Smart Innovation, Systems and Technologies 201

Yuzo Iano · Rangel Arthur · Osamu Saotome · Guillermo Kemper · Reinaldo Padilha França *Editors*

Proceedings of the 5th Brazilian Technology Symposium

Emerging Trends, Issues, and Challenges in the Brazilian Technology, Volume 1

International

Springer

Smart Innovation, Systems and Technologies

Volume 201

Series Editors

Robert J. Howlett, Bournemouth University and KES International, Shoreham-by-sea, UK

Lakhmi C. Jain, Faculty of Engineering and Information Technology, Centre for Artificial Intelligence, University of Technology Sydney, Sydney, NSW, Australia

The Smart Innovation, Systems and Technologies book series encompasses the topics of knowledge, intelligence, innovation and sustainability. The aim of the series is to make available a platform for the publication of books on all aspects of single and multi-disciplinary research on these themes in order to make the latest results available in a readily-accessible form. Volumes on interdisciplinary research combining two or more of these areas is particularly sought.

The series covers systems and paradigms that employ knowledge and intelligence in a broad sense. Its scope is systems having embedded knowledge and intelligence, which may be applied to the solution of world problems in industry, the environment and the community. It also focusses on the knowledge-transfer methodologies and innovation strategies employed to make this happen effectively. The combination of intelligent systems tools and a broad range of applications introduces a need for a synergy of disciplines from science, technology, business and the humanities. The series will include conference proceedings, edited collections, monographs, handbooks, reference books, and other relevant types of book in areas of science and technology where smart systems and technologies can offer innovative solutions.

High quality content is an essential feature for all book proposals accepted for the series. It is expected that editors of all accepted volumes will ensure that contributions are subjected to an appropriate level of reviewing process and adhere to KES quality principles.

Indexed by SCOPUS, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST), SCImago, DBLP.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/8767

Yuzo Iano · Rangel Arthur · Osamu Saotome ·
Guillermo Kemper · Reinaldo Padilha França
Editors

Proceedings of the 5th Brazilian Technology Symposium

Emerging Trends, Issues, and Challenges in the Brazilian Technology, Volume 1

Springer

Editors
Yuzo Iano
School of Electrical and Computer
Engineering (FEEC)
UNICAMP
Campinas, São Paulo, Brazil

Osamu Saotome
Divisão de Engenharia Eletrônica
Instituto Tecnológico de Aeronáutica
São José dos Campos, São Paulo, Brazil

Reinaldo Padilha França
School of Electrical and Computer
Engineering (FEEC)
UNICAMP
Campinas, São Paulo, Brazil

Rangel Arthur
School of Technology (FT)
UNICAMP
Limeira, São Paulo, Brazil

Guillermo Kemper
Universidad Peruana de Ciencias Aplicada
Lima, Peru

Associate-Editors
Ana Carolina Borges Monteiro
School of Electrical and Computer
Engineering
UNICAMP
Campinas, São Paulo, Brazil

Diego Pajuelo Castro
School of Electrical and Computer
Engineering
UNICAMP
Campinas, São Paulo, Brazil

Gabriel Gomes de Oliveira
School of Electrical and Computer
Engineering
UNICAMP
Campinas, São Paulo, Brazil

ISSN 2190-3018 ISSN 2190-3026 (electronic)
Smart Innovation, Systems and Technologies
ISBN 978-3-030-57547-2 ISBN 978-3-030-57548-9 (eBook)
https://doi.org/10.1007/978-3-030-57548-9

© Springer Nature Switzerland AG 2021
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Committees

Organizing and Executive Committee

Yuzo Iano, LCV/DECOM/FEEC/UNICAMP, General Chair BTSym & WSGE
Osamu Saotome, ITA, Associate-General Chair BTSym
Rangel Arthur, FT/UNICAMP, Vice-General Chair BTSym
Carlos Arturo Raymundo Ibáñez, UPC, Associate-General Chair BTSym
Guillermo Leopoldo Kemper Vásquez, UPC, Associate-General Chair BTSym
Fernando Hashimoto, DMU/UNICAMP, Associate-General Chair BTSym
Rogério Bazi, PUC-Campinas, Associate-General Chair BTSym
Gabriel Gomes de Oliveira, LCV/DECOM/FEEC/UNICAMP, Technical Program and Finance Chair
Ana Carolina Borges Monteiro, LCV/DECOM/FEEC/UNICAMP, Proceedings Chair
Reinaldo Padilha França, LCV/DECOM/FEEC/UNICAMP, Proceedings Chair
Telmo Cardoso Lustosa, LCV/DECOM/FEEC/UNICAMP, Local Arrangements Chair
Alysson Gomes de Oliveira, LCV/DECOM/FEEC/UNICAMP, Marketing Chair
Diego Arturo Pajuelo Castro, LCV/DECOM/FEEC/UNICAMP, Proceedings Chair
David Minango, LCV/DECOM/FEEC/UNICAMP, Registration Chair
Lisber Arana, LCV/DECOM/FEEC/UNICAMP, Institutional Relationship Chair
Alex R. Ruelas, LCV/DECOM/FEEC/UNICAMP, Institutional Relationship Chair
Abel Dueñas Rodríguez, LCV/DECOM/FEEC/UNICAMP, Media Chair

Scientific and Academic Committee

Osamu Saotome, ITA
Luiz Cézar Martini, DECOM/FEEC/UNICAMP
David Bianchini, PUC/Campinas

Luis Geraldo Pedroso Meloni, DECOM/FEEC/UNICAMP
Ana Cláudia Seixas, PUC/Campinas Cláudia Seixas
Cristiano Akamine, Universidade Presbiteriana Mackenzie PUC/Campinas
Celso Iwata Frison, PUC/Minas-Poços de Caldas
Luiz Vicente Figueira de Mello Filho, Universidade Presbiteriana Mackenzie
Guillermo Leopoldo Kemper Vásquez, USMP & UNI-INICTEL
Lia Toledo Moreira Mota, PUC/Campinas
Lucas Heitzmann Gabrielli, DECOM/FEEC/UNICAMP
Edgard Luciano Oliveira da Silva, EST/UEA
Talía Simões dos Santos, FT/UNICAMP
Janito Vaqueiro Ferreira, DMC/FEM/UNICAMP
Vlademir de Jesus Silva Oliveira, UNEMAT/Sinop
Navid Razmjooy, Tafresh University, Tafresh, Iran
Hugo Enrique Hernandez Figueroa, DECOM/FEEC/UNICAMP
Antônio José da Silva Neto, IPRJ/UERJ
Bruno Sanches Masiero, DECOM/FEEC/UNICAMP
Marcos Antonio do Nascimento Guimarães, UNIP/CAMPINAS, JUNDIAÍ
Alessandra Akkari, Universidade Presbiteriana Mackenzie
Maria Thereza de Moraes Gomes Rosa, Universidade Presbiteriana Mackenzie
Angela del Pilar Flores Granados, FEA/UNICAMP
Paolo Bacega, Faculdade Anhanguera
Marcos Fernando Espindola, IFSP São Paulo
Polyane Alves Santos, Instituto Federal Da Bahia
Vicente Idalberto Becerra Sablón, USF/UNIMEP
Grimaldo Wilfredo Quispe Santivañez, UNAAT, Tarma, Perú
Jessie Leila Bravo Jaico, UNPRG, Lambayeque, Perú
Ernesto Karlo Celi Arévalo, UNPRG, Lambayeque, Perú
Nestor Adolfo Mamani Macedo, UNMSM, Lima, Perú
Luz María Paucar-Menacho, UNS, Ancash, Perú
Edwin Alberto Valencia Castillo, UNC, Cajamarca, Perú
Christian Del Carpio, UPC, Lima, Perú
Gustavo Huaman Bustamante, Inictel/UNI, Lima, Perú
Itamar Salazar Reque, Inictel/UNI, Lima, Perú

Technical Reviewers Committee

Rangel Arthur, INOVA/FT/UNICAMP
Ana Carolina Borges Monteiro, LCV/DECOM/FEEC/UNICAMP
Reinaldo Padilha França, LCV/DECOM/FEEC/UNICAMP
Adão Boava, Universidade Federal de Santa Catarina, UFSC
Amilton Lamas, PUC-Campinas
Agord de Matos Pinto Júnior, DESIF/FEEC/UNICAMP
Angela del Pilar Flores Granados, FEA/UNICAMP

Joaquim Marcos Santa Rita da Silva, Instituto Nacional de Telecomunicações
Alysson Gomes de Oliveira, LCV/DECOM/FEEC/UNICAMP
José Alexandre Nalon, Centro Universitário Salesiano São Paulo, UNISAL
Murilo Cesar Perin Briganti, LCV/DECOM/FEEC/UNICAMP
Luigi Ciambarella Filho, Universidade Veiga de Almeida/Develop Biotechnology
Lucas Alves, LCV/DECOM/FEEC/UNICAMP
Ingrid Araujo Sampaio, Universidade Estadual de Campinas
Denise Helena Lombardo Ferreira, PUC Campinas
Hermes José Loschi, LCV/DECOM/FEEC/UNICAMP
Daniel Rodrigues Ferraz Izario, LCV/DECOM/FEEC/UNICAMP
Mariana Carvalho, DCA/FEEC/UNICAMP
Diego Pajuelo, LCV/DECOM/FEEC/UNICAMP
Douglas Aguiar do Nascimento, LCV/DECOM/FEEC/UNICAMP
Edson José Gonçalves, LCV/DECOM/FEEC/UNICAMP
Alessandra Akkari, Universidade Presbiteriana Mackenzie
Polyane Alves Santos, Instituto Federal Da Bahia
Kelem Christine Pereira Jordão, LCV/DECOM/FEEC/UNICAMP
Josué Marcos de Moura Cardoso, LCV/DECOM/FEEC/UNICAMP
Giovanni Moura de Holanda, FITec, Technological Innovations, Brazil
Euclides Lourenço Chuma, LCV/DECOM/FEEC/UNICAMP
Gabriel Gomes de Oliveira, LCV/DECOM/FEEC/UNICAMP
José Yauri, DCA/FEEC/UNICAMP
Julío Humberto León Ruiz, LCV/DECOM/FEEC/UNICAMP
Navid Razmjooy, Tafresh University, Tafresh, Iran
Antônio José da Silva Neto, IPRJ/UERJ
Raphael Ronald, CTI (Centro de Tecnologia da Informação)
Raquel Jahara Lobosco, UFRJ/Macaé
Necesio Gomes Costa, UFRJ
Claudia Maria Moraes Santos, UNIVAP
Vitor Chaves de Oliveira, IFSP São Paulo/Mackenzie
Gustavo Mesones, UPC, Lima, Perú
Julio Ronceros, UPC, Lima, Perú
Kalun Lau, UPC, Lima, Perú
Manuel Márquez, UPC, Lima, Perú
Mirko Klusmann, UPC, Lima, Perú
Nikolai Vinces Ramos, UPC, Lima, Perú
Sergio Salas, UPC, Lima, Perú

Sponsor

Beta Telecommunications

Foreword

It is with great satisfaction that I write this foreword to the book, Proceedings of the 5th Brazilian Technology Symposium (BTSym'19), held at the Mackenzie Presbyterian University, SP, Brazil, in October 2019. This event is in its fifth edition and has consolidated to become an excellent opportunity for researchers, teachers, and students to present and discuss the results of their research works.

The 2019 edition of BTSym is characterized by the broad scope of exposed areas, with papers dealing with current and important topics for Brazilian and world technological development, including subjects related to the various branches of human, engineering, smart and sustainable future of cities, architecture, biomedical and computer science.

Events such as BTSym are an essential part of the research and innovation process. Firstly, these events contribute to the promotion of research activities, which are key to a country's technological development. The dissemination of research results, as promoted by BTSym, contributions to the transformation of research findings into technological innovation. In addition, these events facilitate the sharing of findings, eventually leading to the formation of research networks, which accelerate the achievement of new results.

Therefore, I congratulate the BTSym General Chair, Prof. Dr. Yuzo Iano, and his group of collaborators for the important initiative of organizing the BTSym 2019 and for providing the opportunity for authors to present their work to a wide audience through this publication.

Finally, I congratulate the authors for the high-quality research papers and scientific experiments presented in this proceedings.

Best wishes,
Ana Carolina Borges Monteiro
Researcher and Associate-Editor
University of Campinas
Campinas, Brazil

Preface

This book contains the Proceedings of the 5th Brazilian Technology Symposium, an event held in Campinas, SP, Brazil, in October 2019. This book was divided into two parts: the first part on Emerging Trends in Systems Engineering and Telecommunications and the second part on Emerging Trends in Industry 4.0, Smart and Sustainable Future of Cities.

The Brazilian Technology Symposium is an excellent forum for presentations and discussions of the latest results of projects and development research, in several areas of knowledge, in scientific and technological scope, including smart designs, sustainability, inclusion, future technologies, architecture and urbanism, computer science, information science, industrial design, aerospace engineering, agricultural engineering, biomedical engineering, civil engineering, control, and automation engineering, production engineering, electrical engineering, mechanical engineering, naval and oceanic engineering, nuclear engineering, chemical engineering, probability, and statistics.

This event seeks to bring together researchers, students, and professionals from the industrial and academic sectors, seeking to create and/or strengthen the linkages between issues of joint interest. Participants were invited to submit research papers with methodologies and results achieved in scientific-level research projects, completion of course work for graduation, dissertations, and theses.

The 60 full chapters accepted for this book were selected from 210 submissions, and in each case, the authors were shepherded by an experienced researcher, with a rigorous peer review process. Among the main topics covered in this book, we can highlight natural language processing, artificial neural network, physical ergonomics, blockchain, communication technologies, Industry 4.0, computational intelligence, fuzzy logic, mentoring process, molecular interactions, fermentation process, simulation, digital image processing, pattern recognition, machine learning, magnetic field visualization, frameworks developments, physiological parameters applications, computational vision, data science, spatial-temporal parameters, measuring system, sensory analysis, precision agriculture, electrocardiogram, autonomic nervous system, content analysis, statistical process control, measurement uncertainty, model predictive control, IoE, IoT, smart cities, electrical installations, photocatalysis, X-ray, lexical tone, industrial environment, workstations, sensors, autonomous cars, PCB, data privacy, fluid dynamics, HPV, quantum computing, and much more.

We hope you enjoy and take advantage of this book and feel motivated to submit your papers in the future to Brazilian Technology Symposium.

Campinas, Brazil

Best wishes,
Reinaldo Padilha França
Proceedings Chair of Brazilian
Technology Symposium

Acknowledgments

Our appreciation goes to a lot of colleagues and friends who assisted in the development of this book, Proceedings of the 5th Brazilian Technology Symposium (BTSym'19).

First of all, I would like to thank all the members of the Organizing and Executive Committee, for the commitment throughout the year, several meetings were held, and many challenges were overcome for the accomplishment of the BTSym 2019.

Also, and with great merit, I would like to thank all the members of the Scientific and Academic Committee and Technical Reviewers Committee, for their excellent work, which was essential to ensure the quality of our peer review process, and collaborating with the visibility and technical quality of the BTSym 2019.

The Brazilian Technology Symposium is an event created by the Laboratory of Visual Communications of the Faculty of Electrical and Computer Engineering of the University of Campinas (UNICAMP). In this way, I would like to thank the Mackenzie Presbyterian University, especially for the support and hosting of the BTSym 2019, which was critical for the success of their accomplishment.

Beta Telecommunications played a crucial role in holding the BTSym 2019, in the same way as the Pro-Rectory of Extension and Culture at the University of Campinas, especially Prof. Fernando Augusto de Almeida Hashimoto, due to financial support from it, and it was possible to consolidate with quality many BTSym 2019 organization aspects, which ensured the quality to support the authors and speakers.

Finally, I thank all the authors for their participation in the BTSym 2019, I sincerely hope to have provided an experience that was very useful and enriching in the personal and professional lives of everyone, and my special thanks go to Prof. Rangel Arthur. Due to the continuous support in activities that enrich the academic environment through research directed to innovative technological development.

He is a generous, charismatic, dedicated, determined, and fair person, being an excellent fellow in academic research that certainly contributes a lot to BTSym, who has a lot to thank Prof Rangel Arthur for.

Best wishes,
Prof. Yuzo Iano
General Chair of Brazilian
Technology Symposium

Contents

Emerging Trends in Systems Engineering and Telecommunications

HoG Multi-face Detection

Leandro Duarte and Cláriton Bernadelli

Abstract Human face detection is important in many applications, such as human–machine interface, automatic surveillance, and facial recognition. This work exposes a solid and general face detection system capable of detecting multiple faces in the same image, even in low light situations and chaotic backgrounds. The detection system uses a representation of the Gaussian pyramid and evaluates it in all scales the existence of faces using descriptors of HoG characteristics and linear classifiers SVM. The system shows that the gradient distribution in the face contours is sufficiently discriminative to distinguish faces and non-faces and the use of cascade detectors improves overall system performance by decreasing the number of false positives. Employing experimental tests, the methodology was applied to facial and non-facial test images, allowing the evaluation of the effectiveness of the face detection system and the influence of adjustable parameters on the accuracy and performance of the system.

Keywords Face detection · HoG · Image processing · SVM

1 Introduction

The detection of human faces in generic images is a great challenge. The variability of facial features requires a reinforced set of attributes to clearly describe the shape of the human face, especially in situations of variable or precarious lighting and chaotic backgrounds. The faces have a wide range of aspects, colors, textures, and many other intrapersonal characteristics, yet their contours are similar and can be sufficiently discriminative to distinguish faces from non-faces, as evidenced by [1].

L. Duarte (✉)
Federal University of Uberlândia, Uberlândia, MG, Brazil
e-mail: leandro.duarte@ufu.br

C. Bernadelli
Federal University of Triângulo Mineiro, Uberaba, MG, Brazil
e-mail: clariton.bernadelli@uftm.edu.br

© Springer Nature Switzerland AG 2021
Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_1

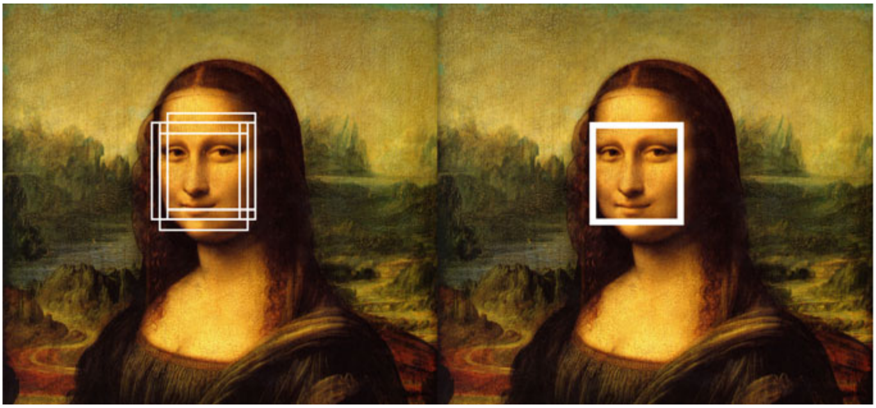


Fig. 1 Example detection: Left image without suppression of multiple detections and the right image with suppression of multiple detections

The face detection algorithms that use the characteristic descriptor histogram of oriented gradients (HoG) and the linear classifier support vector machine (SVM) proved to be effective and generalist, as they focus on the contours of the face within the image. The HoG characteristic descriptor does not directly detect the contours of the face, but the normal vector of the hyperplane obtained in the training phase of the SVM puts great weights on the characteristics of the contours of the face [2, 3]. For the face detection system to detect faces of varying sizes and anywhere in an image, a Gaussian pyramid representation is adopted in order to make the system invariant to scale [4]. A sliding window selects each portion of the pyramid images to check for the presence or not of a face [3].

It is common during the detection process that a face is detected several times in the same image, as shown in the overlays in Fig. 1. It is also common that there are false positives during detection. This work uses a methodology of suppression of multiple detections based on the proximity of the centroids of the detections, in addition to using a second set HoG and SVM cascade. The first classifier has a lower computational cost and performs its inferences from a reduced set of HoG characteristics, searching for possible faces in the image. The second classifier, possessing a more robust set of HoG characteristics, performs an empirically more assertive inference on the possible faces, indicated in the first classifier.

2 Method Summary

The basic idea for using the HoG characteristic descriptor and the SVM linear classifier is that the overall appearance of the faces and their shape can be characterized by the distribution of gradients and directions of the edges [3, 5].

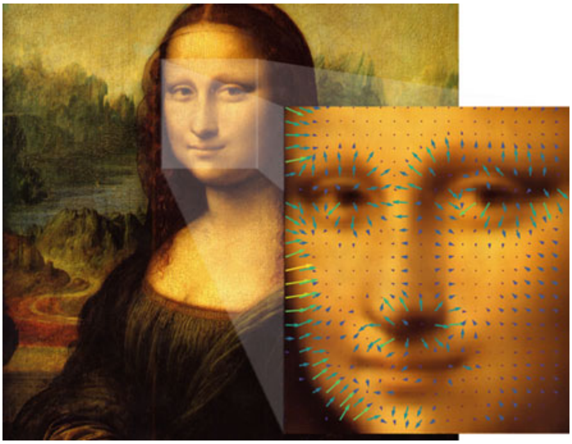


Fig. 2 Representation of the magnitude and direction of the gradient of an image cell. The highlighted rectangle delimits the region of the test image

The practical implementation is done by dividing the image contained in the region delimited by the sliding window into small spatial regions called cells. For each of the cells, a 1-D histogram is created that discriminates through the gradients of magnitude and direction the orientation of the edges in that cell [6], as shown in Fig. 2.

