

As future work, a priori is find a fitness function that makes assertive binary PSO both in reducing characteristics as for classification. In consequence, the use of more complex dataset and using other cameras also, such as the Microsoft Kinect.

ACKNOWLEDGMENT

The author would like to thank the support of the Brazilian agencies: CNPQ, Facepe and Capes.

REFERENCES

[1] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 41, no. 6, pp. 1064–1076, 2011.

[2] G. Lu, L.-K. Shark, G. Hall, and U. Zeshan, "Immersive manipulation of virtual objects through glove-based hand gesture interaction," *Virtual Reality*, vol. 16, no. 3, pp. 243–252, 2012.

[3] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan, "Vision-based hand-gesture applications," *Communications of the ACM*, vol. 54, no. 2, pp. 60–71, 2011.

[4] G. Dewaele, F. Devernay, and R. Horaud, "Hand motion from 3d point trajectories and a smooth surface model," in *Computer Vision-ECCV*

2004. Springer, 2004, pp. 495–507.

[5] J. Schlecht, B. Carqué, and B. Ommer, "Detecting gestures in medieval images," in *2011 18th IEEE International Conference on Image Processing (ICIP)*. IEEE, 2011, pp. 1285–1288.

[6] P. V. A. Barros, N. T. M. Júnior, J. M. M. Bisneto, B. J. T. Fernandes, B. L. D. Bezerra, and S. M. M. Fernandes, "An effective dynamic gesture recognition system based on the feature vector reduction for SURF and LCS," in *Artificial Neural Networks and Machine Learning–ICANN 2013*. Springer, 2013, pp. 412–419.

[7] Z. Ren, J. Yuan, J. Meng, and Z. Zhang, "Robust part-based hand gesture recognition using kinect sensor," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1110–1120, 2013.

[8] H. Sakoe, S. Chiba, A. Waibel, and K. F. Lee, "Dynamic programming algorithm optimization for spoken word recognition," *Readings in speech recognition*, vol. 159, 1990.

[9] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," in *Readings in speech recognition*, A. Waibel and K.-F. Lee, Eds. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990, pp. 267–296. [Online]. Available: http://dl.acm.org/citation.cfm?id=108235.108253

[10] S. Meena, "A study on hand gesture recognition technique," Master's thesis, National Institute of Technology, Rourkela, 2011.

[11] A. El-Sawah, C. Joslin, N. D. Georganas, and E. M. Petriu, "A framework for 3D hand tracking and gesture recognition using elements of genetic programming," in *Fourth Canadian Conference on Computer and Robot Vision, 2007. CRV'07*. IEEE, 2007, pp. 495–502.

[12] S. Calinon and A. Billard, "Recognition and reproduction of gestures using a probabilistic framework combining PCA, ICA and HMM," in *Proceedings of the 22nd international conference on Machine learning*. ACM, 2005, pp. 105–112.

[13] J. Bao, A. Song, Y. Guo, and H. Tang, "Dynamic hand gesture recognition based on SURF tracking," in *2011 International Conference on Electric Information and Control Engineering (ICEICE)*. IEEE, 2011, pp. 338–341.

[14] P. J. Burt and E. H. Adelson, "The laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. 31, no. 4, pp. 532–540, 1983.

[15] S. G. Wysoski, M. V. Lamar, S. Kuroyanagi, and A. Iwata, "A rotation invariant approach on static-gesture recognition using boundary histograms and neural networks," in *Proceedings of the 9th International Conference on Neural Information Processing, ICONIP'02*, vol. 4. IEEE, 2002, pp. 2137–2141.

[16] R. Grzeszcuk, G. Bradski, M. Chu, and J.-Y. Bouguet, "Stereo based gesture recognition invariant to 3D pose and lighting," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, 2000, pp. 826–833 vol.1.

[17] E. Keogh, L. Wei, X. Xi, S.-H. Lee, and M. Vlachos, "LB_Keogh supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures," in *Proceedings of the 32nd international conference on Very large data bases*. VLDB Endowment, 2006, pp. 882–893.

[18] P. V. A. Barros, N. Junior, J. M. M. Bisneto, B. J. T. Fernandes, B. L. D. Bezerra, and S. M. M. Fernandes, "Convexity local contour sequences for gesture recognition," in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*. ACM, 2013, pp. 34–39.

[19] L. Medsker and L. C. Jain, *Recurrent neural networks: design and applications*. CRC press, 1999.

[20] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, 2007.

[21] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[22] A. M. Day, "Planar convex hull algorithms in theory and practice," in *Computer graphics forum*, vol. 7, no. 3. Wiley Online Library, 1988, pp. 177–193.

[23] J. Kennedy, R. Eberhart *et al.*, "Particle swarm optimization," in *Proceedings of IEEE international conference on neural networks*, vol. 4, no. 2. Perth, Australia, 1995, pp. 1942–1948.

[24] X. Zhang, W. Wang, Y. Li, and L. Jiao, "PSO-based automatic relevance determination and feature selection system for hyperspectral image classification," *Electronics Letters*, vol. 48, no. 20, pp. 1263–1265, September 2012.

[25] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A k-means clustering algorithm," *Applied statistics*, pp. 100–108, 1979.

[26] T. L. Baum, G. Peterie, and N. W. Souled, "A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains," in *Proceedings of Annals of Mathematical Statistics*, vol. 41, 1995, pp. 164 –171.

Design of High Accuracy Tracking Systems with H_∞ Preview Control

Antonio Moran Cardenas, Javier G. Rázuri, Isis Bonet, Rahim Rahmani, and David Sundgren

***Abstract*—Positioning and tracking control systems are an important component of autonomous robot applications. This paper presents the design method of tracking control systems based on H_∞ preview control where the present and future desired positions of the robot are used to determine the control actions to be applied so that the robot describes the desired trajectory as close as possible. The performance improvements achieved with H_∞ preview control have been examined in the frequency and time domains for different types of reference signals when applied to a one-dimensional positioning system. It was found that preview control improves the tracking performance by improving the phase response of the tracking system.**

***Index Terms*—Robotics, planning and scheduling, predictive control, H_∞ control, tracking control, control H_2, frequency-domain analysis, time-domain analysis.**

I. Introduction

PRECISE positioning/tracking control is being studied in many manufacturing fields in order to improve the accuracy and performance of manufacturing process and manipulator driving systems which every time demand more precise, robust and efficient control systems [1]. Certain behavior is desired in positioning/tracking control systems: fast response and convergence, zero tracking error and robustness against changes in the system itself and/or its environment.

The classical way to solve the tracking control problem for linear time-invariant systems has been to design a one-degree-of-freedom, or better, a two-degrees-of-freedom controller which will achieve the desired performance as close as possible. The inherent shortcoming of the classical approach is the overdesign that is entitled in requiring a performance that is independent of the specific measurable signal to be tracked. In the case of standard H_∞ control, technique introduced by Zames [2] a controller is designed for a worst case reference signal which may be different from the one which is actually encountered. Obviously such an approach may introduce a considerable emphasis on the tracking properties of the system which may lead, in some cases, to insufficient disturbance attenuation and robustness properties of the closed-loop system [3].

In order to avoid the shortcomings of classical control when the reference signal is known a priori (for some interval of time), the concept of preview tracking control has been introduced. According to this control strategy, the control input is calculated using not only information corresponding to the present state of the system but also using the known future value of the reference signal. The advantages of preview control has been studied by several authors [4]–[6]. An updated review of preview control is presented in [7]. The most of preview control methods are based on classical LQ/LQG control theory where preview control is used for further reduction of the minimum value achievable by the cost function being optimized [8]–[10].

However, LQG control methods present the limitation of how to include robustness specifications for designing the controller which are usually defined in terms of H_∞ norms. Several preview control methods based on H_∞ control theory, which take into acount the unknown disturbances to design control systems with better performance measures, have been developed. In [11]–[13], the preview performance in terms of H_∞ criterion was investigated. Some authors have been presented solutions for discret-time cases [14], [15], and others for continuous-time problems [16]–[18].

This paper presents the solution to the H_∞ preview tracking control problem. The analytical formulation of the H_∞ preview tracking control law has been derived in the continuous time domain. Preview control laws have been derived for two tracking problems: (a) the reference signal is perfectly known in advance for the total working time and (b) the reference signal is previewed for a fixed interval of time. For both cases the structure of the preview controller corresponds to a two-degrees-of-freedom controller with feedback and feedforward (preview) parts. The benefits and limitations of the proposed control strategies are analyzed theoretically and experimentally in the time and frequency domains for a wide range of operating conditions and different types of reference signals.

This paper is structured as follows: Section II presents the structure of the experimental positioning/tracking system, while Section III describes the State Equation Model of the system. In Section IV, we described the formulation process related to the generalized plant. Experimental results and discussion are presented in Section V, while Section VI presents some conclusions of the research.

Manuscript received on July 29, 2014, accepted for publication on September 22, 2014, published on November 15, 2014.

Antonio Moran Cardenas is with Pontifical Catholic University of Perú PUCP, Lima, Perú (e-mail: amoran@pucp.edu.pe).

Isis Bonet is with the Antioquia School of Engineering (EIA–GISMOC), Colombia (e-mail: pfibonet@eia.edu.co).

Javier G. Rázuri (corresponding author), Rahim Rahmani, and David Sundgren are with the Stockholm University (DSV), Department of Computer and Systems Sciences, Sweden (e-mail: {javier, rahim, dsn}@dsv.su.se).

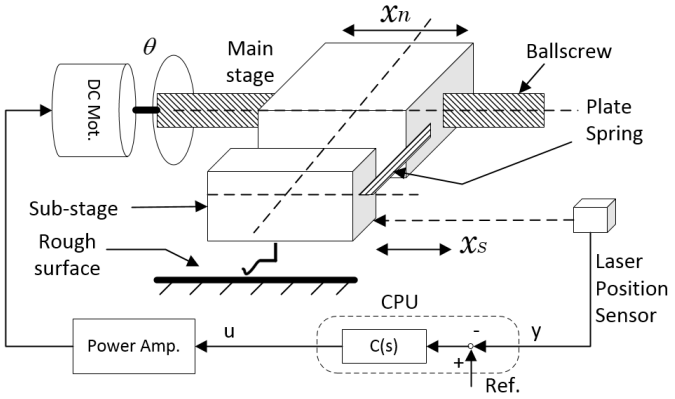


Fig. 1. Experimental tracking system

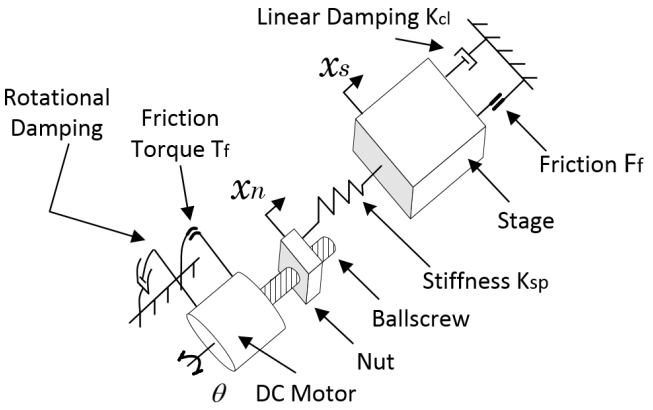


Fig. 2. Vibrational model of experimental system

II. Experimental System

A. Experimental System

Figure 1 shows the structure of the experimental positioning/tracking system. A DC motor rotates a ballscrew to longitudinally move a positioning stage which is composed of a main stage and a sub-stage connected through two flexible plate springs. The control objective is to place the sub-stage in an arbitrary desired position or to follow any desired trajectory. To do that, a laser position sensor measures the position of the sub-stage and sends the measured signal to a computer which calculates a control voltage according to the algorithm of the H_∞ controller. Using a power amplifier, the control signal is applied to the DC motor to control the rotational motion of the ballscrew in order to achieve the desired motion of the sub-stage. The DC motor and main stage are affected by friction torque and friction forces, respectively. The sub-stage moves on a rough surface whose degree of roughness (friction) may be varied to analyze the robustness of the system against external friction forces.

B. Vibrational Model

Figure 2 shows the equivalent vibrational model of the positioning/tracking mechanism. The flexible plates are modeled as a linear spring with stiffness K_{sp}. The rotational damping of the DC motor is represented by K_{cr} and the friction torque by T_f. External forces actuating on the sub-stage are represented by linear damping K_{cl} and friction force F_f.

III. State Equation Model

A. Nominal Model

Neglecting the inductance of the DC motor, its dynamics can be described by the following equation:

$$\left(\frac{R_a}{K_t}\right) T = K_{pa} u - K_e \dot{\theta}, \tag{1}$$

where T and u are the torque and voltage of the DC motor and $\dot{\theta}$ is the angular velocity of the ballscrew. The rotational motion equation of the driving mechanism (motor shaft and ballscrew) is given by

$$T - T_f - T_{qr} = J\ddot{\theta} + K_{cr}\dot{\theta}, \tag{2}$$

where T_f is the Coulomb friction torque and T_{qr} is an equivalent torque representing the rotational effect of the inertia of the main stage and is equal to

$$T_{qr} = K_{sp}K_{bs}(x_m - x_s), \tag{3}$$

where x_m and x_s represent the position of the main stage and sub-stage, respectively. The linear motion equation of the sub-stage is given by

$$(x_m - x_s)K_{sp} - F_f = M\ddot{x}_s + K_{cl}\dot{x}_s, \tag{4}$$

where F_f is the friction force affecting the sub-stage and M is the mass of the sub-stage. Defining the state vector:

$$\bar{x} = \left[\dot{\theta}, \theta, \dot{x}_s, x_s\right]^T, \tag{5}$$

and combining Equations 1 to 4, the following state-space equation of the nominal system P can be formulated:

$$\dot{\bar{x}} = \bar{A}\bar{x} + \bar{B}_1\bar{w} + \bar{B}_2 u, \tag{6}$$

where $\bar{w} = [T_f, F_f]^T$. The parameters specifications are shown in Table I.

TABLE I
List of Parameters of Experimental Model

Name	Simbol	Value
Power amplifier gain	K_{pa}	10
DC motor resistance	R_a	1.1 ohm
DC motor torque constant	K_t	0.0573 $N.m/A$
Back electromotive force constant	K_e	5.665 x 10^{-2} $v.s$
Moment of inertia	J	4.326 x 10^{-5} $kg.m^2$
Rotational damping factor	K_{cr}	4.550 x 10^{-3} $N.m.s$
Ballscrew transducing coefficient	K_{bs}	1.509 x 10^{-3} m/rad
Spring stiffness	K_{sp}	264 N/m
Linear damping factor	K_{cl}	0.747 $N.s/m$
Mass of sub-stage	M	0.244 kg

B. Generalized Plant with Integral Action

In order to design a controller with integral action to eliminate steady-state tracking errors, the generalized plant used for designing the controller is structured considering integral action. To do that, Equation 6 representing the nominal plant, is integrated with the equation of the tracking error e given by the expression:

$$e = r - x_s, \tag{7}$$

where r is the reference signal to be tracked. Defining the state vector

$$x = \left[\dot{\theta}, \theta, \dot{x}_s, x_s \int (r - x_s) dt\right]^T, \tag{8}$$

the state-space equation of the generalized plant with integral action is:

$$\dot{x} = Ax + B_{1p}\bar{w} + B_2 u + B_r r, \tag{9}$$

where it can be noted that the reference signal r is not included in the disturbance vector $\hat{w}$ Matrices A, B_{1p}, B_2 and B_r are obtained by combining Equations 6 and 7.

IV. H_∞ Preview Control Law

As it is well known, the first step for designing an H_∞ controller is the formulation of the generalized plant. By defining the controlled output vector z as

$$z = \left[\rho_1 \dot{x}_s, \rho_2 x_s, \rho_3 \int (r - x_s) dt, u\right]^T, \tag{10}$$

and the measured output vector y as:

$$y = \begin{bmatrix} x_s + \eta_1 \\ \int (r - x_s) dt + \eta_2 \end{bmatrix}, \tag{11}$$

the equations describing the generalized plant P are:

$$\dot{x} = Ax + B_1 w + B_2 u + B_r r \tag{12}$$

$$z = C_1 x + D_{12} u \tag{13}$$

$$y = C_2 x + D_{21} w \tag{14}$$

where the disturbance vector $w = [T_f, F_f, \eta_1, \eta_2]^T$ and $B_1 = \left[B_{1p}^T 0\right]^T$. Factors ρ_1, ρ_2 and ρ_3 in Equation 10 are weighting coefficients selected to tailor the performance and robustness specifications. η_1 and η_2 in Equation 11 represent measurement noise which are assumed to be independent of the plant disturbances $\bar{w}$. Given this assumption, matrices B_1 and D_{21} turn to be orthogonal. Also, given the components of the controlled output z, matrices C_1 and D_12 are orthogonal. The structure of the generalized plant is shown in Figure 3 (to gain clarity, the controlled output z is not shown).

According to H_∞ control theory, given the generalized plant P, a controller is designed so that the H_∞ norm of the transfer function T_{zw} from disturbances w to controlled output z is less than a given scalar γ:

$$\|T_{zw}\| < \gamma. \tag{15}$$

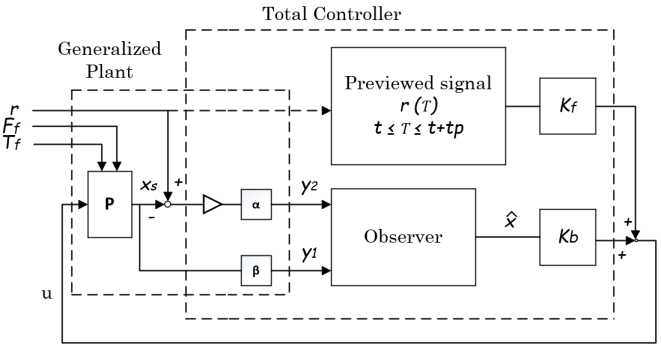


Fig. 3. Block diagram of generalized plant and preview controller.

Optimal H_∞ controllers are designed for the minimum value of γ and H_2 controllers are designed for $\gamma = \infty$. By analogy with Game Theory, the H_x control problem can be expressed as the minimization of the following cost function:

$$J = \int \left[z^T z - \gamma_2 w^T w\right] dt, \tag{16}$$

which can be expanded as:

$$J = \int_0^\infty \left[x^T C_1^T C_1 x + \begin{bmatrix} w^T u^T \end{bmatrix} \begin{bmatrix} -\gamma^2 & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} w \\ u \end{bmatrix}\right] dt, \tag{17}$$

where $R = D_{12}^T D_{12}$. The conventional approach to design H_∞ tracking systems has been to formulate a generalized plant which includes the reference signal as external disturbance and design a one or two-degrees-of-freedom controller which will achieve the desired specifications. The inherent shortcoming of this approach is the overdesign that is entitled when designing a controller independently of the reference signal to be tracked which is known a priori.

$$\hat{w} = [T_f, F_f, \eta_1, \eta_2, r]^T.$$

In other words, a controller is designed for the worst reference signal r_{worst} which is, for the most of cases, different to the desired reference signal r. To overcome the shortcoming of the classical H_∞ approach, this paper proposes H_∞ preview control where the known reference signal is used, as it is, for designing the controller. The design of the H_∞ preview controller has been divided in two parts: first, H_∞ preview control laws have been derived for the case of state feedback, and afterwards an observer was designed considering the conventional approach to design H_∞ output feedback controllers. The structure of the H_∞ preview controller is shown in Figure 3 where it can be noted the feedback part K_f and preview part K_p of the controller. H_∞ control laws have been derived for two cases: (1) the reference signal is known for the total working time and (2) the reference signal is previewed for a fixed interval of time shorter than the total working time.

A. Reference signal known for the total working time

When the reference signal is known for the total working time $[t, t+T]$, the H_∞ preview controller is designed considering the generalized plant of Equations 12 and 13 and the following cost function:

$$J = \int_t^{t+T} \left[z^T z - \gamma_2 w^T w\right] dt \tag{18}$$

Using the Hamiltonian approach of Optimal Control Theory [19], the preview control law which minimizes the cost function of Equation 18 subject to the constraint of Equation 12 is given by:

$$u(t) = -R^{-1}B_2^T \left[P(t)x(t) + q(t)\right], \tag{19}$$

where $P(t)$ is the solution of the following differential Riccati equation:

$$\begin{aligned} -\dot{P(\tau)} = P(\tau)A - P(\tau)(B_2R^{-1}B_2^T - \gamma^{-2}B_1B_1^T)P(\tau) + \\ + A^T P(t) + C_1^T C_1, \end{aligned} \tag{20}$$

with $t \leq r \leq t+T$ and $P(\tau)$ satisfying the terminal constraint:

$$P(t+T) = 0;$$

$q(t)$ in Equation 12 is the solution of the following differential equation:

$$\begin{aligned} -\dot{q(t)} = (A^T - P(\tau)(B_2 - R^{-1}B_2^T - \gamma^{-2}B_1B_1^T))q(\tau) + \\ + PB_r r(\tau), \end{aligned} \tag{21}$$

satisfying the terminal constraint:

$$q(t+T) = 0.$$

Since to compute $q(t)$ and $P(t)$, Equations 20 and 21 should be integrated backwards in time, the computational amount could turn cumbersome and impractical so that an approximated simplification easy to compute is necessary. Assuming the working time T is long enough so that $\dot{P}) = 0$ and $P = const$, $q(t)$ can be calculated from the following equation:

$$\begin{aligned} q(t) = exp\left[-(A_{cl}^T + M)t\right] \text{x} \\ \text{x} \int_t^{t+T} \left[exp\left[(A_{cl}^T + M)\tau\right] PB_r r(\tau)\right] d\tau, \end{aligned} \tag{22}$$

where

$$A_{cl} = A - B_2R^{-1}B_2^T P \tag{23}$$

and

$$M = \gamma^{-2}B_1B_1^T P. \tag{24}$$

A_{cl} in Equation 23 represents the state matrix of the closed-loop system for the case of only feedback control. It is important to note that the control law of Equation 19 has the structure of two-degrees-of-freedom controllers with feedback and feedforward (preview) parts so that preview control improves the tracking performance of the system without affecting its feedback characteristics (stability, disturbance attenuation, etc.).

B. Reference signal previewed for a fixed interval of time

In this case it is assumed the reference signal is known for a fixed interval of time T which is much shorter than the total working time assumed to be long enough. In this case, the cost function to be optimized is

$$J = \int_t^{\infty} \left[z^T z - \gamma^2 w^T w\right] d\tau, \tag{25}$$

which can be decomposed in two parts:

$$\begin{aligned} J = \int_t^{t+T} \left[z^T z - \gamma^2 w^T w\right] d\tau + \\ + \int_{t+T}^{\infty} \left[z^T z - \gamma^2 \hat{w}^T \hat{w}\right] d\tau, \end{aligned} \tag{26}$$

where

$$\begin{aligned} J_1 = \int_t^{t+T} \left[z^T z - \gamma^2 w^T w\right] d\tau \text{ and } J_2 = \\ = \int_{t+T}^{\infty} \left[z^T z - \gamma^2 \hat{w}^T \hat{w}\right] d\tau \end{aligned}$$

Since the reference signal r is only known for the interval $[t, t+T]$, it is not included as a component of the disturbance vector w for this interval (cost function J_1), but it is included in $w = \hat{w}$ for the interval $[t+T, \infty]$. Since in this interval r is not known it should be replaced by r_{worst} as in standard H_∞ tracking control. From Linear Optimal Control Theory it is well known that the minimum value of the cost function J_2 is

$$J_2 = x^T(t+T)\hat{P}x(t+T), \tag{27}$$

where $\hat{P}$ is the solution of the following algebraic Riccati equation

$$\begin{aligned} \hat{P}A - \hat{P}(B_2R^{-1}B_2^T - \gamma^{-2}(B_1B_1^T + B_rB_r^T))\hat{P} + \\ + A^T\hat{P} + C_1^T C_1 = 0 \end{aligned} \tag{28}$$

From Equations 26 and 27, the cost function J can be expressed as:

$$J = x(t+T)\hat{P}x(t+T) + \int_t^{t+T} \left[z^T z - \gamma^2 w^T w\right] dt \tag{29}$$

Since the cost functions of Equations 29 and 18 are similar with the only difference being the term $x(t+T)\hat{P}x(t+T)$ in Equation 29, the control law which minimizes the cost function J of Equation 29 subject to the constraint of Equation 12, is also given by Equation 19 with $P(t)$ calculated from Equation 20 and $q(t)$ calculated from Equation 21. The only difference is the terminal value of P which for this case is

$$P(t+T) = \hat{P},$$

where $\hat{P}$ is already known from the Riccati equation 28. Since the computation of $q(t)$ is cumbersome when P varies with time, an approximation for the easy computation of q is necessary. Similarly as in subsection IV-A, assuming that $\dot{P} = 0$ then $P = const$ and takes the value for time $t+T$, e.g., $P = \hat{P}$. With this simplification, the control signal u can be computed easily at every stage of control.

V. Results and Discussion

In order to evaluate the tracking performance of the H_∞ preview controller, the response of the system has been examined in the time and frequency domains for different types of reference signals. The response of the system with H_∞ preview control has been compared with the response for H_∞ feedback controller with 1.5 degrees-of-freedom (DOF) and H_2 preview controller [20]. The H_∞ preview controller has been designed for the case when the reference signal is previewed for a fixed interval of time shorter than the total working time which is assumed to be long enough (sub-section IV-B). The preview time is chosen to be 1 second since longer times do not yield significant performance improvements.

A. Frequency Response

In order to compute the frequency response of the system with preview control, the Laplace transform $Q(s)$ of the preview part $q(t)$ of the control law of Equation 19 is required to be known. Using Equation 22 and recalling the definition of Laplace transform, where $R(s)$ is the Laplace transform of the reference signal $r(t)$ and $Q(s)$ is calculated as:

$$Q(s) = [Q_1(s) - Q_2(s)]\, R(s), \tag{30}$$

$$Q_1(s) = exp\,[Ts]\left[sI + A_{cl}^T + M\right]^{-1} \times exp\left[(A_{cl}^T + M)T\right] PB_r, \tag{31}$$

$$Q_2(s) = \left[sI + A_{cl}^T + M\right]^{-1} PB_r. \tag{32}$$

Figure 4 shows the gain and phase of the frequency response of the tracking system for H_2 preview control (solid line) and 1.5DOF H_2 feedback control (broken line). It can be noted that although the gain of the response is higher at high frequencies for 1.5DOF H_2 feedback control, the phase of the response for H_2 preview control is zero for frequencies up to 10 Hz. Zero phase response is desired in tracking systems since it is always desirable that the system responds to the reference input without delays. Figure 5 compares the gain and phase of the frequency response for H_∞ preview control (solid line) and 1.5DOF H_∞ feedback control (broken line). Similarly as for H_2 control, the gain of the response is higher for 1.5DOF H_∞ feedback control but the phase of the response for H_∞ preview control is close to zero for frequencies up to 1 Hz and is positive for the medium frequency range up to 10 Hz. This positive phase characteristics indicate that the system with H_∞ preview control responds in advance to the reference input. Since in the medium frequency range the gain of the frequency response for H_∞ preview control is low, the controller tries to compensate this deficiency by responding in advance with positive phase. Figure 6 shows the gain and phase of the frequency response for H_∞ preview control (solid line) and H_2 preview control (broken line). It can be noted that the gain of the response is higher for H_∞ preview control than for H_2 preview control. The phase response shows the zero-phase characteristics of H_2 preview control and the positive characteristics of H_∞ preview control discussed before.

B. Time Response

The time response of the tracking system with H_∞ preview control has been analyzed for sinusoidal and step reference inputs.

a) **Sinusoidal Response**
Figure 7 shows the time response of the system for a sinusoidal reference input of 0.5 Hz. Figure 7 (a) corresponds to 1.5DOF H_∞ feedback control and Figure 7 (b) to H_∞ preview control. In Figure 7 (a) it is noted that although the amplitude of the response of the system (solid line) is almost the same as the reference input (broken line), the system responds with delay and is unable to track the reference input with zero error. In Figure 7 (b) it can be noted that the sinusoidal reference input and the response for H_∞ preview control practically overlap from the initial time. It is clear that the tracking error is almost zero for the total working time.

b) **Step Response**
It is usually said that preview control improves the tracking performance of the system especially in situations where the signal to be tracked varies with time, and no significant improvement can be achieved for step reference signals for situations when the control action starts just when the reference signal is applied. However and as it is shown in Figure 8, it has been found that although the step response for the system with 1.5DOF H_∞ feedback control (broken line) and H_∞ preview control (solid line) are almost the same, Figure 8 (a), the control signal u is different for both controllers, Figure 8 (b). Preview control demands lower control signals and therefore requires less control energy to achieve the same tracking performance than 1.5DOF H_∞ feedback control. Calculated results show that the control energy can be reduced even by 10% when using H_∞ preview control.

VI. Conclusions

This paper has presented a novel design method of positioning/tracking systems based on H_∞ preview control using the known future value of the reference input. Analysis of the frequency response shows that preview control improves the tracking performance by improving the phase response of the tracking system so that the system responds to the reference input without delay. The step response of the tracking system shows that H_∞ preview control requires less control energy than 1.5DOF H_∞ feedback control to achieve the same positioning performance.