The contrast of the local histograms is normalized to make the representation less susceptible to variations in lighting and shading. This is done by normalizing the histograms of each cell in a larger spatial region, called a block. These normalized blocks are called histogram of oriented gradients descriptors (HoG). When these descriptors are combined with an SVM classifier, it is possible to perform the detection of frontal faces in generic images [7].

2.1 Invariant Representation at Scale

The Gaussian pyramid is shown in Fig. 3 allows the face detection system to become invariant to scale. In the pyramid, subsequent images are weighted and reduced using a Gaussian average. Each pixel is formed by a local average that corresponds to a neighborhood of pixels at a lower level of the pyramid. In [4], Burt and Andelson proposed a pyramidal structure in which the analysis operators are composed of two stages. The first stage corresponds to the low-pass filtration, using a 5×5 size kernel with a format like that of the probability Gaussian distribution, in order to eliminate high frequency. The second stage is sampling with bilinear interpolation.

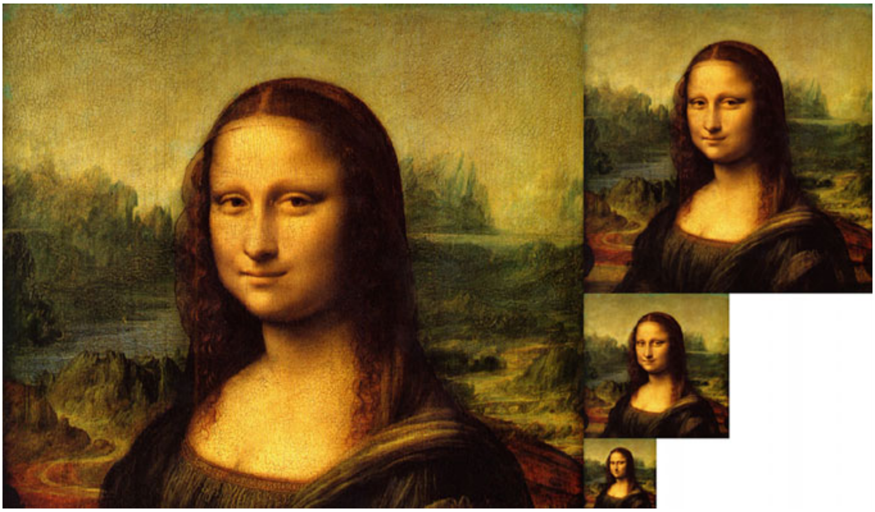


Fig. 3 Representation of the Gaussian pyramid where subsequent images are weighted using a Gaussian average and reduced

2.2 Computation of Gradients

The gradient of an image represents the directional change in the intensity of the pixels, Fig. 4. At each point of the image, there is a 2D vector with the components given by the derivatives in the horizontal and vertical directions. The direction of the resulting vector points to the direction of increased intensity, while its module shows the magnitude of the intensity change in this direction. Thus, the gradient of an image is the vector of its partial directional derivatives [6], as shown in Eqs. 1 and 2. The direction and magnitude gradient can be calculated as shown in Eqs. 3 and 4.

$$\frac{\delta f}{\delta x} = \begin{matrix} -1 \\ 0 \\ +1 \end{matrix} * \boldsymbol{A} \tag{1}$$

$$\frac{\delta f}{\delta y} = \left[-1 \ 0 \ +1 \right] * \boldsymbol{A} \tag{2}$$

$$\theta = \tan^{-1} \frac{g_y}{g_x} \tag{3}$$

$$|\nabla f(x, y)| = \sqrt{g_x^2 + g_y^2} \tag{4}$$

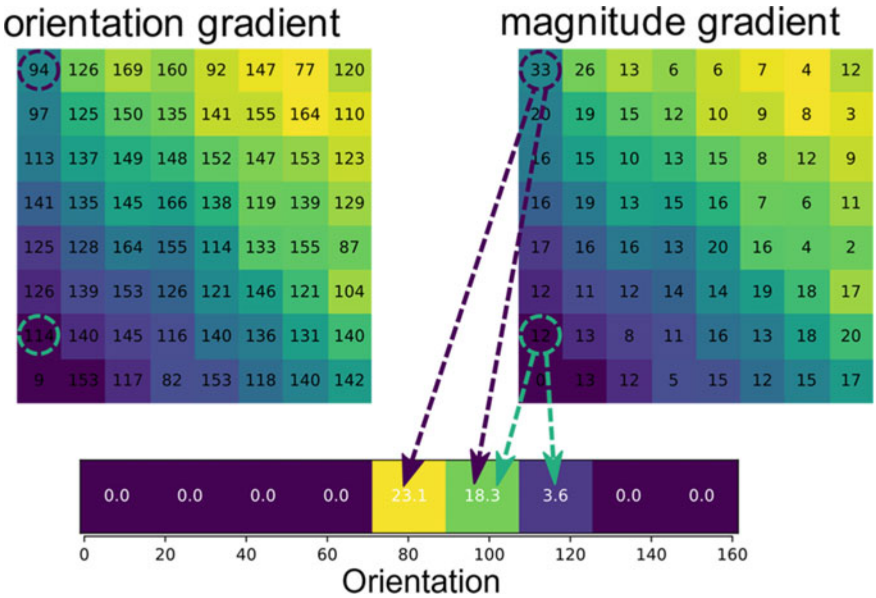


Fig. 4 Weighted voting of magnitude and direction gradients of a cell over an unsigned histogram

2.3 Creation of Histograms

Gradient intensities should be locally normalized to improve descriptor invariability to changes in illumination and contrast. This normalization requires the grouping of cells into larger and spatially connected blocks. The HoG descriptor is then the concatenated vector of the histogram components [1, 2].

This work explores four different methods for the normalization of blocks as described in (5)–(8).

$$L_2 - \text{norm} : f = \frac{u}{\sqrt{\|u\|_2^2 + e^2}} \tag{5}$$

$$L_2 - \text{hyst}: f \text{:norm } L_2 \tag{6}$$

$$L_1 - \text{norm} : f = \frac{u}{\|u\|_1 + e} \tag{7}$$

$$L_1 - \text{sqrt} : f = \sqrt{\frac{u}{\|u\|_1 + e^2}} \tag{8}$$

where: u is the non-normalized vector that contains the histograms of a given block, u_k is its K-norm and is a small constant in order to avoid numerical errors.

2.4 Support Vector Machine

The support vector machine (SVM) is supervised learning models that analyze data for classification [7, 8]. Given a set of labeled training $\{\boldsymbol{x}_i, y_i\}$, $i = 1, \ldots, N$, $y_i \in \{+1, -1\}$, $x_i \in \mathbb{R}^d$, a linear SVM seeks to find an ideal hyperplane $\boldsymbol{w}^T \boldsymbol{x} + b = 0$ that creates a maximum geometric margin γ:

$$\gamma = \frac{1}{\|\langle \boldsymbol{w} \rangle\|_2} = \frac{1}{\langle \boldsymbol{w}.\boldsymbol{w} \rangle} \tag{9}$$

where $\langle . \rangle$ is the internal product between two vectors.

If the training samples are linearly separable the problem of SVM optimization can be formulated as $\min_{w,b} \langle \boldsymbol{w}.\boldsymbol{w} \rangle$, subject to $y_i(\langle \boldsymbol{w}.\boldsymbol{x_i} \rangle + b) \geq 1,\ i = 1, \ldots, N$. If the training samples have noise or outliers, they may not be linearly separable. To tolerate noise and outliers, the adjustment variables ξ_i are introduced, and the corresponding soft margin of the SVM becomes $\min_{2w,b} \langle \boldsymbol{w}.\boldsymbol{w} \rangle + C \sum_{i=1}^{N} \xi_i^2$, subject to $y_i(\langle \boldsymbol{w}.\boldsymbol{x_i} \rangle + b) \geq 1 - \xi_i,\ i = 1, \ldots, N$. Where C is a free parameter determined by a separate validation set or by a cross-validation technique.

2.5 Suppression of Multiple Detections

Multiple detections of the same face are inevitable when using the sliding window method [9]. In fact, it is expected and desired that a true face has several detections because it indicates a greater probability of the existence of a face in that region. However, it is necessary to suppress multiple detections. Suppression can be based on the Euclidean distance of the detection's centers, in the dimensions of the box delimitating the detection and probability indicated by the SVM [2, 9].

Each detection creates a structure that contains the position, dimension, and probability of the inferred window, using this information to create collations that gather all detections, whose detection boxes are overlapped. The number of collations is the number of faces in the image. To determine the position and dimensions of the detection box, that is, the region of the image that contains the face, the weighted average of the parameters of all overlapping detections are performed, as shown (10)–(14).

$$x_i = \frac{\sum_{k=1}^{N} x_k.p_k}{\sum_{k=1}^{N} p_k} \tag{10}$$

$$y_i = \frac{\sum_{k=1}^{N} y_k.p_k}{\sum_{k=1}^{N} p_k} \tag{11}$$

$$w_i = \frac{\sum_{k=1}^{N} w_k . p_k}{\sum_{k=1}^{N} p_k} \tag{12}$$

$$h_i = \frac{\sum_{k=1}^{N} h_k . p_k}{\sum_{k=1}^{N} p_k} \tag{13}$$

$$p_i = \frac{\sum_{k=1}^{N} p_k}{N} \tag{14}$$

where (x_i, y_i) are the coordinates resulting from the center of the i-th grouping, (w_i, h_i) are the width and height of the detection box of the i-th grouping and p_i is the average probability of i- the group is a face.

3 Methodology and Image Database

The main proposal of this work is to show that the adoption of cascade detectors increases the performance of the detection system in relation to the computational cost while maintaining the efficiency and robustness of detection. In this scenario, two cascading detectors are proposed, as shown in Fig. 5. The first of them with a lower computational cost search for regions in the image that may contain faces. The second detector, cascading with the first, will analyze only the faces previously detected.

3.1 Training Database

To train SVM classifiers, we used the HoG features of 2432 front face images from the Yale Database Face B (cropped) database [10]. In addition, 500 non-faces images are used from the USCD Campus Images stock which, after cropping, total 2500

Fig. 5 Scheme implementation of cascading detectors

Table 1 Training and evaluation image database

Image database	Images number
Yale Face Database B(cropped)	2432
Yale Face Database B(original)	5600
USCD Campus Images	500

Table 2 HoG features extractor's attributes

Attributes	HoG1	HoG2
Pixels per cell	16	8
Cells per block	2	4
Histogram channels (bins)	9	9
Direction gradient	(0° to 180°)	(0° to 180°)
Normalization	L_2-norm	L_2-norm

Table 3 SVM classifier attributes

Attributes	SVM1	SVM2
HoG face samples	1216	1216
HoG non-face samples	1250	1250
C	1	1
Kernel	Linear	Linear

images. For the performance appraisal process, a set of 5600 images from the Yale Face Database B (original) image database were used as shown in Table 1.

In the first stage of training, two training sets were created, one for each detector. Each set has 1216 face samples and 1250 non-face samples. Each sample from both sets goes through the HoG feature extraction process, and each set has a different feature extractor according to Table 2.

Thus, the result of the first step is two HoG feature training sets, each with 2466 samples. These sets are separated into two parts, the first consisting of 90% of the samples that are used to update the hyperplane with the SVM algorithm and the second part consisting of the other 10% of the samples used to test model convergence. The summary of SVM classifiers is presented in Table 3.

3.2 Cascade Detectors

The parameters chosen to create the HoG feature extractor are based on the work of Dalal and Triggs [2] and the system performance itself.

The flow of execution of the proposed methodology, shown in Fig. 5, can be presented as follows: Any image of any size is received by the face detection system. Then, its Gaussian pyramid is built. A fixed-size sliding window runs through all the images in the pyramid. The region delimited by the window is submitted to the lowest computational cost HoG 1 descriptor. The feature vector from this feature extraction is subject to classification by SVM 1. If this region is labeled as a face, a similar process is performed again by extracting a new feature set by the HoG 2 extractor which has a higher computational cost, and again, this new feature set is submitted to the second classifier, SVM 2.

All windows that pass through the two detectors are considered faces and undergo the multi-detection suppression process where repeated face detections are eliminated.

4 Implementation and Performance

This section presents the details of the implementation and the systematic study of the effects of the various possible detector configurations. Several studies discuss the best settings for the HoG characteristic descriptor and for the SVM classifier [2, 7, 8]. Therefore, this work focuses mainly on the specific characteristics of the detection system and how they influence the detector result.

Figure 6a shows the sensitivity of the system to change in the number of orientation channels (bins), while Fig. 6b shows the influence of the kernel on the SVM classifier.

For purposes of comparison between the various system configurations possible will be used as a metric the wrong detection (ratio sum of false positives and false negatives with the total number of images of faces and not faces) and correct detection

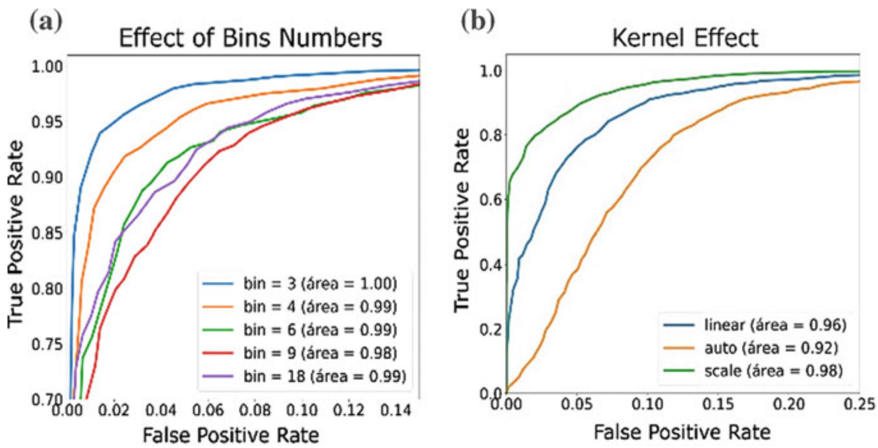


Fig. 6 System sensitivity **a** Number of orientation bins **b** SVM kernel

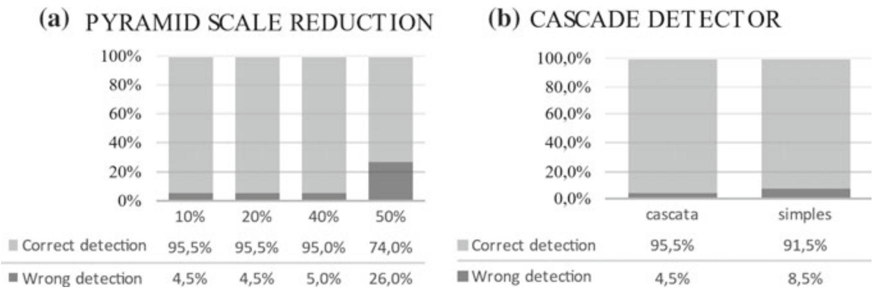


Fig. 7 System sensitivity **a** Cascade detector, **b** pyramid scale reduction

(ratio sum of detections positive and correct negatives with the total number of face and non-face images).

4.1 Image Pyramid Scale Reduction

The system proved to be very sensitive to the step of reducing the scale of the image pyramid. When the reduction step is very large, the scale invariability is reduced, that is, the image pyramid will have fewer layers and it may happen that some faces do not find a sliding window that fits.

In contrast, a very small detection step creates excessive layers, increasing the computational cost of the system. Figure 7a shows that for the reduction steps 10, 20, and 40%, there is no significant difference in performance. However, when contrasting with a reduction of 50% or more, among the layers of the pyramid, the performance losses proved significant.

4.2 Cascade Detector

To make the system more immune to the presence of false positives, this work proposes the implementation of a second HoG descriptor and SVM classifier set that reinforces or discards detections made by the first detector. The deployment of this second cascade detector is justified by reducing more than half of the wrong detections or by improving the correct detection compared to the single detector system, as shown in Fig. 7b.

5 Conclusions

The human face detection system using two cascade detectors, HoG feature puller, and SVM classifier presented and analyzed for their performance in this work proved to be viable, robust, and reliable. The simple methodology enables other applications to use this method as a step-in larger system or in embedded systems for a wide range of applications. Even with the current trend of using deep learning methods for object classification in general, the method presented here is still viable mainly when the number of training samples is small.

References

1. Li, J., Li, B., Xu, Y., Lu, K., Yan K., Fei, L.: Disguised face detection and recognition under the complex background. In: IEEE Symposium on Computational Intelligence in Biometrics and Identity Management (CIBIM). Orlando, FL, USA, pp. 87–93 (2014)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). San Diego, CA, USA (2005), pp. 886–893
3. Shu, C., Ding X., Fang. C.: Histogram of the oriented gradient for face recognition. Tsinghua Sci. Technol. **44**, 2122–2133 (2016)
4. Adelson, E., Anderson, C., Bergen, J., Burt P.J., Ogden, J.M.: Pyramid Methods in Image Processing. RCA Laboratories (1984)
5. Pang, Y., Zhang, K., Yuan, Y., Wang, K.: Distributed object detection with linear SVMs. IEEE Trans. Cybern. **16**, 216–224 (2014)
6. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn. Pearson Education, Upper Saddle River, New Jersey, USA (2002)
7. Chang, C., Lin, C.: LIBSVM. ACM Trans. Intell. Syst. Technol. 2(3), 1–27 (2011)
8. Mathur A., Foody, G.M.: Multiclass and binary SVM classification: implications for training and classification users. IEEE Geosci. Remote Sens. Lett. **5**(2), 241–245 (2008)
9. Viola P., Jones, M.: Robust real-time object detection. In: Second International Workshop on Statistical and Computational Theories of Vision, Canada (2001)
10. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: illumination cone models for face recognition under variable lighting and Pose. IEEE Trans. Pattern Anal. Mach. Intell. **23**, 643–660 (2008)

Glasses with Artificial Vision for Reading Texts

Ernesto Chalaco, David Basantes, Juan Carlos Molina, and Gustavo Caiza

Abstract Glasses with artificial vision for reading texts allow people with visual disabilities to perform much better activities in this society, accessing all kinds of written information and obtaining an adequate inclusion in the daily activities and the education. Glasses are equipped with an embedded system, with the software of text recognition and audio playback. This prototype uses a word recognition system with a Pi Cam that distinguishes the word electronically by means of a simple board computer, Raspberry Pi 3B+ and emits the sound of the word to headphones. Audio-assisted text reading glasses were designed to be a support tool for non-sighted people, and it is a low-cost device with easy functionality, which allows independence in the user. The performance tests for the validation of the prototype were carried out with different sizes and font letters; in addition, it was obtained that font sizes higher than 16 and print letters must be used.

Keywords Artificial vision · Glasses · Visual disabilities

1 Introduction

According to the data of the World Health Organization refers to visual impairment, the estimated number of people living with a type of visual impairment globally is approximately 1.3 billion, out of these people, 217 million have moderate to severe visual problems, and 36 million are blind. Most people with poor vision are over

E. Chalaco · D. Basantes · J. C. Molina · G. Caiza (✉)
Universidad Politécnica Salesiana, UPS, 170146 Quito, Ecuador
e-mail: gcaiza@ups.edu.ec

E. Chalaco
e-mail: echalaco@est.ups.edu.ec

D. Basantes
e-mail: dbasantes@est.ups.edu.ec

J. C. Molina
e-mail: jmolinaa1@est.ups.edu.ec

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_2

50 years old [1], the main causes of chronic eye diseases that provoke vision loss are refractive errors and cataracts that did not undergo surgery.

In Ecuador, the National Council for Equality of Disabilities (CONADIS) identifies 461.687 people with some type of disability registered until May 2019, from which 11.81% have visual impairment [2]. Braille code is a primary system for non-sighted who read and write using the sense of touch. The lack of braille text for studying and training and the high costs involved have been a limiting factor so that non-sighted can study.

With figures set out and based on the words of Professor Stephen W. Hawking, "disability should not be an obstacle to success" taken from [3], and several commercial devices have been developed such as IrisVision [4], eSight [5], OrCam [6], which are intended for people who are blind or visually impaired, based on artificial vision and which perform reading processes and other functions depending on the manufacturer. The international price of these devices ranges from \$3,000 to \$6,000 in 2019 based on [4–6], being expensive and difficult to buy for visually impaired Ecuadorians.

Currently, an artificial vision has had a great impact and development. Detection of objects can be found in industrial manufacturing processes, applications in the social field as is the case of visually impaired people, artificial vision allows them to have a hearing vision of their environment or text they want to know [7]. Based on the high number of visually impaired Ecuadorians, the high prices of commercial devices, low-cost glasses supported with artificial vision for text reading and audio are developed.