Acknowledgment

The authors greatly appreciate the financial support provided by the institution VINNOVA – Swedish Governmental Agency for Innovation Systems through the ICT

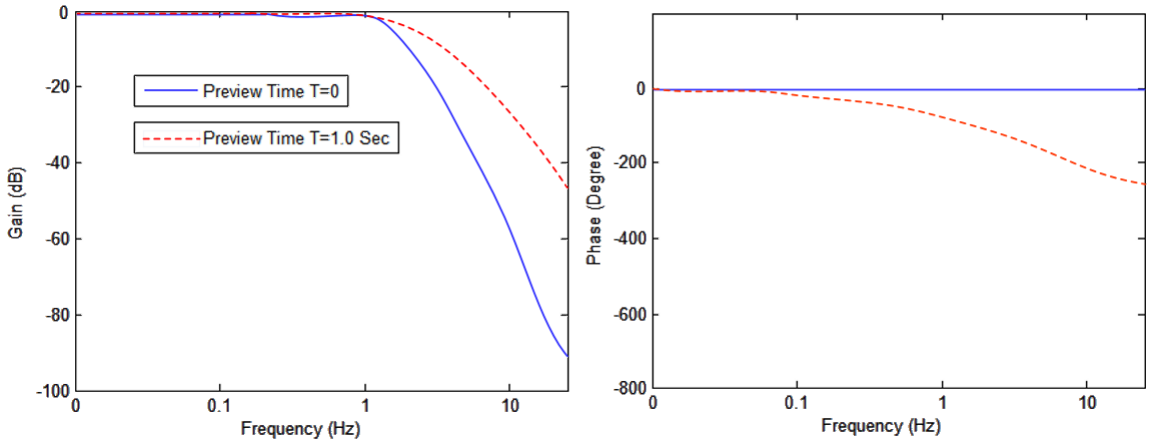


Fig. 4. Frequency response of H_2 preview control

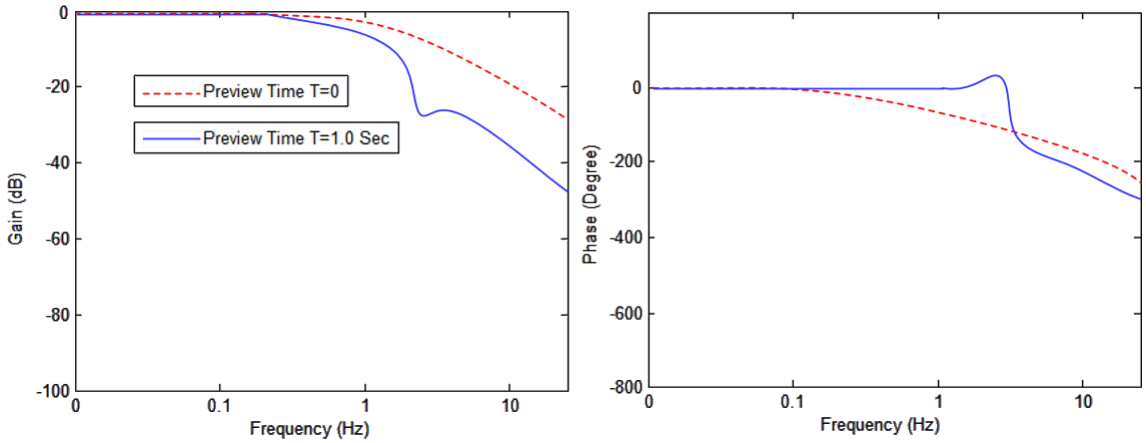


Fig. 5. Frequency response of H_∞ preview control

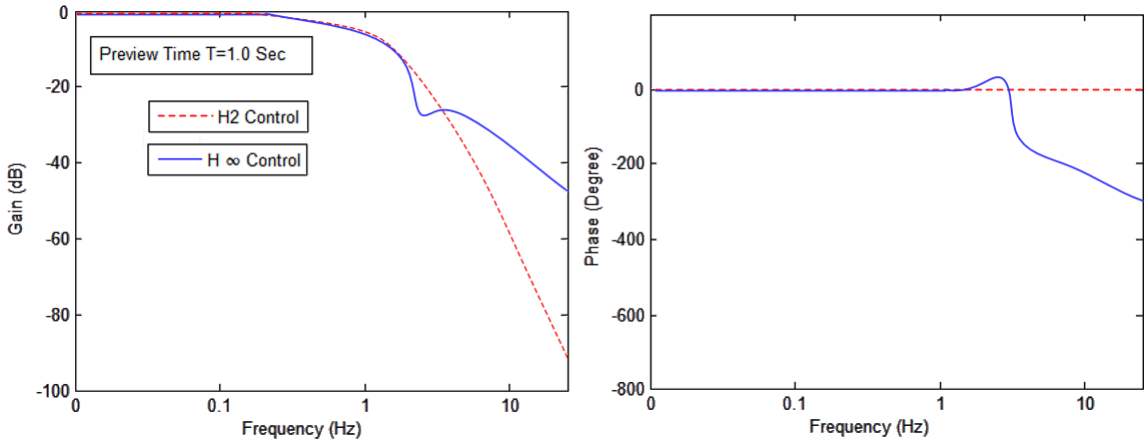


Fig. 6. Frequency response of H_2 and H_∞ preview control

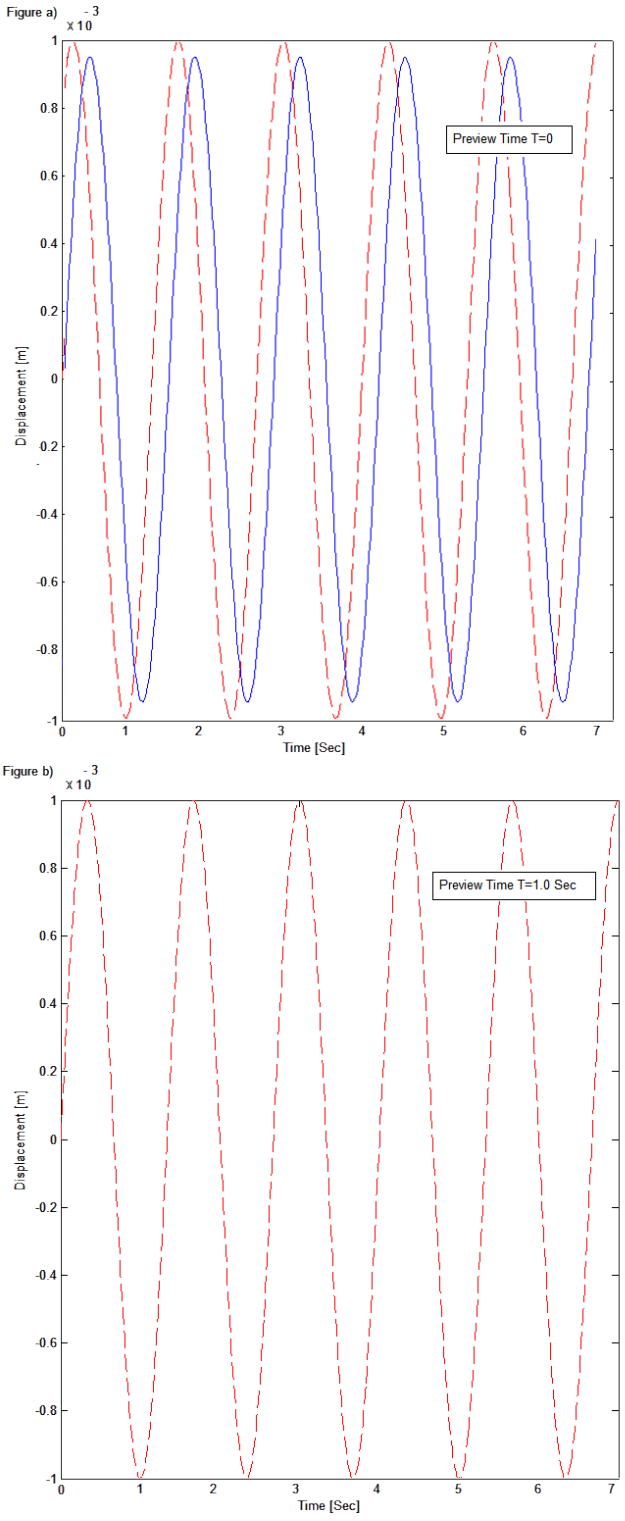


Fig. 7. Time response for sinusoidal input (0.5Hz)

project The Next Generation (TNG). We also grateful to Pontifical Catholic University (Perú) and Antioquia School of Engineering "EIA - GISMOC" (Colombia) in a joint effort for collaborative research.

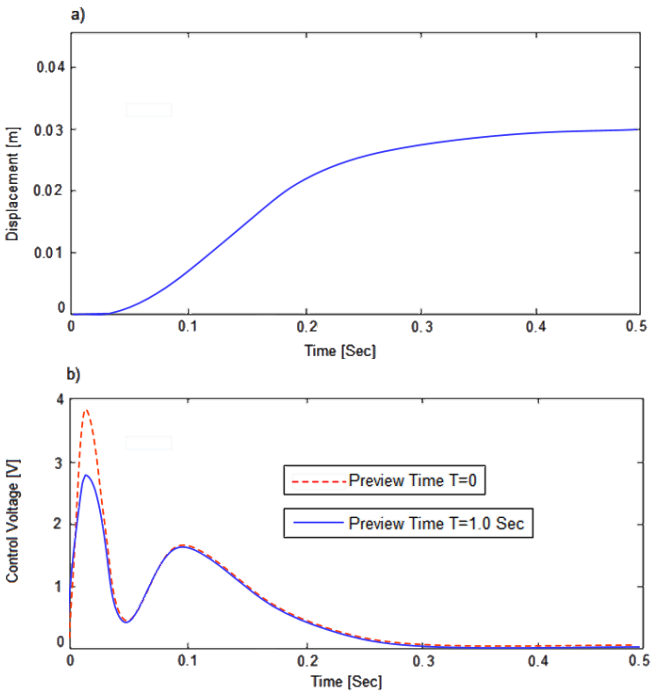


Fig. 8. Step response for H_∞ preview control

REFERENCES

[1] B. Chen, "Robust and perfect tracking of continuous-time systems," in *Robust and H∞ Control*, ser. Communications and Control Engineering. Springer London, 2000, pp. 215–248. [Online]. Available: http://dx.doi.org/10.1007/978-1-4471-3653-8_9

[2] G. Zames, "Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Transactions on Automatic Control*, vol. 26, no. 2, pp. 301–320, 1981.

[3] B. M. Chen, *Robust and H∞ control*. Springer, 2000.

[4] M. Tomizuka, "Optimal continuous finite preview problem," *IEEE Transactions on Automatic Control*, vol. 20, no. 3, pp. 362–365, Jun 1975.

[5] A. Hac, "Optimal linear preview control of active vehicle suspension," in *Proceedings of the 29th IEEE Conference on Decision and Control*, Dec 1990, pp. 2779–2784 vol.5.

[6] A. J. Hazell and D. J. N. Limebeer, "A framework for discrete-time H_2 preview control," *Journal of dynamic systems, measurement, and control*, vol. 132, no. 3, p. 031005, 2010.

[7] N. Birla and A. Swarup, "Optimal preview control: A review," *Optimal Control Applications and Methods*, pp. n/a–n/a, 2014. [Online]. Available: http://dx.doi.org/10.1002/oca.2106

[8] M. O. Cole and T. Wongratanaphisan, "Optimal {lq} feedforward tracking with preview: Practical design for rigid body motion control," *Control Engineering Practice*, vol. 26, no. 0, pp. 41 – 50, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0967066113002463

[9] I. Youn, R. Tchamna, S. Lee, N. Uddin, S. Lyu, and M. Tomizuka, "Preview suspension control for a full tracked vehicle," *International Journal of Automotive Technology*, vol. 15, no. 3, pp. 399–410, 2014. [Online]. Available: http://dx.doi.org/10.1007/s12239-014-0042-6

[10] C.-C. Luo, R.-F. Liu, C.-D. Yang, and Y.-H. Chang, "Helicopter H∞ control design with robust flying quality," *Aerospace science and technology*, vol. 7, no. 2, pp. 159–169, 2003.

[11] A. Kojima and S. Ishijima, "H∞ performance of preview control systems," *Automatica*, vol. 39, no. 4, pp. 693 – 701, 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005109802002868

[12] R. H. Middleton, J. Chen, and J. S. Freudenberg, "Tracking sensitivity and achievable H_∞ performance in preview control," *Automatica*, vol. 40, no. 8, pp. 1297–1306, Aug. 2004. [Online]. Available: http://dx.doi.org/10.1016/j.automatica.2004.02.019

[13] S. Ryu, Y. Kim, and Y. Park, "Robust H∞ preview control of an active suspension system with norm-bounded uncertainties," *International Journal of Automotive Technology*, vol. 9, no. 5, pp. 585–592, 2008.

[14] A. Hazell and D. J. N. Limebeer, "An efficient algorithm for discrete-time preview control," *Automatica*, vol. 44, no. 9, pp. 2441–2448, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0005109808001143

[15] A. Cohen and U. Shaked, "Linear discrete-time H_∞-optimal tracking with preview," *IEEE Transactions on Automatic Control*, vol. 42, no. 2, pp. 270–276, Feb 1997.

[16] G. Tadmor and L. Mirkin, "H∞ control and estimation with preview-part I: matrix ARE solutions in continuous time," *IEEE Transactions on Automatic Control*, vol. 50, no. 1, pp. 19–28, 2005.

[17] A. Kojima and S. Ishijima, "H∞ preview tracking in output feedback setting," *International Journal of Robust and Nonlinear Control*, vol. 14, no. 7, pp. 627–641, 2004. [Online]. Available: http://dx.doi.org/10.1002/rnc.897

[18] H. Wang, H. Zhang, and L. Yu, "H_∞ control for continuous-time systems with multiple preview channels," in *2013 32nd Chinese Control Conference (CCC)*, July 2013, pp. 2311–2315.

[19] H. Weinert, "A. E. Bryson and Y.-C. Ho, Applied Optimal Control, Optimization, Estimation, and Control. New York-London-Sydney-Toronto. John Wiley & Sons. 1975. 481 S." *ZAMM – Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik*, vol. 59, no. 8, pp. 402–402, 1979. [Online]. Available: http://dx.doi.org/10.1002/zamm.19790590826

[20] Y. Mikami, A. Moran, and M. Hayase, "Design of precise positioning system considering mechanical resonance and coulomb friction," in *IPEC'95*, Yokohama, Japan, 1995, pp. 1429–1434.

A More Efficient Representation of Obscuration for VRCC-3D+ Relations

Nathan Eloe, Chaman L. Sabharwal, and Jennifer L. Leopold

***Abstract*—VRCC-3D+ is an implementation of a region connection calculus that qualitatively determines the spatial relation between two 3D objects in terms of connectivity and obscuration. The eight connectivity relations are conceptually the same as RCC8, but calculated in 3D rather than 2D. The fifteen obscuration relations are calculated using the projection of the 3D objects on a particular 2D plane and the distance of the objects from the viewpoint. Herein we present a smaller, more precise set of VRCC-3D+ obscuration relations that retains the qualities of being jointly exhaustive and pairwise disjoint. However, this new set of relations overcomes two problems that existed in the previous set of fifteen relations: (1) lack of a precise mathematical definition for a key predicate, *InFront*, and (2) lack of an intuitive mapping of converse relations.**

***Index Terms*—Computer vision, qualitative spacial reasoning, VRCC-3D, region connection calculus, spatial relations.**

I. Introduction

QUALITATIVE spatial reasoning (QSR) in two dimensions is a well-studied field, and includes models such as the connectivity-based RCC systems [1], [2], [3], and obscuration-based systems such as LOS-14 [4], OCS-14 [5], and OCC [6]. These systems, while expressive, do not accurately portray the real world wherein objects exist and are perceived in three dimensions, not two. As computing power increases and the need to analyze three-dimensional data (e.g., stereoscopic video, robotic vision, etc.) increases, two-dimensional reasoning systems can be inefficient, or even inadequate, for sophisticated applications.

To ameliorate the shortcomings of two-dimensional QSR systems, Albath et al. developed RCC-3D [7], which eventually evolved into VRCC-3D+ [8]. VRCC-3D+ uses composite relations that express both connectivity and obscuration from a given perspective. The connectivity-based relations are the RCC8 relations (DC, EC, EQ, PO, TPP, TPPc, NTPP, NTPPc) defined in three dimensions; these relations have been an ongoing focus of optimization and refinement in the implementation as a QSR system [9]. The obscuration portion of the composite relations are refinements on the basic concepts of no obscuration (nObs), partial obscuration (pObs), equal obscuration (eObs), and complete obscuration (cObs). Over time these relations have been enhanced to improve their expressive power.

There are three criteria that the relations must meet to maintain the quality of the QSR system: the set of relations must be Jointly Exhaustive and Pairwise Disjoint (JEPD), every relation should map to exactly one converse relation, and the relations should have an intuitive mapping to natural language. If the relations are not jointly exhaustive, there will be physical configurations of objects that simply cannot be expressed by any relation. Relations that are not pairwise disjoint will result in ambiguous classification of object configurations. An intuitive mapping of the relations to natural language aids in human usefulness and usability of the system, and ensures that the expressive power of the system does not become needlessly complex; relations that cannot be differentiated in natural language typically do not add to the reasoning power of the system and overburden the computational complexity. Herein the authors focus on refining the obscuration terms of the composite VRCC-3D+ relations.

Manuscript received on August 6, 2014, accepted for publication on September 22, 2014, published on November 15, 2014.

The authors are with Missouri University of Science and Technology, Computer Science Department, Rolla, MO-65409, USA (e-mail: {nwe5g8, chaman, leopoldj}@mst.edu). Corresponding author: Chaman L. Sabharwal (e-mail: chaman@mst.edu).

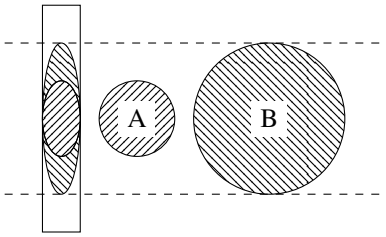
II. Background and Related Work

A. Occlusion

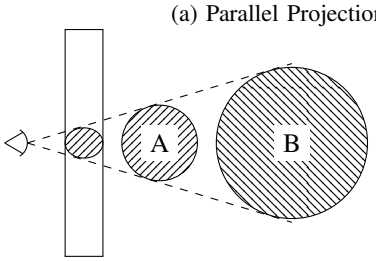
Occlusion of one object by another object is contextually dependent on the observer's location (usually called the view point, the perspective reference point, or the center of perspective projection) relative to the objects. It follows that the occlusion decision can be made from the projection on a view plane. QSR applications are interested in deriving spatial obscuration relations and classification from projection of 3D objects on a 2D projection plane.

There are two types of projections as shown in Fig. 1: parallel and perspective. Both have their advantages and disadvantages. The parallel projection is easier to compute, but loses the concept of depth. With the perspective projection, the object is scaled by the distance from the view point then projected; depth information is preserved. Obscuration predicates are based on two parameters: the perspective projection in a plane and depth (distance of the object from the perspective reference point).

The terms *in front*, *occulusion*, and *closer* are closely related. In natural language, the term *in front* between two objects A and B is synonymously interpreted as "A is in front of B", "A occludes B", and "A is closer than B".



(a) Parallel Projection



(b) Perspective Projection

Fig. 1. The differences between parallel and perspective projections. In both cases, object A obscures object B.

B. QSR and Occlusion

One of the best known obscuration-based QSR systems is LOS-14 [4] which was introduced by A.P. Galton in 1994. It classifies regions based on what can be seen in the Lines Of Sight (LOS) from a given perspective. Fourteen relationships are defined based on obscuration (or the lack thereof) from a given viewpoint.

Another occlusion-based calculus is ROC-20 [10]. It is similar to LOS-14, but extends it to add support for concave objects, which allows for mutual obscuration. Every spatial relationship in ROC-20 is defined in terms of the occlusion that is present and an RCC8 relationship. This system is significantly more expressive than LOS-14 and can apply to a greater number of cases, as it correctly handles concave regions.

The Occlusion Calculus (OCC) was introduced by Kohler in 2002 [6] and characterizes relationships between objects by their respective projections into an image plane. The author states that the information obtained is from one perspective, and as such, this system should be paired with other QSR methods to get a fuller picture. The system sacrifices expressiveness for reduced computational and reasoning complexity.

Guha et al. introduced OCS-14 in 2011 [5]. This model was designed to correct for insufficiencies in earlier occlusion models that made them infeasible for use in computer vision. Earlier methods had not accounted for whether the occluder was a moving object or part of the static background, and whether or not the visible part of an object was a connected blob or a fragment. As OCS-14 is designed for computer vision, feasibility of computation is a concern, but not expressive power.

C. RCC-3D

RCC-3D [7] was designed by Albath et al. to consider three dimensions, be computationally feasible, and give the most comprehensive spatial information about the system possible. Initially designed for use in analyzing the evolution of skeletal structures and other physical attributes, RCC-3D used the concepts of connectivity and obscuration to accomplish the design goals of completeness and computational feasibility. Because RCC-3D was to be used in visualizing physical changes over time, a GUI was deemed necessary. The resulting implementation was named VRCC-3D [11]. However, conceptual ambiguities that were uncovered in the implementation resulted in an evolution of the system, resulting in a revised model called VRCC-3D+ [8].

D. VRCC-3D+

Initially the obscuration portion of the VRCC-3D+ relationships simply were determined by overlapping boundaries and interiors of the projections of the objects in an image plane; the relations were limited to no obscuration (nObs), partial obscuration (pObs), complete obscuration (cObs), and equal obscuration (eObs). As the implementation of the system progressed, it became clear that a vital piece of information was missing; there was no concept of which object was obscuring the other. As such, an additional ternary predicate was added called *InFront*. For two objects A and B, possible values for $InFront(A, B)$ were: YES (A is in front of B), NO (B is in front of A), and E (A and B are equidistant).

The ternary *InFront* predicate was used to refine the concepts of nObs, pObs, eObs, and cObs to express whether an object obscured the other, whether an object was obscured by the other (thereby adding _c to the relation name), or whether they obscured each other (thereby adding _e to the relation name). Some of these relations had an ambiguous combination of predicate values. As such, some of the relations were split, expanding the total number of relations to 15, as shown in Table I.

TABLE I
THE 15 CURRENT VRCC-3D+ OBSCURATION RELATIONS.

	IntInt	IntExt	ExtInt	InFront
nObs	F	T	T	YES
nObs_c	F	T	T	NO
nObs_e	F	T	T	EQUAL
pObs1	T	T	T	YES
pObs2	T	F	T	YES
pObs_c1	T	T	T	NO
pObs_c2	T	T	F	NO
pObs_e	T	T	T	EQUAL
cObs	T	T	F	YES
cObs_c	T	F	T	NO
cObs_e1	T	T	F	EQUAL
cObs_e2	T	F	T	EQUAL
eObs_e	T	F	F	EQUAL
eObs_c	T	F	F	NO
eObs	T	F	F	YES

III. Improvements to the Relations

The first step in improving the obscuration relations is dealing with the mathematically imprecise predicate *InFront*. The ternary nature of this predicate and lack of rigorous mathematical definition led to different interpretations of the same scene by different entities. To replace this predicate, two new predicates are proposed: *Obscures* ($o(A, B)$) and *ObscuredBy* ($o_c(A, B)$). The *Obscures* predicate is defined as follows:

Let $f_O(x, y)$ be a function that maps the point (x, y) on the image plane back to the point (x', y', z') in object O that projects to the point (x, y) and is closest to the image plane. If no point in object O projects to point (x, y), then $f_O(x, y) = (\infty, \infty, \infty)$. Also, let C be the location of the camera in world coordinates. The *Obscures* predicate for objects A and B is defined in Eq. (1):

$$o(A,B) = \begin{cases} T: & \exists\, x, y\ s.t. |C - f_A(x,y)| < \\ & |C - f_B(x,y)| < \infty \\ F: & otherwise \end{cases} \quad (1)$$

In natural language, the meaning of this predicate is that it evaluates to true if there is a point at which the projections overlap and, within that projection, the first object hides some part of the second object. The definition of the converse relation $o_c(A, B)$ is simply $o_c(A, B) = o(B, A)$.

Note that the *Obscures* predicate only considers points at which the projection overlaps. This ameliorates cases such as that shown in Fig. 2. If we remove the condition that the distance between the camera and each of the two points be finite, object B would be reported to obscure object A at a point where the object A does not have a projection in the image plane.

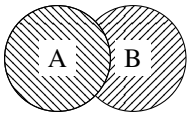


Fig. 2. Partial Obscuration: Object A obscures object B.

Table II shows the mapping of the original *InFront* predicate values to the values of the new *Obscures* and *ObscuredBy* predicates. Note that a value of EQUAL for the *InFront* predicate will map to either both o and o_c being true (T), or both being false (F).

The table of obscuration relations is rewritten as shown in Table III. The first simplification of the relation set follows directly from the predicate extension: when there is no obscuration between two objects, the projections do not overlap. As such o and o_c will always be false. It follows that nObs and nObs_c are impossible relations; only nObs_e is allowed by the predicate set, and only when the values of the new predicates are both false. This leads to the obscuration characterizations in Table IV.

TABLE II
Mapping of *InFront* to o, o_c

InFront	$o(A, B)$	$o_c(A, B)$
YES	T	F
NO	F	T
EQUAL	F(or T)	F(or T)

TABLE III
The 15 current VRCC-3D+ obscuration relations with o and o_c.

	IntInt	**IntExt**	**ExtInt**	o	o_c
nObs	F	T	T	T	F
nObs_c	F	T	T	F	T
nObs_e	F	T	T	F/T	F/T
pObs1	T	T	T	T	F
pObs2	T	F	T	T	F
pObs_c1	T	T	T	F	T
pObs_c2	T	T	F	F	T
pObs_e	T	T	T	F/T	F/T
cObs	T	T	F	T	F
cObs_c	T	F	T	F	T
cObs_e1	T	T	F	F/T	F/T
cObs_e2	T	F	T	F/T	F/T
eObs_e	T	F	F	F/T	F/T
eObs_c	T	F	F	F	T
eObs	T	F	F	T	F

TABLE IV
The reduced set of VRCC-3D+ obscurations (nObs and nObs_c removed).

	IntInt	**IntExt**	**ExtInt**	o	o_c
nObs_e	F	T	T	F	F
pObs1	T	T	T	T	F
pObs2	T	F	T	T	F
pObs_c1	T	T	T	F	T
pObs_c2	T	T	F	F	T
pObs_e	T	T	T	F/T	F/T
cObs	T	T	F	T	F
cObs_c	T	F	T	F	T
cObs_e1	T	T	F	F/T	F/T
cObs_e2	T	F	T	F/T	F/T
eObs_e	T	F	F	F/T	F/T
eObs_c	T	F	F	F	T
eObs	T	F	F	T	F

A. Handling Pathological Cases

Consider the projection shown in Fig. 3a. In this image, object B is partially obscuring object A, and A is partially obscuring object B. Under the previous set of 15 obscuration relations, the only way to express this would be cObs_e, when pObs_e is more intuitively correct. The reason for this is that the interior of the projection of object A does not intersect with the exterior of the projection of object B, resulting in an (intuitively incorrect) identification of cObs as the base obscuration type.

To address this issue, the structure of the relations themselves are examined. Currently, relations have a base of either nObs, pObs, eObs, or cObs. Appended to this base are the refinements of converse (_c) and equality (_e). To simplify this definition, a consistent structure is proposed: obscurations will have the form *x*Obs*y*, where *x* correlates to the *extent* of obscuration (Table V) and *y* corresponds to refinements on the obscuration (Table VI). To clarify the meaning of mutual

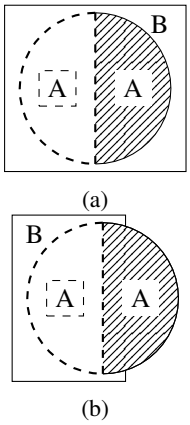


Fig. 3. Two examples of Mutual Obscuration. (a) A's projection completely contained in B's projection, (b) A's projection partially overlaps B's projection.

obscuration, a new suffix is introduced: _m.

A Cartesian product of the prefixes and suffixes show that there are 16 possible obscuration relations. However, it has already been shown that there can only be a single version of nObs; there is not a way to map suffixes directly to prefixes. As such, each relation must be individually handled.

TABLE V
PREFIX AND EXTENT OF OBSCURATION.

Prefix (*x*)	Meaning
n	No Obscuration
p	Partial Obscuration
e	Equal Obscuration
c	Complete Obscuration

TABLE VI
SUFFIX AND OBSCURATION REFINEMENT, WITH MAPPING TO o AND o_c.

Sufffix (y)	**Meaning**	$o(A,B)$	$o_c(A,B)$
[none]	Obscures	T	F
_c	Is Obscured By	F	T
_e	Equally Obscure Each Other	F	F
_m (new)	Mutually Obscure Each Other	T	T

1) Partial Obscuration (pObs): For partial obscuration, all cases where both objects are visible must be considered. The definitions of pObs and pObs_c remain unchanged. The characterization of pObs_e does not change, and has values of false for both o and o_c.The mutual refinement for pObs must handle the case shown in Fig. 3a as well as that shown in Fig. 3b. The case shown in Fig. 3b is straightforward. Figure 3a is more complicated: it must handle when object A is either object. As such, it maps to two characterizations of pObs_m. Table VIIa shows the new characterizations for all pObs relations.

2) Equal Obscuration (eObs): Equal obscuration, by definition, occurs when the size and shape of the projections are identical; the values of the IntInt, IntExt, and ExtInt predicates will always be T, F, and F, respectively. The eObs_e obscuration should only occur if two objects are identical. Mutual equal obscuration can occur (Fig. 4), so that case must be handled. Table VIIb shows the characterizations of the new eObs relations.

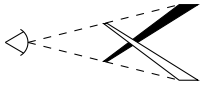


Fig. 4. Equal Mutual Obscuration

3) Complete Obscuration (cObs): By definition, complete obscuration means that one object cannot be seen. As such, there is no cObs_m relation. Table VIIc shows the cObs characterizations.

B. Identification of Converse Obscurations

One of the problems with the old set of VRCC-3D+ obscuration relations was that there were cases where there was no consistent intuitive mapping from a relation to its converse. For example, nObs_e, pObs_e, and eObs_e map to themselves as converses, which is logical; if A and B obscure each other, then B and A should also obscure each other. However, there was not a single cObs_e relation but two. These two relations were the converse of each other. This inconsistency hindered both the implementation and reasoning with the system; it muddled the meaning of the _e suffix. Under this new relation set, every normal (no suffix) obscuration's converse relation is the converse obscuration (named with the _c suffix). The mutual (_m) and equal (_e) relations map to themselves as converse. Table VIII shows the full set of obscurations and their identified converse relation.

TABLE VIII
FULL OBSCURATION RELATION SET WITH IDENTIFIED CONVERSE RELATIONS.