The article has been organized as follows; Sect. 2 describes the elements used for the prototype, Sect. 3 shows the design stages and their operation, Sect. 4 presents the system evaluation and validation results, and Sect. 5 shows the conclusions of the article.

2 Materials and Methods

2.1 Visual Impairment

Visual impairment is the difficulty related to a reduction or loss of visual functions and the barriers present in the environment in which the person is involved. Some barriers that can be cited are absence of auditory signals in replacement of visual information, as is the case of sound emitted by a pedestrian traffic light to alert that the pedestrian can cross; absence of braille literature in libraries, if a visually impaired person attends a library for information and does not find adapted texts there, this person will not have any chance to integrate in the society; difficulties in accessing alternative writing systems [8], such as audiobooks, which are mostly purchased electronically and prior to payment.

2.2 Raspberry Pi Camera

Raspberry Pi Camera Module is also known as raspicam [9], which is a camera that connects in the CSI port of the SBC, as its main features are a resolution of 8 megapixels and the ability to take video at 1080p (pixels) and 30 fps (frames per second), 720p and 60 fps or 640 × 480p and 60 or 90 fp [9]. Its small size allows it to occupy little space in a case that protects it, and the junction between the port and the camera is flexible, allowing it to be placed in almost any position, giving it the opportunity to be placed on the rod of glasses [10]. RPi UPS powerpack is a card makes it easy and safe to efficiently and safely obtain the right voltage and current for the system operation. It is an expansion card that has two slots to place up to two lipo batteries with a maximum of 5 V, and this module provides from 3.7 V to 3800 mA power supply with a maximum rated output current of 2A [11, 12].

3 Implementation

The implementation of glasses that allows the reading of a text is created with the aim that visually impaired people can access texts more easily by using the auditory sense, in this way the contents do not necessarily must be translated into the Braille system, which facilitates learning. The construction and design of the prototype are shown in Fig. 1.

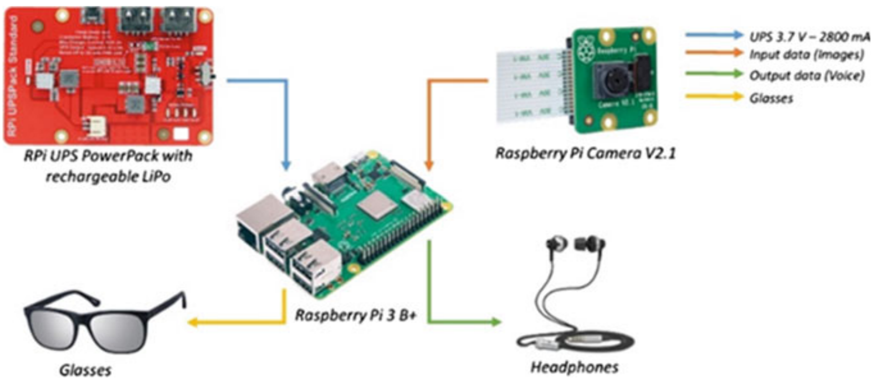


Fig. 1 General component diagram, adapted from [11]

3.1 System Stages

To obtain the best results in the machine vision system, it is necessary to install some important libraries, these are

GaussianBlur. This function allows blurring the image in addition to eliminating high-frequency noises in each pixel. The Kernel number is specified as positive and odd from 0, considered in width and height, by increasing this value also increases the blur proportionally [13].

Numpy. This library is necessary to make the arrangements for the recognition of an image by means of a specific color. For this reason, it has the function cv2. COLOR_BGR2HSV, which uses the hue, saturation, value (HSV) method, the hue or tone ranges from 0 to 179, the saturation from 0 to 255 and the value of brightness from 0 to 255, with this range of values, the color that is wanted to identify in the image is obtained. The data were obtained and analyzed for the low and high shades of green color, and these variables are stored to have a color mask that allows establishing the appropriate range that detects the color efficiently [14].

Canny. The Canny function performs a process divided into three specific parts: First edge detection is performed by Sobel, which consists of calculating the first derivative to detect intensity changes; second the suppression of pixels outside the edge, known as non-maximum, which is a technique that allows to thin the edges by means of the gradient, and these edges will be those that have a thickness equal to 1; third the application of the threshold per hysteresis, which allows to segment an image using a minimum and a maximum threshold so that it can be detected if a pixel is part of a border or not [15].

Minimum closing circle. This function determined as cv2.minEnclosingCircle(), which encloses the edges found in the image in a circle with a given radius, and in this way, the size of the found border can be set. By establishing the appropriate radius values, it is possible to set the distance at which the text should be recognized, delimiting the focus space and preventing other unwanted objects from being detected.

3.2 Capture

At this stage, the moment of capturing the image is set by sound signals, the same that occurs when the text is readable and it is at a predetermined distance. At that moment, the image is processed into the embedded system, cameras with high resolution are generally used in order to correctly digitize the image. And the process is verified when the confirmation message “PHOTO TAKEN” is heard. This is shown in Fig. 2.

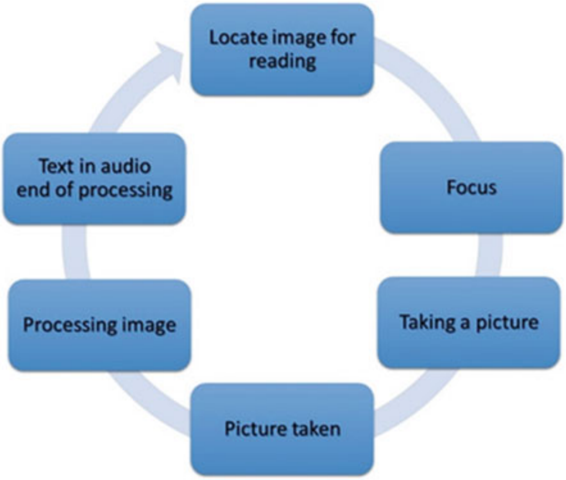


Fig. 2 Block diagram of the capture stage

3.3 Processing

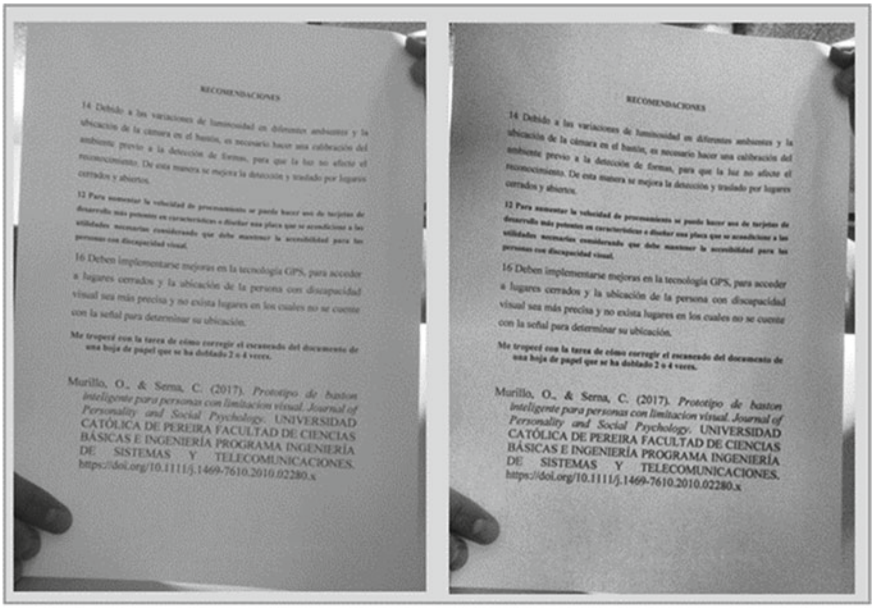
Once the capture of the image is done, a treatment is performed by means of filters. There are a variety of techniques for this, such as grayscale transformation, which obtains the gradient of the image and a process of high contrast using the CLAHE function included in the OpenCV, in order to have more black tones and avoid errors in the translation of the text. Grayscale filters and high contrast processing are presented in Fig. 3.

3.4 Transformation to Text

This stage focuses on transforming the image into plain text (.txt) where all characters are recognized. The image is converted to text by using optical character recognition (OCR) that allows reading image text, as shown in Fig. 4. The processing time can take a few seconds, depending on the number of words in the image, the higher the number of words the greater the processing time.

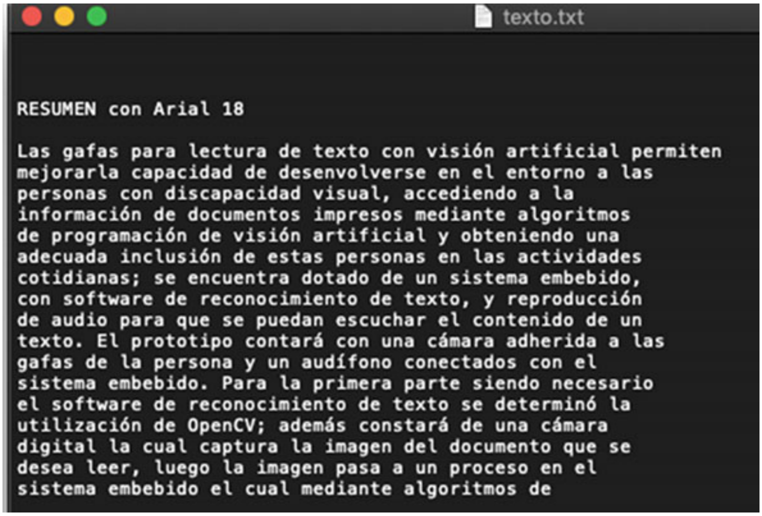
3.5 Audio Transformation

It corresponds to the last stage, which finishes with the playback of audio text through a headphone (Jack 3.5 mm) or speaker, where the words found in the image captured at the beginning of the process are heard. The text to speech (TTS) converter is used,



(a) (b)

Fig. 3 Grayscale filters (**a**) and high contrast filters (**b**)


texto.txt

RESUMEN con Arial 18

Las gafas para lectura de texto con visión artificial permiten
mejorarla capacidad de desenvolverse en el entorno a las
personas con discapacidad visual, accediendo a la
información de documentos impresos mediante algoritmos
de programación de visión artificial y obteniendo una
adecuada inclusión de estas personas en las actividades
cotidianas; se encuentra dotado de un sistema embebido,
con software de reconocimiento de texto, y reproducción
de audio para que se puedan escuchar el contenido de un
texto. El prototipo contará con una cámara adherida a las
gafas de la persona y un audífono conectados con el
sistema embebido. Para la primera parte siendo necesario
el software de reconocimiento de texto se determinó la
utilización de OpenCV; además constará de una cámara
digital la cual captura la imagen del documento que se
desea leer, luego la imagen pasa a un proceso en el
sistema embebido el cual mediante algoritmos de

Fig. 4 Text resulting from applying Tesseract OCR which is an open CV library that converts the information into audio so that it can be played by an artificial voice.

4 Results

Figure 5 shows the 3D design of the prototype with its components. To perform the tests, some parameters of the documents to be read are taken into account. First, two green lines of at least 5 mm thick must be drawn, one line at the top or bottom and the other on the left or right side of the sheet. It was decided to place the rectangular margin due to its practicality when drawing it around a text and with green color, because of the ease of detection when using artificial vision.

4.1 Results of the Margin Detection Stage

The distance at which the best results can be obtained is between 20 and 30 cm from the glasses, where the camera and the sheet containing the text are located. Rectangular margin has a great advantage in discarding any type of contours of different shapes and sizes existing in the image. Several sheets were tested, and the average distance with better results is 20.6 cm.

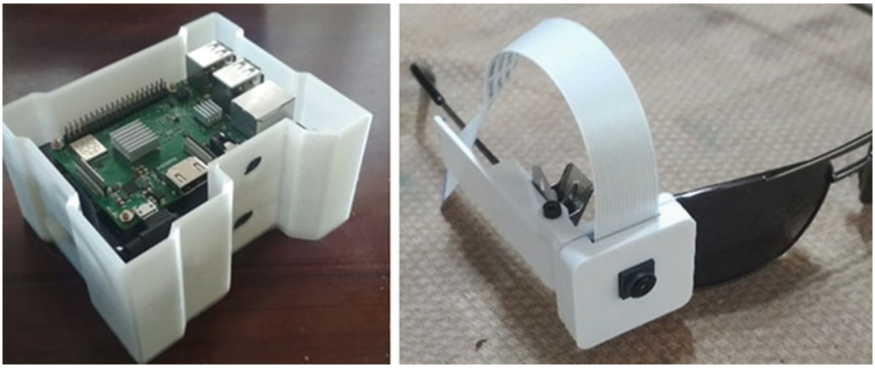


Fig. 5 Final design of the prototype

4.2 Results of the OCR Character Detection System

The tests were performed using multiple sheets to measure the distance, and in this step, what is expected is to know the number of words that the system can translate from the captured image as shown in Fig. 6.

It can be seen in Fig. 7 the test results obtained from each of the photos in Fig. 6, where the words recognized by the system are discovered.

These results indicate the corresponding operations to obtain the percentage of system efficiency based on the font size.

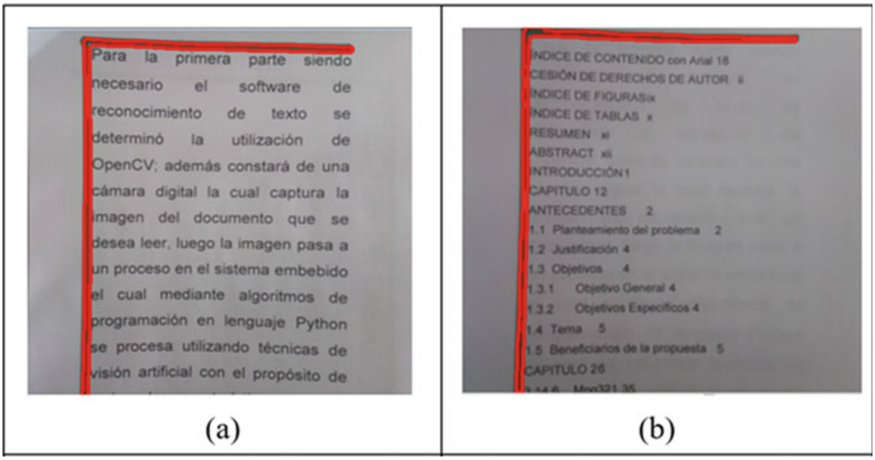


Fig. 6 Images were taken by Pi Camera V2.1

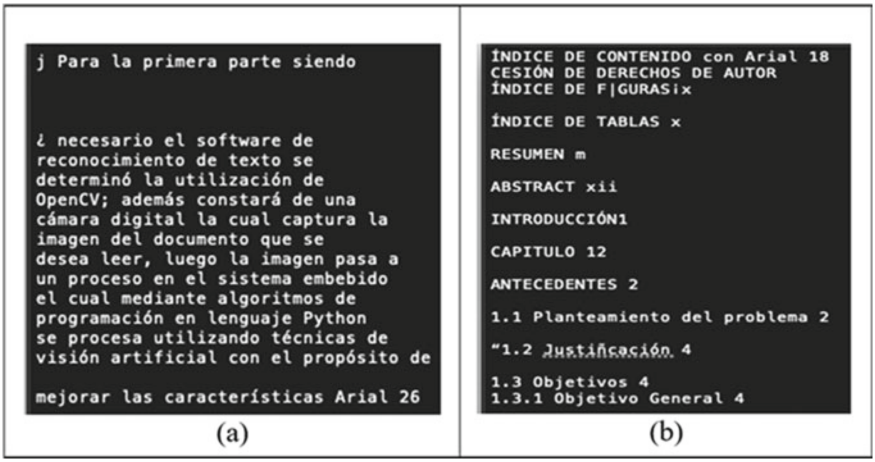


Fig. 7 Percentage of validation

Table 1 Percentages of system efficiency

Image	Font size	Words in the image	Words in the text	Efficiency percentage (%)
A	26	71	71	100
B	18	65	63	96.9
C	16	274	257	93.7
D	12	514	445	86.5
E	11	514	17	3.3
F	12–14–16	207	131	63.3

Table 1 shows the individual and total percentage of successful words with the program; these are the ones that TTV will recognize and pronounce. It should be clarified that the efficiency percentage of the system depends on many factors such as the angle of inclination of the sheet or the distance from the glasses, but two important factors are the light level and the font size because the bigger, the greater the number of words translated, fact that is consistent with the results obtained, where the bigger the font size, an error that approaches 0% is obtained.

4.3 Results of the Validation Stage

The prototype validation of visually impaired glasses was performed with several people from the UNESCO subject of the Salesian Polytechnic University, who experienced its portability and ease of operation.

A direct survey was conducted where several questions were asked to obtain the validation of the prototype with a value of 1–5 points, in which 1 is the lowest value and 5 is the highest value of satisfaction when using the prototype. The parameters evaluated were portability, utility, functionality, audio and implementation. The results are shown in Fig. 8.

The results show the acceptance of the prototype which can be improved by decreasing the weight of the prototype and the familiarization with the device.

5 Conclusions

The audio-assisted text reading glasses were designed to be a support tool for non-sighted; in addition, it is a device of low cost and easy functionality that allows independence in the user. The performance tests for the validation of the prototype were carried out with different sizes and font letters, in the latter factor, the most common ones were used: Arial, Times New Roman and Calibri, useful for the translation to plain text. In addition, it is noted that a font size higher than 16 gets an

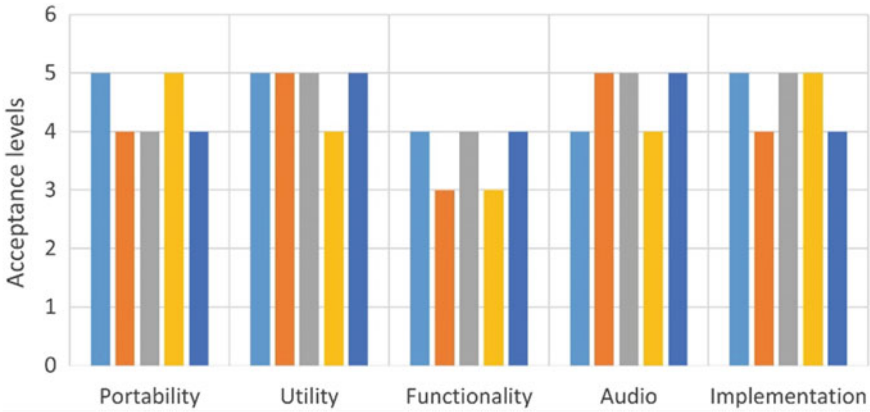


Fig. 8 Acceptance levels of the parameters according to the performed survey

efficiency of 93%, and this performance decreases considerably when using a font size of 12, with an efficiency of 86%.

To reduce errors when taking and processing the photograph, some libraries such as GaussianBlur, Numpy, Canny and minimum locking circle were used in the machine vision system. These libraries allow to improve the image quality and correct any defect scan in order to obtain an appropriate result before translating to text and later by voice.

References

1. Organizacion mundial de la salud: Ceguera y discapacidad visual. Organizacion mundial de la salud (2017)
2. CONADIS: Personas con discapacidad registradas. Consejo Nacional para igualdad de discapacidades (2019)
3. Organización Mundial de la Salud and Banco Mundial: Resumen de Informe mundial sobre la discapacidad. Educación p. 27 (2011)
4. IrisVision Global: IrisVision. [Online]. Available: https://irisvision.com/product/. Accessed 17 Sep 2019
5. eSight Corp: eSight. [Online]. Available: https://esighteyewear.com. Accessed 17 Sep 2019
6. OrCam Technologies: OrCam. [Online]. Available: https://www.orcam.com/en/about/. Accessed 17 Sep 2019
7. Wang, C., Miao, W., Zhao, J.: Vision-based curvature model for artificial intelligence in vehicles. In: Proceedings—2012 International Conference on Control Engineering and Communication Technology, ICCECT 2012, pp. 245–248 (2012)
8. Ministerio de Educación de Chile: Necesidades educativas especiales asociadas a Discapacidad Visual. Primera, Santiago de Chile (2007)
9. Raspberry Pi Foundation: Camera module. [Online]. Available: https://www.raspberrypi.org/documentation/hardware/camera/. Accessed 19 Sep 2019
10. Salcedo Tovar, M.L.: Minicomputador educacional de bajo costo Raspberry Pi: Primera parte, vol. 7, Venezuela, pp. 28–45 (2015)

11. Raspberry Pi Foundation: Raspberry Pi 4 bumper edition! [Online]. Available: https://www.raspberrypi.org/weekly/raspberry-pi-4-bumper-edition/. Accessed 19 Sep 2019]
12. Rodrigo, A.: Sistema de sonorización haciendo uso de raspberry pipara su uso e implementación en un entorno inteligente. España, p. 23 (2014)
13. Smoothing Images—OpenCV 2.4.13.7 documentations." [Online]. Available: https://docs.opencv.org/2.4.13.7/doc/tutorials/imgproc/gausian_median_blur_bilateral_filter/gausian_median_blur_bilateral_filter.html. Accessed 04 Apr 2019
14. Pawlik, A., Segal, J., Sharp, H., Petre, M.: Crowdsourcing scientific software documentation: a case study of the NumPy documentation project. Comput. Sci. Eng. **17**(1), 28–36 (2015)
15. Hao, G., Min, L., Feng, H.: Improved self-adaptive edge detection method based on Canny. In: Proceedings—2013 5th International Conference on Intelligent Human-Machine Systems Cybernetics IHMSC 2013, **2**(2), 527–530 (2013)

Implementation of a Fashion Virtual Assistant with the Use of a Kinect v2 Camera and Image Processing

Christopher Vizcarra, Gabriel Medina, and Alfredo Barrientos

Abstract This article is about the problem and development of a fashion virtual assistant proposed by using a Kinect v2 camera and image processing, for fashion retail stores. It comes up mainly as a response to the inability of providing unique experiences during the shopping process through the use of diverse devices. Because of this, similar virtual assistant solutions, oriented to provide clothing recommendations, were analyzed to be able to provide software that could give a more personalized suggestion for the users based on their physical characteristics.