	IntInt	**IntExt**	**ExtInt**	o	o_c	**Converse**
nObs_e	F	T	T	F	F	nObs_e
pObs	T	T F	T	T	F	pObs_c
pObs_c	T	T	T F	F	T	pObs
pObs_e	T	T	T	F	F	pObs_e
pObs_m	T	T F T	F T T	T	T	pObs_m
eObs	T	F	F	T	F	eObs_c
eObs_c	T	F	F	F	T	eObs_c
eObs_e	T	F	F	F	F	**eObs_e**
eObs_m	T	F	F	T	T	eObs_m
cObs	T	T	F	T	F	cObs_c
cObs_c	T	F	T	F	T	cObs
cObs_e	T	T F	F T	F	F	cObs_e

IV. EFFECT ON VRCC-3D+

The new set of obscurations directly benefits the implementation of VRCC-3D+ on two fronts: it is easier to verify that the implementation of the predicates is correct (leading to more correct results), and the reduced

TABLE VII
THE NEW VRCC-3D+ OBSCURATION CHARACTERIZATIONS.

(a) New pObs characterizations.

	IntInt	IntExt	ExtInt	o	o_c
pObs	T	T F	T	T	F
pObs_c	T	T	T F	F	T
pObs_e	T	T	T	F	F
pObs_m	T	T F T	F T T	T	T

(b) New eObs characterizations.

	IntInt	IntExt	ExtInt	o	o_c
eObs	T	F	F	T	F
eObs_c	T	F	F	F	T
eObs_e	T	F	F	F	F
eObs_m	T	F	F	T	T

(c) New cObs characterizations.

	IntInt	IntExt	ExtInt	o	o_c
cObs	T	T	F	T	F
cObs_c	T	F	T	F	T
cObs_e	T	T F	F T	F	F

set of obscurations and direct mapping to a converse relation improves the computational complexity of relation determination.

Migration from using the ternary *InFront* to two binary predicates immediately improved the unit test pass rate for obscuration relations from 29% to 77%. The expected output for the unit tests was determined by visual inspection; moving to the more mathematically precise predicates caused the implementation to more closely emulate human perception; several of the common errors reported stem from the perspective point being different for the person analyzing the file and the implementation, and more are due to floating point rounding errors. The majority of the errors that were fixed stemmed from an incorrect suffix on the base obscuration. Also, the use of two binary predicates instead of a single ternary predicate makes it trivial to implement a decision tree predicate picker similar to that presented in [12], which has been shown to improve the speed of computation.

The primary improvement in computational complexity is due to the unique mapping of a relation to a converse relation. In order to fully describe a scene containing objects A and B using VRCC-3D+, the system must compute the relationship between objects A and B, but also between B and A (the converse relation). To calculate an RCC8 relation in three dimensions, the computational complexity is in the worst case $O(f_a \times f_b)$, where f_a and f_b are the number of faces in objects A and B, respectively.

Generating the projections for objects A and B is $O(f_a)$ and $O(f_b)$, respectively; calculating the values of the predicate values is dependent on the (non-constant) complexity of performing intersection operations on the projections and the cost of ray casting to determine which object is closer to the view point. Regardless of the implementation of these operations, the cost of computing the values of the predicates is more than linear. In contrast to this, if the relationship between objects A and B is known, and both parts of the composite relation have well-defined converses, the relationship between objects B and A can be determined to be the converse of each part of the composite relation: a lookup operation with complexity $O(1)$.

Every QSR system is designed as a balance of three criteria: ease of reasoning with the system, computational complexity, and expressive power of the system. An improvement to one aspect of the system comes at the cost of another. RCC-3D (the system that over time became VRCC-3D+) was initially designed to balance the three: by using compound relations, high expressive power and low computational complexity could be obtained without sacrificing too much in the way of ease of reasoning with the system. Herein, by reducing the 15 obscuration relations to 12 and introducing the concept of mutual obscuration as a refinement, both computational complexity and ease of reasoning have been improved.

Table IX shows the new set of 34 composite relations present in VRCC-3D+ (a reduction from the 46 relations stemming from the old set of 15 obscuration relations), and also serves to illustrate the importance of working to minimize the number of obscuration relations; if the four obscuration relations with multiple characterizations were expanded to 9 separate obscurations, the number of VRCC-3D+ relations would grow to 50 relations. Usage of a QSR system becomes increasingly more complex as the set of relations in the system increases.

An illustrative example of this is the composition table for the system which increases the speed of classifying relations by reducing the number of possible relations between two objects; in a scene with three objects A, B, and C, if the relationships between A and B, and B and C are known, the composition table reduces the set of relationships that are possible between objects A and C. This can be used in conjunction with a decision tree [12] to speed up computation. Calculation of this table has a non-constant polynomial complexity in the number of relationships.

TABLE IX
MAPPING OF RCC8 RELATIONS TO OBSCURATION RELATIONS

	nObs_e	pObs	pObs_c	pObs_e	pObs_m	eObs	eObs_c	eObs_e	eObs_m	cObs	cObs_c	cObs_e
DC	X	X	X		X	X	X			X	X	
EC	X	X	X		X	X	X			X	X	
PO		X	X	X	X	X	X		X	X	X	
EQ								X				
TPP							X				X	X
TPPc						X				X		X
NTPP											X	
NTPPc										X		

V. FUTURE WORK

Future work will focus on mapping the revised VRCC-3D+ relations to natural-language terms suitable for end-user applications involving spatial querying. Preliminary efforts in this direction have commenced for the VRCC-3D+ connectivity relations [13]. Given the ambiguity of natural-language terms such as *in front/behind*, *occludes*, and *closer/nearer*, it may prove difficult to find unambiguous mappings for the mathematically precise VRCC-3D+ obscuration relations. Extensive human experiment studies will need to be conducted, and likely domain-specific ontologies will have to be developed for the relation-to-term associations.

VI. CONCLUSIONS

In this paper, simplifications to the 15 obscuration relations present in VRCC-3D+ have been analyzed and presented. This change in the mathematical set of relations improved the computational correctness from 27% to 77%. The VRCC-3D+ obscuration relations are now easier to understand and computationally easier to implement because of the introduction of a new predicate for classification and a new class of obscuration.

REFERENCES

[1] B. Bennett, "Spatial reasoning with propositional logics," in *Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning*, ser. KR'94, J. Doyle, E. Sandewall, and P. Torasso, Eds. San Francisco, CA.: Morgan Kaufmann, 1994.

[2] J. Ouyang, Q. Fu, and D. Liu, "A model for representing topological relations between simple concave regions," in *Proceedings of the 7th International Conference on Computational Science*, ser. ICCS'07, vol. 1. Berlin, Heidelberg: Springer-Verlag, 2007, pp. 160–167.

[3] D. A. Randell, Z. Cui, and A. Cohn, "A spatial logic based on regions and connection," in *Proceedings of the 3rd Conference on Principles of Knowledge Representation and Reasoning*, ser. KR'92, B. Nebel, C. Rich, and W. Swartout, Eds. San Mateo, California: Morgan Kaufmann, 1992, pp. 165–176. [Online]. Available: citeseer.ist.psu.edu/randell92spatial.html

[4] A. Galton, "Lines of sight," in *AISB Workshop on Spatial and Spatio-Temporal Reasoning*, 1994.

[5] P. Guha, A. Mukerjee, and K. S. Venkatesh, "OCS-14: You can get occluded in fourteen ways," in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence*, ser. IJCAI'11, T. Walsh, Ed., 2011, pp. 1665–1670.

[6] C. Kohler, "The occlusion calculus," in *Cognitive Vision Workshop*, ser. ICVW'02, 2002.

[7] J. Albath, J. L. Leopold, C. L. Sabharwal, and A. M. Maglia, "RCC-3D: Qualitative spatial reasoning in 3D," in *Proceedings of the 23rd International Conference on Computer Applications in Industry and Engineering*, ser. CAINE'10, 2010, pp. 74–79.

[8] C. Sabharwal, J. Leopold, and N. Eloe, "A more expressive 3D region connection calculus," in *Proceedings of the 2011 International Workshop on Visual Languages and Computing (in conjunction with the 17th International Conference on Distributed Multimedia Systems)*, ser. DMS'11, 2011, pp. 307–311.

[9] N. Eloe, J. L. Leopold, and C. L. Sabharwal, "Efficient computation of object boundary intersection and error tolerance in VRCC-3D+," in *Proceedings of the 18th International Conference on Distributed Multimedia Systems*, ser. DMS'12, 2012, pp. 67–70.

[10] D. Randell, M. Witkowski, and M. Shanahan, "From images to bodies: Modelling and exploiting spatial occlusion and motion parallax," in *Proceedings of the 17th International Joint Conference on Artificial Intelligence*, ser. IJCAI'01, vol. 1. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 57–63.

[11] J. Albath, J. L. Leopold, and C. L. Sabharwal, "Visualization of spatio-temporal reasoning over 3D images," in *Proceedings of the 2010 International Workshop on Visual Languages and Computing (in conjunction with the 16th International Conference on Distributed Multimedia Systems)*, ser. DMS'10, 2010, pp. 277–282.

[12] N. Eloe, J. L. Leopold, C. L. Sabharwal, and D. McGeehan, "Efficient determination of spatial relations using composition tables and decision trees," in *Proceedings of the IEEE Symposium on Computational Intelligence for Multimedia, Signal, and Vision Processing*, ser. CIMSIVP'13, 2013, pp. 1–7.

[13] J. L. Leopold, C. L. Sabharwal, and K. Ward, "Spatial relations between 3D objects: The association between natural language, topology, and metrics," in *Proceedings of the 2014 International Workshop on Visual Languages and Computing*, ser. VLC'14, 2014, (To appear).

Modelo computacional del diálogo basado en reglas aplicado a un robot guía móvil

Grigori Sidorov, Irina Kobozeva, Anton Zimmerling, Liliana Chanona-Hernández, Olga Kolesnikova

***Resumen*—En este artículo se presenta la descripción formal detallada del módulo de control del diálogo para un robot móvil que funciona como guía (en un museo). El módulo incluye el modelo proposicional del diálogo, la especificación de los actos de habla y los bloques del habla, así como el inventario de los patrones de habla correspondientes a todos los actos de habla en el modelo. El modelo del diálogo se implementa como una red de estados y transiciones entre estados condicionados por reglas que estipulan los factores verbales y visuales. La arquitectura del módulo no depende de un idioma particular y puede ser adaptado a cualquier lenguaje natural.**

***Palabras clave*—Modelo de diálogo, reglas, robot guía móvil, actos de habla, bloques de habla.**

Computational Model of Dialog Based on Rules Applied to a Robotic Mobile Guide

***Abstract*—This paper presents a formal detailed description of the dialogue management module for a mobile robot functioning as a guide. The module includes a propositional dialogue model, specification of speech acts and speech blocks as well as the inventory of speech patterns corresponding to all speech acts of the model. The dialogue model is implemented as a network of states and transitions between states conditioned by rules, which include verbal and visual factors. The architecture of the module is language independent and can be adapted to any natural language.**

***Index Terms*—Model of dialog, rules, mobile robotic guide, speech acts, speech blocks**.

Manuscrito recibido 6 de julio de 2014, aceptado para su publicación 3 de noviembre de 2014, publicado 15 de noviembre 2014.

Grigori Sidorov (autor correspondiente) trabaja en el Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), México, DF, México (correo electrónico: sidorov@cic.ipn.mx, página web: www.cic.ipn.mx/~sidorov).

Irina Kobozeva es catedrática de la Universidad Estatal Lomonosov de Moscú, facultad de filología, Departamento de Lingüística Teórica y Aplicada, Moscú, Rusia.

Anton Zimmerling trabaja en la Universidad Estatal para Humanidades Sholokhov de Moscú, siendo jefe del Laboratorio de Lingüística General y Computacional, Moscú, Rusia.

Liliana Chanona-Hernández trabaja en el Instituto Politécnico Nacional (IPN), Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME), México D.F., México.

Olga Kolesnikova trabaja en el Instituto Politécnico Nacional (IPN), Escuela Superior de Cómputo (ESCOM), México D.F., México.

I. Introducción

La posibilidad de manejar un robot móvil dándole comandos directamente por voz humana es muy atractiva, ya que es mucho menos natural teclear los comandos. Aún mejor se percibe la posibilidad de dialogar con este dispositivo móvil, cuando el robot no solamente entiende un comando, sino también puede contestar dentro de su ámbito de competencia, porque tal actividad se asemeja mucho a la interacción con un ser humano. Sin embargo, el modelado de diálogo es un problema muy difícil y a pesar de ser investigado por más de cincuenta años sigue siendo un problema abierto.

La dificultad de manejo de diálogo se relaciona con la información multimodal que el robot tiene que "entender", tomando en cuenta las diferencias fundamentales entre los lenguajes de representación del espacio en la "mente" del robot y en la mente humana. También a eso se anuda el problema de comprensión del lenguaje natural a un nivel tan profundo que permita interpretar correctamente las intenciones del interlocutor [10, 11, 13].

Durante los últimos veinte años se han realizado muchos trabajos de modelado de la comunicación humana con robots móviles, pero por lo general es de carácter exploratorio. Normalmente se analiza en profundidad alguno de los aspectos particulares de este problema bastante complejo, tales como la capacidad de influir en el éxito de la comunicación con gestos, la importancia de la dirección de la mirada y el conocimiento de la situación para la interacción efectiva [6, 25]. Por otro lado, cabe mencionar que existen algunos desarrollos de robots móviles que ya pueden realizar ciertas funciones y entablar un diálogo con el usuario en un nivel aceptable de naturalidad en unas áreas muy específicas, por ejemplo, [1, 2, 7, 17, 19]. Sin embargo, no existe un modelo estándar para el modelado de diálogo e interpretación de los actos de habla.

II. Enfoque algorítmico vs. aprendizaje automático por computadora

En este artículo se presenta un enfoque algorítmico para el problema del modelado de diálogo en lenguaje natural tomando como ejemplo un diálogo guiado por un robot móvil. Cabe señalar que en la lingüística computacional moderna, cuando hablan del enfoque algorítmico en general, a menudo se utiliza la metodología basada en el aprendizaje automático por computadoras. Por ejemplo, en el caso del diálogo, un método basado en el aprendizaje automático fue aplicado al

problema central de su procesamiento: detección de los actos de habla, como se describe en [21]. Sin embargo, allá no se presentaron las evidencias convincentes de que este aprendizaje permite una buena detección de los actos de habla en un diálogo.

Cabe mencionar que los actos de habla constituyen claramente un buen material para el uso de este tipo de aprendizaje automático, pero el problema de utilizar las computadoras para esa tarea se dificulta dado las características de los actos de habla en un diálogo, que tienen la naturaleza semántica y pragmática, y en su etapa actual del desarrollo de la teoría de procesamiento de diálogo y de texto en general es muy difícil detectarlas confiablemente. Entonces, para el manejo computacional del diálogo normalmente se usan los enfoques y modelos algorítmicos más tradicionales, y no los métodos basados en el aprendizaje automático.

En este artículo nos basamos en un enfoque algorítmico para la descripción del módulo de manejo de diálogo que se usa para interactuar con un robot guía móvil siguiendo el enfoque general basado en reglas descrito en [15].

III. Módulos para implementación de un guía robótico autónomo

El esquema general de procesamiento de diálogo incluye dos módulos lingüísticos auxiliares: el módulo de reconocimiento de voz (en nuestro caso hemos usado las librerías de *Dragon Naturally Speaking*) y el módulo de análisis lingüístico (de niveles morfológico y sintáctico) utilizando el sistema Freeling [9]. En nuestro caso, el sistema que maneja el diálogo está siendo desarrollado para el idioma español, sin embargo, al cambiar los módulos lingüísticos de análisis y los módulos de procesamiento de voz, el sistema puede funcionar para cualquier idioma, es decir, el conjunto de actos de habla es universal: independiente de un idioma específico.

Ya que estamos hablando del diálogo con un robot, se necesita tener un robot físicamente. La "mente" del robot (donde se procesa toda la información) es nada más que una computadora estándar. En este sentido, todos los procedimientos consisten en el desarrollo y aplicación de programas de computadoras, siguiendo la metodología tradicional de desarrollo de software. Si usamos un robot móvil, es cómodo usar una computadora portátil (y no de escritorio), la cual puede ubicarse directamente sobre el robot.

En nuestro caso, el "cuerpo" del robot es un robot móvil Pioneer 3DX. El robot sabe navegar en un espacio conocido de un punto a otro evadiendo obstáculos que se encuentran en su camino. Es un robot autónomo, es decir, nadie lo está guiando paso por paso, sino el robot calcula su ruta por sí mismo. El robot está equipado con un sensor láser de gran precisión para medir distancias hasta posibles obstáculos en su camino: esta manera de conocer los posibles obstáculos sustituye la visión de los humanos. Digamos, de manera similar se mueven los murciélagos, usando principalmente su sonar biológico. Note que este tipo de robot es el estándar de facto para el desarrollo de aplicaciones de los robots móviles autónomos en el mundo real —sus dimensiones son comparables a los de un ser humano.

El robot se mueve utilizando un mapa de su entorno preparado con anterioridad. La preparación del mapa es un procedimiento muy sencillo y rápido, simplemente el usuario debe mover, en este caso manualmente, al robot en el espacio del mapa y después un programa recalcula el mapa correspondiente. En el mapa se puede marcar las áreas prohibidas, es decir, lugares donde el robot no debe pasar. También se puede marcar algunos puntos específicos como metas provisionales. Por ejemplo, en este artículo consideramos, dado que estaremos describiendo un posible escenario de uso del robot como guía de un museo, por decir algo, las obras maestras del museo de Louvre (Francia): la Mona Lisa (Mona_Lisa, M_L), la Venus de Milo (Venus_de_Milo, V_M), el Escriba Sentado (El_Escriba, E_S). Más adelante, estos nombres se utilizarán en nuestra descripción del modelo de diálogo, pero en cada aplicación particular de este modelo, deben ser sustituidos por otros objetos materiales que se encuentren en el entorno particular.

El robot se mueve con mucha precisión en el espacio representado en el mapa, dado que se conoce su propia posición en cada momento. Más aún, si aparece un obstáculo nuevo (ausente en el mapa), el robot puede entender esa situación y procesarla correctamente, es decir, evadir este obstáculo si se le obstaculiza su camino. Normalmente los obstáculos son los humanos que van y vienen, o los muebles que cambiaron de lugar.

En una situación muy particular, cuando los obstáculos (humanos), tapan todo el espacio disponible para el robot, el robot puede perderse. Cuando se despejara el ambiente, un módulo de localización tiene que volver a calcular su posición.

La idea de conocer su posición y la posición de otros objetos (posibles metas en el desplazamiento) en el mapa, permite al robot saber cuándo cada objeto aparece en su campo de "visión". Esto sucede si no hay ningún obstáculo entre el robot y el objeto (o la posición conocida del objeto) y la distancia entre ellos es la suficientemente pequeña. Proponemos considerar el umbral de 3 metros para considerar un objeto como "visible", es decir, dentro del campo de "visión" del robot. En nuestro modelo, cuando un objeto X se encuentra en el "campo de visión" (recordemos que no hay cámaras en el robot, sino se usa el sensor laser), eso se refleja como la proposición Visual_Act(X), donde X es la variable para un objeto.

Para el desarrollo de aplicaciones del robot Pioneer 3DX, se usa la arquitectura Cliente-Servidor. Es decir, existe un servidor que procesa e interpreta los comandos de un cliente y los transforma a los comandos de bajo nivel para el robot, y existe un programa cliente que manda al servidor los comandos de alto nivel (tipo "ejecuta el bloque 1" o "desplázate hacia Venus_de_Milo"). Note que el programa cliente puede correr en la misma computadora que el servidor

o en alguna otra computadora. En este caso el cliente se comunica con el servidor, por ejemplo, a través del protocolo TCP-IP.

IV. Escenarios de uso de un robot móvil

Entre los escenarios de uso de un robot móvil (es decir, las funciones que puede ejercer un robot móvil) de manera muy general se puede mencionar los siguientes:

1) La función de un guía (por ejemplo, un guía en un museo, empresa o exposición). Es el escenario que estamos considerando en este artículo.
2) La función de un mecanismo transportador, como, por ejemplo, un usuario en silla de ruedas que está controlada con comandos de voz.
3) La función de un mensajero, para pasar un objeto a cierto lugar de manera autónoma o siguiendo comandos enviados por el usuario. En este caso hay dos posibilidades:
   a) el robot "conoce" el camino;
   b) el robot no conoce el camino y para pasar debe comunicarse por el acceso remoto con el usuario [2].
4) La función de un reportero que informa de lo que percibe en cada momento.
5) La función de un discípulo que está capacitado para comprender su entorno (véase [4, 5, 20]).
6) La función de un camarero en un restaurante.
7) La función de un ayudante en selección de objetos, por ejemplo, ayudar al comprador en una tienda.
8) La función de un manipulador que ejecuta los comandos del usuario para mover los objetos (por ejemplo, el primer robot manipulador de T. Winograd, SHRDLU).

Por supuesto, el robot puede configurarse para ejecutar una combinación de esas funciones. A menudo se combina las funciones de un mensajero con la función de un manipulador: el robot debe moverse hacia el usuario en la posición adecuada y hacer algo con unos objetos especificados, por ejemplo, un mesero manipularía los platillos. Las funciones del estudiante y periodista pueden combinarse entre sí y con la función de un mensajero, como por ejemplo, en [18], donde el robot navega recibiendo los comandos del usuario e informa qué él percibe y además aprende los nombres de los objetos percibidos, preguntándolo al usuario.

Cada función impone ciertos requisitos para el diálogo y un repertorio de actos comunicativos (actos de habla: réplicas y/o reacciones no verbales). Dependiendo de las características del diálogo, el iniciador y el controlador de desarrollo de un diálogo puede ser tanto el usuario como el robot. Las funciones del robot a veces lo obligan ser el iniciador y el controlador que ejecuta todo el diálogo, pero en otros casos el robot solo inicia el diálogo y es el usuario quien lo controla posteriormente. También ambos comunicantes (el robot y el usuario) pueden asumir la función de iniciar el diálogo.

Otros aspectos del diálogo que están fuera del alcance de este artículo son: el repertorio más amplio de los actos comunicativos que el robot debe producir y entender, el módulo conceptual intencional (información enciclopédica y modelos de comportamiento racional). Es decir, en nuestro caso, el módulo de diálogo está desarrollado para el sistema que no tiene el módulo de "conocimiento amplio" y por lo tanto modela solamente una función relativamente simple del robot guía, cuando el robot asume el control en el transcurso de diálogo. En este caso, se reduce el repertorio de los actos comunicativos necesarios (*speech acts*, SA).

V. Particularidades de diálogo de un robot guía con visitantes

El término de robot guía en este caso, se refiere no sólo a una determinada ocupación (profesión), sino a cualquier agente que en determinadas situaciones da a conocer a los visitantes un lugar concreto y realiza una visita guiada en el lugar y proporciona la información sobre aquellos objetos que se encuentran en este lugar. Ese lugar puede ser alguna localidad (por ejemplo, de la ciudad), la propiedad de una persona (por ejemplo, una casa) o el interior de un edificio o parte del mismo (por ejemplo, un museo, un piso de alguna escuela, o departamentos de las instituciones). En su caso, el guía puede ser el dueño de la casa, un agente inmobiliario, un empleado de una institución, un residente local, y en general cualquier persona que se compromete a ejecutar la función social indicada: dar a conocer.

En todo caso, la ejecución de funciones un guía comienza cuando el agente acepta este papel, que a su vez implica que la contraparte acepta la función de un "visitante de lugares desconocidos" y tiene la intención correspondiente. En el caso de un robot, es el ser humano el responsable de la conveniencia de su aceptación en el papel del guía, dado que es la persona que inicia el proceso en el momento adecuado, es decir, una persona puede, por ejemplo, llamar al programa apropiado y el robot comienza a ejecutar el diálogo.

El objetivo general de un guía —dar a conocer el lugar a los visitantes— se divide en una serie de subtareas, cada una de las cuales es traerlos al objeto específico (en el sentido más amplio) de interés para ellos y, posiblemente, contar una determinada cantidad de información sobre este objeto. Durante la ejecución de cada una de las subtareas, el guía puede, además, llamar atención a lo largo de la ruta de acceso al siguiente objetivo a los objetos "secundarios", por ejemplo, en un museo durante el camino hacia un obra de arte, mencionar donde está la cafetería.

A su vez, la ejecución de cada tarea implica la ejecución de actos de comunicación (verbales y/o no verbales) y el desplazamiento al objeto de interés preestablecido por el camino conocido de antemano para el guía.

VI. Los actos de habla y los bloques del habla

En el caso del diálogo del robot guía con el visitante vamos a distinguir los actos de habla y los bloques del habla. En

nuestro caso, el modelo se desarrolla para el diálogo de un robot guía en un museo y el repertorio de las acciones del robot incluye tres tipos de actos: los actos de habla, los bloques del habla y los bloques multimodales.

A. *Actos de habla*

Los actos de habla entendemos de la manera estándar como en la teoría de los actos de habla: como actos de pronunciar una expresión lingüística, generalmente simple, que tienen una cierta fuerza ilocutiva, es decir, expresa (directa o indirectamente) la intensión del hablante, junto con varios otros componentes del significado pragmático [13].

El conjunto de actos de habla para el robot en esta situación es:

- Pregunta sobre la aceptación de la sugerencias de visita (Tour-Question),
- Sugerencia para seleccionar el primer objeto de la visita de un conjunto de alternativas, en este caso, Venus_de_Milos, la_Mona_Lisa, el_Escriba_Senta-do, etc. (Question_Altern(V_M, E_S, M_L)),
- Finalización del diálogo en caso del rechazo de la propuesta (Closure) o cuando se acaban los objetos de interés.

Cada acto de habla corresponde a un conjunto de réplicas que permiten a evitar que se repita la misma frase. Se dan ejemplos de algunas réplicas a continuación.

B. *Bloques de habla*

Los bloques del habla son actos de habla compuestos que están organizados como una secuencia en la cual los actos se interconectan a través de la misma intención comunicativa general de tal manera que la respuesta se espera a toda la secuencia y no a los actos de habla individuales.

En el caso que estamos considerando se utilizan tres tipos de bloques del habla: a) los bloques fronterizos (Borderline Blocks, BB); b) los bloques narrativos (Narrative Blocks, NB) y c) los bloques secundarios (Secondary Blocks, SB).

1) *Bloques fronterizos*

Los bloques fronterizos incluyen el bloque introductorio (Introductory Block, IB) y el bloque terminal (Terminal Block, TB). Cada bloque consta de una secuencia de actos de habla elementales.

El bloque introductorio (IB) es una secuencia de la siguiente forma (en el orden dado):

Saludo (Greeting) + Presentación de la persona (Introduction) + Propuesta de la visita (Tour-Proposal).

El bloque introductorio comienza su operación en la fase inicial, la cual pertenece al tipo de fases de habla (los tipos de fases se considerarán en adelante).

El bloque terminal (TB) es una secuencia de la siguiente forma:

El anuncio del fin de la inspección (Announce_End) + Autoevaluación (Evaluation) + Despedida (Farewell).

El bloque terminal se inicia en la fase final, la cual se determina con base en el historial del diálogo. Luego, en la representación formal del diálogo, la transición a la fase de pronunciación del bloque terminal depende de la limitación contextual del tipo NB(X) ∈ HD, donde HD es el historial del diálogo en transcurso (*history of the dialogue*), y X obtiene uno de los tres valores —V_M, M_L o E_S— es decir, se identifica el hecho de que los objetos planeados para la inspección ya han sido vistos todos.

2) *Bloques narrativos*

Los bloques narrativos (NB) son textos o presentaciones que contienen una cierta cantidad de información acerca de los objetos de demostración (en nuestro caso, los objetos son la_Venus_de_Milos, la_Mona_Lisa, el_Escriba_Sentado). Los bloques narrativos se ponen en operación al alcanzar el robot el objeto de demostración, o diciendo con más precisión, después de que el robot completa el bloque multimodal del tipo MMB2 (se considerará en adelante). A los bloques narrativos es posible subir la información necesaria con anticipación.

3) *Bloques secundarios*

El bloque secundario (Secondary Block, SB) es pronunciado por el robot en una situación relacionada con la posibilidad durante la inspección de las localidades utilizar los servicios (baño y fuente de agua fría), o visitar la cafetería y funciona en conjunto con el bloque multimodal secundario SMB (se considerará a continuación).

El bloque secundario es una desviación del objetivo principal: es una declaración de la intención de esperar hasta que el usuario termine su uso del servicio *(Promise_Wait) + la petición dirigida al usuario de informar de su regreso (Request_Return)*.

El bloque secundario se inicia cuando el usuario responde positivamente a la propuesta de utilizar el servicio.

C. *Bloques multimodales*

Los bloques multimodales son secuencias de actos de habla (SA) y acciones no verbales realizados en un orden fijo; se clasifican en los principales y los secundarios.

1) *Bloques multimodales principales*

El sistema incluye dos tipos de bloques multimodales principales (Main Multimodal Blocks, MMB).

MMB1: *la petición (comando) de seguir al robot pronunciada por él (Imper_Move) + el desplazamiento del robot al objetivo (Move(X))*, donde X ∈ {V_M, M_L, E_S}.

Cabe mencionar que estamos suponiendo que el usuario seguirá el robot, es decir, realizará el comando (Imper_Move), porque él estuvo de acuerdo con realizar la visita. Si es necesario verificar que el usuario está siguiendo al robot, es

posible introducir la confirmación adicional en el modelo: el robot puede pedir la confirmación para la continuación de la visita cada 5-10 minutos, y en el caso de ausencia de la respuesta o de la respuesta negativa finalizar la visita.

Se supone que el robot guía comienza su operación siempre desde el punto inicial fijo L (en nuestra aplicación particular, este punto es la entrada al museo pasando las escaleras) y se mueve hacia el primer objetivo aceptado por la ruta más corta, luego se desplaza al segundo objetivo por la ruta más corta, etc.; aquí recordamos que el robot es capaz de evitar los obstáculos que no están en su mapa. También es importante recordar que el robot del cual estamos hablando no puede ni subir ni bajar las escaleras. Aún más, él no puede detectar la escalera hacia abajo, porque su sensor laser solo funciona en un plano. Por lo que es necesario marcar las zonas con escaleras como áreas prohibidas para el robot.

Hay dos bloques de este tipo en nuestro modelo de diálogo. En el transcurso del diálogo, el primer bloque se inicia a través de la respuesta del usuario al acto de habla (SA) del robot del tipo Question_Altern(V_M, M_L, E_S) (véase más arriba). El segundo bloque se pone en ejecución al obtener la respuesta negativa del usuario a la pregunta del robot acerca del deseo de usar un servicio y se pronuncia al finalizar el primer bloque narrativo.