Keywords Fashion virtual assistant · Recommendation system · Smart mirror

1 Introduction

Technology has evolved over time, having a greater presence in practically each of the activities carried out daily. This is mainly due to the existence of the Internet, which allows having a great variety of information quickly and easily through the highly acclaimed intelligent devices. Nowadays, it is common to find smart applications or devices in stores whose main objective is to provide a unique user experience.

According to the study carried out by [1], the user's experience in the premises of the sale directly affects their level of satisfaction for their purchases along with their desire to return to the stores for future purchases. Its study concludes that a user experience that allows performing the desired tasks in a simple and efficient way, together with the easy access to additional information on the details of the products offered, allows reaching a higher level of user satisfaction at the end of the process.

C. Vizcarra · G. Medina · A. Barrientos (✉)
Universidad Peruana de Ciencias Aplicadas (UPC), Lima 15023, Peru
e-mail: pcsiabar@upc.edu.pe

C. Vizcarra
e-mail: u201513977@upc.edu.pe

G. Medina
e-mail: u201512091@upc.edu.pe

© Springer Nature Switzerland AG 2021
Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_3

On the other hand, a user experience which provides complementary multimedia information as well as a set of offered products and developed functionalities allow increasing user willingness in order to return to future purchases.

Our proposal aims to improve the user experience in fashion retail stores using a fashion virtual assistant that calculates personal colorimetry values by using a Kinect v2 camera and a webcam to obtain a personalized clothing recommendation. The webcam requires the best possible resolution because this will influence the values obtained from the process, so that is why it was chosen one with high resolution. The match between clothing is due to a series of previously defined rules. This assistant has been contemplated to be used in a smart mirror, and due to this, its design has been considered according to its average measurements resulting in a futuristic and intuitive design, which will allow users to orient themselves anywhere in the process.

2 Related Work

The approaches that you want to give to the recommendation of the clothing directly influence the result. The existing solutions have as main objective to achieve the optimum recommendation between garments or sets.

In [2], the author proposed a system of recommendations for fashion electronic commerce. For its implementation, three threads were used, which are the preparation of historical data, the calculation of product scores and the generation of recommendations. With this, every time a customer selects a product, and he recommends the most suitable products that are not in the same category, based on their calculated score.

In [3], the author proposed another recommendation approach, which is based on faster F-R-CNN for the detection of the clothing region. Each image is separated by its lower and upper body, eliminating cluttered backgrounds. Its compatibility is measured by means of an energy function, and it is qualified if it is a compatible or incompatible pair. For its qualification, it is considered a previous training, where attributes such as style and category are considered.

In [4], the author carried out the implementation of a system of recommendations generating datasets consisting of a variable number of clothing. They were classified into categories of the exterior, upper, lower, full-body, footwear and accessories. In addition, variations of invalid outfits were generated. The system was trained on an average of 400,000 iterations, which resulted in an established outfit structure and compatibility between the garments based on their characteristics.

In [5], the author introduces a new variable in the recommendation of the garment, the location. To resolve the relationship between locations and garments, the images must have the variables of visual content, garment attributes, location attributes and the denotation of the different locations. The attributes of the garment consist mainly of categories. The location attributes are summarized in colors and attributes such as the season, the weather and the type of attraction. Finally, through a scoring function, where the highest score is given to the categories of garments that appear most for the given location, the system learns the correlation between the garments and the locations. In [6], the author proposes a new data model for garments based on the storage of garment images, a dataset of garment attributes and the second dataset of garment meanings, values extracted from the investigated fashion theory [7] presents a similar approach, a model that allows generating sets of items considering their different attributes and combination of these, the results are then evaluated in order to refine the subsequent results generated based on the reception of the previous ones, [8, 9] proposes an alternative to the traditional way, with respect to the recommendation of garments based on the description of the texts of the products.

3 Methodology

According to the research carried out on the different factors that influence the personalized recommendation of clothing and the functionalities necessary for the correct interaction with the mirror by the customers of a clothing store, the development of the intelligent mirror was divided into two stages: conception and implementation.

3.1 Conception

For this stage, it was defined as a .Net-based work environment used by the intelligent mirror integrated system that makes use of the Kinect SDK's measurement capture and images analysis features by OpenCV, SBT and Python. The development environment used for the programming of this system was Visual Studio 2019. On the other hand, the work environments for the administrative system and the Web service based on React + Redux and Node, respectively, were defined, both using TypeScript as language and Visual Studio Code as a development environment.

Subsequently, the necessary color analysis modules were defined for the recommendation of garment colors based on the user's natural colors. The natural colors to obtain for the correct recommendation are the colors of eyes, skin and hair, according to the research carried out by [10, 11], taking into consideration the theory of colors described by [9]. Additionally, the body and face forms to be considered for the recommendation based on the form of the user were defined, recovered from the investigation by [12, 13], together with the information provided by [14, 15].

3.2 Implementation

Once all the functionalities necessary for the development of the system integrated into the intelligent mirror and the Web administrative system were defined, the organization and design of the logical and physical architecture of both systems began.

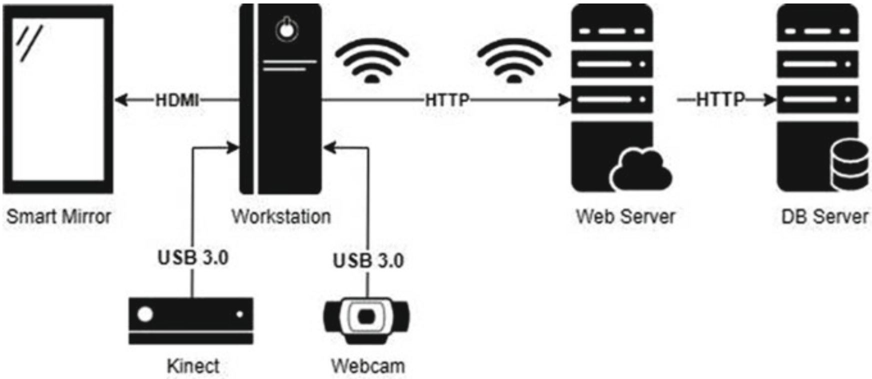


Fig. 1 System physical architecture

Web Service. Both the mirror and the Web system require the main Web service to obtain the necessary information to fulfill their management and consultation functions. This information is hosted on a REST server programmed in the TypeScript language, using the Node v8.12 development environment, in Visual Studio Code, forming the architecture shown in Fig. 1.

Its internal structure is based on modules, each in charge of different service sectors. Among these are the authorization modules, in charge of the validation of the credentials of users of the system; compatibility, responsible for the recommendation of garments based on the characteristics of the current user received by the mirror system; users, responsible for the management of system users; garments, in charge of the management of garments of the system and its details; sets, responsible for the management of the current sets of the system and colors, families, subfamilies, occasions and sizes, responsible for consulting the current values of their respective entities.

The database used is a MySQL Schema, for the connection with the Schema, the ORM is used for Node, TypeORM, which allows simplifying the general queries for the entities of only consulting and adding an additional level between the system and database.

Considering the development flow, a logical architecture was designed, shown in Fig. 2, where the interaction between the service and the Web systems and the smart mirror is broadly considered, considering the platforms of these systems.

Web Administrative System. The Web system was developed using React + Redux for Google Chrome and Mozilla Firefox browsers and is intended to allow system administrators to manage and consult the registered entities.

Integrated Smart Mirror System. It was developed using WPF, a .Net-based Microsoft technology designed for the creation of graphical interfaces, in the C # language for the Windows platform. The design and distribution of screen components were organized in order to take advantage of the space available in the mirror

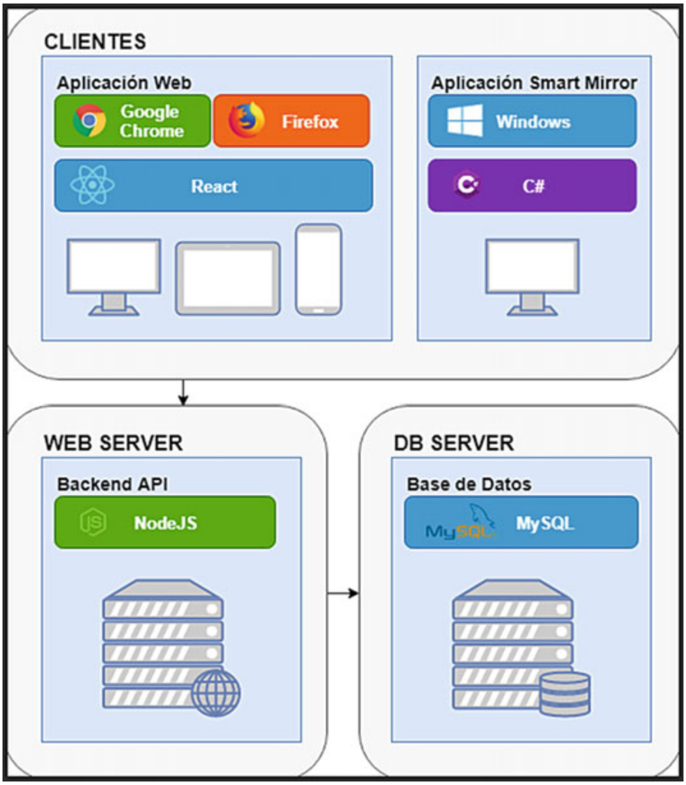


Fig. 2 System logic architecture

and allow user reflection without components that obscure it. The restrictions considered during system design were the interaction with the system would be through a touch interface of the device, and the device will always be in the vertical orientation.

For the functions of measuring the user's natural colors and body and face shapes, the system uses the OpenCV functionalities in conjunction with SBT and Python to obtain the results from a photograph of the user captured during the initial steps of the interaction with the system.

OpenCV and SBT are used to capture the eye colors to segment the initial image and separate the person from the background, next step the face is obtained by filtering the hair and the rest of the body, then, the area is segmented of each eye separately, and the iris from each eye obtained. In order to obtain these areas, the analysis sub-module is trained with different face captures that allow recognizing the valid areas for each step of the segmentation of the original capture. Once the user's iris is obtained, the sub-module obtains the iris color by analyzing the color range to which each pixel belongs, the total amount and the percentage of the total area of the iris by color range. These values are compared to determine the dominant color in the user's eyes.

$$C_D = A_1 * P_1 - A_2 * P_2. \tag{1}$$

The formula for calculating the color of eyes is based on the results of the colors with a greater area in both iris of the user. Where A_1 and A_2 are the areas of the first color and second color, respectively, and P_1 and P_2 are the percentages of the total area corresponding to each area.

OpenCV and Python are used to capture the hair color to perform the segmentation of the user's original capture. Using masks, the largest area of the capture is obtained, which allows the person to be obtained, and similarly, another mask is used to obtain the second largest area to obtain the person's hair. An additional mask segmentation is performed to obtain the largest area again and avoid double tone problems, and once the segmentation is finished, the average color is obtained. To obtain the average color of the segmented area, the method of random vertices is used, which is based on obtaining the average of the values of N vertices randomly selected within the image. This color value is used to obtain the closest color to it.

Similarly, OpenCV and Python are used for skin color capture to segment the original capture by means of masks. In this case, once the person's face is segmented, the largest area of the face is taken, an area corresponding to the skin. Similarly, to the previous analysis, random points and the color proximity formula are used to obtain the skin color closest to the average obtained.

On the other hand, the Kinect SDK and HD Face functionalities are used to obtain the vertices of the forehead, cheek and jaw end to calculate the shape of the user's face and the vertices of the shoulders, elbows and wrists to calculate the shape of the body. From these measurements, the face and body shape of the user is obtained.

The personal station relevant to the combination of the three values is obtained from the user's natural colors, and together with the body and face shapes, the initial catalog segmented into subfamilies of valid garments and compatible colors is obtained with the user's natural palette.

Once the initial catalog, shown in Fig. 3, of the user is obtained, the system flow is based on the garments chosen by the user. The determined flow for the formation of the set is top, bottom, exterior, footwear, belt and tie.

Where the outer category garments, such as belt and tie, are optional and can be omitted if the user wishes. For the first selection, the initial category has no restrictions; however, the following selections must follow the established order.

4 Conclusions

The previous investigations carried out were useful for choosing the initial configuration of personal colorimetry and obtaining the type of face and body. This configuration made it possible to make recommendations for types of garments based on the type of face and body and the colors suggested based on the user's personal colors.

The configuration steps and the process of selecting clothes to create an outfit allowed an easier user interaction with the system's functions. Further improvements

Fig. 3 Catalog screen interface

can be achieved by refining the spacing and sizing of elements in the display in order to show the provided information more clearly.

Further research toward the application of the user's physical characteristics for recommendations can be prosecuted following the use of image processing and proportion calculation applied in the system. Refining the existing processes to obtain results with less variation or adding additional characteristics to be measured will allow a more refined recommendation process.

References

1. Poushneh, A., Vasquez-Parraga, A.Z.: Discernible impact of augmented reality on retail customer's experience, satisfaction and willingness to buy. J. Retail. Consum. Serv. **34**, 229–234 (2017). https://doi.org/10.1016/J.JRETCONSER.2016.10.005
2. Hwangbo, H., Kim, Y.S., Cha, K.J.: Recommendation system development for fashion retail e-commerce. Electron. Commer. Res. Appl. **28**, 94–101 (2018). https://doi.org/10.1016/j.elerap.2018.01.012
3. Liu, Y.J., Gao, Y.B., Bian, L.Y., Wang, W.Y., Li, Z.M.: How to wear beautifully? Clothing pair recommendation. J. Comput. Sci. Technol. **33**(3), 522–530 (2018). https://doi.org/10.1007/s11390-018-1836-1
4. Tangseng, P., Yamaguchi, K., Okatani, T.: Recommending outfits from personal closet. In: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), pp 2275–2279 (2017). https://doi.org/10.1109/ICCVW.2017.267
5. Zhang, X., Jia, J., Gao, K., Zhang, Y., Zhang, D., Li, J., Tian, Q.: Trip outfits advisor: location-oriented clothing recommendation. IEEE Trans. Multim. **19**(11), 2533–2544 (2017). https://doi.org/10.1109/TMM.2017.2696825

6. Guan, C., Qin, S., Long, Y.: Apparel-based deep learning system design for apparel style recommendation. Int. J. Cloth. Sci. Technol. **31**(3), 376–389 (2019). https://doi.org/10.1108/IJCST-02-2018-0019
7. Li, Y., Cao, L., Zhu, J., Luo, J.: Mining fashion outfit composition using an end-to-end deep learning approach on set data. IEEE Trans. Multim. **19**(8), 1946–1955 (2017). https://doi.org/10.1109/TMM.2017.2690144
8. Zhou, W., Mok, P.Y., Zhou, Y., Zhou, Y., Shen, J., Qu, Q., Chau, K.P.: Fashion recommendations through cross-media information retrieval. J. Vis. Commun. Image Represent. **61**, 112–120 (2019). https://doi.org/10.1016/j.jvcir.2019.03.003
9. Itten, J.: Arte del Color. Aproximación subjetiva y descripción objetiva del arte. Editorial Bouret (1992). https://eacvvcae.files.wordpress.com/2014/02/l-arte-del-color_johannes-itten.pdf
10. Bobzin, L.: Rediseño de ambos médicos. Materialidades, colores y moldería femenina. Universidad de Palermo (2017). https://fido.palermo.edu/servicios_dyc/proyectograduacion/archivos/4544.pdf
11. Valencia, M.: Diseño de una colección de trajes de línea casual para madres e hijas afroecuatorianas en la parte céntrica de guayaquil, con el fin de aplicar la colorimetría. Instituto Superior Tecnológico Sudamericano (Tecsu) de Guayaquil (2018). http://repositorio.tecsu.edu.ec:8080/xmlui/bitstream/handle/123456789/153/Moda_05102018.pdf
12. Vaca, P.: Estudio Antropometrico del personal que labora en el departamento de producción de la empresa textil Wellman S.A. Universidad Politécnica Salesiana (2013). https://dspace.ups.edu.ec/bitstream/123456789/6743/1/UPS-QT04506.pdf
13. Espinosa, F.: Aplicación de procedimientos antropométricos para la selección de obras pictóricas con miras a su autentificación: Caso de estudio Miguel de Santiago. Universidad Tecnológica Equinoccial (2012). http://repositorio.ute.edu.ec/bitstream/123456789/4560/1/50942_1.pdf
14. Valero: Canon del cuerpo humano (2013). https://valero7.webnode.es/tecnicas/a16-canon-del-cuerpo-humano/
15. Moreno, G.: Dibuja de la Figura Humana. Universidad Autónoma Metropoitana (1991). https://core.ac.uk/download/pdf/83080021.pdf

Suggestion Mining for Generating New Services and New Product Ideas from YELP Reviews

Daniel Collucci, Alexandre Junqueira, and Carlos Forster

Abstract YELP reviews strongly influence the business's reputation and revenue as well as customers and new-users attitudes toward the acquisition of products and services. Due to the overwhelming number of customer reviews available on Web sites, usually, the decision-making process of customers, new-users, and business owners, is not associated with the daunting task of reading reviews online. Text mining can address the issue of summarizing huge volumes of unstructured data; furthermore, text mining can extract emotions, sentiments, and insights from texts written by customers. This study proposes a text mining approach to unravel suggestions from YELP reviews. This study shows that suggestion mining differentiates reviews with suggestions from reviews without suggestions and therefore can identify new services and new product ideas from customer reviews.

Keywords Natural language processing · Recurrent neural network · Customer reviews

1 Introduction

1.1 Background

YELP is a word-to-mouth platform that connects customers, new-users, businesses, and business owners [1]. Even though YELP reviews are readily available and easily accessible online, bridging the gap between unstructured and structured data in order to extract information and knowledge is an overpowering task [2]. Text mining

D. Collucci (✉)
Itaú-Unibanco, Data Science Team, São Paulo, Brazil
e-mail: dacollucci@mix.wvu.edu

A. Junqueira
Itaú-Unibanco, Advanced Analytics Team, São Paulo, Brazil

C. Forster
Computer Science Department, Aerospace Technology Institute, São Paulo, Brazil

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_4

bridges this gap and allows the extraction of insights from textual data, including suggestions for new services and new product ideas from customer reviews [3].

Suggestions from customers are commonly expressed in the form of recommendations, complaints, and feedback. Suggestion mining unravels and extracts structured and numerical data from customer feelings, expressions, and behaviors. Mining for suggestions can guide the development and the improvement of products [4] and improve customer experience and satisfaction [5]. The goal of this study is to utilize text mining techniques to differentiate customer reviews, which suggest new services and products from customer reviews, which do not include such suggestions.

1.2 Related Work and Relevant Concepts

Even though scientific contributions on feature extraction to perform opinion and sentiment analysis are extensive, there is a lack of information on suggestion mining [6]. Suggestion mining involves the identification of sentences which propose improvements in services as well as the identification of sentences and specific words which suggest new product ideas [7]. There is a growing interest in suggestion mining because reviewers often suggest improvements about their experiences [8]. Therefore, suggestion mining provides information that is worth unraveling within the context of strategic business management [9]. Previous studies employed suggestion mining techniques, essentially, machine learning techniques applied to textual data in order to provide strategic insights to businesses, including suggestions for product improvement [10], and tweets about Microsoft Windows phone improvements [11].

Machine learning algorithms use features during the learning process; features are pieces of information extracted from the data [12]. The extraction of features from textual data involves the numerical representation of words; the transformation of words into numbers is commonly executed using a bag-of-words (BOW) or a word-embedding approach [13]. The BOW generates vectors for each text by considering each word as a counting point [14]. Word-embedding also generates text as vectors; nevertheless, it allows words with similar meanings to have similar representation in vector space [15].

The Naive Bayes algorithm is commonly used in natural language processing problems; nevertheless, recurrent neural networks (RNNs), including, long short-term memory (LSTM) networks have gained a lot of attention due the ability to consider the influence of the past on future events for solving textual learning tasks [16]. The Naive Bayes classifier has its roots on the Bayes' theorem, which describes the probability of a label based on prior knowledge of features that may be associated with that label [17]. The LSTM network is equipped with internal units, also known as hidden layers or as memory blocks. These units allow these types of RNNs to process sequential information [18]. A memory block is composed of a neuron, an input gate to control the flow of numerical information into the neuron, a forget gate to control how much of the input will be forgotten from the memory block, and an output gate to control the impact of the neuron activation function on the other neurons [19].

2 Methodology

2.1 Data

The data used in this study was downloaded from the Kaggle YELP Academic Dataset Challenge [20]. The original YELP data consisted of 229,907 reviews of 11,537 businesses in the Phoenix metropolitan area (AZ-USA) written by 48,873 users. In this study, only one individual rated YELP reviews as $y = 1$ (review displays suggestions about new services and products) or as $y = 0$ (review does not display suggestions about new product ideas). Therefore, in order to make the labeling process feasible, this study utilized reviews between 40 and 100 words, essentially 12,759 reviews.

2.2 Feature Selection and Machine Learning Methods

Prior to feature selection, preprocessing was used and consisted of stemming, removal of punctuation and stopwords, and conversion of uppercase characters to lowercase. Preprocessed textual data was represented as a bag-of-words and as word-embeddings; these representations were fed into a multinomial Naive Bayes and an LSTM network, respectively. Two types of word-embeddings were used: word-embedding developed using the 40–100 words long reviews and 300 dimensions pre-trained GloVe embedding [21]. The data used in the multinomial Naive Bayes algorithm was partitioned into training (80%) and a validation set (20%) using a 10-fold cross-validation approach. The data used to train the LSTM networks was partitioned into training (80%), a validation set (10%), and a test set (10%). The two LSTM networks were defined with 60 neurons each and a recurrent dropout rate of 0.4. The "rmsprop" optimizer was used, and the LSTM algorithms were trained using a batch size of 64 in 20 epochs. Accuracy was used to evaluate the performance of the algorithms. The baseline accuracy consisted of the prediction of the most common class in the dataset (zero-rule algorithm).

3 Results

3.1 Exploratory Data Analysis

Figures 1 and 2 display word-clouds of reviews with suggestions and without suggestions, respectively. These two word-clouds were generated with the most frequent

Fig. 1 Word-cloud of reviews with suggestions using high-frequency words

Fig. 2 Word-cloud of reviews without suggestions using high-frequency words

words. Both word-clouds appear to display only positive words. Figure 3 displays a word-cloud of reviews with suggestions, while Fig. 4 displays a word-cloud of

Fig. 3 Word-cloud of reviews with suggestions using low-frequency words

Fig. 4 Word-cloud of reviews without suggestions using low-frequency words

reviews without suggestions. Both of these word-clouds used only infrequent words. The word-cloud of Fig. 3 highlights negative and discontent words as well as miscellaneous legal terms. On the other hand, the word-cloud of Fig. 4 displays mostly positive words, including friendly adjectives describing the business atmosphere and positive customer feedback associated with business services.

3.2 Predictive Algorithm Performance

The accuracy of the multinomial Naive Bayes algorithm was slightly higher compared with the LSTM equipped with the pre-trained word-embedding. The LSTM wired with the non-pre-trained word-embedding displayed lower accuracy compared with the multinomial Naive Bayes and the LSTM GloVe; nevertheless, it displayed an improvement in accuracy in comparison with the zero-rule algorithm (Table 1).

Table 2 shows the precision, recall, and F1-score of the multinomial Naive Bayes classifier. The combination of high precision and high recall for the false class (reviews without suggestions) indicates the low occurrence of false negatives and

Table 1 Algorithm accuracy evaluation

Algorithm	Accuracy
Multinomial Naïve Bayes	0.76
LSTM GloVe 300 dimension embedding	0.75
LSTM	0.72
Zero-rule algorithm	0.60

Table 2 Multinomial Naive Bayes predictive performance

Class	Precision	Recall	F1-score
False	0.78	0.85	0.82
True	0.72	0.61	0.66

false positives, respectively. The high precision of the true class (reviews with suggestions) indicates a low incidence of false positives. However, the slight low recall may indicate a high incidence of false negatives.

Figures 5 and 6 show the accuracy and the loss of the LSTM wired with the pre-trained word-embedding, respectively. There are no signs of over-fitting, and the validation accuracy is improved during the training process (Fig. 5). The accuracy and the loss of the LSTM non-equipped with the pre-trained word-embedding are

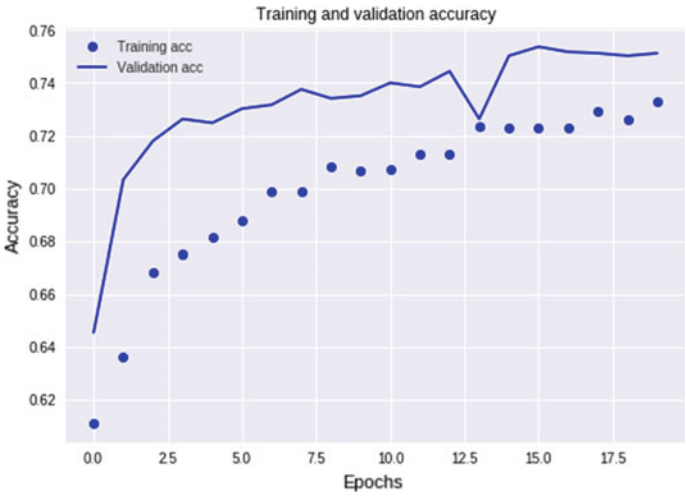


Fig. 5 LSTM GloVe embedding train and validation accuracy

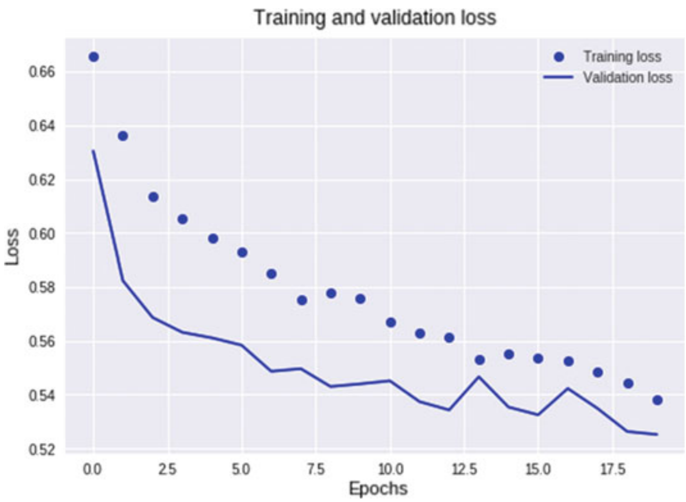


Fig. 6 LSTM GloVe embedding train and validation loss

displayed in Figs. 7 and 8, respectively. There is evidence of over-fitting. The over-fitting initiates early during the training process. As a result, the validation accuracy, which reaches 74%, is not constant and degrades.

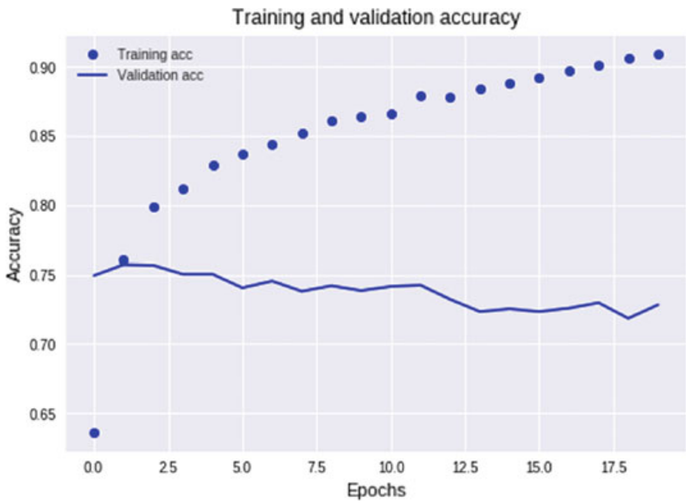


Fig. 7 LSTM train and validation accuracy

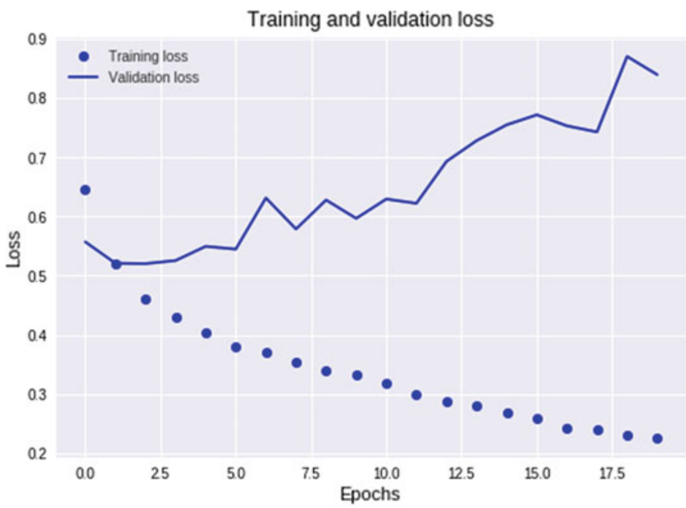


Fig. 8 LSTM train and validation loss

4 Discussion

A simple bag-of-words and two word-embeddings were used in this study to extract features from the text. Supervised suggestion mining and two machine learning algorithms, a multinomial Naive Bayes and two LSTM networks, were used to distinguish reviews which had suggestions about new services and new product ideas from reviews which did not have such suggestions. These algorithms were able to distinguish reviews which had suggestions from reviews which did not have suggestions.

In terms of accuracy, the multinomial Naive Bayes classifier outperformed the LSTM networks (Table 1). The comparison of the accuracy in the training and validation sets between the LSTM GloVe embedding (Fig. 5) and the LSTM not wired with the pre-trained word-embedding indicates evidence of over-fitting only for the LSTM not wired with the pre-trained word-embedding (Fig. 7). The cross-entropy loss in the training and validation sets of the LSTM wired with the pre-trained word-embedding decreased and indicates this algorithm is learning with the input and therefore is making good predictions (Fig. 6). However, while the error on the training set of the LSTM not wired with the pre-trained embedding decreased, the validation loss which is the error obtained by running the validation set through the trained LSTM increased (Fig. 8). This is a clear indication the LSTM not wired with pre-trained word-embedding is not learning from the data.

There are advantages of using word-embedding over BOW, more specifically, the representation of similar words closes together in the word-embedding space results in lower dimensionality of the extracted features [22]. The task of learning word-embedding together with the classification of reviews task resulted in accuracy quite similar compared with the use of the pre-trained word-embedding (Table 1). Previous studies also show whenever corpora size is large, LSTM networks equipped with non-pre-trained word-embedding exhibit good generalization [21]. However, it appears the corpora size used, as an input in this study was not large enough to avoid over-fitting. It appears the LSTM not wired with the pre-trained word-embedding did not learn efficiently how to extract corpora semantic and syntactic features. Furthermore, it appears this network also did not display a higher ability to capture word dependencies in comparison with LSTM equipped with pre-trained GloVe embedding.

5 Conclusions

This study shows that supervised suggestion mining and machine learning algorithms including a multinomial Naive Bayes classifier and an LSTM network can distinguish reviews, which had suggestions from reviews, which did not include such suggestions. Furthermore, this study also shows these algorithms have the ability to identify new services and new product ideas from YELP reviews. In the near future, the authors plan to increase the numbers of texts in the analysis, essentially, incorporate texts with the length between 200 and 300 words. Long reviews are likely to contain a detailed description of the business, and therefore, may provide insights for parsing dependency analysis short reviews cannot, including the assignment of a suggestion to a dependency word. Other directions will focus on hyper-parameters tuning of the LSTM network and utilization of ensemble modeling and regularization to reduce over-fitting and improve accuracy.

Acknowledgements Any opinions, findings, and conclusions expressed in this manuscript are those of the authors and do not necessarily reflect the views, official policy or position of the Itaú-Unibanco. The authors wish to thank the Itaú-Unibanco for the time-release in conducting this study.

References

1. Luca, M.: Reviews, reputation, and revenue: The case of yelp.com. (2016)
2. Gandomi, A., Haider., M.: Beyond the hype: Big data concepts, methods, and analytics. Int. J. Inf. Manag. **35**(2):137–144 (2015)
3. Weiss, S., Indurkhya, N., Zhang, T.: Fundamentals of Predictive Text Mining. Springer (2015)
4. Chen, Y., Lin, C., Huang, C.: Application of data mining techniques on new product development decision-making. In: International Conference on Frontier Computing, pp. 596–600. Springer (2018)
5. Berezina, K., Bilgihan, A., Cobanoglu, C., Okumus, F.: Understanding satisfied and dissatisfied hotel customers: text mining of online hotel reviews. J. Hos. Mkgt. Manag. **25**(1), 1–24 (2016)
6. Christensen, K., Nørskov, S., Frederiksen, L., Scholderer, J.: In: Search of new product ideas: Identifying ideas in online communities by machine learning and text mining. Creat. Innov. Manag. **26**(1), 17–30 (2017)
7. Brun, C., Hagege, C.: Suggestion mining: detecting suggestions for improvement in users comments. Res. Comput. Sci. **70**(79.7179), 5362–5379 (2013)
8. Hoornaert, S., Ballings, M., Malthouse, Van den Poel. E.D.: Identifying new product ideas: waiting for the wisdom of the crowd or screening ideas in real time. J. Prod. Inno. Manag. **34**(5), 580–597 (2017)
9. Golchha, H., Gupta, D., Ekbal, A., Bhattacharyya, P.: Helping each other: a framework for customer-to-customer suggestion mining using a semi-supervised deep neural network. arXiv preprint arXiv:1811.00379 (2018)
10. Ramanand, J., Bhavsar, K., Pedanekar, N.: Wishful thinking: finding suggestions and 'buy' wishes from product reviews. In: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in text, pp. 54–61 (2010)
11. Dong, L., Wei, F., Duan, Y., Liu, X., Zhou, M., Xu, K.: The automated acquisition of suggestions from tweets. In: AAAI (2013)
12. Guyon, I., Elisseeff, A.: An introduction to feature extraction. In: Feature Extraction, pp. 1–25. Springer (2006)
13. Steven, B., Klein, E., Loper, E.: Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc." (2009)
14. Wallach. H.M.: Topic modeling: beyond bag-of-words. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 977–984. ACM (2006)
15. Chollet, F.: Deep Learning with Python. Manning Publications Co. (2017)
16. Goodfellow, I., Bengio, Y., Courville, A., Bengio, Y.: Deep learning, vol. 1. MIT Press, Cambridge (2016)

17. Chen, J., Huang, H., Tian, S., Qu, F.: Feature selection for text classification with naive bayes. Expert Syst. Appl. **36**(3), 5432–5435 (2009)
18. Lipton, Z., Berkowitz, J., Elkan, C.: A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019 (2015)
19. Pennington, J., Socher, R., Manning, C.D.: Glove: global vectors for word representation. In: Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543 (2014)
20. Yelp Dataset Challenge: Yelp dataset challenge, 2015. https://www.kaggle.com/yelp-dataset/yelp-dataset (2015)
21. Wang, Y., Huang, M., Zhao, L.: Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 606–615 (2016)
22. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781 (2013)

Reduction AWGN from Digital Images Using a New Local Optimal Low-Rank Approximation Method

Sadegh Kalantari, Mehdi Ramezani, Ali Madadi, and Vania V. Estrela

Abstract In this paper, image noise reduction has been formulated as an optimization problem. The target image is denoised using a low-rank approximation of a matrix. Considering the fact that the smaller pieces of the picture are more similar (more dependent) in natural images, therefore, it is more logical to use low-rank approximation on smaller pieces of the image. In the proposed method, the image corrupted with additive white Gaussian noise (AWGN) is locally denoised, and the optimization problem of low-rank approximation is solved on all fixed-size patches (Windows with pixels needing to be processed). Therefore, for practical purposes, the proposed method can be implemented in parallel. This is one of the advantages of such methods. In all noise reduction methods, the two factors, namely the amount of the noise removed from the image and the preservation of the edges (vital details), are very important. In the proposed method, all the new ideas including the use of training image (TI image) obtained from the noisy image, the use of SVD adaptive basis, iterability of the algorithm, and patch labeling have all been proved to be efficient in producing sharper images, good edge preservation, and acceptable speed compared to the state-of-the-art denoising methods.

Keywords Optimal low-rank approximation · SVD · Signal denoising · Image denoising · Signal processing

1 Introduction

Signal denoising is one of the most important issues in the field of digital image processing. Noise might destroy or damage vital information and details of a signal. Therefore, denoising is one of the initial stages of the process of feature extraction,

S. Kalantari (✉) · M. Ramezani · A. Madadi
Department of Electrical and Control Engineering, Tafresh University, 39518 79611 Tafresh, Iran
e-mail: sadeghkalantari@tafreshu.ac.ir

V. V. Estrela
Departamento de Engenharia de Telecomunicações, Universidade Federal Fluminense, Rio de Janeiro, Brazil

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_5

object recognition, image matching, etc. Several methods have been proposed to denoise digital images and signals until now [1–11]. AWGN is one of the most significant noises studied by researchers. In each pixel of the noisy image built by AWGN, a value is added to its gray level. This value is sampled independently from Gaussian distribution. AWGN denoising methods can be divided into three categories including correlation-based (or spatial filter), transform-based, and hybrid methods. Correlation-based methods work directly with the spatial values of the image. The advantages of correlation-based methods include the simplicity of its algorithms and their acceptable speed. In transform-based methods, the image is first transferred to the desired transform domain and then denoised. Therefore, applying a transformation to a signal is equivalent to changing the axes of the coordinates. Changing the coordinates of the signal makes it possible to effectively separate the noisy and non-noisy part of the signal which itself is one of the advantages of these methods. The hybrid methods take advantage of both these methods simultaneously.

In the process of image denoising, the amount of eliminated noise, as well as the preserved edge and textures, is considered as the most important issues. Moving average and Gaussian filter are regarded as the easiest correlation-based methods. These methods like a low-pass filter soften the high-frequency parts of the image (including noise and edges). Despite their good denoising capability, the mentioned methods eliminate some of the tiny yet vital information. Bilateral filter (BF) has been used to improve edge preservation [11]. In this filter, the value of each pixel is estimated by the average weight of the neighbors. The intensity and spatial similarity were the criteria by which the weight of the neighbors was determined. This preserves sharp edges during noise cancelation; however, the staircase effect and gradient reversal are the disadvantages of this method. NLM method is a non-local version of BF which estimates the value of each pixel using the weighted average of the similar pixels. In this method, the weights are determined based on the similarity between the pixels. The simplicity of the algorithm and relatively convenient speed are among the advantages of this method.

In transform-based methods, the images can be represented using some sparse bases like wavelet, curvelet, and contourlet [12]. Some of these bases are fixed, but some are selected adaptively and adapted to the signal information. Because of complex singularities in many images, using a fixed basis such as wavelet will not always yield acceptable results. Two methods [13, 14] have proposed an adaptive representation method called K-SVD. In these methods, an optimization problem is solved using greedy algorithms. Then, a dictionary is trained in order to denoise the image. The columns (atoms) of this dictionary can be used as an adaptive basis for sparse representation. Today, the use of sparse representation in many applications such as denoising, super-resolution, image reconstruction, inpainting, etc., has provided acceptable results [15–17]. Noise spreads over all transform coefficients. However, most of the basic image information is focused only on a few of the largest coefficients. Hence, in such methods, image denoising can be done using many shrinkage methods such as [18]. In general, methods such as K-SVD, [19, 20], which solve an optimization problem in order to denoise an image, are called optimization-based methods. These methods are considered as transform-based methods. Since these methods solve optimization problems, they usually have a lot of computational complexity and are slow.

The BM3D method is one of the hybrid methods which use spatial filtering and transform-based methods [21]. This method denoises the image by sparse filtering and grouping similar patches into 3d arrays. Today, more complete versions of the BM3D method have been presented. These new methods have improved the results using a shape-adaptive principal component analysis [22]. Some of the hybrid methods like ASVD and SAIST take advantage of useful properties of an SVD basis. ASVD method uses SVD for a training basis in order to represent the image patches. SAIST method denoises image using SVD and sparse representation of image patches [23, 24]. The unique feature of these methods is that they combine the most useful properties of both correlation-based and transform-based methods; therefore, the output images provide better results in terms of speed and quality.

This paper intends to provide an adaptive local denoising method using an adaptive SVD basis. In this method, a training image (TI) is created using the noisy image and Gaussian filter. The resulted TI data is used in signal denoising. Based on the locality of the algorithm, SVD is computed for all image patches (with overlapping) and then each patch is denoised individually and adaptively using TI information. In order to prevent artifacts, the aggregation phase is done after computing estimated patches. In this phase, the average obtained values are replaced in overlapping areas. Since these methods can consider each patch both individually and in parallel with other patches, they can be implemented for practical purposes. Considering the ideas in the proposed method, it belongs to the family of hybrid methods.

In this study, using the above-mentioned ideas, a denoising method having appropriate quantitative factors (PSNR and FSIM) and acceptable computational complexity is proposed. In addition to its sharp and high-quality images, another benefit of this method is that it solves the noise problem and increases the efficiency by combining some of the common image processing methods and using the information of the noisy image.