MMB2: *la detención del robot junto al objetivo (Stop_H) + el comando de detención (Imper_Stop)*. Este bloque se inicia mediante un acto visual representado en el modelo como Visual_Act(X), donde X ∈ {V_M, M_L, E_S}.

2) *Bloques multimodales secundarios*

Los bloques multimodales secundarios (Secondary Multimodal Blocks, SMB) son la secuencia de acciones relacionada con los objetos secundarios en la ruta (en nuestro caso son los servicios del baño, la cafetería y fuente de agua fría representados en el modelo como WC, Café y Cooler respectivamente).

SMB: *la detención junto al objeto secundario (Stop(X)) + la petición de llamar la atención al objeto secundario Announce_Utility(X) + la pregunta sobre el deseo de usar el servicio X, Question_Utility(X),* donde X ∈ {Cooler, Café, WC}. SMB, igual como MMB2, se inicia mediante el acto visual representado en el modelo como Visual_Act (X), donde X ∈ {Cooler, Café, WC}, es decir, el robot "ve" el objeto que está marcado en el mapa y se encuentra suficientemente cerca de él.

VII. Actos de habla y su fuerza ilocutiva

Según la descripción anterior, el robot por un lado es capaz de realizar las acciones ilocutivas independientes, y por otro lado las acciones ilocutivas dependientes en el sentido de Baranov y Kreidlin (1992) [23].

Las acciones ilocutivas independientes, es decir, iniciados por la "iniciativa" del robot, son bloques del habla de los tipos BB, NB, los bloques multimodales MMB2 (Stop_X + Imper_Stop) y SMB.

Las acciones ilocutivas dependientes, es decir, iniciados con la condición de que el usuario realizó algún acto de habla (SA), son el bloque del habla SB, los actos de habla Question_Altern y Ask_Confirm, y el bloque MMB1 (Imper_Move + Move_X).

Cada tipo de los actos de habla o los bloques del habla, así como el componente de habla del bloque multimodal, se relaciona en el módulo de síntesis de texto con un conjunto de los patrones de habla (réplicas) de los cuales el robot hace la selección memorizando cual opción ya fue seleccionada para elegir en la siguiente iteración otro patrón del mismo tipo y guardarlo en la memoria, etc. de elegir otra plantilla del mismo tipo y recordarlo, etc. Con esta estrategia el comportamiento del robot es más natural.

Cada acto de habla ilocutivo independiente y el bloque del habla del robot en un diálogo del tipo que estamos modelando, impone cierto compromiso comunicativo (*committment*) para el usuario, bajo la condición que por lo general él debe respetar las máximas de comunicación de Grice [3], es decir, él adopta un propósito común en el diálogo y respeta la dirección del desarrollo del diálogo y no tiene la intención de confundir al robot, etc. Todos los actos de habla del usuario en este tipo de diálogo son dependientes desde el punto de vista ilocutivo: por las limitaciones consideradas anteriormente no se supone que el usuario tome la iniciativa en el diálogo, por ejemplo, hacer una pregunta al robot aun acerca de la algo relacionado con la visita. El usuario debe responder de acuerdo con la finalidad ilocutiva de las acciones del robot, pero si él toma "libertades", el robot lo tratará como un error e intentará de "obtener" la reacción esperada. En el futuro, se planea eliminar esta restricción para el diálogo.

VIII. Actos de habla del usuario

En virtud de las restricciones anteriores, el repertorio de actos de habla que se esperan del usuario, consta de los siguientes SA, a las siglas de las cuales se agrega el prefijo uSA (user Speech Act):

- Consentimiento (uSA(Yes)) o Denegación (uSA(No)) como reacciones a la propuesta de inspección, que forma parte del bloque IB, y corresponde a la propuesta de usar un objeto secundario, que está incluida en el bloque SB1.
- La respuesta con la alternativa seleccionada a la pregunta del robot acerca de con cual objeto se empieza la visita (véase más arriba Question_Altern(X)), en nuestro caso particular).
- El mensaje del regreso a su lugar después de usar el objeto secundario uSA(Ready).
- Todo SA no previsto en el sistema para este tipo del diálogo con el usuario se interpreta como error: Inesperado (no interpretable) acto de habla del usuario uSA(Error).

Si el sistema no puede asociar un acto de habla del usuario con alguno de esos cinco tipos ilocutivos, el robot vuelve a la fase inmediata anterior al acto de habla reconocido como

uSA(Error), y repite aquel acto de habla después del cual el usuario generó el acto de habla uSA (Error).

Si en relación con el tipo ilocutivo de enunciados el usuario se limita al ámbito de las obligaciones de comunicación impuestas por los actos de habla del robot, entonces en relación con la expresión de los actos de habla permitidos, el usuario tiene cierta libertad. El analizador de voz identifica unidades marcadas en el diccionario del sistema en correspondencia con su contenido proposicional o función ilocutiva. El módulo del control del diálogo asigna a la réplica del usuario en un momento dado del diálogo aquel tipo ilocutivo del conjunto de los posibles tipos, que corresponde a los marcadores de las unidades reconocidas. Por ejemplo, las lexemas *Venus_de_Milos* etc. en la réplica generada por el usuario en respuesta al acto de habla del robot del tipo "Propuesta de seleccionar el primer objeto de inspección del conjunto de las alternativas Question_Altern(V_M, M_L, E_S)" serán el marcador del tipo ilocutivo "Selección del objeto Venus_de_Milos uSA(V_M)". En el futuro, es posible realizar un análisis más complejo, porque el analizador sintáctico permite trabajar con el árbol de análisis sintáctico de cada frase del diálogo.

IX. Modelo del diálogo

El modelo del diálogo se puede representar como una red de transiciones del robot de un estado del diálogo al otro dependiendo de la información verbal o visual recibida, el historial del diálogo y las intenciones del usuario. Los estados se representan como reglas, otra posibilidad sería construir un autómata finito [11].

En nuestro modelo los estados del diálogo son de los siguientes tipos:

- el estado de habla (talk, talking state),
- el estado de percepción (perc., perception state),
- el estado de movimiento (move, moving state),
- el estado final (Final, Final State).

La transición de una fase a otra se realiza después de ejecutar las acciones (de habla y/o de movimiento) del robot. Las acciones se ponen en ejecución como reacción al acto de habla del usuario uSA(X), donde X toma un valor del conjunto de los tipos ilocutivos SA del usuario, o un acto de la percepción visual del robot de un objeto Y, Visual_Act(Y), donde Y toma valores del conjunto de los objetos marcados en el mapa interno del robot.

En consecuencia, la acción del robot dependiendo de su estado actual está dada por el par "(el acto de habla de usuario como entrada, el estado del robot) → (Acción del robot, el estado del robot como salida)".

Por lo tanto, el modelo del diálogo se puede representar en la forma de un conjunto de fases y acciones las cuales llevan el diálogo de una fase a otra. También se puede utilizar la recursividad, es decir, invocar una función dentro de sí misma, pero evitar ciclos infinitos el posible nivel de recursividad debe ser limitado.

En la tabla 1 se presenta la lista de reglas que describen la gran parte de nuestro modelo. Las reglas tienen la forma:

el estado del robot & condición (si se cumple), la acción (qué hacer) → un nuevo estado del robot

Entonces, para aplicar una regla, el robot debe estar en cierto estado y la condición se debe cumplir. En este caso, el robot realiza la acción correspondiente y pasa a un nuevo estado. Después de cada uso de una regla, esta regla se agrega al historial del diálogo HD (*history of dialogue*). A veces, la condición puede estar ausente: esto significa que la condición es verdadera, es decir, la acción debe ser ejecutada en todo caso. A veces la acción no es necesaria, es decir, no existe la necesidad para que el robot haga algo. Tales casos se simbolizan con ∅. Cabe mencionar que después de la aplicación de una regla, el robot siempre pasa a un nuevo estado. En la tabla 1 se dan ejemplos de las reglas. En esa tabla no se representan los actos de habla correspondientes a objetos M_L y CAFÉ, pero se hacen de manera similar. En el futuro, se planea usar actos de habla con parámetros, para no depender de la lista concreta de objetos.

Por ejemplo, las dos primeras reglas se interpretan de la siguiente manera: mientras el robot se encuentra en la fase inicial de habla, él realiza dos actos de habla por su "iniciativa", es decir, estos actos no se evocan por las causas externas: primero, el robot ejecuta el "bloque introductorio (IB)", luego pronuncia la pregunta acerca de la intención de hacer el recorrido Tour-Question, y después pasa a la fase de percepción (de la información de entrada) $perc_1$, etc.

X. Conclusiones

En este artículo, se presentó el módulo de control del diálogo en la comunicación con un robot guía móvil. El módulo consta del modelo del diálogo, la descripción de los actos de habla y los bloques del habla, a través de los cuales se puede construir cada componente del modelo. Todo el modelo, salvo los patrones de respuestas del robot, es independiente del idioma.

Aunque el artículo describe sólo una aplicación limitada del modelo, el modelo propuesto es universal en su arquitectura y proporciona los medios para el modelado formal del diálogo con el robot móvil autónomo usando el material lingüístico diverso, ya que el inventario de los actos de habla y las estrategias comunicativas del hablante, a diferencia de las restricciones que imponen diferentes idiomas a la estructura sintáctica [22], son comunes para todas las lenguas naturales.

Agradecimientos

Trabajo realizado con el apoyo de gobierno de la Ciudad de México (proyecto ICYT-DF PICCO10-120), el apoyo parcial del gobierno de México (CONACYT, SNI) e Instituto Politécnico Nacional, México (proyectos SIP 20144274, 20144534; COFAA), proyecto FP7-PEOPLE-2010-IRSES: "Web Information Quality – Evaluation Initiative (WIQ-EI)" European Commission project 269180, y Ministry of

Tabla 1. Lista de reglas de manejo de diálogo para dos objetos.

Estado previo		Actos de habla	Acción		Estado nuevo
$talk_1$	&	∅	IB	→	$talk_2$
$talk_2$	&	∅	Tour-Question	→	$perc_1$
$perc_1$	&	uSA(Yes)	Question_Altern(V_M, E_S, M_L)	→	$perc_2$
$perc_1$	&	uSA(No)	Closure	→	Final
$perc_1$	&	uSA(Error)	∅	→	$talk_2$
$perc_2$	&	uSA(V_M)	MMB1(V_M)	→	$move_1$
$perc_2$	&	uSA(E_S)	MMB1(E_S)	→	$move_2$
$perc_2$	&	uSA(E_S)	MMB1(M_L)	→	$move_3$
$perc_2$	&	uSA(Error)	∅	→	$talk_3$
$talk_3$	&	∅	Question_Altern(V_M, E_S, M_L)	→	$perc_2$
$move_1$	&	visual_act(V_M)	MMB2(V_M)	→	$talk_4$
$move_1$	&	visual_act(Cooler)	SMB(Cooler)	→	$perc_3$
$perc_3$	&	uSA(Yes)	SB	→	$perc_4$
$perc_3$	&	uSA(No)	MMB1(V_M)	→	$move_1$
$perc_4$	&	uSA(Ready)	MMB1(V_M)	→	$move_1$
$talk_4$	&	∅	NB(V_M)	→	$talk_5$
$talk_5$	&	NB(E_S) ∉ HD	MMB1(E_S)	→	$move_2$
$talk_5$	&	NB(E_S) ∈ HD	TB	→	Final
$move_2$	&	Visual_Act(E_S)	MMB2(E_S)	→	$talk_6$
$move_2$	&	Visual_Act(WC)	SMB(WC)	→	$perc_5$
$perc_5$	&	uSA(Yes)	SB	→	$perc_6$
$perc_5$	&	uSA(No)	MMB1(E_S)	→	$move_2$
$perc_6$	&	uSA(Ready)	MMB1(E_S)	→	$move_2$
$talk_6$	&	∅	NB(E_S)	→	$talk_7$
$talk_7$	&	NB(V_M) ∉ HD	MMB1(V_M)	→	$talk_7$
$talk_7$	&	NB(V_M) ∈ HD	TB	→	Final

education and Science of Russian Federation, federal project 2685 “Parametric description of grammar systems”.

REFERENCIAS

[1] “Dialog with Robots,” in *AAAI 2010, Fall Symposium*, November 2010, Arlington VA, 2010.

[2] S. A. Green, M. Billinghurst, C. X. Qi, G. J. Chase, “Human-Robot Collaboration: A Literature Review and Augmented Reality Approach in Design,” *International Journal of Advanced Robotic Systems*, vol. 5, pp. 1–18, 2007.

[3] P. Grice, “Logic and conversation,” in *Syntax and Semantics, 3: Speech Acts*, P. Cole & J. Morgan (eds.), New York: Academic Press, 1975 (Reprinted in: *Studies in the Way of Words*, ed. H. P. Grice, Cambridge, MA: Harvard University Press, 1989, pp. 22–40).

[4] G. J. M. Kruijff, H. Zender, P. Jensfelt, H. I. Christensen, “Clarification dialogues in human-augmented mapping,” in *Human Robot Interaction’06*, Salt Lake City, Utah, USA, 2006.

[5] G. J. M. Kruijff, P. Lison, T. Benjamin, H. Jacobsson, H. Zender, I. Kruijff-Korbayova, N. Hawes, “Situated dialogue processing for human-robot interaction,” *Cognitive Systems*, pp. 311–364, 2010.

[6] H. Kuzuoka, K. Yamazaki, A. Yamazaki, J. Kosaka, Y. Suga, C. Heath, “Dual ecologies of robot as communication media: Thoughts on coordinating orientations and projectability” in *CHI’04 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2004, pp 183–190.

[7] S. Lemaignan, R. Ros, R. Alami, “Dialogue in situated environments: A symbolic approach to perspective-aware grounding, clarification and reasoning for robot,” in *Proc. Robotics, Science and Systems, Grounding Human-Robot Dialog for Spatial Tasks workshop*, 2011.

[8] M. Marge, A. Pappu, B. Frisch, Th. K. Harris, A. Rudnicky, “Exploring Spoken Dialog Interaction in Human-Robot Teams,” in *Proceedings of Robots, Games, and Research: Success stories in USA, RSim IROS Workshop*, St. Louis, MO, USA, 2009.

[9] L. Padró, M. Collado, S. Reese, M. Lloberes, I. Castellón, “FreeLing 2.1: Five Years of Open-Source Language Processing Tools,” in *Proceedings of 7th Language Resources and Evaluation Conference (LREC 2010)*, ELRA, La Valletta, Malta, 2010.

[10] P. Pakray, U. Barman, S. Bandyopadhyay, A. Gelbukh, "A Statistics-Based Semantic Textual Entailment System," in MICAI 2011, *Lecture Notes in Artificial Intelligence*, vol. 7094, Springer, pp. 267–276, 2011.

[11] L. A. Pineda, V. M. Estrada, S. R. Coria, J. F. Allen, "The obligations and common ground structure of practical dialogues," *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*, vol. 11, no. 36, pp. 9–17, 2007.

[12] G. Salton, A. Wong, C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[13] J. Searle, "Indirect speech acts," in *Syntax and Semantics, 3:Speech Acts*, P. Cole & J.L. Morgan (eds.), New York: Academic Press, 1975. pp. 59–82. (Reprinted in: *Pragmatics: A Reader*, S. Davis (Ed.), Oxford: Oxford University Press, pp. 265–277, 1991).

[14] G. Sidorov, *Non-linear construction of n-grams in computational linguistics: syntactic, filtered, and generalized n-grams*, 2013, 166 pp.

[15] G. Sidorov, "Desarrollo de una aplicación para el diálogo en lenguaje natural con un robot móvil," *Research in computing science*, vol. 62, pp. 95–108, 2013.

[16] G. Sidorov, F. Velasquez, E. Stamatatos, A. Gelbukh, L. Chanona-Hernández, "Syntactic N-grams as Machine Learning Features for Natural Language Processing," *Expert Systems with Applications*, vol. 41, no. 3, pp. 853–860, 2014.

[17] E. A. Sisbot, R. Ros, R. Alami, "Situation assessment for human-robot interaction," in *Proc. of 20th IEEE International Symposium in Robot and Human Interactive Communication*, 2011.

[18] M. Skubic, D. Perzanowski, A. Schultz, W. Adams, "Using Spatial Language in a Human-Robot Dialog," in *IEEE International Conference on Robotics and Automation*, Washington, D.C., 2002.

[19] *Systems. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vienna, Austria, Association for Computing Machinery, 2004.

[20] E. A. Topp, H. Huttenrauch, H. I. Christensen, E. K. Severinson, "Bringing Together Human and Robotic Environment Representations—A Pilot Study," in *Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS)*, Beijing, China, 2006.

[21] Y. Wilks, N. Webb, A. Setzer, M. Hepple, R. Catizone, "Machine learning approaches to human dialogue modeling," in J.C.J. van Kuppevelt et al. (eds.), *Advances in Natural Multimodal Dialogue Systems*, pp. 355–370, 2005.

[22] P. M. Arkadiev, N. V. Serdobolskaya, A. V. Zimmerling, "Project of a typological database of syntactic constraints on movement," in *Computational linguistics and intellectual technologies, proceedings of the international conference "Dialogue. 2010"*, Moscow, vol. 9, no. 16. pp. 437–441, 2010 (in Russian).

[23] A. N. Baranov, G. E. Kreidlin, "Illocutive forcement in dialogue structure," *Problems of Linguistics [Voprosy jazykoznania]*, vol. 2. pp. 84–99, 1992 (in Russian).

[24] I. M. Kobozeva, N. I. Laufer, I. G. Saburova, "Modelling conversationtion in man-machine systems," in *Linguistic support of information systems [Lingvisticheskoe obespechenie informacionnyh system]*, INION of Academy of Sciences of USSR, Moscow, 1987 (in Russian).

[25] M. V. Prischepa, "Development of the user profile based on the psychological aspects of human interaction with the informational mobile robot," *SPIIRAS proceedings [Trudy SPIIRAN]*, vol. 2, no. 21, pp. 56–70, 2012 (in Russian).

NoXperanto: Crowdsourced Polyglot Persistence

Antonio Maccioni, Orlando Cassano, Yongming Luo, Juan Castrejón, and Genoveva Vargas-Solar

***Abstract*—This paper proposes NoXperanto, a novel crowdsourcing approach to address querying over data collections managed by polyglot persistence settings. The main contribution of NoXperanto is the ability to solve complex queries involving different data stores by exploiting queries from expert users (i.e. a crowd of database administrators, data engineers, domain experts, etc.), assuming that these users can submit meaningful queries. NoXperanto exploits the results of "meaningful queries" in order to facilitate the forthcoming query answering processes. In particular, queries results are used to: (i) help non-expert users in using the multi-database environment and (ii) improve performances of the multi-database environment, which not only uses disk and memory resources, but heavily rely on network bandwidth. NoXperanto employs a layer to keep track of the information produced by the crowd modeled as a Property Graph and managed in a Graph Database Management System (GDBMS).**

***Index Terms*—Polyglot persistence, crowdsourcing, multi-databases, big data, property graph, graph databases.**

I. Introduction

BIG datasets are not uniform collections of homogeneous data and, for this reason, they are not logically stored in one single database. In addition, the size of big datasets is such that it is not possible to adopt data cleaning, dataset preparation and integration using classic methods. Otherwise, data would never be ready for being analyzed and exploited in reasonable time.

Often, these datasets are sharded across distributed persistence supports that adopt different access and management policies, different degrees of availability, fault tolerance and consistency.

Polyglot Persistence [1] is a brand new term to indicate the combination of approaches that cope with a variety of persistence systems and that are used for "integrating" heterogeneous datasets.

As part of the emerging polyglot persistence movement [1], the simultaneous use of multiple SQL, NoSQL and NewSQL data stores is gradually becoming a common practice in modern application development [2], [3]. They enable the integration of these data stores for managing big datasets in a scalable and loosely coupled way. This approach seems to be a pertinent strategy to deal with big datasets integration. Nonetheless, the combination of these heterogeneous data stores, flexible schemas and non-standard APIs, represent an added complexity for application developers. For instance, considering that data are spread across multiple data stores, each of which possibly relies on distinct data models (graph, document, etc.), developers must be familiar with a high number of implementation details, in order to effectively work with and maintain the overall data entities.

Providing an integrated view of the underlying data collections for enabling querying, is still an open issue, often solved at the application code level. Given that there is no uniformization of schemas and there is a lack of meta-information available about the data (i.e. mappings, meta-data, semantic equivalences, data similarities), query results are not integrated collections of data, often, they are bags of non related data collections. It follows that, the quality of querying in such polyglot systems is poor.

Contribution. This paper proposes NoXperanto, a novel crowdsourcing[1] approach to address querying over data collections managed by polyglot persistence settings.

We avoid the pursue of uniformity, but rather, we preserve the variety of original systems and languages. The main contribution of NoXperanto is the ability to solve complex queries involving different data stores by exploiting queries from expert users (i.e. a crowd of database administrators, data engineers, domain experts, etc.), assuming that these users can submit meaningful queries.

Crowds have been revealed to be effective to solve queries over a single database [4], but they have never been used to query a multi-database system, where the required expertise (about schemas, models, data formats and instances of the database) cannot be superficial. In fact, from the structure of the queries, we can infer who is expert.

NoXperanto exploits the results of such "meaningful queries" in order to facilitate the forthcoming query answering processes. In particular, the results of queries are used to: (i) help non-expert users in using the multi-database environment and (ii) improve performances of the multi-database environment, which not only uses disk and memory resources, but heavily rely on network bandwidth.

[1]The modality to fulfil a target by relying on the contributions coming from a large group of people.

Manuscript received on May 3, 2014; accepted for publication on June 10, 2014, published on November 15, 2014.

Antonio Maccioni is with the Università Roma Tre, Rome, Italy (e-mail: maccioni@dia.uniroma3.it).

Orlando Cassano is with the Université Libre de Bruxelles, Brussels, Belgium (e-mail: orlando.cassano@cetic.be).

Yongming Luo is with the Eindhoven University of Technology, the Netherlands (e-mail: y.luo@tue.nl).

Juan Castrejón is with the Université de Grenoble, Grenoble, France (e-mail: juan.castrejon@imag.fr).

Genoveva Vargas-Solar (corresponding author) is with the Centre national de la recherche scientifique (CNRS), LIG-LAFMIA labs, Saint Martin d'Hères, France (e-mail: genoveva.vargas-solar@imag.fr).

NoXperanto employs a *layer* to keep track of the information produced by the crowd. This is modeled as a Property Graph and thus, it is persisted by a Graph Database Management System (GDBMS).

Outline. The remainder of the paper is organized as follows. While Section III introduces the concepts that we need throughout the paper, Section II discusses works that are related to our. Section IV explains our approach in detail. Section V describes implementation issues. Section VI concludes the paper with future perspectives of research.

II. Related Works

Most information systems are supported by distributed and heterogeneous data sources: multi-databases, data warehouses and Web portals. Such systems are, in fact, mediation infrastructures that emerged more than 20 years ago to enable transparent access to multiple data sources through querying, analysis, navigation and management facilities.

With the advent of NoSQL stores and polyglot persistence approaches, problems of data integration emerge with some specifities: the high heterogeneity of data models (e.g., NoSQL underlying data models are not standardized), the absence of semantic descriptions, the volume of data to integrate, just to name a few. Transformation and other well-known techniques to bridge different databases seem hard to employ at large scale [5]; they would bring data duplication that is unfeasible when the data size is huge. In order to deal with heterogeneity, the majority of information systems define an integrated schema. However, the integration process, generally controlled by the administrator of the system, is hardly suited for managing data in highly distributed and evolving environments such as the Web. Hardly coupled multi-sources storing huge datasets with semantic information (e.g. global schemas, mappings, rewritting rules) can be restrictive and expensive for applications that require simplicity, performance and large-scale dynamicity.

Advances in polyglot persistence querying are addressing the integration of heterogeneous data by providing pivot querying languages like UnQL that provide some operators for integrating graph, document, key value and relational data. There exists equivalent works which consider a unified language, proposing some SQL extension [6], [7], [8], [9], [10], or consider a metamodel approache such as the Apache's MetaModel project[2].

To overcome existing problems of data integration, we propose a solution to exploit a crowd of experts in a transparent way. Experts are all the users who are able to submit meaningful queries. We do not assume having a unified query language, but rather we follow the idea of the polyglot persistence movement to support multiple languages.

The idea of using crowdsourcing for answering queries over a database has been covered in previous works [7], [8], [11], [9], [4], [10], [12]. In particular, CrowdDB [4] integrates human inputs within a query plan that contains operations that usual database systems cannot always answer correctly (i.e. matching, ranking, etc.). RandomDB [12] answers non deterministic queries using social network crowds with the aim to replace existing random number generators.

[2]http://metamodel.eobjects.org/

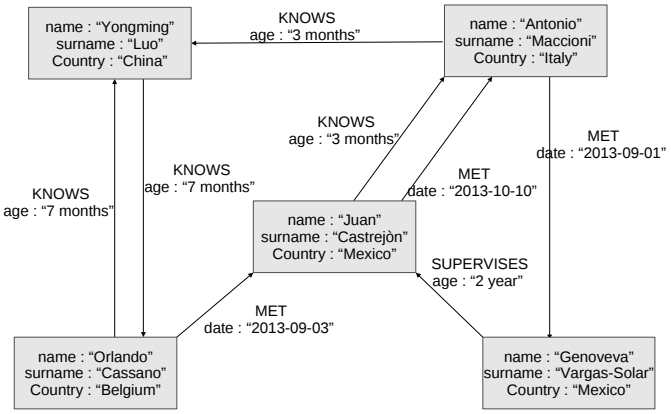


Fig. 1. An example of property graph.

III. Querying a polyglot multi-store

This section introduces basic concepts that are needed to introduce our approach in the next section.

Property Graph. Graph data models are able to generalize other data models:

- *relational data:* both schema and instance of relational data can be modeled with graphs [13];
- *XML data:* XML document is modeled as a tree and a tree is an acyclic, undirected graph;
- *document stores:* each document consists of nested key-value pairs (therefore a tree), so a document store can be modeled as a set of trees;
- *key-value stores:* they corresponds to a set of nodes of the property graph.

NoXperanto does not generalize, nor integrates, other models; rather, it adopts a graph data model in order to link more easily the parts of the entities that are spread across different databases. In particular, we use the *property graph* model [14]. Briefly, a property graph is a multi-graph (a graph where two nodes can be connected by more than one edge) where every node and every edge has associated a set of *properties*. A property is a key-value pair, denoted in the format of $< key, value >$.

An instance of property graph is reported in Figure 1. This represents people (i.e. the nodes) and their relationships. For example the node representing a person whose name is *Genoveva* supervises the node representing the person whose

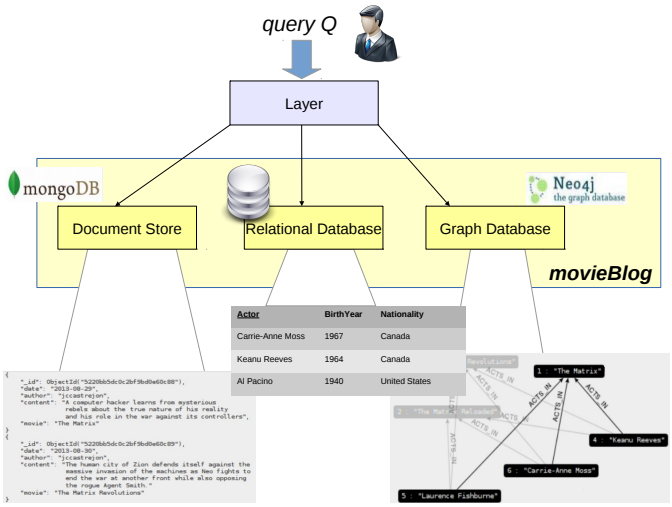


Fig. 2. The movie blog environment.

name is *Juan*. In this case *¡name : Genoveva¿* is a property where *name* is the key and *Genoveva* is the value.

Though the property graph is directed, we simplify it using the undirected version. The property graph is the "de-facto" standard model adopted by the GDBMSs (i.e. Neo4J, OrientDB, etc.). We will see in the Section V more details about it, since we make use of a GDBMS.

Running Example. To facilitate the understanding of the approach, let us consider the following running example. We have an infrastructure of databases, called *movieBlog*, behind a big blog website talking about movies. Information about entities such as the movies are splitted into different databases. The infrastructure consists in a multi-database, where we have a document store containing blog entries, a relational database storing information about actors, and a graph database keeping track of the movies' cast. Figure 2 depicts our scenario.

The definition of the databases of our running example are in Figure 3.

```
[DEFINITIONS]
define "blogEntries" as document
define "movies" as graph
define "actors" as table
```

Fig. 3. The database definition of the databases in Figure 2.

The example defines three databases of different kind: a relational database (i.e. table *actors*), a graph database (i.e. graph *movies*) and a document store (i.e. document *blogEntries*). The keys of these databases are the attributes *Actor*, *id* and *_id*, respectively for the databases *actors*, *movies* and *blogEntries*.

Polyglot Environment. Databases contain *entities* (e.g., a tuple in a relational database or a document in a document store). Every entity can be referred with a *key* (e.g., a primary key in a relational database or a document id in a document store). Queries are expressions that aim at retrieving: (i) an entity, (ii) a part of an entity or (iii) a combination of entities. Without referring to a particular language or algebra, we can refer to these tasks with general operator of *selection* (σ), *projection* (π) and *join* ($\bowtie$), respectively.

A σ query over a polyglot environment still involves separate answering processes, where database instances are independent from each other. In this case, the query can be dispatched to all the underlying databases, after an eventual syntactic translation of the query expression. A π is straightforward since it operates locally by extracting a sub-part of a result. While the semantics of σ and π in a polyglot environment are well-defined, the semantics of the $\bowtie$ needs to be clarified. In fact, if we join heterogeneous databases we can hardly rely upon a single algebra. We define such semantics as follows. Let us suppose to have two entities e_i and e_j belonging, respectively, to two heterogeneous databases. Among the others, the two entities have attributes a_i and a_j, respectively. Let us also suppose that there exists a comparing operator $\diamond$. A join $e_i \bowtie_{e_i.a_i \diamond e_j.a_j} e_j$ is the union between the information content of e_i and the content of e_j iff the predicate $e_i.a_i \diamond e_j.a_j$ is satisfied.