Various sections of the article are as follows. In Sect. 2, the linear representation of the image using SVD will be examined and the low-rank approximation problem will be formulated. In Sect. 3, the proposed method is presented and the applied ideas are explained. In Sect. 4, the results of the proposed method are presented and conclusions are made in Sect. 5.

2 SVD and Low-Rank Approximation

2.1 SVD

Suppose A represents a gray-level image. The basic principle of linear representation of the image is that the matrix A can be represented as the sum of the weighted basis shown in (1). In this equation, a_i are the coefficients and ϕ_i are the corresponding bases. These bases can be chosen from well-known bases like wavelet, curvelet, bandlet, contourlet, etc. [25–28].

$$A = \sum_{i=1}^{N} a_i \phi_i \tag{1}$$

These bases are similar to the Fourier series. Each periodic 2D function can be represented in terms of exponential basis and C_{mn} coefficients are according to Eq. (2).

$$\begin{gathered} f(x, y) = \sum_{n=-\infty}^{+\infty} \sum_{m=-\infty}^{+\infty} C_{mn} \mathrm{e}^{\frac{in\pi x}{a}} \cdot \mathrm{e}^{\frac{im\pi y}{b}}, \quad -a < x < a, -b < y < b \\ C_{mn} = \frac{1}{ab} \int_{-a}^{a} \int_{-b}^{b} f(x, y) \cdot \mathrm{e}^{\frac{-in\pi x}{a}} \cdot \mathrm{e}^{\frac{-im\pi y}{b}} \mathrm{d}x \mathrm{d}y \end{gathered} \tag{2}$$

According to the expansion of the Riemann–Lebesgue lemma, C_{mn} coefficients are descending and the result will be as Eq. (3) [29].

$$\lim_{m \to \infty} \lim_{n \to \infty} C_{mn} = 0 \tag{3}$$

This lemma states that the high harmonic coefficients are negligible and they have very little effect on reconstructing the signal. So, if the coefficients are truncated from a specified frequency onward, the original signal can be reconstructed with a fairly good approximation. In the next section, the idea of truncation is used for denoising the signal with SVD.

In SVD theory, each matrix can be decomposed as Eq. (4):

$$A = U \Sigma V^t \tag{4}$$

In which $U_{m \times m} = [u_1 \cdots u_m]$ (left singular vector) and $V_{n \times n} = [v_1 \cdots v_n]$ (right singular vector) are orthogonal matrix. So, according to Eq. (5),

$$V V^t = V^t V = I_n, \ U U^{t} = U^t U = I_m. \tag{5}$$

It should be noted that the columns of $U_{m \times m}$ and $V_{n \times n}$ matrices are composed of AA^t and $A^t A$ orthonormal eigenvectors matrices, respectively. $\Sigma_{m \times n}$ is a semi-diagonal matrix in which the values on its diagonal are the singular values of the $A^t A$ or AA^t matrices. So, we will have:

$$\begin{gathered} \Sigma_{m \times n} = \mathrm{diag}(\sigma_1, \ldots, \sigma_p), \ p = \min\{m, n\} \\ \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > 0, \sigma_{k+1} = \cdots = \sigma_p = 0 \end{gathered} \tag{6}$$

in which σ_1 is the largest and σ_k is the smallest nonzero singular value of the matrix A.

2.2 Formulation of Low-Rank Approximation Problem

Approximation of a matrix with a lower-rank one can be done using SVD. The goal in this section is to estimate a low-rank matrix B using a matrix A. According to Eq. (7), in singular value decomposition of the matrix A with rank r, we will have:

$$A = [U_1|U_2]\begin{bmatrix} \sigma_1 & 0 & 0 & \\ 0 & \ddots & 0 & 0 \\ 0 & 0 & \sigma_r & \\ & 0 & & 0 \end{bmatrix}\begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}, \quad \text{rank}(A) = r \tag{7}$$

According to the definition of the two matrices U and V, the singular value decomposition of the matrix A can be represented as shown in Eq. (8):

$$A = u_1\sigma_1 v_1^t + u_2\sigma_2 v_2^t + \cdots + u_r\sigma_r v_r^t, \quad \sigma_1 > \sigma_2 > \cdots > \sigma_{r.} \tag{8}$$

In general, we represent matrix A linearly in the form of (9) or (10) equations according to SVD of A.

$$A = [U_{1a}|U_{1b}|U_2]\begin{bmatrix} \sigma_1 & 0 & & & & 0 & \\ 0 & \ddots & 0 & & & & \\ & 0 & \sigma_k & 0 & & & \\ & & 0 & \sigma_{k+1} & 0 & & \\ & & & 0 & \ddots & 0 & \\ & & & & 0 & \sigma_{r+1} & 0 \\ & 0 & & & & & 0 \end{bmatrix}\begin{bmatrix} V_{1a}^T \\ V_{1b}^T \\ V_2^T \end{bmatrix} \tag{9}$$

$$\begin{aligned} A &= u_1\sigma_1 v_1^t + u_2\sigma_2 v_2^t + \cdots + u_k\sigma_k v_k^t + u_{k+1}\sigma_{k+1} v_{k+1}^t + \cdots u_r\sigma_r v_r^t \Rightarrow \\ A &= \sigma_1 u_1 v_1^t + \sigma_2 u_2 v_2^t + \cdots + \sigma_k u_k v_k^t + \sigma_{k+1} u_{k+1} v_{k+1}^t + \cdots \sigma_r u_r v_r^t \Rightarrow \\ A &= \sum_{i=1}^{r} a_i \phi_i \end{aligned} \tag{10}$$

Comparing Eqs. (1) and (10), we can see that σ_i are coefficients and $u_i v_i$ the corresponding bases, which unlike the constant bases of Fourier series, are selected adaptively using signal information. According to Eq. (11):

$$\sigma_1 > \sigma_2 > \cdots > \sigma_r \tag{11}$$

Primary coefficients play a greater role in the reconstructing matrix A. This also applies to represent a signal by the Fourier series based on Riemann–Lebesgue lemma.

Now, the main problem is to determine B (the low-rank approximation of the matrix A() as an optimization problem in the form of (12):

$$B = \arg\min_{z} \|A - z\|_2^2 \quad \text{st,} \quad \text{rank}(z) = k. \tag{12}$$

In the above statement, k is regarded as "sparsity." Assuming applying SVD on matrix A, and considering Eckart–Young–Mirsky's theorem which is available in [30], and also Eq. (9), the closed-form answer for the optimization problem in (13) is as follows:

$$\begin{aligned}
&\Sigma_B = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_k, 0, 0, \ldots, 0) \\
&B = U\Sigma_B V^t = u_1\sigma_1 v_1^t + u_2\sigma_2 v_2^t + \cdots + u_k\sigma_k v_k^t \Rightarrow \\
&B = U_{1a} \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_k \end{bmatrix} V_{1a}^t = U_{1a}\Sigma_B V_{1a}^t.
\end{aligned} \tag{13}$$

The theorem in the above reference describes a criterion for calculating the optimal value of k as (14):

$$\sum_{i=1}^{k} \sigma_i^2 \geq \sum_{i=k+1}^{p} \sigma_i^2, \quad p = \min\{m, n\} \tag{14}$$

In accordance with the above criterion, the sum of the k values of retained singular values must be greater than (or equal to) the sum of the truncated values so that an appropriately low-rank approximation of A can be obtained. Now, considering the discussions in the next section, we will discuss how to eliminate noise using SVD.

2.3 Signal Denoising Using SVD

Signal denoising is one of the applications of low-rank approximation. Assume that we have sampled the continuous-time signal $x(t)$ and represented $x = \begin{bmatrix} x_1 \; x_2 \; x_3 \; \cdots \; x_{m\times n} \end{bmatrix}$ as a vector. Now, we can classify the samples in an appropriate order and represent them as a matrix A.

$$A_{m\times n} = \begin{bmatrix} x_1 & x_{m+1} & x_{2m+1} & \cdots \\ x_2 & x_{m+2} & x_{2m+2} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ x_m & x_{2m} & x_{3m} & \cdots \end{bmatrix}$$

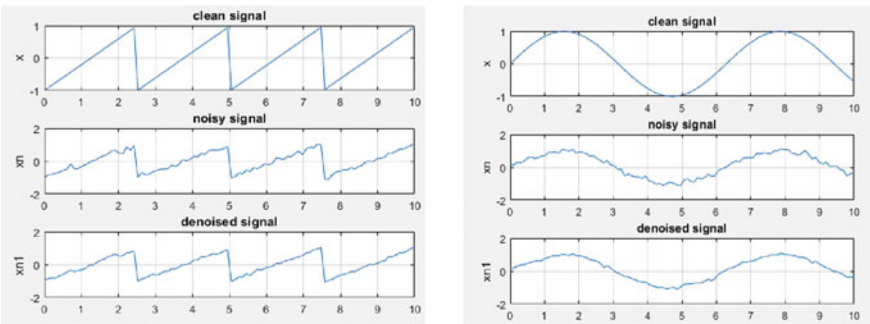
According to Eq. (15), if we compute the singular values of a matrix A, some of them will be much larger than the other singular values and smaller coefficients will have fewer roles in creating the matrix structure [12].

$$A = \sigma_1 u_1 v_1^t + \sigma_2 u_2 v_2^t + \cdots + \sigma_k u_k v_k^t + \sigma_{k+1} u_{k+1} v_{k+1}^t + \cdots \sigma_r u_r v_r^t \tag{15}$$

Now, consider the following noisy signal:

$$x_n = x + n \tag{16}$$

In the above equation, x_n it is a noisy signal, x is the noise-free signal and n is noise. Noise increases the significance of the last sentences by increasing the smaller singular values of the matrix (or signal), thereby destroying the original structure of the matrix. If these sentences can be truncated (or at least their impact in reconstructing the signal is minimized), signal denoising will be possible using low-rank approximation. In order to test the denoising method using the Eckart–Young–Mirsky criterion, two one-dimensional noisy signals are denoised and the results are shown in Fig. 1. It should be noted that the number of samples in these tests was 100 and they were arranged in a 10×10 matrix. In the following tests, k is the sparsity of the semi-diagonal matrix Σ that is determined using Eckart–Young–Mirsky criterion. For example, $k = 3$ means the signal is reconstructed (denoised) by 3 of the largest singular values. The following results show that the truncation of smaller singular values has improved the noisy signals.



a) $k = 3$ b) $k = 2$

Fig. 1 Results of applying noise reduction method on two noisy signals using SVD and Eckart–Young–Mirsky criteria

3 The Proposed Method

Before explaining the proposed method, we need to review some basics used in the proposed method in order to better understand the effects of the applied ideas.

3.1 Review of the Gaussian filter, High- and Low-Frequency Component and Canny Edge Detection Method

3.1.1 2D Gaussian Filter

2D Gaussian filter is a low-pass filter based on the probability distribution function of (17).

$$f(x, y) = \frac{e^{-\frac{x^2+y^2}{2\sigma_g^2}}}{2\pi\sigma_g^2} \tag{17}$$

In Fig. 2, there is a view of the 2D Gaussian filter. In the above equation, σ_g is a standard deviation; x and y represent the filter locations. Gaussian filter is a low-pass filter. Due to the fact that noise appears at high frequencies, image denoising using a Gaussian filter is one of the easiest ways. High-frequency parts of the image (edges + details) and noise are somewhat destroyed after applying the Gaussian filter to the image. Therefore, what remains constitute the smooth parts of the image. In other words, the resulted image is smoothed (blurred). Using this filter, we can decompose the high- and low-frequency components of the image.

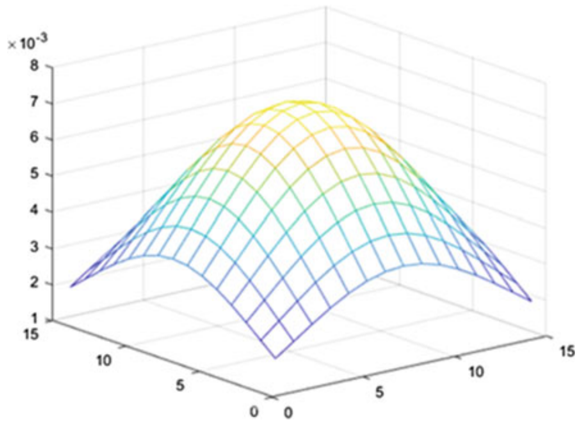


Fig. 2 2D Gaussian filter with window size = 15 and the standard $\sigma_g = 6$

3.1.2 Decomposition of High-Frequency and Low-Frequency Components of Image Using Gaussian Filter

In general, each image can be decomposed into a sum of a low-frequency (general details) and a high-frequency image (including edges and fine details). For example, Figs. 3 and 4 illustrate the decomposition of a noisy and non-noisy image into two high-frequency and low-frequency images, respectively.

By separating the high-frequency and the low-frequency components using the Gaussian filter, we will be able to separate the noise + edge image from the noisy one (the high-frequency component of the image). Then, using low-frequency component data that contains general information, the noise + edge image is denoised. Finally, integrating the denoised high-frequency component and the low-frequency component, we obtain a denoised image. Using this idea, we can use the information contained in the image itself and obtain higher accuracy in the process of truncating the SVD coefficients. Based on the analysis, the larger SVD coefficients (low-frequency ones) contain the coarse details of the image and smaller coefficients (high-frequency ones) include fine details of the image and noise [31–33]. Therefore, in case the coefficients can be truncated adaptively using low-frequency component information, the output denoised image will have an acceptable sharpness as well.

Original Image | Low frequency image (low frequency component of image, standard deviation = 2) | High Frequency Image (High Frequency Component)

Fig. 3 Decomposing the low-frequency and high-frequency components of the image using the Gaussian filter

Original Noisy Image | Low frequency Noisy Image (standard deviation = 4) | High Frequency Image (High Frequency Component)

Fig. 4 Decomposing low-frequency and high-frequency components of noisy image using Gaussian filter

3.1.3 Canny Edge Detection Method

Canny edge detection is one of the most common methods for extracting the image edges. Derivation (edge detection) of a noisy image results in the creation of another noisy image [34–37]. For a better understanding of this issue, consider Fig. 5.

If a noisy signal is derived for the edge detection (structures and textures areas which have sudden changes), the above result will be obtained. So, the derivation of the resulting signal will be noisy and not useful for identifying the edges. In order to prevent this problem, the signal must be smoothed. That is, it must be passed through

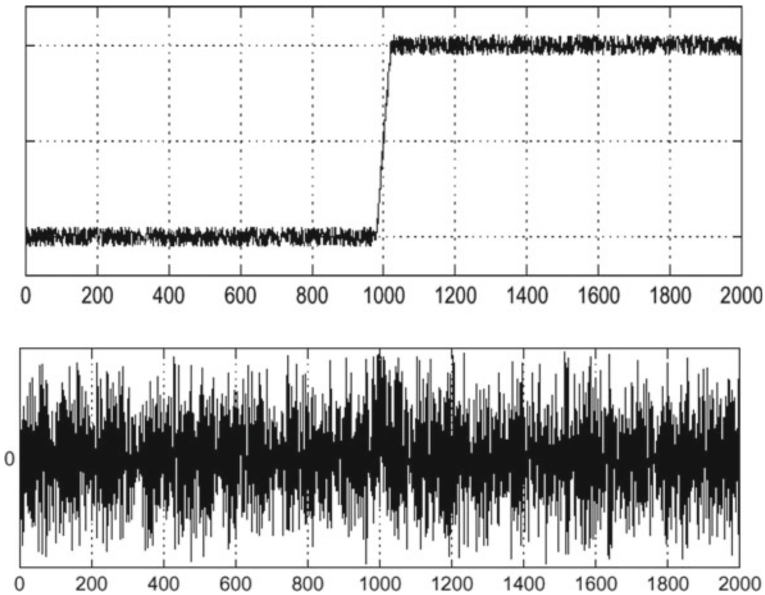


Fig. 5 Derivation of a noisy signal

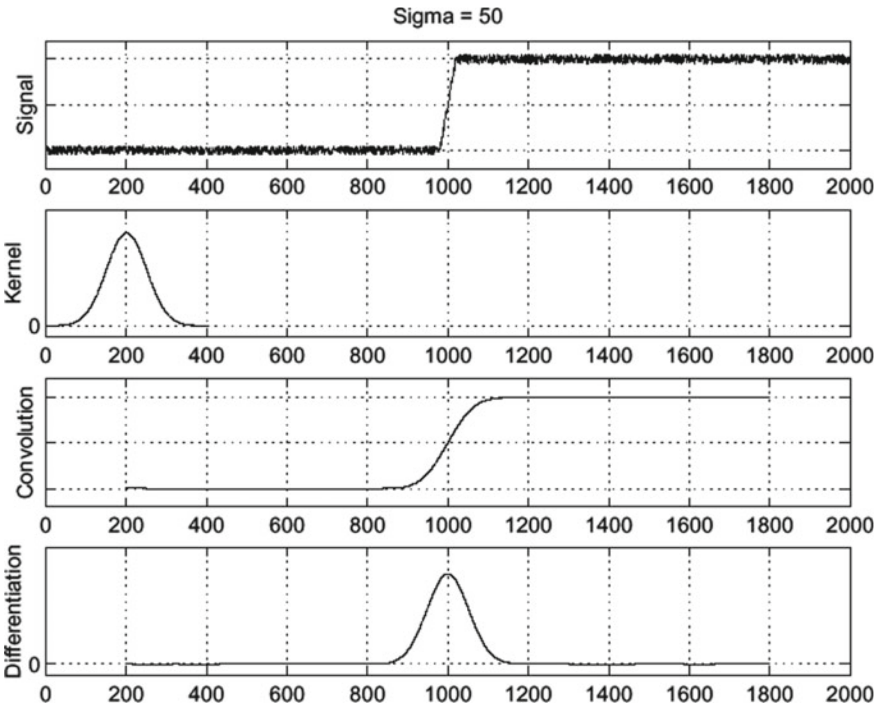


Fig. 6 How to use a derivative of a noisy signal to identify the edge

a low pass filter (Gaussian filter) so that the high-frequency components become less effective. Then, it can be derived to reveal the edges. In Fig. 6, this process is depicted for a noisy signal. In Fig. 6, f is the noisy signal and h refers to a Gaussian filter. Given the AWGN noise and the definition of the standard deviation calculation, in this paper, the standard deviation of the Gaussian filter for each image is estimated using Eq. (18).

$$\sigma = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (a_{ij} - \mu_A)^2}{m \times n}, \quad \mu_A = 0. \tag{18}$$

For the reasons given, to identify the edges and structured areas in the image, first, we pass the image from a low-pass filter to minimize the effects of noise, then it is derived (canny algorithm) and the edges of the image are found. In Fig. 7, these steps are shown.

As we can see from the above pictures, we were able to detect the critical edges of the image (training image) with an acceptable accuracy using the idea of filtering the image with Gaussian filters and extracting the edge using the low-frequency component of the image. We use the information in the training image to remove

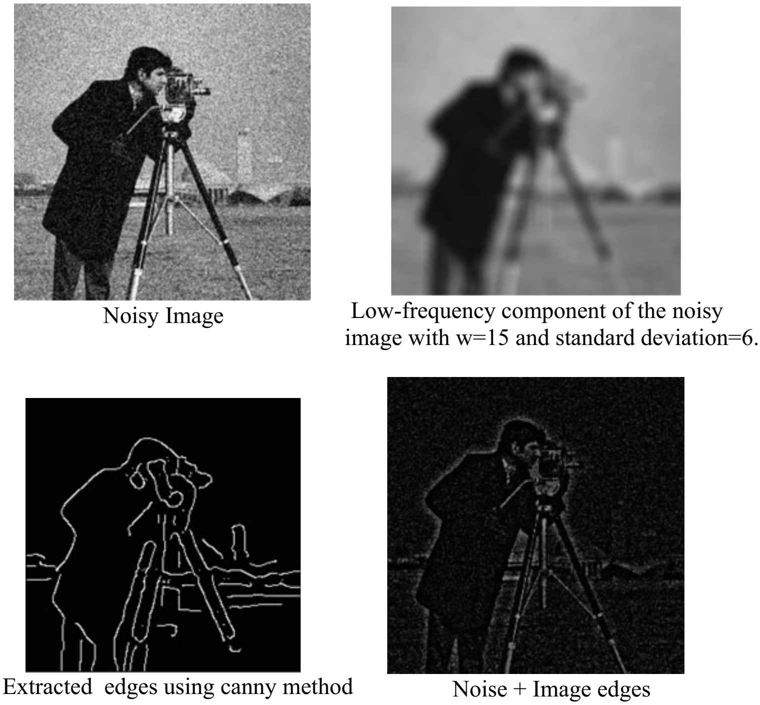


Fig. 7 Production of TI and noise + edges image from a noisy image

noise from the noise + edge image. After denoising the noise + edge image, it will be combined with the low-frequency component of the image and a denoised image will be created.

3.2 Using Training Image and Algorithm of the Proposed Method

In the previous section, we were able to determine the image edge (important parts of the image structures) using a Gaussian filter from the noisy image. The result can be used as a training image or as a model for the detection of textured areas. The resulting TI makes it possible to use the information in the noisy image to eliminate the noise adaptively. In this section, based on the reviewed basics and ideas, a new algorithm is proposed which, in addition to eliminating noise in smooth areas, improves the output image sharpness. The proposed method is shown in Table 1 and Fig. 9. As can be deduced from Fig. 8, some noise reduction methods like [20] eliminate many critical details and blur the output image.

Table 1 Proposed method

Input: Noisy image A
Output: Denoised image H
Parameters: Patch size, n: The number of groups in patch labeling, $iternum$: The number of iteration
algorithm:
Set ws and n
For $iter = 1$ to $iternum$ do
1. Decompose A for the first iteration and H for other iteration to A_L (low-frequency component) and A_H (high-frequency component) using part 3
2. Get the TI image using Canny edge detection and A_L
3. Do patch labeling on the TI with n Group using Part 3.2
4. For all of the patches with ws size in A_H, Extract the corresponding patch in the TI image. For patches with the same label, set the value of k equal to specified value for their agent and go to step 6
5. If the mean gradient of the patch is greater than the average gradient of the image, set k with ECKERT method, otherwise set k equal to 1
6. Reconstruct the patch with specified k and place it in the corresponding location in denoised image H_H
7. Do aggregation phase
8. Compute denoised image with the following equation: $H = H_H + A_L$

Fig. 8 Result of the method [20]

As described in previous sections, we aimed to apply low-rank approximation to smaller pieces of the image. Since the smaller parts in a natural image are more similar to each other, the low-rank approximation of smaller pieces is more justifiable. For this purpose, for each piece extracted from the image (the high-frequency component of the noisy image), the singular value decomposition of that slice is calculated. In order to better preserve the edges and important details of the image,

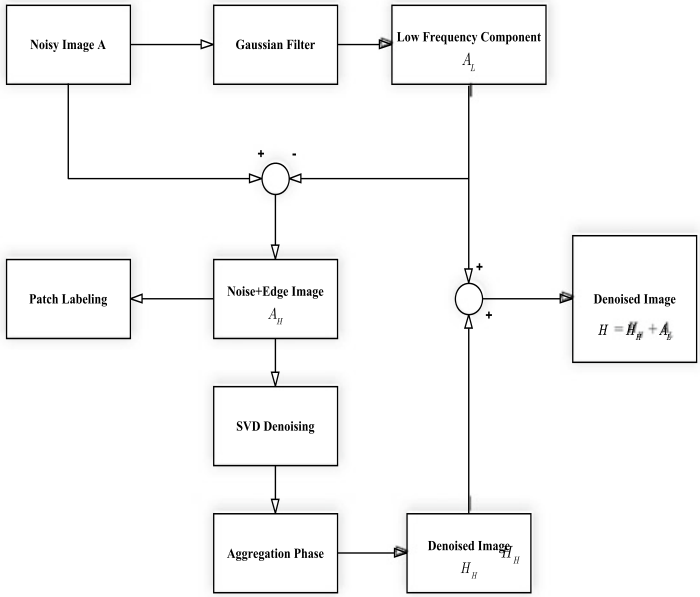


Fig. 9 Block diagram of the proposed denoising algorithm

the corresponding patches in the TI image were used. The pieces of the noisy image having low edges (the part of the image which is almost smooth) must be reconstructed with smaller singular values. The parts with more details must also be made with a greater number of singular values. Therefore, the corresponding patch in the TI image is also extracted for each patch of the noisy image in the proposed algorithm. If the average gradient in each patch exceeds the mean gradient of the total TI image, it indicates that the patch belongs to the textured parts of the image. So, we use the ECKERT method to determine k. If the average gradient of the patch is less than the mean gradient of the total TI image, it indicates that the patch belongs to the smooth areas of the image. Therefore, in order to reduce the computational complexity, the k value is considered to be 1. In addition to reducing the computational complexity, this idea preserves important and vital information in structured areas and prevents blurring. This operation is performed on all the extracted overlapping patches of the noisy image, and the denoised image is created eventually. Since there is an overlap between patches, multiple estimates can be obtained for different parts of the image. In those parts, the mean of those values is applied. This is called the aggregation process. For example, if two different values are obtained for a pixel (due to the overlap of windows), their average is considered as the final pixel value.

Since an optimization problem is solved in the noise reduction process, the proposed method can be considered as an optimization-based combination method. In order to reduce the computational complexity and increase the speed of the proposed algorithm, the patches of the image are labeled in terms of their similarity (patch labeling process). Therefore, in the preprocessing stage, before entering the low-rank approximation process, the similarity of different patches of the TI image is examined using Euclidean distance, and the patches that are similar to each other carry the same label. Using Eq. 19, we can examine the similarity of two patches:

$$s(p_i, p_c) = \|p_i - p_c\|_2^2 \tag{19}$$

In the above equation, $\|\|_2$ refers to Euclidean distance; p_i refers to vectorized reference patch; and p_c refers to vectorized candidate patch. The smaller the value of $s(p_i, p_c)$, the more similar the two patches. The reference patch with n a number of the same patches together make up a group all patches of which have the same label. The computational complexity of the proposed method is reduced by labeling different TI patches and using the above idea. For all patches of a given group, k value is considered to be the same. In fact, k value is only calculated for the agent of each group; the same value is considered for the rest of the patches of the same group.

Iterability of the algorithm is another idea which can improve the image quality. The output image is refined and the noise effects will be minimized by performing the above steps on the noisy image (in the first iteration) and the denoised image (in the next iteration). It should be noted this idea will increase the quality of the output image only for a limited number of iterations.

4 The Results of Proposed Method

In this section, the results of the proposed method are compared with the other methods including [3, 12, 19–21]. Takeda et al. [3], Zhang et al. [19], and Jiang et al. [20] are among the optimization-based methods and [12, 21] are among the hybrid ones. The denoising problem is formulated in an optimization framework using kernel regression in [6], using joint statistical modeling in method [19] and using sparse representation in [20]. These methods are usually slow due to the difficulty of some optimization problems. Method [12] is similar to the proposed method in using SVD, and we expect its results to be close to those of the proposed method. Method [21] is expected to have good results in terms of PSNR and FSIM. Its low speed is due to the use of the sparse filtering of the patches and three-dimensional patch grouping.

The above methods are implemented on the standard test images of CAMERAMAN, LENA, PEPPERS, MANDRILL, BOATS, and BARBARA, and the performance of each method will be investigated. The reason for using these standard images is their variability in texture and detail. Most articles in this field actually evaluate the effectiveness of the proposed method on these images. Implementation of all algorithms is done using a computer with an Intel Corei7, 1.8GH processor, 6 GB of RAM, and with MATLAB software run under the Windows 8.1 operating system. In all experiments, the size of the CAMERAMAN image is 256×256, the size of the rest of images is 512×512, and the noisy image is obtained by adding AWGN with various σ_n to test images. To test the performance of the proposed method, peak signal-to-noise ratio (PSNR) and feature similarity index (FSIM) criteria have been used. These two criteria are fully explained in the reference method [12]. Although the PSNR criterion is calculated in all denoising methods, the quantity obtained does not correspond to human visual perception in some cases. Therefore, in addition to the PSNR, another measure called FSIM is also used. This criterion is obtained by measuring the similarity between the two images and combining the phase correlation property and the gradient magnitude. The higher the value of FSIM, the higher the quality of the images will be.

Appropriate parameters are selected by trial and error using numerical experiments. For proper selection of algorithm parameters, like n, number of iterations and patch size, the proposed method has been tested on the CAMERAMAN image and its results are reported in Table 2. According to the results of Table 2 and Fig. 10, the following values are considered for the proposed method: $ws = 25$, $n = 1000$ and $iternum = 3$. The results of the proposed method, as well as those of other methods, are presented in Table 3 and Fig. 11. As indicated in the tests, the parameters selected in the other images in terms of PSNR, IEF, and complexity of the algorithm also have acceptable results. In this paper, for the simple set of parameters, they are considered equal for different images. For this reason, in some tests, the performance of the other methods has been better than the proposed method. However, the average of all tests shows the superiority of the proposed method over the other methods. However, the

Table 2 Sensitivity analysis of the proposed method parameters on the CAMERAMAN image with $\sigma_n = 10$

N	Patch size (ws)	PSNR (dB)
500	9	34.17
500	17	34.32
500	25	34.22
500	35	34.07
1000	9	34.37
1000	17	34.44
1000	25	34.54
1000	35	34.28
2000	9	34.2
2000	17	34.31
2000	25	34.15
2000	35	34.11
3000	9	34.11
3000	17	34.12
3000	25	34.04
3000	35	34.01

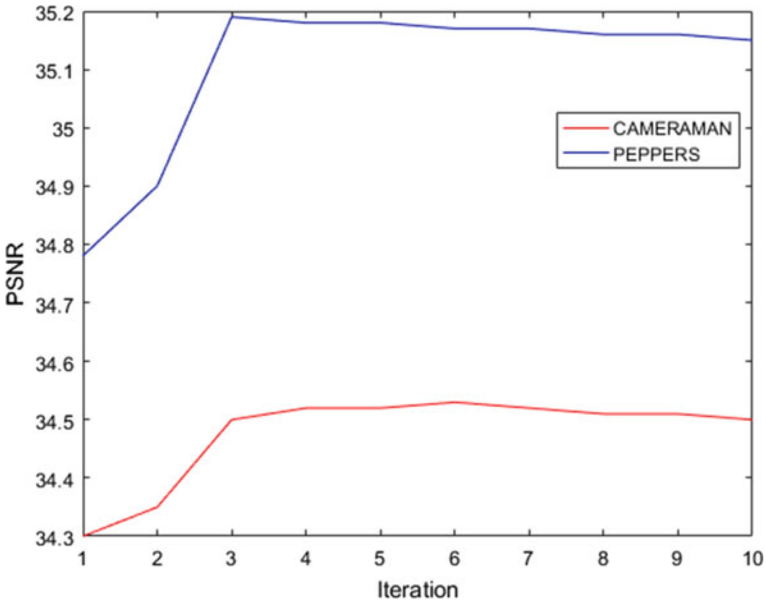


Fig. 10 Effects of the number of iterations on the CAMERAMAN and PEPPERS images

tests show that for more detailed images such as Mandrill, it is better to have smaller groups and smaller patch sizes. For less detailed images, we need larger groups and larger window sizes.

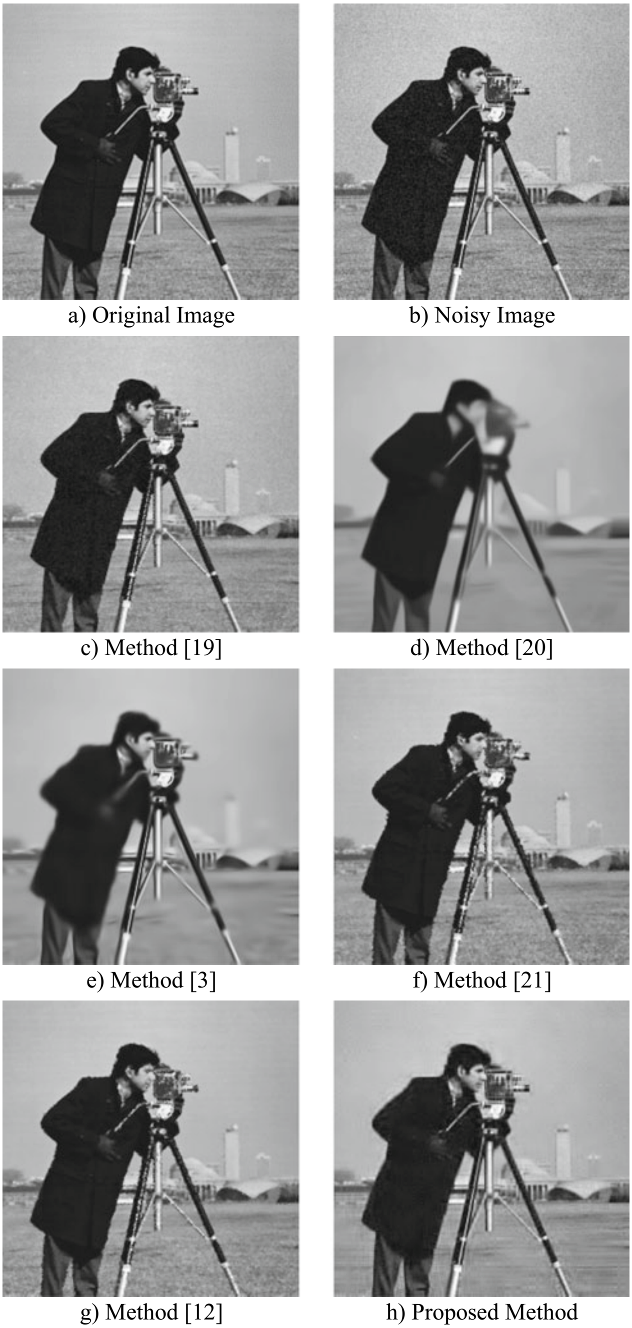
The results of the proposed method are presented globally and with various sparsities in Fig. 12. The effectiveness of the ideas like locality and an adaptive optimal *k* in the proposed method can be investigated using Fig. 12. It is clear that the above ideas have improved the algorithm in terms of visual and PSNR criteria. Also, the results obtained using the above ideas are significantly superior to the other comparisons in terms of visual quality (sharpness of image and denoising) and PSNR criteria.

To test the speed, different methods are compared in terms of runtime in the same conditions and the results are presented in Table 4. As expected, the reference methods [3, 19, 20] are slower than the proposed method because they are among optimization-based methods. The runtime of the method [12] is closer to that of the proposed method; and that is because of its similarity to the proposed method. Method [21] is also slower than the proposed method. That is due to the technique used to group similar patches into three-dimensional arrays.

Also, according to the results obtained in this section, the results of the proposed method are comparable and even superior to the state-of-the-art denoising methods in terms of PSNR, FSIM, and edge preservation. This is due to the fact that the TI obtained from the noisy image has been used to determine optimally k. It is noteworthy that k it helps to maintain useful details in the textured parts of the image. Since the proposed method is considered as one of the optimization-based hybrid methods, it is expected that the execution time of the algorithm will take longer than

Table 3 Results of different methods on standard test images

Image	σ_n (noise level)	Method [19]		Method [20]		Method [3]		Method [21]		Method [12]		Proposed	
		PSNR	FSIM	PSNR	FSIM	PSNR	FSIM	PSNR	FSIM	PSNR	FSIM	PSNR	FSIM
CAMERAMAN	10	34.20	0.87	33.36	0.77	34.23	0.83	34.33	0.87	34.35	0.88	**34.50**	**0.89**
	30	30.80	**0.79**	31.00	0.77	**31.40**	0.80	31.05	0.75	31.22	0.77	31.20	0.77
	50	28.80	0.61	30.00	**0.76**	28.90	0.57	29.62	0.66	29.60	0.66	**29.90**	0.68
LENA	10	34.80	0.96	35.10	0.94	35.60	0.96	**36.07**	**0.98**	36.00	0.98	36.03	0.97
	30	31.00	0.88	31.30	0.91	30.50	0.85	31.39	0.95	31.35	0.95	**31.70**	**0.95**
	50	28.50	0.72	29.3	0.84	28.10	0.68	29.07	0.92	28.96	0.92	**29.15**	**0.93**
PEPPERS	10	33.30	0.95	34.30	0.92	35.00	0.97	35.03	0.95	35.01	0.95	**35.10**	**0.97**
	30	30.80	0.87	30.40	0.89	**31.20**	0.87	30.40	0.87	30.45	0.89	30.50	**0.89**
	50	28.50	0.70	**29.50**	**0.88**	28.83	0.70	29.33	0.81	29.34	0.82	29.30	0.81
MANDRILL	10	32.01	0.83	32	0.82	32.10	0.85	32.15	0.86	**32.20**	0.86	32.19	**0.87**
	30	29.50	0.75	29.45	0.74	29.78	0.75	29.85	0.76	29.95	0.76	**29.96**	**0.76**
	50	27.85	0.64	27.71	0.64	27.86	0.66	27.90	0.69	27.92	0.69	**27.95**	**0.70**
BOATS	10	34.20	0.95	35	0.92	35.45	0.95	35.99	0.97	35.84	**0.97**	**36**	0.96
	30	30.56	0.87	31	0.90	30.35	0.83	31.35	**0.94**	31.21	0.93	**31.50**	**0.95**
	50	28.10	0.70	29.10	0.82	28.05	0.67	**29**	0.94	28.64	0.90	28.61	0.89
BARBARA	10	34.63	0.94	34.93	0.92	35.57	0.93	36	0.96	35.93	0.96	**36.00**	**0.96**
	30	30.56	0.85	31.03	0.89	35.45	0.82	31.31	0.92	31.31	**0.94**	**31.62**	0.93
	50	28.31	0.70	29.02	0.81	27.98	0.66	29.01	0.90	28.92	0.90	**29.08**	**0.90**
Average	-	30.96	0.80	31.29	0.83	31.45	0.80	31.60	0.86	31.55	0.87	**31.68**	**0.87**



a) Original Image
b) Noisy Image
c) Method [19]
d) Method [20]
e) Method [3]
f) Method [21]
g) Method [12]
h) Proposed Method

Fig. 11 Results of different methods on CAMERAMAN with a standard deviation of 10 in dB


a) Original Image


b) Noisy Image


c) $k = 1$,PSNR=30.09


d) $k = 2$, PSNR=31.8


e) $k = 3$,PSNR=31.4


f) $k = 4$,PSNR=31.02


g) Global, PSNR=30.75


h) Proposed, PSNR=33.61

Fig. 12 Results of the proposed method with different sparsities and in general on a noisy CAMERAMAN image with a standard deviation of 25

Table 4 Runtime of different methods

Time (s)	Method
11.21	Proposed method
13.85	Method [20]
38.66	Method [19]
27.25	Method [3]
33.56	Method [21]
11.43	Method [12]

the other methods. However, the computational complexity of the proposed method has been greatly reduced, and the speed of the proposed algorithm is acceptable compared to the other methods, and that is because of patch labeling and using the same sparsity for similar patches. Finally, the proposed method was able to provide an acceptable performance in terms of quantitative and visual evaluation as well as algorithm speed, and it was because of implementing the ideas related to local adaptive SVD basis, using the training image, employing patch labeling and the repeatability of the method.

5 Conclusion

In this paper, a localized AWGN noise reduction method is presented based on an SVD basis. In this method, a TI is created using a low-frequency component of noisy image and image gradient. The data pertaining to the TI is applied to the adaptive optimal low-rank approximation of each patch. The high-frequency component of the noisy image will be the input of the proposed algorithm. To eliminate the noise of each patch from the high-frequency component image, the corresponding patch is extracted from the TI image. If the average of the gradient of the patch exceeds the average gradient of the total TI image, it means that the patch belongs to the textured parts of the image. Therefore, it can be reconstructed using the answer gained from the optimization problem. Conversely, in case the average of the gradient of each patch is not larger than that of the total TI image, then the patch belonging to the smooth areas must be reconstructed with less sparsity. This is the set procedure for all the image patches in order to create a high-frequency component of the denoised image. Combining the high-frequency component of the denoised image with the low-frequency component of the noisy image will make the output image. In the proposed method, using an adaptive basis for signal representation and ideas like patch labeling and the adaptive determination of the sparsity in the SVD domain has resulted in acceptable speed and computational complexity.