Clearly, we can develop a join operator in a polyglot environment at a code-level exploiting existing join algorithms (e.g., nested-loop style join). In this context, such operations are very expensive as we cannot rely upon the optimizations of the DBMSs but we are compelled to load all the data in memory and, sometimes, to transfer them over a network.

Next section explains how NoXperanto is able to answer queries using a crowd of experts and, in particular, it will focus on how the join operators can be computed efficiently.

IV. NoXperanto

This section explains our approach in detail. We first give an overview of the overall approach. Then, we conclude the section describing some use case in order to point out the advantages of the system.

A. Crowdsourcing Approach

In NoXperanto we aim at solving complex queries over a system containing several heterogeneus databases. To perform such queries we have to keep track of the relationships among entities stored in different databases. We employ two ways to indicate these relationships: one is *explicit* and the other is *implicit*.

Explicit Working Mode. In the *explicit* manner, the user can define how two classes of entities in different databases are related. For example, in Figure 4 we define that an entity of the database *blogEntries* is the same of an entity in the database *movies* if the value of *blogEntries.movie* is equal to the value of *movies.titles*.

```
[DEFINITIONS]
define "comments"
  on "blogEntries.movie" = "movie.titles"
  as link
```

Fig. 4. The explicit definition of relationships.

In this case we can exploit several techniques (i.e. ontology matching, schema matching, string matching, etc.) to find the instances of such definitions. They work at schema level and are very expensive to be performed at run-time. We can mitigate this complexity by using an hybrid approach between this explicit mode with the implicit mode that is explained next, but this lies outside the scope of this paper.

However, in many cases the administrator does not explicitly specify such definitions. It turns out that an automatic discovery of the relationships might be very useful. NoXperanto provides such a mechanism through the *implicit* modality.

Implicit Working Mode. The *implicit* working mode is driven by a crowdsourcing approach. Crowdsourcing is employed in different contexts to solve difficult problems and to build up knowledge bases relying on the effort of many people. Usually, a problem is split into many sub-problems that are solved through the so called microtasks. A microtask is a complicated task for a computer but it is easy for a person. The final problem is solved by considering all the microtasks that people of the crowd have processed. Often, these people are unaware of the problem or do not even know that they are processing a microtask.

The *implicit* working mode expects the system to be able to recognize the relationships. In this case, the system uses the knowledge of a crowd of experts, which appears when they submit complex queries. If those queries are also meaningful, in the sense that they produce a non empty set of results, we persist the relationships between entities of the two (or more) databases in our *property graph layer*.

More in detail, we extract the predicate of the join conditions from such queries to define a *crowd link*. For example, let us imagine an expert submitting $\sigma_{id,Nationality}$ $(actors \bowtie_{Actor=id} movie)$. In a SQL-like language, the query will look like the following:

```
SELECT id, Nationality
FROM actors, movies
WHERE actors.Actor = movies.id
```

The result of the query is {*¡Keanue Reeves, Canada¿, ¡Carrie-Anne Moss, Canada¿*}, thus, it is a non empty result set and the query is meaningful. In the answering process we have identified some relationships between entities in *movies* and entities in *actors*. For each of them, NoXperanto persists a crowd link on the property graph layer as in Figure 5.

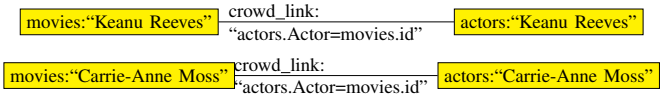


Fig. 5. Crowd links explicitly produced.

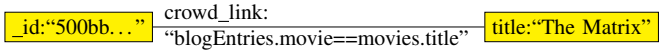


Fig. 6. Crowd links produced in Use case 2.

B. The Approach at Work

In this section we describe a sequence of four use cases to show the real benefits of a crowdsourcing approach in this context. The cases alternate queries from a non expert and expert users. For the sake of simplicity we consider a SQL-like syntax but of course, it is just the semantics of the language and does not refer to any relational database. We refer the multi-database managed by NoXperanto with *movieBlog*. In this way, we query our datasets in a transparent way. The four cases are defined as follows:

– *Use case 1*: a non-expert user asks for all information about “The Matrix”. It is a simple keyword search-like query (either submitted by a human being or by a programmatic access), so we cannot assume that the user is an expert.

```
SELECT *
FROM movieBlog
WHERE movieBlog.title == ``The Matrix''
```

The result of the query is

```
{year=1999}
```

since we have found an attribute *title* within the databases in *movieBlog* and an entity where the value of such attribute is “The Matrix”.

– *Use case 2*: a domain expert perform the following query:

```
SELECT *
FROM movies
JOIN blogEntries
  ON blogEntries.movie == movies.title
```

Note that the expert does not generally refer to *movieBlog* but to a precise database. The result of the query is:

```
{year=1999,
content="A computer hacker learns from
         mysterious rebels about the
         true nature...''
author="jccastrejon''}
```

We can say that there is a relationship between *blogEntries.movie* and *movies.title*. Since a non-expert user could not write such a join with meaningful

results, we determine that this user was an expert user. Consequently, she can be included within our crowd. In fact, the system stores a crowd link for each of the results as in Figure 6. We have bridged the document store with the graph databases at runtime.

– *Use case 3*: a non-expert user submits the query of Use case 1, but in this case the result is different.

```
{year=1999,
content="A computer hacker learns from
         mysterious rebels about the
         true nature...''
author="jccastrejon''}
```

We provided information to the user that we were not able to retrieve before. This additional information is provided by answering a join-free query. This is due to the information within the *layer*, which allows to bridge sharded parts of the same entity.

– *Use case 4*: another expert user submits the query of Use case 2. The scenario is similar to use case 2, but the query answering is much more efficient. We do not have to perform an expensive join operation. We can directly exploit the presence of the crowd links to determine a pairs of entities to form the final results.

V. Implementation Issues

This section provides and overview of the implementation concerns that are required to develop the NoXperanto approach based on the requirements outlined in the previous sections.

Data Layer Issue. We manage the heterogeneity of multiple data stores by relying on a data layer based on the property graph data model (see Section III). We implemented this layer on an emerging GDBMS, that is Neo4J.[3] It provides a REST interface so that the interaction with the applications running in the polyglot environment is facilitated. Moreover, this interface would provide operations to manage data entities and links between them. For example, the specific syntactic sugar to specify when to consider the crowd links in the query answering.

Language Issue. Our approach does not consider a unified query language for polyglot persistence applications, but rather rely on the existing language support provided by scalable data stores. Thus, we propose to provide extensions for each of these languages, based on the general query semantics described in Section III. In particular, these extensions rely on the REST interface of the property graph model to manage the link and join operations described in Section IV.

To implement these language extensions we intend to follow a model-driven approach, as proposed in our previous work.[4] In particular, the language definitions would be implemented using the Xtext framework [15], while the implementation of the link and join operations would be conducted with the Acceleo project,[5] by relying on text templates that refer to the graph model REST interface. The native query mechanisms of each of the supported data stores would be used to retrieve the data in each of the systems.

Consistency Issue. Our approach assumes an eventual consistency model [16] in the persistence of the crowd links among the distributed entities. As a consequence, even when a link between data entities has been identified, users executing the same query may not always receive the same result. Nonetheless, each of the data stores in the environment has its consistency semantics.

To handle this heterogeneity in the consistency semantics we propose to extend our language support with operators to allow the user to specify the level of consistency that he expects from each of the supported data stores. We intend to implement this functionality based on the operators that few systems (e.g., Riak,[6] a key-value data store) already provide to trade availability for consistency on a per-request basis.

Crowd Management Issue. We have developed a small utility to manage the crowd of experts. This consists, basically, on a query parser and a small interface where the administrator can check the current state of the crowd links.

VI. Conclusions and Perspectives

In this paper we have presented NoXperanto, an approach to solve queries over an heterogeneus environment of databases using the knowledge of a crowd of experts. This knowledge is extracted from the results of the queries. In NoXperanto we avoid expensive pre-processing. As a result, we are able to scale with respect to the number of the databases within our environment.

Our future work will be devoted to finish the system, implementing an interface for setting up the multi-database environment and allowing the users to specify weather or not the system has to use the crowd links. The approach opens several research direction. In particular, we will investigate other scenarios where the results of a query can be exploited to facilitate forthcoming query answering.

Acknowledgments

The ideas within this paper were developed during the EDBT Summer School 2013 in Aussois (France). The authors of this paper are grateful to the EDBT association for the organization of the Summer School and to all the speakers for their helpful suggestions.

[3] http://www.neo4j.org/
[4] https://github.com/jccastrejon/edbt-unql/
[5] http://www.eclipse.org/acceleo/
[6] http://basho.com/riak/

REFERENCES

[1] P. J. Sadalage and M. Fowler, *NoSQL distilled: a brief guide to the emerging world of polyglot persistence*. Addison-Wesley, 2012.

[2] B. F. Cooper, "Spanner: Google's globally-distributed database," in *Proceedings of the 6th International Systems and Storage Conference*. ACM, 2013, p. 9.

[3] D. Borthakur, "Petabyte scale databases and storage systems at facebook," in *SIGMOD Conference*, 2013, pp. 1267–1268.

[4] M. J. Franklin, D. Kossmann, T. Kraska, S. Ramesh, and R. Xin, "CrowdDB: answering queries with crowdsourcing," in *SIGMOD Conference*, 2011, pp. 61–72.

[5] R. De Virgilio, A. Maccioni, and R. Torlone, "Converting relational to graph databases," in *GRADES*, 2013.

[6] A. Marcus, E. Wu, D. Karger, S. Madden, and R. Miller, "Human-powered sorts and joins," *Proc. VLDB Endow.*, vol. 5, no. 1, pp. 13–24, Sep. 2011.

[7] J. Selke, C. Lofi, and W.-T. Balke, "Pushing the boundaries of crowd-enabled databases with query-driven schema expansion," *Proc. VLDB Endow.*, vol. 5, no. 6, pp. 538–549, Feb. 2012.

[8] A. Bozzon, M. Brambilla, and S. Ceri, "Answering search queries with CrowdSearcher," in *Proceedings of the 21st international conference on World Wide Web*, ser. WWW'12, 2012, pp. 1009–1018.

[9] G. Demartini, B. Trushkowsky, T. Kraska, and M. J. Franklin, "CrowdQ: Crowdsourced query understanding," in *CIDR*, 2013.

[10] H. Park, R. Pang, A. Parameswaran, H. Garcia-Molina, N. Polyzotis, and J. Widom, "An overview of the deco system: data model and query language; query processing and optimization," *SIGMOD Rec.*, vol. 41, no. 4, pp. 22–27, jan 2013.

[11] J. Wang, T. Kraska, M. J. Franklin, and J. Feng, "CrowdER: crowdsourcing entity resolution," *Proc. VLDB Endow.*, vol. 5, no. 11, pp. 1483–1494, Jul. 2012.

[12] R. De Virgilio and A. Maccioni, "Generation of reliable randomness via social phenomena," in *MEDI*, 2013, pp. 65–77.

[13] R. Angles and C. Gutiérrez, "Survey of graph database models," *ACM Comput. Surv.*, vol. 40, no. 1, 2008.

[14] M. A. Rodriguez and P. Neubauer, "Constructions from dots and lines," *CoRR*, vol. abs/1006.2361, 2010.

[15] M. Eysholdt and H. Behrens, "Xtext: implement your language faster than the quick and dirty way," in *Proceedings of the ACM international conference companion on Object oriented programming systems languages and applications companion*, ser. SPLASH'10. New York, NY, USA: ACM, 2010, pp. 307–309. [Online]. Available: http://doi.acm.org/10.1145/1869542.1869625

[16] D. Pritchett, "BASE: An Acid alternative," *Queue*, vol. 6, no. 3, pp. 48–55, May 2008. [Online]. Available: http://doi.acm.org/10.1145/1394127.1394128

Haar Wavelet Neural Network for Multi-step-ahead Anchovy Catches Forecasting

Nibaldo Rodriguez, Gabriel Bravo, and Lida Barba

***Abstract*—This paper proposes a hybrid multi-step-ahead forecasting model based on two stages to improve pelagic fish-catch time-series modeling. In the first stage, the Fourier power spectrum is used to analyze variations within a time series at multiple periodicities, while the stationary wavelet transform is used to extract a high frequency (HF) component of annual periodicity and a low frequency (LF) component of inter-annual periodicity. In the second stage, both the HF and LF components are the inputs into a single-hidden neural network model to predict the original non-stationary time series. We demonstrate the utility of the proposed forecasting model on monthly anchovy catches time-series of the coastal zone of northern Chile (18^oS-24^oS) for periods from January 1963 to December 2008. Empirical results obtained for 7-month ahead forecasting showed the effectiveness of the proposed hybrid forecasting strategy.**

***Index Terms*—Neural network, wavelet analysis, forecasting model.**

I. Introduction

In order to develop sustainable exploitation policies, forecasting the stock and catches of pelagic species off northern Chile is one of the main goals of the fishery industry and the government. However, fluctuations in the environmental variables complicate this task. To the best of our knowledge, few publications exist on forecasting models for pelagic species. In recent years, linear regression models [1], [2] and artificial neuronal networks (ANN) [3], [4] have been proposed for forecasting models. The disadvantage of models based on linear regressions is the supposition of stationarity and linearity of the time series of pelagic species catches. Although ANN allow modeling the non-linear behavior of a time series, they also have some disadvantages such as slow convergence speed and the stagnancy of local minima due to the steepest descent learning method. To improve the convergence speed and forecasting precision of anchovy catches off northern Chile, Gutierrez [3] proposed a hybrid model based on a multilayer perceptron (MLP) combined with an autoregressive integrated moving average (ARIMA) model.This forecaster obtained a coefficient of determination R^2 of 82%, which improved slightly when combining the MLP model with the ARIMA model, reaching an R^2 of 87%.

In this paper, the proposed forecasting model is based on single-hidden neural network combined with Haar stationary wavelet transform (SWT). The stationary wavelet decomposition was selected due to its popularity in hydrological [5], [6], electricity market [7], financial market [8] and smoothing methods [9]–[11]. This SWT technique is based on the discreet wavelet transform (DWT) or the stationary wavelet transform (SWT) [12]. The advantage of these wavelet transforms in non-stationary time series analysis is their capacity to separate low frequency (LF) from high frequency (HF) components. On the one hand, the LF component reveals long-term trends, while the HF component describes short-term fluctuations in the time series. Being able to separate these components is a key advantage in proposed forecasting strategies since the behavior of each frequency component is more regular than the raw time series.

In this paper, Haar stationary wavelet decomposition is applied to build a hybrid multi-step-ahead forecasting model to achieve more accurate models than conventional single-hidden neuronal network. Our proposed multi-step-ahead anchovy catches forecasting model is based on two phase. In the first phase,the Fourier power spectrum is used to analyze variations within a time series at multiple periodicities, while the stationary wavelet transform is used to extract a high frequency (HF) component of annual periodicity and a low frequency (LF) component of inter-annual periodicity. In the second stage, both the HF and LF components are the inputs into a single-hidden neural network model with N_i input nodes, N_h hidden nodes and two output nodes to predict the original non-stationary time series.

This paper is organized as follows. In the next section, we present hybrid multi-step-ahead forecasting model. The simulation results are presented in Section 3 followed by conclusions in Section 4.

II. Proposed forecasting model

This section presents the proposed forecasting model for one-month-ahead anchovy catches in northern Chile, which is based on the Haar stationary wavelet transform and single-hidden neural network model.

A. Stationary wavelet decomposition

A signal $x(n)$ can be represented at multiple resolutions by decomposing the signal on a family of wavelets and

Manuscript received on August 7, 2014, accepted for publication on September 22, 2014, published on November 15, 2014.

Nibaldo Rodriguez (corresponding author) is with the Pontificia Universidad Católica de Valparaíso, Av. Brasil 2241, Chile (e-mail: nibaldo.rodriguez@ucv.cl).

Gabriel Bravo is with the Universidad San Sebastián, Concepción, Chile (e-mail: gabo.bravoro@hotmail.com).

Lida Barba is with the Universidad Nacional de Chimborazo, Av. Antonio Jose de Sucre, Riobamba, Ecuador (e-mail: lbarba@unach.edu.ec).

scaling functions [9]–[11]. The approximation (scaled) signals are computed by projecting the original signal on a set of orthogonal scaling functions of the form:

$$\phi_{jk}(t) = \sqrt{2^{-j}}\phi(2^{-j}t - k) \tag{1}$$

or equivalently by filtering the signal using a low pass filter of length r, $h = [h_1, h_2, ..., h_r]$, derived from the scaling functions. On the other hand, the detail signals are computed by projecting the signal on a set of wavelet basis functions of the form

$$\psi_{jk}(t) = \sqrt{2^{-j}}\psi(2^{-j}t - k) \tag{2}$$

or equivalently by filtering the signal using a high pass filter of length r, $g = [g_1, g_2, ..., g_r]$, derived from the wavelet basis functions. Finally, repeating the decomposing process on any scale J, the original signal can be represented as the sum of all detail coefficients and the last approximation coefficient.

In time series analysis, discrete wavelet transform (DWT) often suffers from a lack of translation invariance. This problem can be tackled by means of the un-decimated stationary wavelet transform (SWT). The SWT is similar to the DWT in that the high-pass and low-pass filters are applied to the input signal at each level, but the output signal is never decimated. Instead, the filters are up-sampled at each level.

Consider the following discrete signal $x(n)$ of length N where $N = 2^J$ for some integer J. At the first level of SWT, the input signal $x(n)$ is convolved with the $h_1(n)$ filter to obtain the approximation coefficients $a_1(n)$ and with the $g_1(n)$ filter to obtain the detail coefficients $d_1(n)$, so that:

$$a_1(n) = \sum_k h_1(n-k)x(k) \tag{3a}$$

$$d_1(n) = \sum_k g_1(n-k)x(k), \tag{3b}$$

because no sub-sampling is performed, $a_1(n)$ and $d_1(n)$ are of length N instead of $N/2$ as in the DWT case. At the next level of the SWT, $a_1(n)$ is split into two parts by using the same scheme, but with modified filters h_2 and g_2 obtained by dyadically up-sampling h_1 and g_1.

The general process of the SWT is continued recursively for $j = 1, ..., J$ and is given as:

$$a_{j+1}(n) = \sum_k h_{j+1}(n-k)a_j(k) \tag{4a}$$

$$d_{j+1}(n) = \sum_k g_{j+1}(n-k)a_j(k), \tag{4b}$$

where h_{j+1} and g_{j+1} are obtained by the up-sampling operator inserts a zero between every adjacent pair of elements of h_j and g_j; respectively.

Therefore, the output of the SWT is then the approximation coefficients a_J and the detail coefficients $d_1, d_2, ..., d_J$, whereas the original signal $x(n)$ is represented as a superposition of the form:

$$x(n) = a_J(n) + \sum_{j=1}^{J} d_j(n) \tag{5}$$

The wavelet decomposition method is fully defined by the choice of a pair of low and high pass filters and the number of decomposition steps J. Hence, in this study we choose a pair of haar wavelet filters [12].

B. Neural network forecasting model

In order to predict the future value $\hat{x}(n-h)$, we can separate the original time series $x(n)$ into two components by using Haar stationary wavelet decomposition. The first extracted component x_H of the time series is characterized by slow dynamics, whereas the second component x_H is characterized by fast dynamics. Therefore, in our forecasting model a time series is considered as nonlinear function of several past observations of the components x_L and x_H as follows:

$$\hat{x}(n+h) = f\big(x_L(n-m), \dots\ x_H(n-m)\big); \tag{6}$$

the h value represents forecasting horizon and m denotes lagged values of both the LF and HF components.

A single-hidden neural network with two output nodes is used to estimate the nonlinear function $\hat{f}(\cdot)$, which is defined as

$$y_k(n) = \sum_{j=1}^{N_h} b_j\phi_j(u_i, v_j), k = 1, 2 \tag{7a}$$

$$\hat{x}(n+h) = y_1(n) + y_2(n), \tag{7b}$$

where N_h is the number of hidden nodes, $u = [u_1, u_2, \dots\ u_{2m}]$ denotes the input regression vector containing $2m$ lagged values, $[b_1, \dots\ b_{N_h}]$ represents the linear output parameters, $[v_1, v_2, \dots\ v_{N_h}]$ denotes the nonlinear parameters, and $\phi_j(\cdot)$ are hidden activation functions, which are derived as

$$\phi_j(u_i) = \phi\Big(\sum_{i=1}^{m} v_{j,i}u_i\Big) \tag{8a}$$

$$\phi(u) = \frac{1}{1 + exp(-u)}. \tag{8b}$$

In order to estimate both the linear and nonlinear parameters of the MLP, we use the Levenberg-Marquardt (LM) algorithm [13]. The LM algorithm adapts the $\theta = [b_1, \dots\ b_{N_h}, v_{j,1}, \dots\ v_{j,m}]$ parameters of the neuro-forecaster minimizing mean square error, which is defined as:

$$E(\theta) = \frac{1}{2}\sum_{i=1}^{N_s}\Big(x(n+h) - \hat{x}(n+h))\Big)^2. \tag{9}$$

Finally, the LM algorithm adapts the parameter θ according to the following equations:

$$\theta = \theta + \Delta\theta \tag{10a}$$

$$\Delta\theta = (\Upsilon\Upsilon^T + \mu I)^{-1}\Upsilon^T E, \tag{10b}$$

where Υ represents the Jacobian matrix of the error vector evaluated in θ_i and the error vector $e(\theta_i) = d_i - y_i$ is the

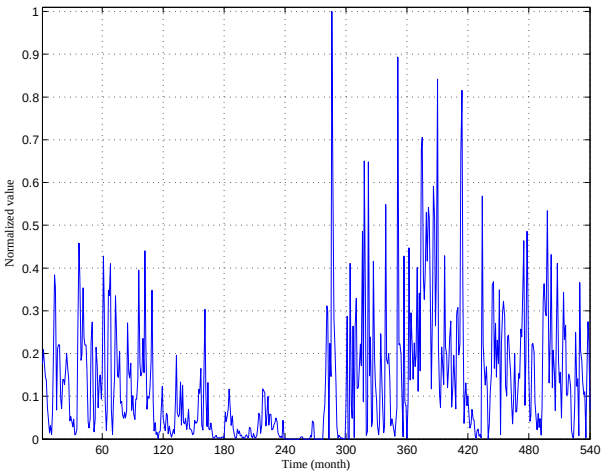


Fig. 1. Monthly anchovy catches time series

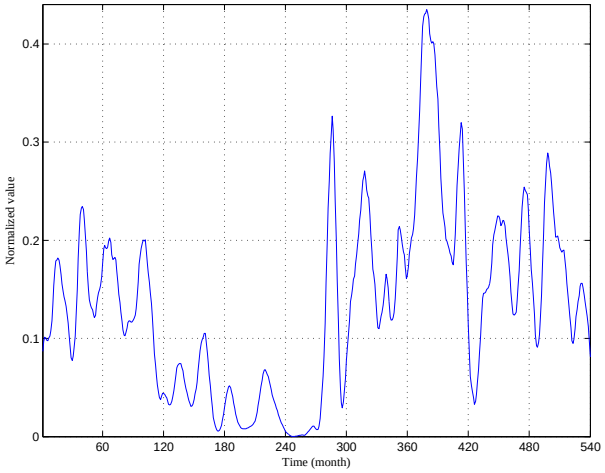


Fig. 3. Low frequency anchovy catches time series

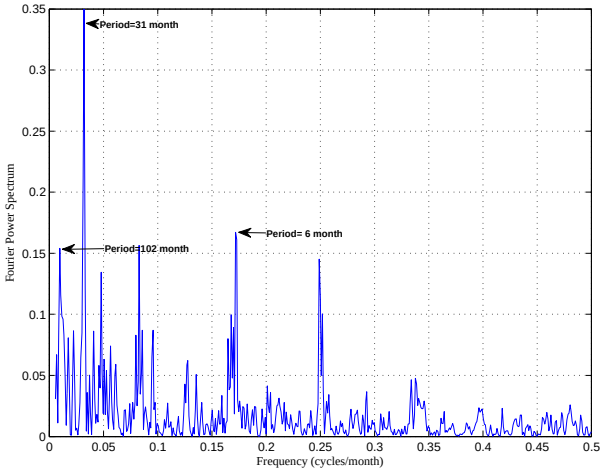


Fig. 2. Fourier power spectrum of catches time series

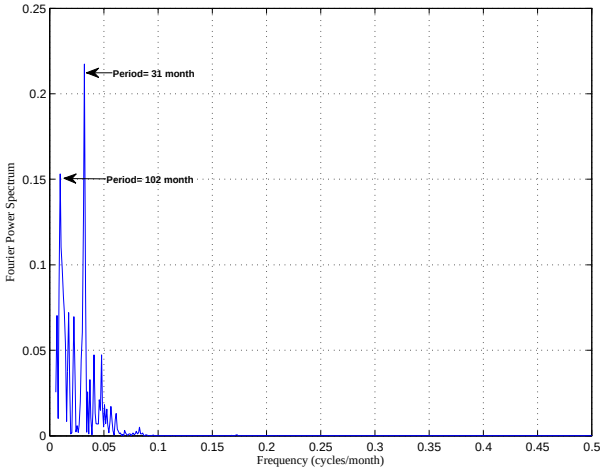


Fig. 4. Fourier power spectrum of LF time series

error of the MLP neural network for i patter, I denotes the identity matrix and the parameter μ is increased or decreased at each step of the LM algorithm.

III. Experiments and Results

In this section, we apply the proposed hybrid model for 7-month-ahead anchovy catches forecasting. The data set used corresponded to anchovy landings off northern Chile. These samples were collected monthly from 1 January 1963 to 30 December 2008 by the National Fishery Service of Chile (www.sernapesca.cl). The raw anchovy data set have been normalized to the range from 0 to 1 by simply dividing the real value by the maximum of the appropriate set. On the other hand, the original data set was also divided into two subsets. In the first subset the 70% of the time series were chosen for the training phase (parameters estimation), whereas the remaining data set were used for the testing phase.

The normalized raw time series and the Fourier power spectrum are present in the Figures 1 and 2; respectively. From Figure 2 it can be observed that there are two peaks of significant power. The first peak has an inter-annual periodicities of 31 months, whereas the second peak has a intra-annual periodicities of 6 months. After we applied the Fourier power spectrum to the raw time series, we decided to use 3-level wavelet decomposition due to the significative peak of 31 months. Both the LF and HF times series are presented in Figures 3 and 5; respectively, whereas the power spectrum of both time series are illustrated in Figure 4 and 6; respectively.

In this study, three criteria of forecasting accuracy called root mean squares error (RMSE), mean absolute percentage error (MAPE) and ralative error (RE) were used to evaluate the forecasting capabilities of the proposed hybrid forecasting models, which are defined as

$$RMSE = \sqrt{MSE}, \tag{11}$$

$$MAPE(\%) = \frac{1}{N_s}\sum_{i=1}^{N_s}\left|\frac{A_i - F_i}{F_i}\right| \times 100, \tag{12}$$

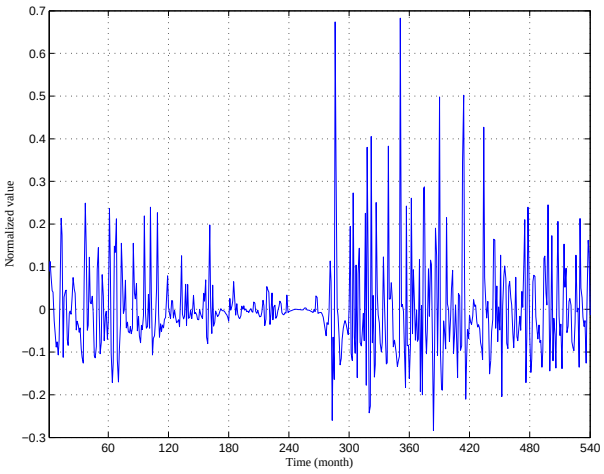


Fig. 5. High frequency anchovy catches time series

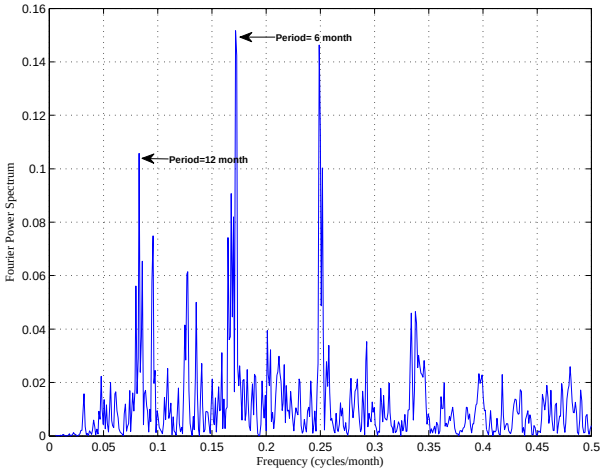


Fig. 6. Fourier power spectrum of HF time series

$$RE(\%) = \frac{A_i - F_i}{F_i} \times 100, \tag{13}$$

where A_i is the actual value at time i, F_i is the forecasted value at time i, and N_s is the number of forecasts.

The MLP neural network was calibrated using 31 previous months as input data plus one bias unit due to the periodicity of 31 months of the raw time series (see Figure 2). Finding the optimal number of hidden nodes is a complex problem, but in all our experiments, the number of hidden nodes is set as $\sqrt{N_i + N_o} = \sqrt{62 + 2}$ (number of input nodes and output nodes). In the training process, overall weights were initialized by a Gaussian random process with a normal distribution $N(0, 1)$ and the stopping criterion was a maximum number of iterations set at 200. Due to the random initialization of the weights, we used 50 runs to find the best MLP neural network with a low prediction error.

The Figures 7 and 8 shows the results of testing data for 50 run and 200 iterations, whose best result was achieved in the run 49. After the training-testing process, the architecture

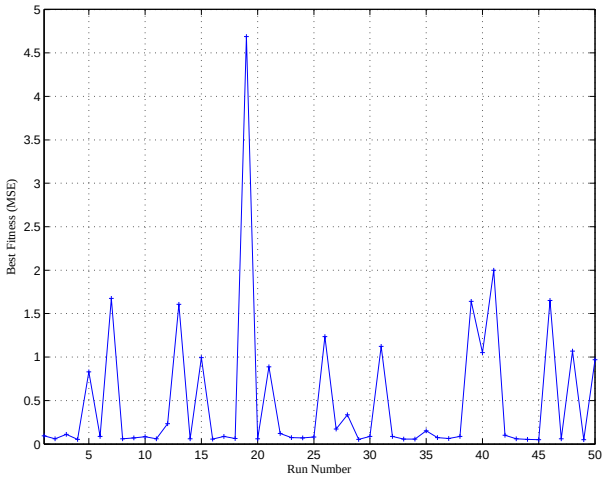


Fig. 7. Run versus MSE

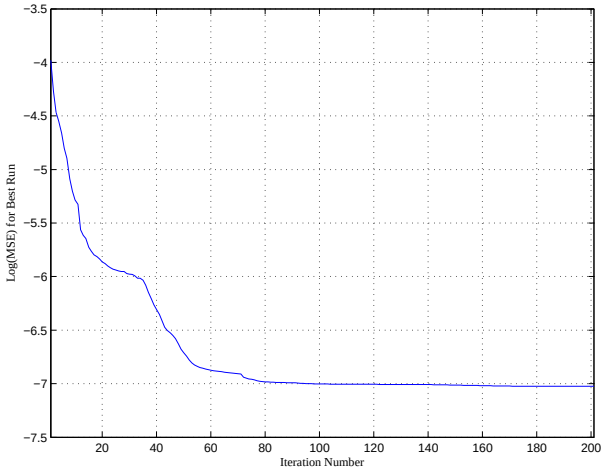


Fig. 8. Iteration number for Best Run

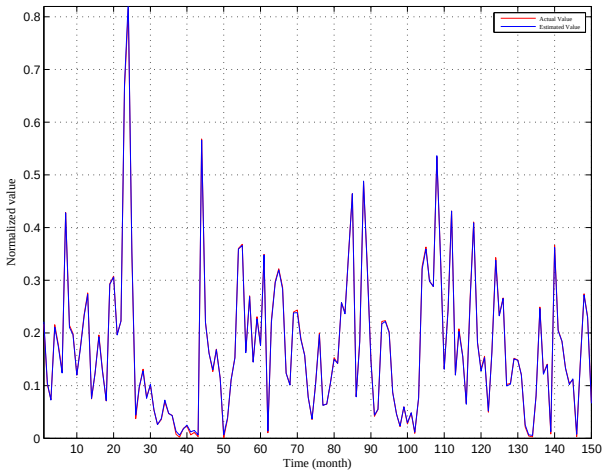


Fig. 9. Seven-month-ahead MLP(31,8,2) forecasting for test data set

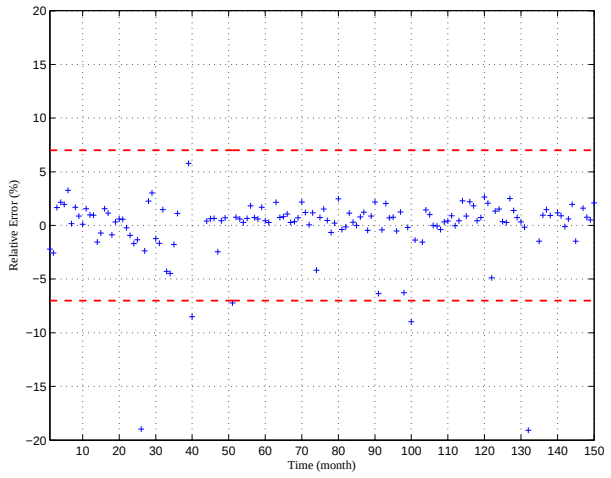


Fig. 10. Retative error for Seven-month-ahead MLP(31,8,2) forecasting during test data set

was calibrated with 31 input nodes, 8 hidden nodes, and two output node; this is denoted as MLP(31,8,2).

Figures 9 and 10 show the results obtained with the MLP(31,8,2) forecasting model during the testing phase. Fig. 9 provides data on observed monthly anchovy catches versus forecasted catches; this forecasting behavior is very accurate for testing data with a MAPE of 10.87% and a RMSE of 0.0028. On the other hand, from Figure 10 it can be observed that an important fraction (over 90%) of the predicted catches values are acceptable with residuals ranging from 7% to -7%.

IV. Conclusions

In this paper was proposed a 7-month-ahead anchovy catches forecasting strategy to improve prediction accuracy based on Haar stationary wavelet decomposition combined with a single-hidden neural network model. The reason of the improvement in forecasting accuracy was due to use Haar SWT to separate both the LF and HF components of the raw time series, since the behavior of each component is more smoothing than raw data set. It was show that the proposed hybrid forecasting model achieves a MAPE of 10.87% and a RMSE of 0.0028. Besides, proposed forecasting results showed that the 31 previous month contain valuable information to explicate a highest variance level for anchovy catches forecasting. Finally, hybrid forecasting strategy can be suitable as a very promising methodology to any other pelagic species.

Acknowledgment

This research was partially supported by the Chilean National Science Fund through the project Fondecyt-Regular 1131105 and by the VRIEA of the Pontificia Universidad Católica de Valparaíso.

References

[1] K. Stergiou, "Prediction of the mullidae fishery in the easterm mediterranean 24 months in advance," *Fisheries Research*, vol. 9, pp. 67–74, 1996.

[2] K. Stergiou and E. Christou, "Modelling and forecasting annual fisheries catches: comparison of regression, univariate and multivariate time series methods," *Fisheries Research*, vol. 25, pp. 105–138, 1996.

[3] J. Gutierrez, C. Silva, E. Yañez, N. Rodriguez, and I. Pulido, "Monthly catch forecasting of anchovy engraulis ringens in the north area of chile: Nonlinear univariate approach," *Fisheries Research*, vol. 86, pp. 188–200, 2007.

[4] S. P. Garcia, L. B. DeLancey, J. S. Almeida, and R. W. Chapman, "Ecoforecasting in real time for commercial fisheries:the atlantic white shrimp as a case study," *Marine Biology*, vol. 152, pp. 15–24, 2007.

[5] J. F. Adamowski, "Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis," *Journal of Hydrology*, vol. 353, no.3-4, pp. 247–266, 2008.

[6] O. Kisi, "Stream flow forecasting using neuro-wavelet technique," *Hydrological Processes*, vol. 22, no. 20, pp. 4142–4152, 2008.

[7] N. Amjady and F. Keyniaa, "Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method," *International Journal of Electrical Power Energy Systems*, vol. 30, pp. 533–546, 2008.

[8] Z. Bai-Ling, C. Richard, A. J. Marwan, D. Dominik, and F. Barry, "Multiresolution forecasting for futures trading using wavelet decompositions," *IEEE Trans. on neural networks*, vol. 12, no.4, 2001.

[9] R. R. Coifman and D. L. Donoho, "Translation-invariant denoising, wavelets and statistics," *Springer Lecture Notes in Statistics*, vol. 103, pp. 125–150, 1995.

[10] G. Nason and B. Silverman, "The stationary wavelet transform and some statistical applications, wavelets and statistics," *Springer Lecture Notes in Statistics*, vol. 103, pp. 281–300, 1995.

[11] J.-C. Pesquet, H. Krim, and H. Carfantan, "Time-invariant orthonormal wavelet representations," *IEEE Trans. on Signal Processing*, vol. 44, no. 8,, pp. 1964–1970, 1996.

[12] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge, England: Cambridge University Press, 2000.

[13] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the marquardt algorithm," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1996.

A Comparison between Two Metaheuristics Applied to the Cell Formation Problem with Alternative Routings

Orlando Durán A., Luis Pérez P., and Felipe Olmos de Aguilera

***Abstract*—This work proposes a genetic algorithm for optimization of the cell formation problem with alternative routings. A series of test problems were generated and used to evaluate the performance of the proposed Genetic Algorithm and a Simulated Annealing algorithm as well. The novelty of the proposed work lies in the representation technique and the transformations that allow treating the original multidimensional problem as a two-dimensional one. That simplified the programming tasks and the resolution method.**

***Index Terms*—Manufacturing systems design, manufacturing cells, computational intelligence, genetic algorithms, simulated annealing.**

I. Introduction

Cellular manufacturing is an organizational approach based on group technology (GT) that structures a plant as a set of manufacturing cells. Each cell consists in a set of production equipment that process similar components. Cellular based organizations provide considerable cost and productivity benefits to practical manufacturing environments. A list of advantages derived of grouping machines into cells can be found in [1]. The main challenge in the design of manufacturing cells is the identification of machines and components that will make part of a cell. This identification process requires an effective approach to form part families so that similarity within a part family can be maximized. Clustering analysis is the most frequently used method for manufacturing cell design. However, there is still the challenge of creating an efficient clustering method because the cellular formation problem (CFP) is a NP-complete problem. Additional difficulties arise if one considers that many components could have alternative processing methods or routings. This last aspect increases significantly the complexity of this type of problems.

Manuscript received on August 16, 2014, accepted for publication on September 22, 2014, published on November 15, 2014.

Orlando Durán A. (corresponding author) is with the Pontificia Universidad Católica de Valparaíso, Chile (e-mail: orlando.duran@ucv.cl).

Luis Pérez P. is with the Universidad Técnica Federico Santa María, Chile (e-mail: luis.perez@usm.cl).

Felipe Olmos de Aguilera is with the Universidad Técnica Federico Santa María, Chile (e-mail: felipe.olmosdeaguilera@gmail.com).

A manufacturing cell is an organizational structure that groups similar machines that have similar design features or processing capabilities to constitute more efficient production systems. Each cell is composed of a number of machines so that they can produce and maintain continuous production flows in order to reduce the time lost in transfer between workstations. In Figure 1, a basic outline of a manufacturing cell is shown, where the input of a part, its processing sequence, and its output are displayed.

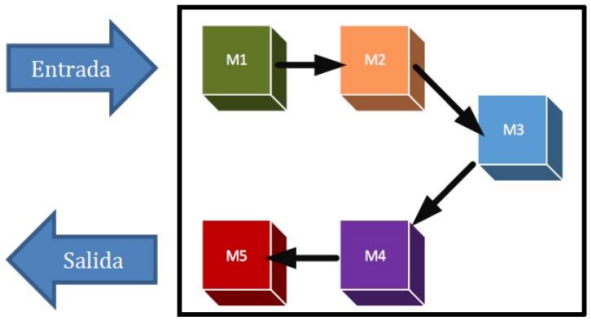


Fig. 1. Magnetization Schematic representation of a manufacturing cell

The most used structure to represent a set of machine-workpiece relationships is the incidence matrix. Figure 2 represents a binary incidence matrix; in which each one of the 1S correspond to the machines that a certain workpiece uses in its production process.

S_{10}	S_{11}	S_{12}	S_{13}	S_{14}	S_{15}	1	2	1	2	3	3
S_{20}	S_{21}	S_{22}	S_{23}	S_{24}	S_{25}	1	1	2	1	2	3
S_{30}	S_{31}	S_{32}	S_{33}	S_{34}	S_{35}	1	2	3	2	1	3
S_{40}	S_{41}	S_{42}	S_{43}	S_{44}	S_{45}	1	2	2	3	1	3

Fig. 2. The incidence matrix

From the incidence matrix in Figure 2 it can be said:

- Workpiece 1 pass through machines 3 and 4.
- Workpiece 2 will go through machines 1 and 2.
- Workpiece 3 will pass through machines 3 and 5.
- Workpiece 4 pass through machines 1 and 2.
- Workpiece 5 pass through machines 3 and 4.
- Workpiece 6 will pass through machines 5 and 6.

To form manufacturing cells it is necessary to apply some clustering method to detect which are the machines and parts to be grouped so as to obtain all the advantages mentioned above. Figure 3 shows how it is an incidence matrix made after the grouping.

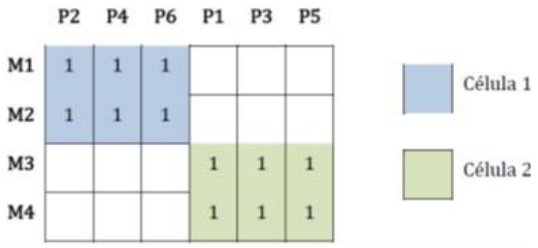


Fig. 3. An illustrative example after the clustering process.

Many works related to manufacturing cell formation assume that each piece has a unique sequence. This is far from reality because any operation on a given workpiece can be performed on alternative machines. This introduces new variables to the cell formation problem. In Figure 4, alternative manufacturing sequences are shown for the same parts and machines of Figure 2.

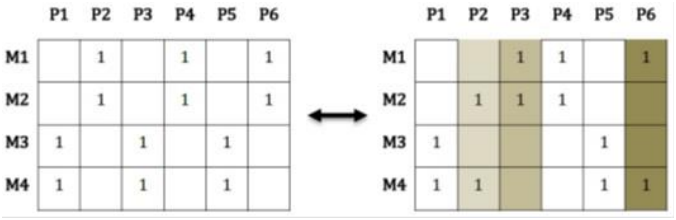


Fig.4. Alternative manufacturing sequences.

From the incidence matrix in Figure 4, it can be said:

- Workpieces 1, 4 and 5 have no alternative routes.
- Workpiece 2 can be processed by machines 1 and 2, or 2 and 4.
- Workpiece 3 can be processed by machines 3 and 4, or 1 and 2.
- Workpiece 6 can be processed by machines 1 and 2, or 1 and 4.

II. Metaheuristics and Machine Grouping

In recent years, different metaheuristic methods have been used to solve the cell-formation problem. An extensive review of the use of metaheuristics in cellular manufacturing is presented by [2]. From the point of view of alternative routings fewer works have been reported in literature. Rodrigues and Weller [3] considered alternative routing to minimize extra-cellular processing of task applying a Tabu search combined with a branch and bound based strategy.

Caux et al. [4] proposed a method that considers alternative routings and machine capacity constraints. The proposed algorithm simultaneously deals with the cell formation problem and the part-routing assignment problem where one of problems was then solved from the solutions of the other. Jayaswal and Adil [5] proposed a SA-based heuristic methodology considering operational sequence, machine replication, alternative process routings to minimize the inter-cell movements and machine investments and operating costs. Wu et al. [6] proposed a hybrid SA method with genetic operation considering alternative process routing and insertion move was utilized in solution improvement stage in order to speed up solution search and to escape from local optima.

Chan et al. [7] presented a multi-objective optimization model using a GA approach to solve the proposed model. Hu and Yasuda [8] addressed the cell formation problem with alternative process routes developing a GA methodology with new chromosome representation, separating crossover heuristic and special mutation technique which produced efficient and optimal solution. Kao and Lin [9] proposed a PSO based algorithm for cell definition. The proposed approach considers a twofold procedure: machine partition and part-routing assignment. Experimental results demonstrated that the algorithm found equal or fewer exceptional elements than existing algorithms for most of the test problems selected from the literature.

III. Chromosome Representation

To model the problem we used a set of vector and matrices that we describe in the following:

- An incidence matrix,(a_{ij}).
- A multilayer matrix where the third dimension is associated to alternative routings for each one of the component (o_{ijn}).
- Machine matrix (y_{ik}).
- Component matrix (z_{jk}).

The multilayer matrix (o_{ijn}) represents the union of a series of alternative routings for every component that make part of the problem, where each layer (n: number of layers) represents an alternative way of manufacture each one of the components. Therefore, the number of layers will be equal to the maximum number of alternative routings that any component has. Figure 5 shows a multilayer matrix with three alternative routings for each component.

Fig. 5. A three-layer incidence matrix.

The algorithm works with a 2-dimension matrix (MxP), thus the initialization process consists in selecting for each one of the component one of the alternative routings from o_{ijn} to construct a 2-dimensional matrix (a_{ij}) that takes part of the

optimization process. Considering, for instance, Figure 5, the matrix shown in Figure 6 is obtained according to the following process: the first row indicates that the component 1 is using the second alternative routing; the component 2 uses the first alternative routing and the components 3 and 4 uses the third alternative routing. The Machine matrix (y_{ik}) represents, with a 1, where each one of the machines (indicated by the row i) is assigned to a specific cell (indicated by the column k). Component matrix (z_{jk}) represents, with a 1, where each one of the components (indicated by the row j) is assigned to a specific cell (indicated by the column k).

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Fig.6. An incidence matrix resulting from the initial transformation.

Therefore, the machine and the component matrices will have many columns as manufacturing cells will be defined. For instance, the representation of a random solution with 2 cells (columns), 4 machines, and 6 components is shown in Figure 7.

1	0
0	1
1	0
0	0
0	0

0	1
1	0
1	0
0	0
0	1
0	1

Fig 7. a) Machine matrix; b) Component matrix.

Figure 7a shows the machine cell and Figure 7b shows the component matrix with 6 components and the same 2 cells. To obtain the incidence matrix (a_{ij}) the union of matrices y_{ik} and z_{jk} is generated. Following, to perform the genetic operators we transformed the matrix into a vector shown in Figure 8.

0	1	0	2	2	1	0	0	2	1	1

Fig.8. Partial chromosome representation

In addition, to the end of the vector a sequence is added with the information corresponding to the columns that were taken of the multilayer matrix to form the incidence matrix. Thus, the starting point will be a vector of length (M + P + P). It will contain the machines and parts locations along with the information of the layer (alternative routing) that it was considered for each one of the workpieces (Figure 9). Concerning the incidence matrix, columns 1, 2 and 5 were taken from the first layer, and the 3, 4 from the second.

0	1	0	2	2	1	0	0	2	1	1	1	1	2	2	1	2

Fig.9. Final chromosome representation

IV. Optimization Model

The fitness function is designed to deliver the number of items that were outside the cells. For this, the incoming vector virtually divided into three parts (see Figure 13) is taken, and from the first M components (M: Number of machines) matrix y is formed (M × C). With the following P components (P: Number of workpieces) the z matrix is formed (P × C), and the last M components determine which columns are taken from the multilayer matrix to form the incidence matrix A.

The cell formation problem with alternative routings can be formulated as follows:

- M be the number of machines,
- P, the number of parts,
- C, the number of cells,
- i, the index of machines ($i = 1, \ldots, M$),
- j, the index of parts ($j = 1, \ldots, P$),
- k, the index of cells ($k = 1, \ldots, C$),
- $A = [a_{ij}]$, M × P binary machine-part incidence matrix,
- M_{max}, the maximum number of machines per cell.

$$y_{ik} = \begin{cases} 1 & \text{if machine } i \in \text{cell } k; \\ 0 & \text{otherwise.} \end{cases}$$

$$z_{jk} = \begin{cases} 1 & \text{if part } j \in \text{family } k; \\ 0 & \text{otherwise.} \end{cases}$$

We selected as the objective function to be minimized the number of times that a given part must be processed by a machine that does not belong to the cell that the part has been assigned to. The problem is represented by the following mathematical model:

Minimize N:

$$N = \sum_{k=1}^{C} \sum_{i=1}^{M} \sum_{j=1}^{P} a_{ij} z_{jk} (1 - y_{ik})$$

Subject to:

$$\sum_{k=1}^{C} y_{ik} = 1 \qquad \forall i,$$

$$\sum_{k=1}^{C} z_{jik} = 1 \qquad \forall j,$$

$$\sum_{i=1}^{M} y_{ik} \leq M_{max} \qquad \forall k .$$

V. Numerical Illustration

The purpose of this section is to show, through numerical examples, how the proposed formulation can be used to design a cellular manufacturing system with alternative routings. To test the performance of the proposed model, we randomly

generated a set of problems with 16 machines, 30 components, and 2 alternative routings.

The genetic algorithm was applied to four randomly generated problems. A set of multilayer matrices filled with 20%, 40%, 60% and 80% of 1s respectively was generated; where 0% represents an empty matrix without operations, and 100% represents a saturated matrix, where each workpiece should go through every machine in the manufacturing system.

A set of tests were performed using the Genetic Algorithm and a Simulated Annealing with the same solution representation. The following control parameters were set to the Simulated Annealing algorithm: Initial Temperature: 8; The logarithmic function was used as the temperature reduction function. In addition, we used the following annealing intervals: 50, 100, 150, y 200. For the Genetic Algorithm, a traditional crossover operator was chosen by which the two parents produce two children and are replaced by them. For the mutation operator, it is applied randomly to 10% of the individuals. Finally, the GA stops when one of two conditions is met: (i) the fitness value of the best individual does not improve after P iterations, or (ii) the total number of iterations exceeds a maximum number ($P \times P$). The use of the number of parts (P) in the stopping criteria draws from the fact that the size of the search space is directly dependent of that number.

TABLE I
COMPARISON OF THE AVERAGE VALUES OBTAINED BY THE GA AND SA.

	$M_{max} = 6$		$M_{max} = 8$		$M_{max} = 10$		$M_{max} = 12$	
	SA	GA	SA	GA	SA	GA	SA	GA
20% of 1s	31,65	25,62	29,65	17,88	26,28	14,3	21,03	8,66
40% of 1s	90,05	82,76	71,4	60,4	58,08	52,86	35,63	33,12
60% of 1s	151,18	144,8	116	108	92,83	89,26	59,65	58,3
80% of 1s	214,4	204,4	163,5	156,7	123,3	125,1	82,4	82

TABLE II
COMPARISON OF THE BEST VALUES OBTAINED BY THE GA AND SA.

	$M_{max} = 6$		$M_{max} = 8$		$M_{max} = 10$		$M_{max} = 12$	
	SA	GA	SA	GA	SA	GA	SA	GA
20% of 1s	29	22	24	13	19	10	14	0
40% of 1s	85	77	63	56	52	49	32	33
60% of 1s	146	137	110	104	90	86	58	58
80% of 1s	210	202	159	154	125	125	82	82

The first test consisted in defining 3 cells with varying the maximum number of machines per cell (Mmax= 6, 8, 10, and 12). In addition, the number of 1s in each initial incidence matrix varied according to the following: 20%, 40%, 60% y 80%; the population size was also tested according to the following: 1,000, 3,000, 5,000, 7,000 and 9,000 individuals. Each test was performed 100 times. In Table I, the average results obtained by both techniques are shown using 3,000 as the population size of the GA. Table II shows the best results obtained by the experiments detailed previously. It can be observed that in every test (except one) GA outperforms the results obtained by the use of SA.

The second test consisted in obtain the arrangements varying the number of cells (2, 3 and 4 cells). In addition, the number of 1s in each initial incidence matrix varied according the same strategy of the first test; the population size was also tested according to the following: 1,000, 3,000, 5,000, 7,000 and 9,000 individuals. Each test was performed 100 times. In Tables 3 and 4 the results obtained by both techniques are shown using 3,000 as the population size of the GA. It can be observed that in every test GA outperforms in average the results obtained by the use of SA (Tables III and IV).

TABLE III
COMPARISON OF THE AVERAGE VALUES OBTAINED BY THE GA AND SA IN THE 2ND TEST.

		2 cells		3 cells		4 cells	
		SA	GA	SA	GA	SA	GA
20% of 1s	Average	17,15	14,14	26,28	14,3	34,18	14,36
40% of 1s	Average	53,88	51,12	57,3	52,34	67,28	52,96
60% of 1s	Average	92,63	87,66	93,93	89,64	100,15	89,96
80% of 1s	Average	129,13	87,66	127,6	125,1	130,93	125,3

TABLE IV
COMPARISON OF THE BEST VALUES OBTAINED BY THE GA AND SA IN THE 2ND TEST ($M_{MAX} = 10$).

2 cells		3 cells		4 cells	
SA	GA	SA	GA	SA	GA
13	11	19	10	25	11
50	48	51	48	61	48
88	86	89	87	94	87
125	125	125	125	125	125

VI. CONCLUSION AND FUTURE RESEARCH DIRECTIONS

A novel representation scheme for solving the cell formation problem with alternative routes is proposed and tested. The proposed technique is feasible and simple. In addition, comparisons were performed. The SA proved that performs well, however the genetic algorithm outperforms the S.A. In addition, we can conclude that this behavior does not depend on the number of 1s in the incidence matrix, i.e. the density of operations or how intensive are the process routings, because in every tested scenario (number of 1s) GA surpasses the SA algorithm. We recommend using genetic algorithm, due to better explore the space on the basis of population, and along with that simulated annealing converges to local minima easily could be better.

The novelty of the proposed work lies in the representation technique and the transformations that allow treating the original multidimensional problem as a two-dimensional one. This simplified the programming tasks and the resolution method.

REFERENCES

[1] H. Selim, R. G. Askin, and A. Vakharia, "Cell Formation in Group Technology: Review, Evaluation, and Direction for Future Research," *Computers and Industrial Engineering*, 34(1), 3–20, 1998.

[2] T. Ghosh, T. S. Sengupta, M. Chattopadhyay, and P. K. Dan, "Meta-heuristics in Cellular Manufacturing: A State-of-the-art Review," *International Journal of Industrial Engineering Computations*, Growing Science Publisher, vol. 2, no. 1, pp. 87–122, 2011.

[3] L. C. A. Rodrigues, and T. R. Weller, "Cell Formation with Alternative Routings and Capacity Considerations: A Hybrid Tabu Search Approach," MICAI 2008, *Lecture Notes in Computer Science*, vol. 5317, pp. 482–491, 2008.

[4] C. Caux, R. Bruniaux, H. Pierreval, "Cell formation with alternative process plans and machine capacity constraints: A new combined approach," *International Journal of Production Economics*, vol. 64, pp. 279–284, 2000.

[5] S. Jayaswal, and G. K. Adil, "Efficient algorithm for cell formation with sequence data, machine replications and alternative process routings," *International Journal of Production Research*, vol. 42, no. 12, pp. 2419–2433, 2004.

[6] T.-H. Wu, J.-Y. Yeh, and C.-C. Chang, "A hybrid simulated annealing algorithm to the cell formation problem with alternative process plans," *International Conference on Convergence Information Technology*, pp. 199–203, 2007.

[7] F. T. S. Chan, K. W. Lau, and P. L. Y. Chan, "A holistic approach to manufacturing cell formation: incorporation of machine flexibility and machine aggregation," *Journal of Engineering Manufacture*, vol. 218, no. B, pp. 1279–1296, 2004.

[8] L. Hu and K. Yasuda, "Minimising material handling cost in cell formation with alternative processing routes by grouping genetic algorithm," *International Journal of Production Research*, vol. 44, no. 11, pp. 2133–2167, 2005.

[9] Y. Kao and Chia-Hsien Lin, "A PSO-based approach to cell formation problems with alternative process routings," *International Journal of Production Research*, vol. 50, no. 15, pp. 4075–4089.

Control de tráfico basado en agentes inteligentes

José A. Castán, Salvador Ibarra, Julio Laria, Javier Guzmán, Emilio Castán

***Resumen*—La tecnología de agentes se ha demostrado ser una ciencia computacional avanzada capaz de lograr mejoras sustanciales en un rango de aplicaciones debido a su paradigma de la estructura de toma de decisiones basado en el razonamiento cognitivo. En este sentido, el artículo presenta el desarrollo de una metodología novedosa que permite incluir un modelo formal basado en agentes autónomos e inteligentes capaces de manipular las fases de los ciclos en una infraestructura de semáforos de acuerdo a las exigencias y limitaciones de la carretera. Este proceso mejora efectiva e inmediata de la calidad del servicio en una intersección, aumentando el rendimiento de la movilidad de los vehículos y mejorando la generación de emisiones, cuando los vehículos se paran en un semáforo rojo. Para corroborar esto, el artículo presenta algunos experimentos con el fin de comparar la metodología propuesta contra una infraestructura pre-programada. Por último, se presentan las conclusiones a destacar la eficacia y la utilidad de la metodología desarrollada con la intención de alcanzar el control de tráfico adecuado de una ciudad en expansión.**

***Palabras Clave*—Sistemas inteligentes, simulación y optimización de vehículos, agentes autónomos.**

Traffic control based on intelligent agents

***Abstract*—Agent technology has been demonstrated to be an advance computational science capable to achieve substantial improvements in a cover range of applications because of its paradigm of decision-making structure based on cognitive reasoning. In this sense, the paper introduces the development of a novel methodology that allows including a formal model founded on autonomous and intelligent agents capable to manipulate the phases of the cycles in a traffic lights infrastructure according to the requirements and constraints of the road. This process improves effectively and immediately the quality of the service in an intersection, increasing the performance of the vehicular mobility and the generation of emissions, when vehicles are stopped in a red light. To corroborate this, the article presents some experiments in order to compare the proposed methodology against a pre-programmed infrastructure. Finally, conclusions are presented to emphasize the effectiveness and usefulness of the developed methodology whit the main intention of achieving an adequate traffic control of an expanding city.**

***Index terms*—Intelligent systems, vehicular simulation and optimization, autonomous agents.**

Manuscrito recibido 18 de marzo de 2014, aceptado para su publicación 27 de abril de 2014, publicado el 15 de noviembre de 2014.

José A. Castán (autor correspondiente), Salvador Ibarra, Julio Laria y Javier Guzmán están afiliados a la Universidad Autónoma de Tamaulipas, Facultad de Ingeniería "Arturo Narro Siller", Tamaulipas, México (correo: {jacastan, sibarram, jlaria, jguzmano}@uat.edu.mx).

Emilio Castán está afiliado con el Instituto Tecnológico de Ciudad Madero, México (correo: ecastan@yahoo.com.mx).

I. Introducción

En la actualidad, las denominadas ciudades de primer mundo enfrentan un sin fin de retos, entre los cuales, el control de tráfico representa uno de los más complejos y significativos. De ahí, que una de las principales inquietudes de los administradores y representantes de las ciudades modernas es lograr el control óptimo de sus vialidades. Por lo tanto, pensar en la tecnología e innovación que les permita conocer dichos retos y al mismo tiempo les represente nuevas oportunidades de mejora, representará un avance significativo en la calidad de vida de sus usuarios.

En este sentido, la innovación que se menciona, se refiere a plataformas que se centran en retos y esfuerzos específicos donde los gobiernos deben asumir un papel para realizar acciones a través de políticas, regulaciones, o medidas fiscales que permitan atacar dichos problemas directamente. Situar la definición de *vialidad* es necesario para ubicar todo lo que esta conlleva, por lo tanto, una vialidad se puede definir como el área o espacio en el que un vehículo puede circular para movilizarse de un lugar a otro. En definitiva, el gran crecimiento demográfico de nuestra civilización sugiere que el consumo y uso de los medios de transporte se incremente. Tal efecto es visible, cuando al realizar una trayectoria que implica poca distancia, el tiempo del viaje es significativamente más elevado del que debiera ser, esto debido a aspectos como lo son: la congestión vial, un mala planeación de las vialidades o unos de los principales problemas estudiados en este artículo, la sincronización semafórica.

Para ofrecer soluciones a estos hechos, en [1] se destaca que durante los últimos 50 años se ha desarrollado una amplia gama de teorías y modelos para optimizar el flujo del tráfico como herramientas para solucionar los problemas económicos y sociales que se originan como consecuencia de una alta demanda del servicio vial.

De hecho, la SCT[1] en México propone un estándar que permite medir la calidad del servicio en las intersecciones

[1] http://www.sct.gob.mx/

viales, el cual es un auxiliar para que los expertos en control de tráfico sean capaces de identificar las áreas de oportunidad de mejora, en la clásica búsqueda de ofrecer al usuario una mejor manera para realizar sus operaciones de transporte. En ese sentido, las intersecciones juegan un papel importante en el desempeño de una red de transporte urbana, ya que representan el punto de convergencia del tránsito vehicular. Además, las intersecciones son las causantes de los cuellos de botella de una red urbana y factores críticos de la capacidad vehicular, la eficiencia y la seguridad de la misma. Esto se debe a que una intersección comprende tanto el área donde dos o más calles se unen (denominado cruce) como todo el espacio destinado a facilitar los movimientos de los vehículos que circulan por ella. En particular, de acuerdo con [3] el estudio de una intersección de dos o más calles, es un ingrediente sustancial para lograr la optimización global del flujo vehicular en una red urbana.

A la luz de esto, debemos prestar atención especial al uso de técnicas revolucionarias como herramienta que permitan solucionar problemas de transporte estático, generando una evolución hacia sistemas de cooperación, los cuales deberían ser suficientemente flexibles para ser aplicados en entornos dinámicos [2, 5, 6]. Tal afirmación es considerada debido a que la dinámica del entorno de una vialidad, es considerada no como un aspecto de movimiento sino en relación a como las condiciones de la vialidad van cambiando conforme el tiempo avanza, efecto que fácilmente se puede apreciar en el comportamiento de una intersección. Por lo tanto, al día de hoy han surgido algunas líneas de implementación que promueven el uso de avanzadas tecnologías de la información para el diseño, estudio y análisis de los sistemas viales.

Así, los *Sistemas Inteligentes de Transporte* (de ahora en adelante *SIT*) aparecen como una nueva oleada de herramientas que permiten optimizar, desde diferentes perspectivas de aplicación, el flujo vehicular. En sí, existen SIT's que conjuntan técnicas computacionales con otras disciplinas como lo son, el control estadístico, modelos de predicción, líneas de espera, entre otras. En esta dirección, se han podido apreciar ciertas ventajas al implementar algún SIT dentro de una red vial, como lo son:

- Mejorar el rendimiento de la vialidad en volumen de servicio.
- Disminuir los tiempos de espera de los vehículos en una luz roja.
- Decremento, por lo tanto, de la generación de gases.
- Evitar las congestiones abrumadoras en horas pico.

Aunque el futuro de los SIT's es prometedor, el campo de aplicación sigue siendo una visión con una perspectiva futurista. Varias aplicaciones y servicios de los SIT actualmente se encuentran trabajando de forma eficaz a lo largo y ancho del mundo, generando una mejora significativa en la seguridad en transporte, la calidad de la movilidad y el nivel de productividad. En este sentido, la amplia y profunda aplicación de los SIT's representa una verdadera revolución en relación al término de la movilidad. Nuevos conceptos, métodos, herramientas y servicios continúan emergiendo en las áreas de comunicación, información, automatización y electrónica, así, los SIT proveen un incremento importante y una plataforma interesante para el campo de la investigación y desarrollo de la Inteligencia Artificial, especialmente aplicando los sistemas inteligentes [2], [4] como herramientas de solución. En esta línea de generación, han aparecido diversos esfuerzos que han permitido evolucionar en la manera de diseñar, simular y analizar las intersecciones viales.

Sin embargo, debido a que las herramientas que actualmente existen presentan algunas carencias, es necesario el surgimiento de nuevos instrumentos que permitan a los analistas de tráfico realizar estudios acerca de los niveles de servicio que una arteria (formando una intersección, ya sea simple o compleja) pudiera presentar y que a la vez sean amigables para el usuario, permitiendo a los SIT's considerar información en tiempo real y trabajar de manera autónoma. Es por este hecho, que surge la necesidad de contar con las herramientas necesarias que permitan a los expertos en control de tráfico, realizar pruebas del mundo real en una herramienta que facilite la integración de múltiples factores y variables del entorno de una vialidad, utilizando métodos, técnicas y modelos fundados y ya bien comprobados en la literatura de control de tráfico. Así, el uso de tecnologías computacionales avanzadas ha surgido como un paradigma de solución en las áreas de los sistemas inteligentes de transporte, monitorizando y optimizando los niveles de servicio del flujo vehicular.

II. TRABAJOS RELACIONADOS

Todas las ciudades en desarrollo deben contar con un sistema de tránsito que presente un control eficiente en sus sistemas de comunicación por tierra, mar o aire. En este caso, el control y el monitoreo del tráfico es uno de los aspectos claves en áreas metropolitanas debido al incremento constante de los vehículos que circulan. Existen numerosos métodos para reducir los retrasos y disminuir los problemas ambientales causados por vehículos automotores. Actualmente la tecnológica que se utiliza para controlar el tráfico esta basada en microcontroladores y microprocesadores. Estos señalamientos tienen ciertas limitantes ya que ellos no son flexibles a las variantes que se presentan durante un día cualquiera. En los sistemas de control de tránsito la detección de las variables y las situaciones que suceden en una intersección son datos de gran importancia para determinar los tiempos de las luces de un semáforo. De acuerdo con esto, en [7] se presenta una aplicación que utiliza lógica difusa para el control multiagente de un semáforo autónomo basado en sensores inalámbricos los cuales permiten evitar problemas de congestión, accidentes, altas velocidades e irregularidades presentadas en el tráfico. De hecho, este enfoque provee una solución minimizando el tiempo de espera de los vehículos de emergencia utilizando un control difuso bajo situaciones que normalmente ocurren en nuestras vialidades.

Asimismo, en [8] se presenta una nueva infraestructura para establecer un control de velocidad inteligente, el cual esta

Fig. 1. Instrumento Virtual para la Simulación y Optimización de Arterias (SiSOA)

utiliza tecnología RFID para la identificación de los requerimientos del tráfico. Este sistema tiene una gran precisión para medir la velocidad de los vehículos ya que utiliza un sensor basado en el efecto Hall. Para lograr una adaptación eficiente de la velocidad de los vehículos este enfoque introduce una adaptación para un controlador difuso. Finalmente los resultados sugieren que siempre existirán situaciones indeseables en las vialidades, observando una mejora en la seguridad de los ocupantes de los vehículos bajo el control inteligente de velocidad. Así en [9] se muestra un control de semáforos inteligentes utilizando un controlador FPGA basado en un sistema neuro-difuso. Tal enfoque es capaz de tomar decisiones para reducir los retrasos en una intersección. En particular la teoría de la lógica difusa en el controlador provee respuestas a los intervalos para las luces verdes en función de la dinámica del tráfico usando las variables de entrada en el flujo del tránsito vehicular de la intersección en cuestión y las vecinas.

Además de las tendencias actuales que se han dedicado a acortar la distancia entre el control de tránsito tradicional y los métodos de computación avanzada, el campo de desarrollo de los entornos de simulación para probar y optimizar la movilidad vehicular es otra interesante y prometedora área para futuros estudios. Hoy en día existen herramientas computacionales que ofrecen la oportunidad de implementar y probar enfoques para corroborar sus ideas. A pesar de esto, dichas herramientas no son funcionales para todas las propuestas.

A la luz de esto, algunos autores tienden a desarrollar sus propios sistemas mientras generan sus propuestas metodológicas. Por ejemplo en [10] se diseño y evaluó un modelo enfocado a las reservaciones para controlar intersecciones. Este enfoque propone el desarrollo de un nuevo simulador para probar dicha metodología. Ciertamente la herramienta integra un simulador microscópico de tráfico con un simulador de redes y un analizador de emisiones de gases de efecto invernadero. Algunos experimentos son presentados utilizando el desarrollo de dicho simulador para comparar la movilidad vehicular y los beneficios medioambientales de la metodología introducida contra métodos tradicionales. Las referencias [11–20] son otros ejemplos en los cuales los investigadores han requerido crear sus propias herramientas computacionales.

Para finalizar, en [21] se presenta una revisión complete de la literatura acerca de enfoques basados en agentes bien situado en el dominio del tráfico y de los sistemas de

transporte. En particular, dicho artículo revisa aplicaciones clasificándolas en cinco categorías:

1) control y administración de tráfico basado en agentes;
2) sistemas basados en agentes para transporte en intersecciones;
3) sistemas basados en agentes para control de tráfico aéreo;
4) sistemas basados en agentes para trenes; y
5) sistemas multiagente para modelar y simular el tráfico vehicular.

Sin embargo, los agentes de software y las técnicas de inteligencia artificial aplicadas al tráfico urbano surgen en la última década como una efectiva solución para proveer altos niveles de efectividad en el control de infraestructuras de transporte.

III. El entorno propuesto

SiSOA (Sistema inteligente para la Simulación y Optimización de Arterias) es una sofisticada herramienta computacional, diseñada e implementada en lenguaje MatLab, la cual nos permite simular el comportamiento de una intersección vial (ver Fig. 1), a partir de datos estadísticos utilizando la teoría de las líneas de espera, para medir los niveles de servicio de dicha intersección con el principal objetivo de optimizar el flujo vehicular.

Como se ha comentado, SiSOA es una herramienta computacional que conjunta la teoría de las líneas de espera con sistemas inteligentes. De esta manera, se puede considerar que cada semáforo dentro de una intersección será "operado" por un agente inteligente, permitiendo de esta manera, que dichos dispositivos reguladores sean capaces de optimizar, de manera autónoma, el flujo vehicular mediante una adaptación de los tiempos en los ciclos de las cajas semafóricas, evitando así todos los problemas implícitos en la congestión vial. En particular, para realizar el estudio de una intersección en SiSOA se requiere ejecutar tres fases antes de simular el servicio vial en una intersección simple de 4 semáforos.

1) Introducir los datos y configuración de la intersección.
2) Establecer el tipo de estudio y el tiempo de la simulación.
3) Comenzar la simulación y verificar el correcto desarrollo de la simulación.

En particular, se define un *paquete* como un convoy de vehículos de diferentes características que arriba a cualquiera de las vías de una intersección. La ideología de utilizar paquetes de vehículos para establecer la densidad de las llegadas a una intersección, se fundamento en base a los resultados obtenidos en los aforos realizados, ya que los tiempos entre una llegada y otra en un análisis individual no era del todo representativo para las cuestiones de estudio de SiSOA. De esto modo, en la tabla 1 se definen los paquetes establecidos de acuerdo a las ocurrencias observadas en el estudio.

Para poder demostrar la utilidad de SiSOA en la optimización del nivel de servicio vial en una vialidad, es necesario efectuar un estudio de las características de la intersección a analizar, por lo tanto, permítanos presentar la estructura geográfica y física de la vialidad, así como los movimientos permitidos para esta zona. En particular, ambas avenidas (Boulevard Adolfo López Mateos y Ejercito Mexicano) son bloques con dos vías de dos carriles cada uno, donde cada una de las vías representa flujo en un solo sentido. Ambas vías presentan 2 tipos de movimientos permitidos, como lo son:

- flujo de vuelta a la izquierda con señalamiento semafórico.
- flujo derecho.

Empíricamente para este estudio, los vehículos que dan vuelta a la derecha, son considerados dentro del conteo de vehículos que van de frente, debido a que dicho servicio de la arteria se puede utilizar en el momento en que el semáforo indica el verde de frente. Para poder generar los datos que se utilizan en el simulador, se realizó un estudio de movimientos en la intersección durante un mes (los 30 días del mes de abril del 2012). Este estudio se efectuó durante una ventana muestral de 7 horas, desde las 7 am hasta las 2 pm. La tabla 2 presenta los datos recolectados y las medias calculadas para uno de los semáforos de la intersección donde:

- μ_H representa la media calculada por hora.
- μ_D representa la media calculada por día.

TABLA 1.
PAQUETES Y VEHÍCULOS QUE LO CONSTITUYEN. (S:SEDAN, C:CAMIONETA, P:PÚBLICOS, T:TAXIS, A, PT: AMBULANCIAS, PATRULLAS)

Id	Clasificación de Vehículos					Total
	S	C	P	T	A y/o Pt	
01	6	3	3	2	0	14
02	7	5	2	2	0	16
03	5	5	3	3	1	17
04	6	4	3	3	0	16
05	5	6	1	4	0	16
06	8	6	1	2	0	17
07	4	4	2	3	0	13
08	6	5	3	3	1	18

TABLA 2.
AFORO DE PAQUETES EN UN SEMÁFORO DE LA VÍA BALM.

HORA	μ_L	μ_{MA}	μ_{MI}	μ_J	μ_V	μ_S	μ_D	μ_H
7-8	93	88	74	81	93	28	5	66
8-9	83	79	81	64	80	20	5	59
9-10	66	67	52	52	58	31	7	48
10-11	67	63	73	58	60	43	26	56
11-12	61	74	71	61	63	30	25	55
12-13	87	77	75	84	71	46	61	72
13-14	108	88	107	88	91	37	55	82
μ_D	81	77	76	70	74	34	26	

TABLA 3.
AFORO DE VEHÍCULOS QUE SALEN DE UN SEMÁFORO.

HORA	TIPO DE MOVIMIENTO	μ_L	μ_{MA}	μ_{MI}	μ_J	μ_V	μ_S	μ_D	μ_H
7-8	V. Izq.	35	42	28	42	28	35	35	35
	Frente	273	210	224	287	259	280	266	257
8-9	V. Izq.	28	28	42	28	42	28	42	34
	Frente	287	252	245	217	210	224	252	241
9-10	V. Izq.	35	28	35	35	35	28	35	33
	Frente	217	287	210	273	280	245	252	252
10-11	V. Izq.	28	28	28	35	42	42	42	35
	Frente	252	224	259	273	266	231	266	253
11-12	V. Izq.	28	42	28	28	42	35	28	33
	Frente	294	252	266	224	252	252	217	251
12-13	V. Izq.	35	35	35	35	42	28	28	34
	Frente	259	273	231	238	266	196	238	243
13-14	V. Izq.	28	35	35	42	28	35	28	33
	Frente	224	245	203	252	238	203	217	226
μ_D		145	142	134	144	145	133	139	

Una vez que se tienen definidas la tasa por día, que es el dato que utilizaremos para esta simulación, para cada esquina de llegada a la intersección, ahora se deben definir las tasas de servicio por cada uno de los semáforos. Para este caso, se consideran 4 semáforos y en base a un estudio de observación se pudo calcular el comportamiento del ciclo semafórico de la intersección. Por ejemplo, uno de los semáforos estudiados presenta verde a la izquierda y verde de frente con una duración de 15 segundos y 30 segundos respectivamente. Es necesario aclarar que en el verde de frente se consideran los 15 segundos del verde a la izquierda. Así, la tabla 3 presenta la información obtenida de uno de los semáforos de la intersección.

Ya que están definidos los paquetes y las llegadas de ellos en base a los aforos, se procede a establecer, para cada uno de los semáforos, las tasas de llegada, las tasas de servicio y los tiempos del ciclo semafórico. Para los alcances de esta investigación, únicamente se utilizarán la distribución normal para la inyección de vehículos en el simulador. En este sentido, la tabla 4 muestra la media y la desviación estándar que se obtuvieron a partir de los datos recolectados durante los aforos.

TABLA 4.
Aforo de Vehículos que salen de un Semáforo.

Datos del Aforo	**Caso de Estudio**			
	Semáforos			
	1	**2**	**3**	**4**
μ	1.70	1.40	1.50	1.10
σ	1.05	1.06	1.07	1.09

El funcionamiento inteligente de SiSOA permite que los semáforos interactúen entre ellos de manera autónoma, en base a un control adaptativo, el cual les permite, a los agentes, intercambiar los tiempos de su ciclo, de acuerdo a los requerimientos y la situación actual de la intersección. Para dejar esto más claro, observe la figura 2 que muestra un esquema general de comunicación entre los semáforos de la intersección para intercambiar tiempo en luz verde.

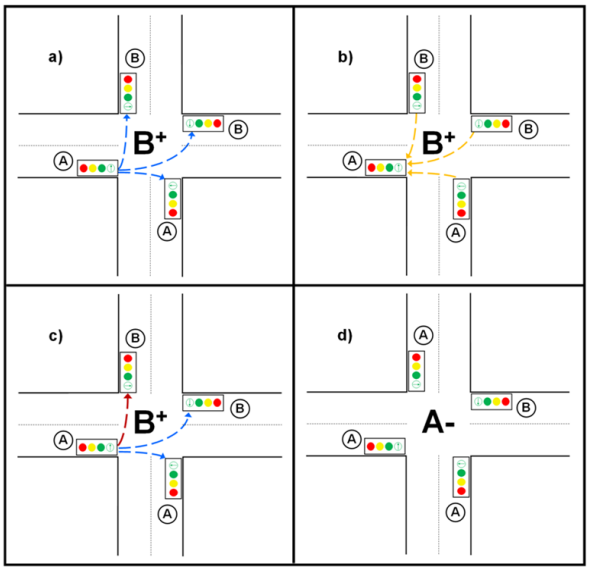


Fig. 2. Esquema General del Proceso de Ajuste del Ciclo Semafórico.

De acuerdo al ejemplo, si un semáforo se percata que en un momento determinado del día el tiempo de su ciclo es muy largo para el nivel de exigencia de la vía, entonces lanza una señal al resto de los semáforos (ver la figura 2a). A su vez, el resto de los semáforos responden su nivel de servicio (de acuerdo a los establecidos por la SCT) al semáforo que ha realizado la oferta (ver la figura 2b). Con esta información, el semáforo que ha realizado la invitación es capaz de decidir

TABLA 5.
RESULTADOS POR SEMÁFORO Y COMPARACIONES DE LAS TASAS DE RENDIMIENTO.

Semáforo	Días del Mes	# de vehículos atendidos por arteria			$RM_{caso} = \left(\frac{RE_{caso} - RE_{comparación}}{RE_{comparación}}\right) * 100\%$		
		RE_{Aforo}	SiSOA				
			RE_{Sin}	RE_{Con}	RE_{Aforo} vs RE_{Sin}	RE_{Aforo} vs RE_{Con}	RE_{Sin} vs RE_{Con}
1	30	15,120	18,553	20,817	22.70	37.67	12.20
2	30	15,212	18,942	21,019	24.52	38.17	10.96
3	30	15,025	18,612	19,972	23.87	32.92	07.30
4	30	15,246	18,103	21,044	18.73	38.02	16.24

(de manera autónoma) cual es el semáforo que requiere más tiempo para su servicio. Ya que ha tomado una decisión, le informa al semáforo en cuestión y al resto de semáforos de la intersección, el tiempo que le cederá y a partir de qué momento lo hará (ver la figura 2c). De esta manera los cuatro semáforos (para este caso en particular) ajustan el tiempo del ciclo semafórico. En tal caso, de manera autónoma los semáforos son capaces de adaptarse a las condiciones y exigencias de la intersección, elevando de esta manera el nivel de servicio de cada una de las arterias, impactando de forma positiva al nivel de servicio promedio de la intersección (ver la figura 2d).

IV. RESULTADOS Y DISCUSIÓN

A partir de los datos obtenidos de las simulaciones realizadas, se han generado algunos resultados preliminares, los cuales permiten apreciar de manera significativa las ventajas que ofrece SiSOA al diseñador y evaluador de control de tráfico. En particular la Tabla 5, muestra el número de vehículos que fueron atendidos en cada uno de los puntos de la intersección analizada durante los treinta (30) días del mes de abril del 2012.

En la tabla se resume la información primeramente por el número de vehículos atendidos por cada vía como el Rango de Efectividad (RE) de cada caso. En segundo punto, podemos apreciar una comparación entre cada uno de os rangos de efectividad para conocer el rango de mejora (RM). Con esta información, se puede enfatizar la aportación al número de vehículos atendidos en cada arteria utilizando la herramienta de simulación SiSOA, más allá de establecer la comunicación entre los semáforos. Sin embargo, también podemos apreciar como el establecer una comunicación e interacción colaborativa entre los dispositivos semafóricos incrementa de manera significativa el rendimiento de las arterias estudiadas.

Además, en la figura 3 se puede observar claramente como la comunicación entre los semáforos resulta efectiva al momento de intentar incrementar el nivel de rendimiento en las tasas de servicio de las vialidades, surgiendo esta tecnología como una herramienta eficaz para solucionar los problemas del sistema vial.

Por otra parte, la figura 4a presenta el consumo de combustible y la generación de tres gases contaminantes (Hidrocarburos HC**;** Monóxido de Carbono CO**;** Oxido Nitrógeno NOx) en un estudio durante 4 diferentes horas de muestreo (1, 4, 6 12 horas de servicio) analizando unidad de gramo por segundo mientras los vehículos están detenidos en una luz roja en donde se compara la metodología propuesta de control inteligente IC contra el servicio ofrecido por una infraestructura pre-programada de control tradicional TC.

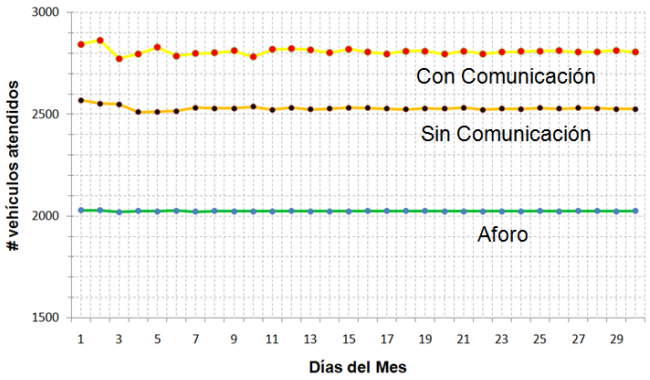


Fig. 3. Resultados preliminares de la Simulación del mes de abril de 2012.

Adicionalmente, la Fig. 4b realiza una comparación de Dióxido de Carbono CO_2 bajo las mismas características que el estudio anterior. Por ejemplo, se puede apreciar en la Fig. 4b la emisión de CO2 bajo IC muestra un mejor rendimiento que TC en alrededor del 78%.

V. CONCLUSIONES Y TRABAJO FUTURO

La congestión en los sistemas de vialidad es una de las causas principales en la baja productividad y en el decremento de los estándares de una ciudad moderna. En este sentido, algunos avances recientes en inteligencia artificial sugieren que algunos vehículos de navegación y sistemas de control de tráfico puedan ser manejados por medio de agentes inteligentes en un futuro muy cercano. Debido a que la

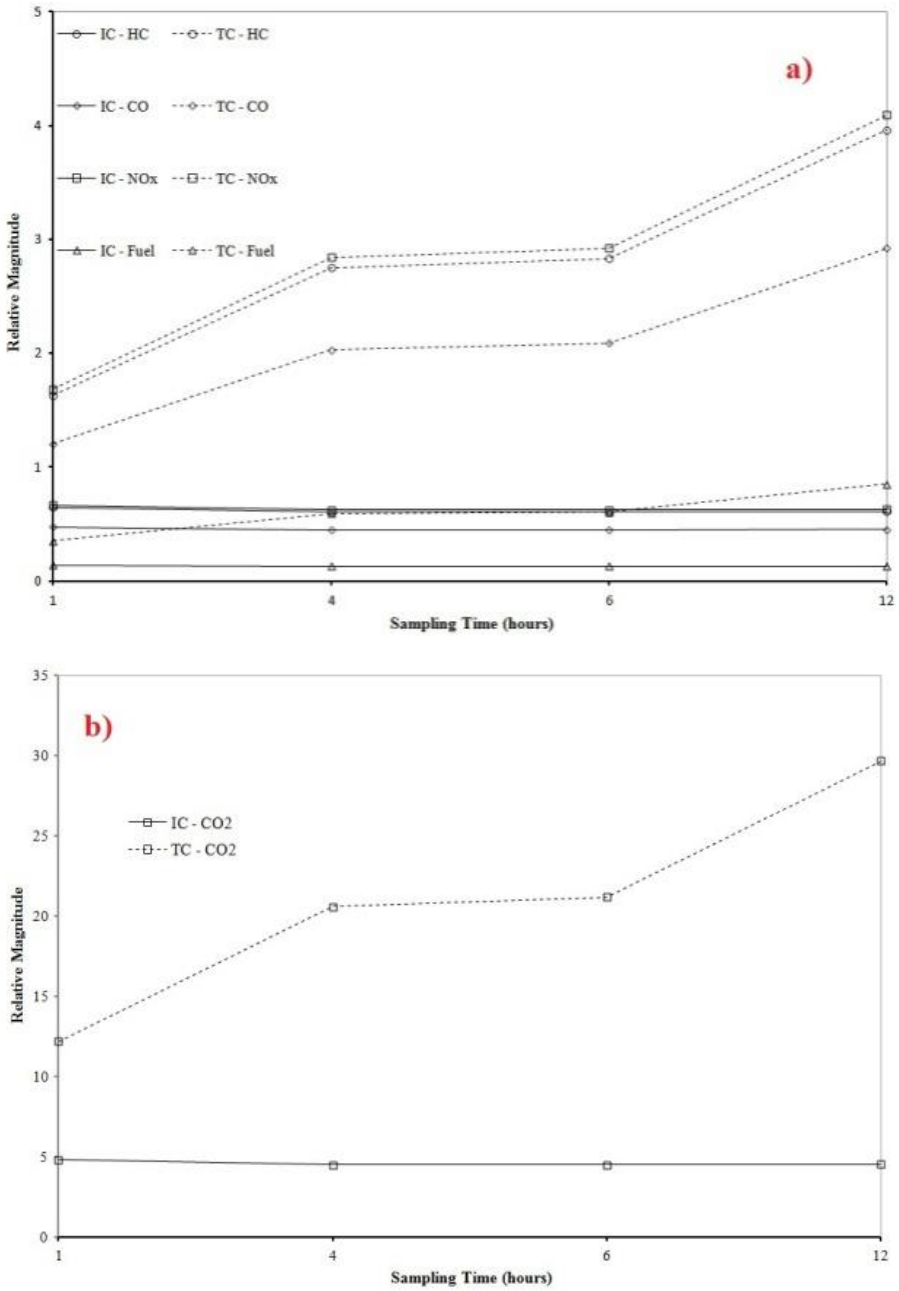


Fig. 4. Magnitud relativa de la emisión de contaminantes y gasto de combustible.

conectividad se vuelve ubicua, el control de tráfico basado en agentes ofrece un enfoque ideal para el manejo de las vialidades, ya que sus características principales de solución de problemas le permiten realizar tareas en áreas geográficamente distribuidas y alternar entre una operación y otra de manera paralela.

Los agentes inteligentes autónomos podrán por lo tanto administrar la información de centros de control de tráfico, caminos, vías de alta velocidad, carreteras, calles, vehículos, casas, oficinas, logrando como resultado una mejora inmediata en el rendimiento de la monitorización de los sistemas de transporte inteligente. Ellos, podrán utilizar el Internet así como redes ad hoc o inalámbricas, para recolectar información en tiempo real con la principal intención de realizar decisiones más seguras.

En este sentido, el surgimiento de SiSOA no es una idea sin fundamentos teóricos o de aplicación, sino que es un aliciente para continuar avanzando realizando esfuerzos significativos por crear herramientas, técnicas, modelos, etc, que permitan optimizar los niveles del servicio de transporte. SiSOA ofrece un nivel de autonomía total por medio de la cual, cada semáforo es capaz de evaluar sus llegadas y sus salidas, con lo

cual un semáforo puede obtener y calcular sus niveles de servicio vial en tiempo real. Más adelante, cada semáforo utiliza dicha información para auto-evaluarse y definir si puede ofrecer tiempo de su ciclo al resto del grupo de semáforos. Si es así, SiSOA permite una interacción directa entre los semáforos de la intersección, estableciendo una comunicación mediante la cual puede alcanzar acuerdos para intercambiar tiempos de su ciclo.

En sí, tal auto-sincronización permite al experto en control de tráfico ser únicamente un monitor del desempeño de las intersecciones viales. En tal caso, SiSOA ha demostrado ser un sistema computacional altamente sofisticado, que conjunta una de las aplicaciones más poderosas de la inteligencia artificial como lo son los agentes inteligentes con la ya multi-probada teoría de líneas de espera, con el principal objetivo de administrar y optimizar los niveles de movilidad y la generación de gases de efecto invernadero en una intersección vial. Finalmente, se puede argumentar que los sistemas inteligentes son una poderosa tecnología computacional que puede ser utilizada de manera eficaz en diversos campos de aplicación, funcionando en conjunto con otras teorías con el principal objetivo de incrementar el rendimiento de sistemas distribuidos situados en ambientes dinámicos trabajando en tiempo real.

AGRADECIMIENTOS

Este artículo ha sido realizado gracias al apoyo otorgado por el Fondo Mixto del Gobierno del Estado de Tamaulipas y el Consejo Nacional de Ciencia y Tecnología (CONACYT) bajo el proyecto TAMPS-2011-C35-186242.

REFERENCIAS

[1] N. A. Chaudhary, Chi-Leung Chu, *Software for Timing Signalized Arterials*, Research Report 4020-1, Texas Transportation Institute, College Station, Texas, September, 2002.

[2] K. Dresner, P. Stone, *A Multiagent Approach to Autonomous Intersecction Management*, 2008.

[3] Nicholas J. Garber, Lester A. Hoel, *Traffic and Highway Enginnering*, West Publishing Company, St. Paul, Mn, 1988.

[4] Y. Hu, P. Thomas, J. Stonier Russel, "Traffic signal control using fuzzy logic and evolutionary algorithms," *IEEE Congress on Evolutionary Computation*, pp.1785–1792, 2007.

[5] K. A. S. Al-Khateeb, J. A. Y. Johari, W. F. Al-Khateeb, "Dynamic Traffic Light Sequence Algorithm Using RFID," *Journal of Computer Science*, vol. 4, no. 7, pp. 517–524, 2008.

[6] W. Wen, "A dynamic and automatic traffic light control system for solving the road congestion problem," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2370–2381, 2008.

[7] Y. Saeed, M. Saleem Khan, K. Ahmed, A. Salam Mubashar, "A Multi-Agent Based Autonomous Traffic Lights Control System Using Fuzzy Control," *International Journal of Scientific & Engineering Research*, vol. 2, no. 6, 2011.

[8] J. Pérez, F. Seco, V. Milanés, A. Jiménez, J. C. Díaz and T. de Pedro, "An RFID-Based Intelligent Vehicle Speed Controller Using Active Traffic Signals," *Sensors*, vol. 10, 2010.

[9] A. R. Zade, D. R. Dandekar, "FPGA Implementation of Intelligent Traffic Signal Controller Based On Neuro-Fuzzy System", *International Conference on Advanced Computing, Communication and Networks*, , Maharashtra, India, 2011.

[10] S. Huang, A. W. Sadek, Y. Zhao, "Assessing the Mobility and Environmental Benefits of Reservation-Based Intelligent Intersections Using an Integrated Simulator," IEEE, *Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, 2012.

[11] R. S. Chen, D. K. Chen, S. Y. Lin, "ACTAM: Cooperative multiagent system architecture for urban traffic signal control," *IEICE Trans. Inf. Syst.*, vol. E88-D, no. 1, pp. 119–126, 2005.

[12] B. Chen, H. H. Cheng, J. Palen, "Mobile-C: A mobile agent platform formobile C/C++ agents," *Softw. Pract. Exp*., vol. 36, no. 15, pp. 1711–1733, 2006.

[13] Mobile-C. [Online]. Available: www.mobilec.org.

[14] A. M. Garcia-Serrano, D. Teruel Vioque, F. Carbone, V. D. Mendez, "FIPA-compliant MAS development for road traffic management with a knowledge-based approach: The TRACK-R agents," in *Proc. Challenges Open Agent Syst. Workshop*, Melbourne, Australia, 2003.

[15] J. Z. Hernandez, S. Ossowski, A. Garcia-Serrano, "Multiagent architectures for intelligent traffic management systems," *Transp. Res. Part C: Emerging Technol*., vol. 10, no. 5/6, pp. 473–506, 2002.

[16] R. T. Van Katwijk, P. Van Koningsbruggen, B. De Schutter, J. Hellendoorn, "Test bed for multiagent control systems in road traffic management," *Transp. Res. Rec*., vol. 1910, pp. 108–115, 2005.

[17] Z. Q. Liu, T. Ishida, H. Y. Sheng, "Multiagent-based demand bus simulation for Shanghai," in *Proc. Massively Multi-Agent Syst. I*, vol. 3446, pp. 309–322, 2005.

[18] F. Y. Wang, "Agent-based control for networked traffic management systems," *IEEE Intell. Syst*., vol. 20, no. 5, pp. 92–96, 2005.

[19] F. Y. Wang, "Toward a revolution in transportation operations: AI for complex systems," *IEEE Intell. Syst*., vol. 23, no. 6, pp. 8–13, 2008.

[20] H. S. Zhang, Y. Zhang, Z. H. Li, D. C. Hu, "Spatial-temporal traffic data analysis based on global data management using MAS," *IEEE Trans. Intell. Transp. Syst*., vol. 5, no. 4, pp. 267–275, 2004.

[21] Bo Chen, Harry H. Cheng, "A Review of the Applications of Agent Technology in Traffic and Transportation Systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 2, 2010.

Acoustic Fingerprint Recognition Using Artificial Neural Networks

Eduardo Zurek, Margarita R. Gamarra, José R. Escorcia, Carlos Gutierrez, Henry Bayona, Roxana P. Pérez, and Xavier García

***Abstract*—This paper presents an implementation of Artificial Neural Networks (ANN) for acoustic fingerprints recognition, applied to the identification of marine vessels. In many cases, the vessel recognition process from an audible signal is performed by human operators, which could lead to failures in the identification process. Before entering the ANN classification process, the signal is filtered and its spectral characteristics are extracted. A comparison of the classification process between three types of neural networks is presented.**

***Index Terms*—Acoustic fingerprint, FFT, PCA, ANN, feed-forward backpropagation, RBF, PNN.**

I. Introduction

The detection of merchant vessels from an audible signal is one of the main tasks within the operations of underwater units. A human operator perform this process, which could lead to failures in the identification process due to factors such as physiological limitations, subjectivity in interpretation, emotional condition and noise present in the signal.

Therefore, poor identification cause misinformation and compromise the integrity of the submarine, the crew aboard it, and even the support staff that is requested for a specific action. Thus, the need for an automatic identification system to assist the operator is imperative to ensure safety and tactical coordination regarding specific operations.

The sound produced by a vessel is mainly caused by its impeller machinery and cavitation caused by the propeller. These sounds are unique to each type of vessel and therefore such information may be used for identification and tracking purposes. This signal is a stationary process given its characteristics of variation and its period [1–3], consequently, signal spectral analysis techniques have proven to be useful to differentiate and classify boats.

Section 2 presents related work. The proposed approach for the recognition of three types of vessels from its acoustic fingerprint is presented in Section 3. It involves the use of the Fast Fourier Transform (FFT) to extract the spectral characteristics of the signal, the Principal Component Analysis (PCA) method to reduce the dimension of extracted features and finally a classifier using ANN. Various tests were performed to evaluate the performance of each type of network in the process and these are presented in Section 4.

II. Related Works

Some techniques implemented for the automatic detection of boats are based on the extraction of features in the frequency domain. Some authors implement FFT, as in [4] to extract features. Using an omnidirectional hydrophone the power spectral density is extracted, for a moving object. With this information, a feed-forward neural network is trained.

Similarly, in [5] FFT is implemented to recognize acoustic fingerprint. This information is compared with a digital soundwaves database, which has been gathered for comparison and identification purposes.

In [6] equipment for the detection of UBA (Underwater Breathing Apparatus) is designed by acoustic signals. The acquired data were processed using Minimum Variance no Distortion Response (MVDR) techniques and the Multiple Signal Classification (MUSIC). Likewise, in [7] information from sonars is processed such that the amplitude spectrum is removed, then PCA is aimed to separate sets of boats, and the most significant components are used as input to a neural network applied for non-linear classification.

However, spectral analysis is not the only method used; autoregressive models are also used. In [8], authors use spectral information to design an autoregressive model or ARMA.

In [9], a method of treatment of the signal is developed to

Manuscript received on August 6, 2014; accepted for publication on October 2, 2014, published on November 15, 2014.

Eduardo Zurek (corresponding author) is with the Universidad del Norte, Department of Systems Engineering, Barranquilla, Colombia (e-mail: ezurek@uninorte.edu.co).

Margarita R. Gamarra is with the Universidad Autónoma del Caribe, Department of Electronic and Telecommunication Engineering, IET group, Barranquilla, Colombia (e-mail: margarita.gamarra@uac.edu.co).

José R. Escorcia is with the Universidad Autónoma del Caribe, Department of Electronic and Telecommunication Engineering, IET group, Barranquilla, Colombia, and with the Escuela Naval de Suboficiales A.R.C. Barranquilla, Colombia (e-mail: jescorcia@uac.edu.co).

Carlos Gutierrez is with the Escuela Naval de Suboficiales A.R.C. Barranquilla, GITIN group, Barranquilla, Colombia (e-mail: ing.cagm@ gmail.com).

Henry Bayona is with the Escuela Naval de Suboficiales A.R.C. Barranquilla, GITIN group, Barranquilla, Colombia (e-mail: elektronikesub@ gmail.com).

Roxana P. Pérez is with the Escuela Naval de Suboficiales A.R.C. Barranquilla, GITIN group, Barranquilla, Colombia (e-mail: roxp64@gmail. com).

Xavier García is with the Escuela Naval de Suboficiales A.R.C. Barranquilla, GITIN group, Barranquilla, Colombia (e-mail: xavier.garciah@ gmail.com).

automatically extract the harmonic structure of the noise radiated by small boats.

Artificial Neural Networks have a particular and important place in the classification stage of many of the strategies described in literature. In particular, [10] uses Kohonen neural network with smoothed spectral data and differences between components of order k as inputs.

In [11], a classification system based on neural networks for object recognition from a sound signal was used. The results suggest that the typical indoor objects can be distinguished on a range of distances with high accuracy based only on the information in the ultrasound echoes.

In [12], a method based on neural networks for detection and classification of underwater mines using sonar images is proposed. The cited articles show a successful classification rate higher than 90% using neural networks.

Likewise, in [13] a neural network scheme with supervised and unsupervised learning is used, so that the performances among these are compared, given the same spectral information in input.

The technological and scientific advances continue developing but despite the high value that acoustic fingerprints vessels have, publications regarding these applications are limited. The researches mentioned in this section present several applications using Neural Networks in the pattern recognition, but the use of techniques in the fingerprint acoustic recognition of vessel is limited. In this article we propose a vessel identification process using spectral features and ANN classifier, from acoustics signals generated by the own vessel.

III. Proposed Approach

In order to identify acoustic fingerprints the stages of the pattern recognition process is developed: filtering, extraction and selection of features and finally classification. In these stages, well-known audio signal processing techniques are implemented. Figure 1 shows the techniques used at each stage and the results.

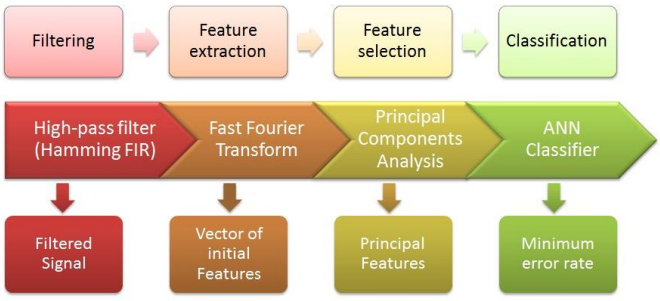


Fig. 1. Scheme of the proposed approach for recognition of acoustic fingerprint.

Initially, a database is generated with the reduced feature space, in order to perform the training of the ANN. Each audio sample is filtered, and then a vector of initial characteristics by applying the FFT to the filtered signal is obtained. The dimension of features is reduced using PCA and finally an ANN performs the classification process. This net is previously trained with the initial database. The next subsections present the techniques used in each stage.

A. *Filtering*

Due to acquisition hardware, acoustic signals have noise in frequency of 60 Hz and its harmonics. However, the representative frequencies generated by each boat are values higher than the 1000 Hz, and then it is necessary to apply a high pass filtering in order to eliminate the low frequency noise. This initial stage is suitable, because the next stages will work only with the representative frequencies of each signal, which facilitates the process of recognition.

In this case, the applied filter is FIR, taking into account the advantage presented in [14]: FIR systems are always stable, they are suited to multi-rate applications and they can design to yield a linear phase response. This filter use a window type Hamming, high-pass, order 100 and a cut-off frequency of 1 kHz. These parameters are selected by testing to the original acoustic signal such that a suitable filtered signal is obtained.

B. *Feature Extraction* [15]

Aim to identify the boat is necessary to extract features of the signal which allows characterize each pattern and classify the boat. The original audio signal is reduced to 3000 samples, which is a representative data frame of the signal. Then convolution is performed with a Hamming window to reduce the undesirable effects of the rectangular window [16]. FFT is applied to the resulting signal; this transform yields real and complex values: the real values represent the distribution of the frequency components while the complex values carry information on the phase of the components [17].

This information is adequate to represent the acquired audio signal. In this stage 512 Fourier coefficients are extracted, but this amount is reduced aim to distinguish the three sources sound using only the discriminant features. Then, it is necessary to use a selecting process to define the most significant features, which allows discriminate the three input signals.

C. *Feature Selection* [15]

In order to reduce the dimensionality of the feature space produced from the FFT algorithm applied to the audio signals, Principal Component Analysis (PCA) is used. PCA is a data analysis technique that is used in order to extract the discriminant features from a large data set [18]. In this case the number of features used in classification is tested from 256 to 64 first principal components aim to select the adequate size.

There are different algorithms for calculating the principal components of a data set, however the one used in this case is based on eigenvalue decomposition.

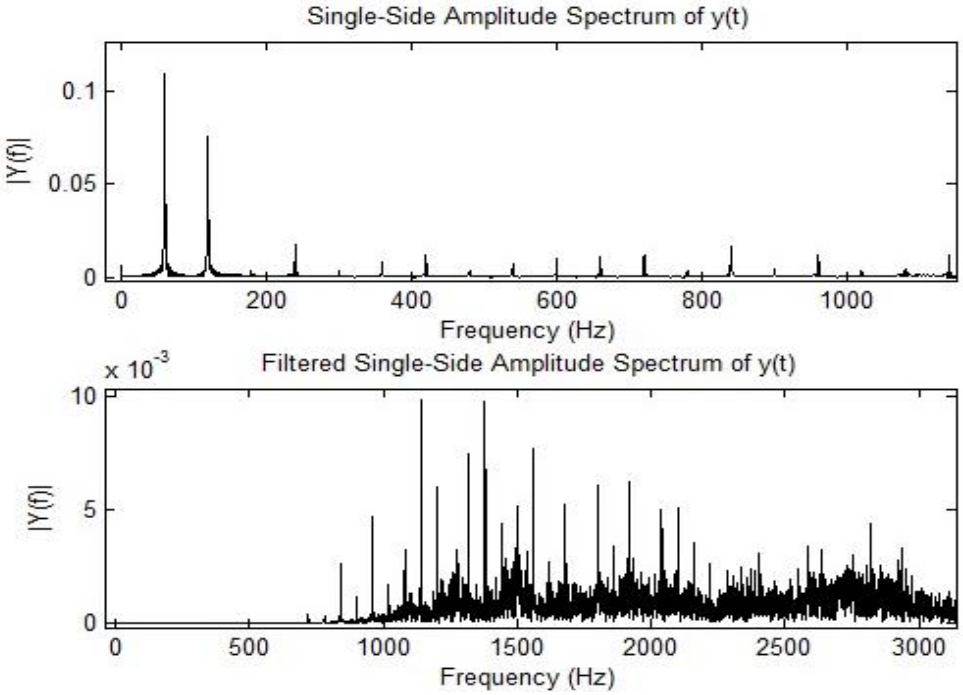


Fig. 2. Spectrum of the acoustic signal from a vessel. Top: original spectrum, below: filtered spectrum

IV. Classification

Once feature selection of all the database signals is completed, these features are used to train the ANN. After several re-training process, the ANN with the highest performance is selected [19].

Other set of acoustic signals are used to measure the performance of the neural network; this set is different to the training set. A comparison of hit rate between three different types of neural network classifier and the k-nearest neighbor (KNN) is performed. These neural networks are Feed-forward Backpropagation, Radial Basis Function (RBF) and Probabilistic Neural Network (PNN).

- Feed-forward networks have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors [20].

- The RBF is a feed forward network, consists of one input layer, hidden layer and the output layer. The RBF is different from the ordinary feed forward networks in calculating the activations of hidden neurons. The activations at the hidden neurons are computed by using the exponential of distance measures [21]. In this case, the net input to the transfer function is the vector distance between its weight vector w and the input vector p, multiplied by the bias *b*. The transfer function for a radial basis neuron is exponential.

- The PNN consists of input layer, two hidden layers, and an output layer. The process based classification that differentiates PNN and RBF is that PNN works on the estimation of probability density function [21]. When an input is presented, the first layer computes distances from the input vector to the training input vectors, and produces a vector whose elements indicate how close the input is to a training input. The second layer sums these contributions for each class of inputs to produce as its net output a vector of probabilities. Finally, a competitive transfer function on the output of the second layer picks the maximum of these probabilities, and produces a 1 for that class and a 0 for the other classes [22].

V. Results

All stages proposed in Figure 1 were performed in Matlab 7.11. In Figure 2, the spectrum of the signal without filtering is shown, where the harmonics of the frequency of 60Hz are predominant and below, the filtered signal is shown where the 60Hz harmonics are eliminated. The used filter was described in subsection III.A.

Although higher frequencies of 1kH, 60Hz harmonics are also present, the dominant signal is typical of each vessel, facilitating the subsequent classification process.

In the training process of the ANN, 1500 samples, 500 of each type of vessel used to identify. To determine the performance of the network created 1500 different samples were used and tested in the classification process.

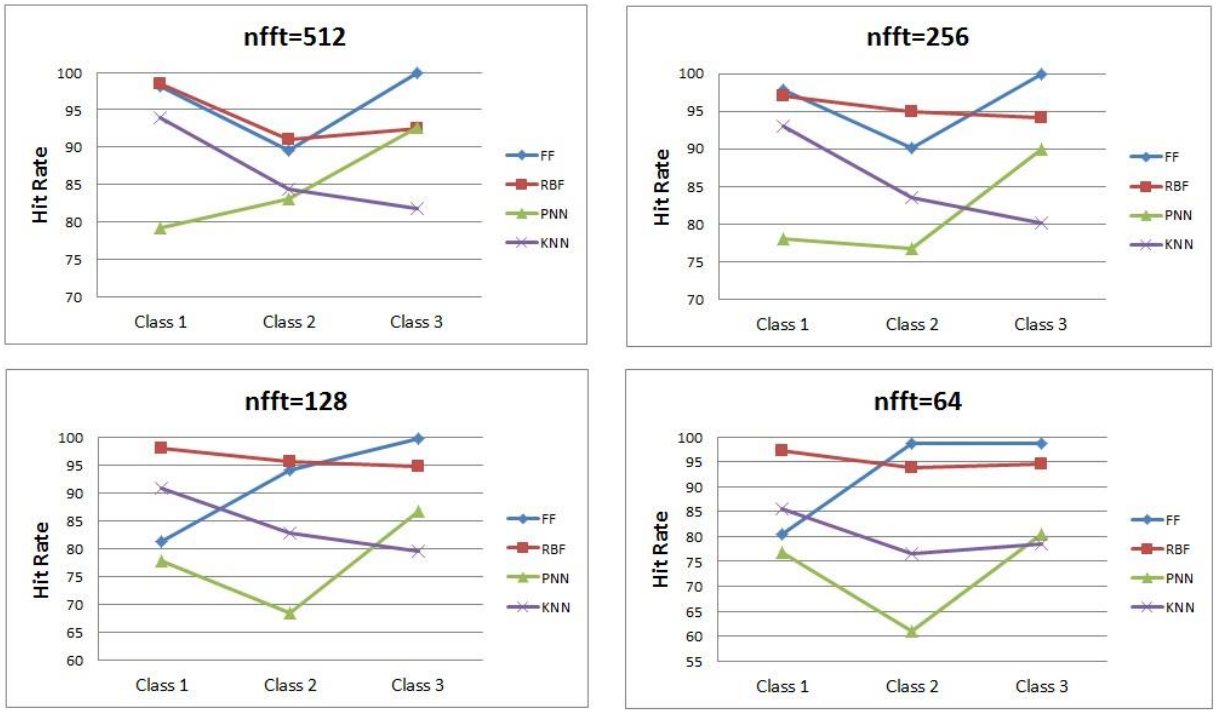


Fig. 3. Graphics of Hit Rate for each classifier

The propagation constant for RBF and PNN was set at 1. The Feed-Forward Backpropagation neural network was set with the following parameters:

- 1 hidden layer sigmoid.
- 20 neurons in the hidden layer.
- Training function: Leveberg-Marquardt optimization.
- Transfer function: Hyperbolic tangent sigmoid.

Table 1 presents a comparison between the hit percentages for each type of neural network classifier and KNN in each case varying the number of Fourier coefficients used (*nfft*).

TABLE 1.
HIT RATE WITH DIFFERENT NFFT

nfft	Feed-Fordward Backpropagation	Radial Basis	Probabilístic	KNN
512	95.9%	93.9%	84.9%	86.73%
256	96.0%	95.4%	81.6%	85.60%
128	91.6%	96.1%	77.6%	84.33%
64	90.7%	95.2%	72.7%	80.27%

Using 256 coefficients Feed-Forward and RBF neural networks presented a higher performance that using 512; this can be explained due to some features (coefficients) are not distinctive and it can conduce to a misclassification.

An adequate threshold for *nfft* can be considered 128 value and upper, but a value higher than 512 could have a lower or equal performance that this value, with more computational cost.

Traditional metrics for the quality of the classification method are the hit rate and the confusion matrix which conveys information relative to this rate [23]. Taking into account the values of the confusion matrix, the graphics of the Figure 3 were obtained. The hit rate of the classifiers for each class of vessel to identify is drawn for different *nfft* values.

This graphics shows that the Feed-forward and Radial Basis classifiers have a higher hit rate that the PNN classifier. For Class 1 and 2, KNN classifier presented better performance that the PNN classifier. In all *nfft* cases, the Feed-forward network had the highest hit rate for the identification of the vessel Class 3.

Then, for this particular recognition case (vessel identification from acoustic fingerprint) using 512 or 256 coefficients, the implementation of Feed-forward and RBF neural networks is suggested; due to these classifiers had a higher hit rate (greater than 90%) than the other ones tested.

VI. CONCLUSION

A spectral analysis was used for boat recognition using their acoustic fingerprint. A previous filtering process is needed in order to reduce the 60Hz noise in the audio signal.

The FFT was calculated using four different numbers of sampling points (nfft) in order to set the value that gets higher performance. The hit rate for four each nfft number was obtained using three different ANN architectures and KNN classifiers.

According to results presented in Table 1 and Figure 3 the Feed-forward and RBF neural networks classifiers have a higher hit rate for three vessel classes, compared with PNN and KNN classifiers.

Using a FIR high-pass filter, 256 coefficients extracted from FFT application and an ANN classifier (Feed-forward or RBF) is possible obtain a good performance in vessel recognition process using the acoustic fingerprint.

It is important to implement the identification proposed approach in an embedded system, which would greatly reduce the processing time and improve the vessel identification process.

ACKNOWLEDGMENT

The work was supported by the Escuela Naval de Suboficiales ARC Barranquilla and Colciencias. The work has been financed by a research grant from Colciencias under call 632, Armada Nacional and Ministerio de Defensa de Colombia: "Desarrollo de un Sistema de Reconocimiento de Huellas Acústicas de Embarcaciones Marinas para la Flotilla de Submarinos de la Armada Nacional Utilizando Técnicas de Inteligencia Artificial".

REFERENCES

[1] R. J. Urick, *Principles of Underwater Sound*. New York: McGraw-Hill, 1983.

[2] C. Regazonni, A. Tesei, and G. Tacconi, "A Comparison Between Spectral and Bispectral Analysis for Ship Detection From Acoustical Timese R. R. Raories," in *IEEE International Conference on Acoustics, Speech and Signl Processing*, Adelaide, 1994.

[3] L. A. Pflug; G. E. Loup, and P. Jackson, "Variability in Higher Order Statistics of Measured Shallow-water Shippin Noise," in *IEEE Signal Processing Workshop on Higher-Order Statistics*, Banff, 1997.

[4] R. H. Baran and J. P. Coughlin, "A Neural Network for Target Classificacion Using Passive Sonar," in *Proceedings of the Conf. on Analysis of Neural Networks Application*, Assoc. Comput. Mach., 1991.

[5] A. Hernandez, *Implementación de un Algoritmo de Generación de Huella Digital Acústica para el Reconocimeinto de Sonidos*, Universidad de la Sierra Juarez, 2013.

[6] L. Fillinger, A. Hunter, M. Zampolli and M. Clarijs, "Passive acoustic detection of closed-circuit underwater breathing apparatus in an operational port environment," *J. Acoust. Soc. Am.*, 2012.

[7] W. Soares-Filho, J. M. Seixas, and L. P. Caloba, "Enlarging Neural Class Deetection Capacity in Passive Sonar Systems," in *IEEE Int. Symp. on Circuits ans Systems*, Scottsdale, 2002.

[8] K. Eom, M. Wellman, N. Srour, D. Hillis, and R. Chellappa, "Acoustic Target Calssification Using Multscale Methods," in *Sensors and Electron Devices Symp*., 1997.

[9] G. Ogden, M. Zurk, M. Jones, M. Peterson, "Extraction of Small Boat Harmonic Signatures from Passive Sonar," *J. Acoust. Soc. Am.*, 2011.

[10] J. Meister, "A Neural Network Harmonic Family Classifier," *J. Acoust. Soc Am*, vol. 93, no. 3, pp. 1488–1495, 1993.

[11] W. Streilein, P. Gaudiano, and G. Carpenter, "A neural network for object recognition through sonar on a mobile robot," in *Proceedings of Intelligent Control (ISIC)*, held jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA), Intelligent Systems and Semiotics (ISAS), 1998.

[12] S. Geethalakshmi, P. Subashini, and S. Ramya, "A Study on Detection and Classification of Underwater Mines Using Neural Networks," *Int. J. Soft Comput. Eng. IJSCE*, vol. 1, no. 5, pp. 150–157, 2006.

[13] B. Howell, S. Wood, and S. Koksal, "Passive sonar Recognition and analysis using hubrid neural networks," *IEEE Oceans*, vol. 4, pp. 1917–1924, 2003.

[14] "Lecture 8. DT Filter Design: IIR Filters," in *Discrete-Time Signal Processing*, Massachusetts Institute of Technology, 2006.

[15] E. Zurek, M. Gamarra, J. Escorcia, C. Gutierrez, R. Pérez, X. García, H.. Bayona, "Spectral Analysis Techniques for Acoustic Fingerprints Recognition," in *2014 XIX Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, IEEE, 2014.

[16] A. C. Sanabria, C. Predraza, J. Vitola, "Quick Algorithm for Searching Audio by Content on GPUs," *Intekhnia*, vol. 6, no. 1, pp. 35–43, 2011.

[17] D. Mitrovic, M. Zeppelzauer, C. Breiteneder, "Features for Contet-Based Audio Retrieval," *Adv. Comput.*, vol. 78, pp. 71–150, 2010.

[18] Z. Burka, *Perceptual Audio Calssification Using Principal Component Analysis*, Golisano College of Computing and Information Sciences, Rochester, 2010.

[19] M. R. Gamarra, C. Quintero, "Using genetic algorithm feature selection in neural classification systems for image pattern recognition," *Ing. E Investig.*, vol. 33, no. 1, pp. 52–58, 2013.

[20] "Neural Networks Toolbox," *Czech Technical University*. [Online]. Available: http://radio.feld.cvut.cz/matlab/toolbox/nnet/backpr52.html.

[21] T. Sitamahalakshmi, A. Vinay, M. Jagadeesh, and K. Chandra, "Performance Comparison of Radial Basis Function Networks and Probabilistic Neural Networks for Telugu Character Recognition," *Glob. J. Comput. Sci. Technol.*, vol. 11, no. 4, 2011.

[22] *Radial Basis Networks*. [Online]. Available: http://www.inf.ufrgs.br/~engel/data/media/file/cmp121/rbf-matlab.pdf.

[23] R. C. Prati, G. Batista, M. C. Monard, "Evaluating Classifier Using ROC Curves," *IEEE Latin America Transactions*, vol. 6, no. 2, pp. 215–222, 2008.

Journal Information and Instructions for Authors

I. Journal Information

Polibits is a half-yearly open-access research journal published since 1989 by the *Centro de Innovación y Desarrollo Tecnológico en Cómputo* (CIDETEC: Center of Innovation and Technological Development in Computing) of the *Instituto Politécnico Nacional* (IPN: National Polytechnic Institute), Mexico City, Mexico.

The journal has double-blind review procedure. It publishes papers in English and Spanish (with abstract in English). Publication has no cost for the authors.

A. Main Topics of Interest

The journal publishes research papers in all areas of computer science and computer engineering, with emphasis on applied research. The main topics of interest include, but are not limited to, the following:

- Artificial Intelligence
- Natural Language Processing
- Fuzzy Logic
- Computer Vision
- Multiagent Systems
- Bioinformatics
- Neural Networks
- Evolutionary Algorithms
- Knowledge Representation
- Expert Systems
- Intelligent Interfaces
- Multimedia and Virtual Reality
- Machine Learning
- Pattern Recognition
- Intelligent Tutoring Systems
- Semantic Web
- Robotics
- Geo-processing
- Database Systems
- Data Mining
- Software Engineering
- Web Design
- Compilers
- Formal Languages
- Operating Systems
- Distributed Systems
- Parallelism
- Real Time Systems
- Algorithm Theory
- Scientific Computing
- High-Performance Computing
- Networks and Connectivity
- Cryptography
- Informatics Security
- Digital Systems Design
- Digital Signal Processing
- Control Systems
- Virtual Instrumentation
- Computer Architectures

B. Indexing

The journal is listed in the list of excellence of the CONACYT (Mexican Ministry of Science) and indexed in the following international indices: LatIndex, SciELO, Periódica, e-revistas, and Cabell's Directories.

There are currently only two Mexican computer science journals recognized by the CONACYT in its list of excellence, *Polibits* being one of them.

II. Instructions for Authors

A. Submission

Papers ready for peer review are received through the Web submission system on www.easychair.org/conferences/?conf=polibits1; see also updated information on the web page of the journal, www.cidetec.ipn.mx/polibits.

The papers can be written in English or Spanish. In case of Spanish, author names, abstract, and keywords must be provided in both Spanish and English; in recent issues of the journal you can find examples of how they are formatted.

The papers should be structures in a way traditional for scientific paper. Only full papers are reviewed; abstracts are not considered as submissions. The review procedure is double-blind. Therefore, papers should be submitted without names and affiliations of the authors and without any other data that reveal the authors' identity.

For review, a PDF file is to be submitted. In case of acceptance, the authors will need to upload the source code of the paper, either Microsoft Word or LaTeX with all supplementary files necessary for compilation. Upon acceptance notification, the authors receive further instructions on uploading the camera-ready source files.

Papers can be submitted at any moment; if accepted, the paper will be scheduled for inclusion in one of forthcoming issues, according to availability and the size of backlog.

See more detailed information at the website of the journal.

B. Format

The papers should be submitted in the format of the IEEE Transactions 8x11 2-column format, see http://www.ieee.org/publications_standards/publications/authors/author_templates.html. (while the journal uses this format for submissions, it is in no way affiliated with, or endorsed by, IEEE). The actual publication format differs from the one mentioned above; the papers will be adjusted by the editorial team.

There is no specific page limit: we welcome both short and long papers, provided that the quality and novelty of the paper adequately justifies its length. Usually the papers are between 10 and 20 pages; much shorter papers often do not offer sufficient detail to justify publication.

The editors keep the right to copyedit or modify the format and style of the final version of the paper if necessary.

See more detailed information at the website of the journal.