References

1. Zhang, X., et al.: Compression artifact reduction by overlapped-block transform coefficient estimation with block similarity. IEEE Trans. Image Process. **22**(12), 4613–4626 (2013a)
2. Khan, A., et al.: Image de-noising using noise ratio estimation, K-means clustering and non-local means-based estimator. Comput. Electr. Eng. **54**, 370–381 (2016)
3. Takeda, H., Farsiu, S., Milanfar, P.: Kernel regression for image processing and reconstruction. IEEE Trans. Image Process. **16**(2), 349–366 (2007)
4. Fan, L., et al.: Nonlocal image denoising using edge-based similarity metric and adaptive parameter selection. Sci. China Inf. Sci. **61**(4), 049101 (2018)
5. Zhang, X., et al.: Gradient-based Wiener filter for image denoising. Comput. Electr. Eng. **39**(3), 934–944 (2013b)
6. Shanthi, S.A., Sulochana, C.H., Latha, T.: Image denoising in hybrid wavelet and quincunx diamond filter bank domain based on Gaussian scale mixture model. Comput. Electr. Eng. **46**, 384–393 (2015)
7. Chen, M., et al.: An OCT image denoising method based on fractional integral. In: Optical Coherence Tomography and Coherence Domain Optical Methods in Biomedicine XXII. 2018. International Society for Optics and Photonics
8. Shamsi, A.: Reconfigurable CT QDSM with mismatch shaping dedicated to multi-mode low-IF receivers. Int. J. Ind. Electron., Control Optim. **2**(3), 257–264 (2019)
9. Smirnov, M.W., Silverstein, D.A.: Noise reduction using sequential use of multiple noise models. Google Patents (2019)
10. Huang, Z., et al.: Progressive dual-domain filter for enhancing and denoising optical remote-sensing images. IEEE Geosci. Remote Sens. Lett. **15**(5), 759–763 (2018)
11. Gavaskar, R.G., Chaudhury, K.N.: Fast adaptive bilateral filtering. IEEE Trans. Image Process. **28**(2), 779–790 (2018)
12. Guo, Q., et al.: An efficient SVD-based method for image denoising. IEEE Trans. Circuits Syst. Video Technol. **26**(5), 868–880 (2015)
13. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process. **54**(11), 4311–4322 (2006)
14. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans. Image Process. **15**(12), 3736–3745 (2006)
15. Mallat, S., Yu, G.: Super-resolution with sparse mixing estimators. IEEE Trans. Image Process. **19**(11), 2889–2900 (2010)
16. Jiang, L., Feng, X., Yin, H.: Structure and texture image inpainting using sparse representations and an iterative curvelet thresholding approach. Int. J. Wavelets Multiresolut. Inf. Process. **6**(05), 691–705 (2008)
17. Kalantari, S., Abdollahifard, M.J.: Optimization-based multiple-point geostatistics: a sparse way. Comput. Geosci. **95**, 85–98 (2016)
18. Pizurica, A., Philips, W.: Estimating the probability of the presence of a signal of interest in multiresolution single-and multiband image denoising. IEEE Trans. Image Process. **15**(3), 654–665 (2006)
19. Zhang, J., et al.: Image restoration using joint statistical modeling in a space-transform domain. IEEE Trans. Circuits Syst. Video Technol. **24**(6), 915–928 (2014)
20. Jiang, J., Zhang, L., Yang, J.: Mixed noise removal by weighted encoding with sparse nonlocal regularization. IEEE Trans. Image Process. **23**(6), 2651–2662 (2014)
21. Dabov, K., et al.: Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans. Image Process. **16**(8), 2080–2095 (2007)
22. Dabov, K., et al.: Image denoising with shape-adaptive principal component analysis. Department of Signal Processing, Tampere University of Technology, France (2009)
23. He, Y., et al.: Adaptive denoising by singular value decomposition. IEEE Signal Process. Lett. **18**(4), 215–218 (2011)
24. Dong, W., Shi, G., Li, X.: Nonlocal image restoration with bilateral variance estimation: a low-rank approach. IEEE Trans. Image Process. **22**(2), 700–711 (2012)

25. Mallat, S.: A Wavelet Tour of Signal Processing: The Sparse Way. Academic, Burlington (2008)
26. Starck, J.L., Candès, E.J., Donoho, D.L.: The curvelet transform for image denoising. IEEE Trans. Image Process. **11**(6), 670–684 (2002)
27. Mallat, S., Peyré, G.: A review of bandlet methods for geometrical image representation. Numerical Algorithms **44**(3), 205–234 (2007)
28. Do, M.N., Vetterli, M.: The contourlet transform: an efficient directional multiresolution image representation. IEEE Trans. Image Process. **14**(12), 2091–2106 (2005)
29. Walter, G.G., Shen, X.: Convergence and Summability of Fourier Series. In: Wavelets and Other Orthogonal Systems, pp. 93–110. CRC Press (2018)
30. Eckart, C., Young, G.: The approximation of one matrix by another of lower rank. Psychometrika **1**(3), 211–218 (1936)
31. Razmjooy, N., Mousavi, B.S., Soleymani, F.: A real-time mathematical computer method for potato inspection using machine vision. Comput. Math. Appl. **63**(1), 268–279 (2012)
32. Razmjooy, N., Mousavi, B.S., Soleymani, F., Khotbesara, M.H.: A computer-aided diagnosis system for malignant melanomas. Neural Comput. Appl. **23**(7–8), 2059–2071 (2013)
33. Razmjooy, N., Naghibzadeh, S.S., Mousavi, B.S.: Automatic redeye correction in digital photos. Int. J. Comput. Appl. **95**(9) (2014)
34. Makaremi, M., et al.: A new method for detecting texture defects based on modified local binary pattern. Signal Image Video Process. 1–7 (2018)
35. Mousavi, B.S., et al.: Digital image segmentation using rule-base classifier. Am. J. Sci. Res. (2011)
36. Mousavi, B.S., et al.: Color image segmentation using neuro-fuzzy system in a novel optimized color space. Neural Comput. Appl. **23**(5), 1513–1520 (2013)
37. Rashid Sheykhahmad, F., et al.: A novel method for skin lesion segmentation. Int. J. Inf., Secur. Syst. Manage. **4**(2), 458–466 (2015)

A Comparison of Machine Learning Classifiers for Water-Body Segmentation Task in the PeruSAT-1 Imagery

Roger Huauya, Fidel Moreno, Jonathan Peña, Erwin Dianderas, Antoni Mauricio, and Jose Díaz

Abstract Water-body segmentation is a high-relevance task inside satellite image analysis due to its relationship with environmental monitoring and assessment. Thereon, several authors have proposed different approaches which achieve a wide range of results depending on their datasets and settings. This study is a brief review of classical segmentation techniques in multispectral images using the Peruvian satellite PeruSAT-1 imagery. The areas of interest are medium-sized highland zones with water bodies around in Peruvian south. We aim to analyze classical segmentation methods to prevent future natural disasters, like alluviums or droughts, under low-cost data constraints. We consider accuracy, robustness, conditions, and visual effects in our analysis.

Keywords PeruSAT-1 · Machine learning · Water-body segmentation

R. Huauya (✉) · F. Moreno · J. Peña · J. Díaz
Universidad Nacional de Ingeniería, Lima, Perú
e-mail: rhuauyam@uni.pe

F. Moreno
e-mail: fmorenoc@uni.pe

J. Peña
e-mail: jonathan.pena.a@uni.pe

J. Díaz
e-mail: jcdiazrosado@uni.edu.pe

E. Dianderas
Instituto de Investigaciones de la Amazonía Peruana, Iquitos, Perú
e-mail: edianderas@iiap.gob.pe

A. Mauricio
Department of Computer Science, Universidad Católica San Pablo, Arequipa, Perú
e-mail: manasses.mauricio@ucsp.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_6

1 Introduction

In Peru, heavy summer rains cause alluviums around highlands which affect riverside inhabitants due to streamflow increase. Furthermore, El Nino's effect soars rain frequency and intensity. Thereon, the Peruvian government develops foresight efforts to control the consequences. In 2017, El Nino's impact affected at least one million people and devastated public infrastructure [1–8]. Current prevention efforts point to environmental monitoring systems based on remote sensing, including satellite solutions.

Water-body segmentation allows to perform a spatial [7–14] and a temporal [15–23] interpretation of the human impact and climate change on aquatic ecosystems. Yang et al. [24] mention that satellite imaging broadens this concept to regional and global scales. Nowadays, the-state-of-the-art manages deep representations from large datasets to overcome troublesome conditions like occlusions or quality constraints. Nonetheless, the availability of datasets limits applications ranges in certain zones and countries, especially in the deep learning era [3, 12]. Thus, classical methods are still active in the majority of practical situations.

This comparative study analyzes six well-known classical classifiers for the water-body segmentation task over the Titicaca lake surroundings due to its ecological relevance. We employ the PeruSAT-1 satellite because it regularly maps the entire Peruvian territory, in medium–high quality and free of charge for Peruvian researchers. Human experts label pixels into two categories (water and non-water) endeavoring the balance. Images undergo a regular preprocessing procedure before segmentation. Finally, we compute accuracy and AUC metrics considering robustness by applying various degrees of mislabeling noise.

2 Related Works

Several authors propose different approaches for the water-body segmentation task since spectral-only [2, 7, 18] to deep learning [9, 19, 20] methods. The first ones do provide outstanding results on low-resolution multispectral images but do fail in complex representations under uncertainties constraints. Nevertheless, deep learning techniques require a tremendous amount of data, and they do not represent a significant improvement in low-quality images [19].

Al-Ahmadi et al. [2] apply the iterative self-organizing data analysis technique, known as ISODATA, for multispectral pattern recognition. ISODATA considers multi-class clusters and removal techniques to label complex classes. Chandrakala et al. [7] determine the threshold values for three different classes: water, urban-land, and green-land. Low-level features are grouped using a k-means classifier and LDA decomposition. Shahbaz et al. [22] use color and intensity features to feed a decision tree in multi-label classification.

Aksoy [1] develop a two-step Bayesian classifier using spatial information in high-resolution images. In the first one, spectral and texture features of each pixel are extracted to train the non-parametric model. Then, the iterative division and fusion algorithms convert pixel-level maps into contiguous regions. Jabar et al. [14] introduce a fuzzy logic classification for high-resolution images to flexible the results for uncertainties, improving the accuracy in localized boundaries. Two levels of segmentation are applied: (1) shadows, vegetation, and roads; and (2) the remaining categories.

Langkvist et al. [18] employ a convolutional neural network (CNN) using the bands as inputs and adding a digital surface model (DSM) to boost accuracy. Height information helps to distinguish categories of similar appearance, like vegetation, soil, etc. Miao et al. [19] propose a restricted-receptive field deconvolution net for high-resolution segmentation. They compress redundant layers to overcome the weak pixel-neighborhood relativity [20] without pre-trained models. Further, an edges weighting loss subdues the blurring boundary problem [9]

3 PeruSAT-1 Satellite Imagery

This study adopts the PeruSAT-1 imaging provided by the Space Agency of Peru (CONIDA). The Peruvian satellite PeruSAT-1 is an Earth observation satellite that captures high-resolution multispectral images for military and civil applications inside the Peruvian territory. Developed applications include agroforestry monitoring, urban planning, illegal crop control, disaster monitoring, and road infrastructure control. Metadata includes geographical coordinates, constant values, reception time, among others. Table 1 presents PeruSAT-1 features.

Table 1 Technical features of the PeruSAT-1 satellite

Item	Description
Dimensions	1.0 m × 1.0 m × 1.7 m
Sweeping width	14.5 km
Spatial resolution	Panchromatic: 0.7 m Multispectral: 2.8 m
Spectral bands	Blue: 0.45–0.52 μm Green: 0.53–0.60 μm Red: 0.62–0.69 μm NIR: 0.76–0.89 μm
Radiometric resolution	12 bits
Mass of camera	

4 Methodology

The pipeline of our procedure has six stages: preprocessing, feature selection, labeling, mislabeling noise, and machine learning classifiers. We analyze six machine learning algorithms, three linear and three nonlinear, implemented in Python.

4.1 Preprocessing

Preprocessing standardizes pixel values in two steps. First, a linear transformation computes the radiance values (rad) using the normalized differences (DN) of the image. Second, the top of the atmosphere reflectance (ToA reflectance) is calculated using a nonlinear transformation from the radiance values. This transformation depends on the satellite orbit, the distance between the Sun and the Earth, and other constants. ToA reflectance allows the spectral analysis of pixels to characterize each element via their spectral signature [11]. After this stage, the satellite imagery is ready for analysis.

4.2 Feature Selection

Since the water-body segmentation task is an end-to-end problem, we adapt inputs by making some operations with the bands. Three indexes were chosen as features and calculated from the four bands: two normalized difference vegetation indexes (NDVI) and a normalized difference water index (NDWI). Therefore, the input data concatenate the three indexes creating an input of three features.

Normalized Difference Vegetation Index. The NDVI, in short, is the most widely used index in remote sensing for detecting vegetation areas [15]. NDVI is simple to compute and varies in a short-range (between –1 and +1), which helps to establish fixed thresholds. Equations 1 and 2 calculate the NDVI indexes.

$$\mathrm{NDVI1} = \frac{N - R}{N + R} \tag{1}$$

$$\mathrm{NDVI2} = \frac{N - 0.8 \times R}{N + 0.8 \times R} \tag{2}$$

Normalized Difference Water Index. The NDWI, in brief, is the most appropriate index for mapping water bodies [10] because it considers the green and near-infrared bands (see Eq. 3).

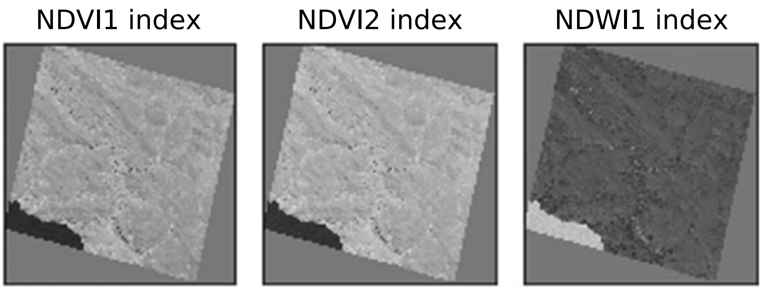


Fig. 1 Accuracy versus noise

$$\text{NDWI} = \frac{G - N}{G + N} \tag{3}$$

After applying the Eqs. 1, 2, and 3, Fig. 1 is obtained.

4.3 Labeling

Labeling was done manually at pixel-level considering two classes: water class and ground class. Then, we generate one mask per class and per sample. Mask sizes are a maximum of 50,000 pixels each to avoid massive computing times. Figure 2 shows the masks created for the training and validation process. Ground areas are inked as orange, while sky blue represents the water bodies.

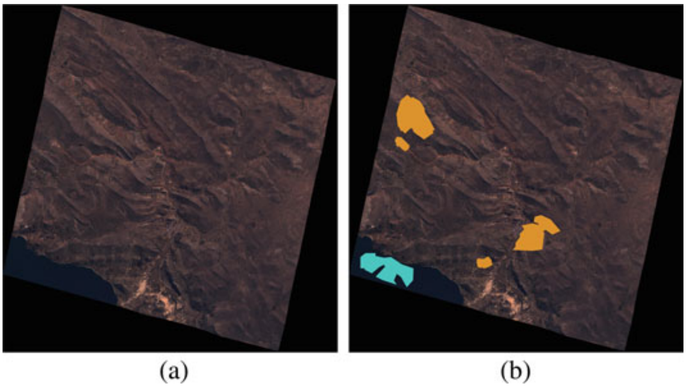
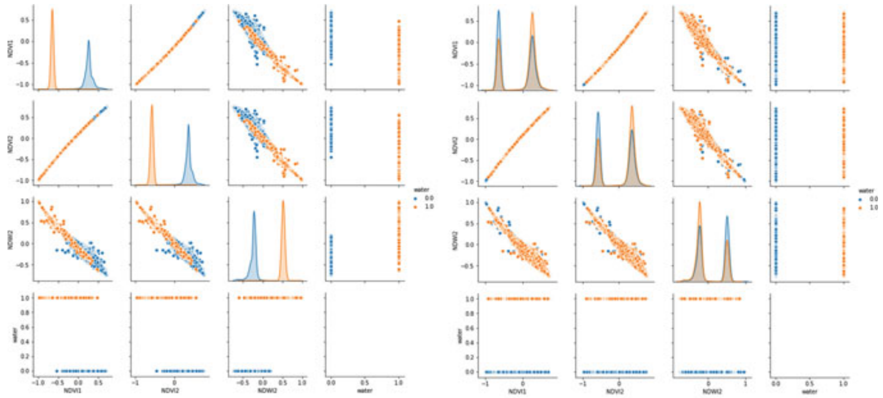


Fig. 2 **a** Original image and **b** Masks created by handcraft



(a) (b)

Fig. 3 Data visualization as a pair plot. **a** 0% of mislabeling noise **b** 60% percent of mislabeling noise

4.4 Mislabeling Noise

Mislabeling noise was added to the training set labels. This noise consists of toggling the label of a percentage of the output training set. Our algorithm adds noise from 0 to 60% with a step of 5%. This procedure can be seen in Fig. 3

4.5 Machine Learning Classifiers

We compare six machine learning classifiers: multilayer perceptron (MLP) [4], logistic regression (LG) [17], support vector machines (SVM) [6], K-nearest neighbors (KNN) [16], decision trees (DT) [13], and random forest (RF) [5]. Logistic regression creates a multivariate regression relationship between a dependent variable and several independent variables. Also, it does not require high computational resources, achieving interpretable solutions. Support vector machines establish a nested hyperplane set to minimize a hinge loss between classes. Then, optimum hyperplane set separates classes in terms of generalization error. Furthermore, SVM uses to perform better in small datasets.

An MLP is a fully connected topological network that correlates patterns and data in an end-to-end solution. MLP methods require a proper amount of data to train as their deep learning counterparts. KNN classification is an instance-based method that includes every new sample into one instance by nearest-neighbor logic. The number of neighbors, the distance criteria, and other hyper-parameters depend tightly on the dataset. KNN regularly achieves good and interpretable results. However, KNN is computationally expensive and has a lack of robustness before noisy datasets. Decision trees are a tree-like graph that learns specific parameters related to the data features in each branch. Random forest is an ensemble method composed of several DT to overcome individual weaknesses.

The implementation of these algorithms was written in Python. The library used for building the classifiers is the Scikit-learn library [21].

5 Results

Table 2 illustrates the validation results considering accuracy and AUC. Accuracy for almost every method is close to 100%. In addition, Fig. 4 shows the inference of the logistic regression on an entire non-trained satellite image.

5.1 Evaluating the Robustness of Classifiers

Although it seems that the six classifiers implemented perform relatively well on the water-body segmentation task, it is also essential to test their performance under disadvantageous conditions. In real-world datasets, there might be some mislabeling or noises on; therefore, we introduce various degrees of mislabeling to test robustness. Figure 5 shows accuracies and AUC versus mislabeling noise percentage. Logistic regression, support vector machines, random forest, and multilayer perceptron are not the most robust because they plummet when the noise rate is high. However,

Table 2 Comparison of the classifiers' accuracy

Metric (%)	MLP	LG	SVM	KNN	RF	DT
Accuracy	99.9489	99.9463	99.9551	99.9548	99.9532	99.9282
AUC	99.9373	99.9373	99.9429	99.9450	99.9406	99.9236

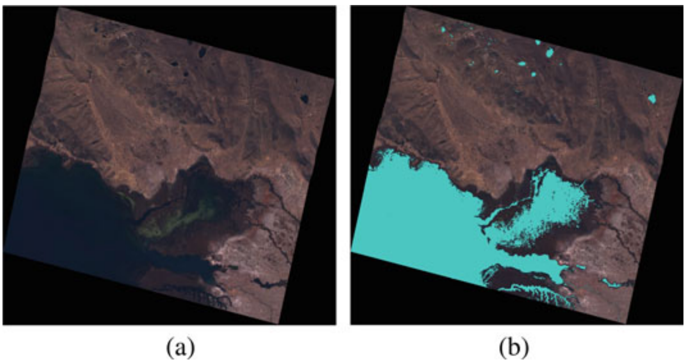
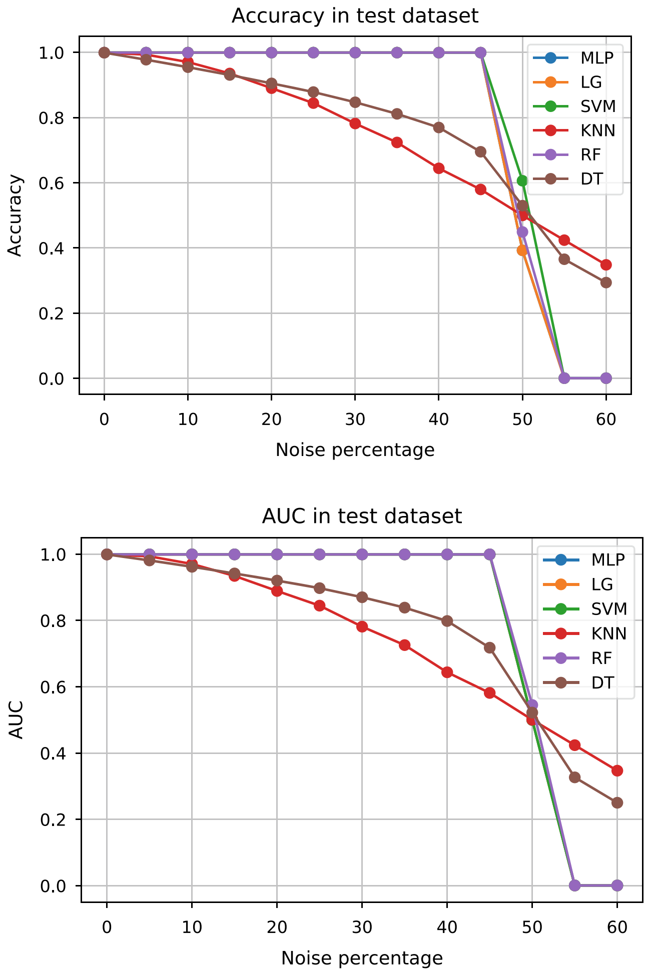


Fig. 4 **a** Original image and **b** Testing image for the trained logistic regression



(a)

(b)

Fig. 5 **a** Accuracy versus noise and **b** AUC versus noise

KNN and decision trees performances in the AUC metric are the best when facing noisy data.

6 Conclusions and Future Works

The SVM model led to the most accurate score, and the KNN led to the most AUC score according to Table 2. Therefore, these models can be used to automate the task of segmenting out water bodies when the dataset has no mislabeling errors. On the other hand, robustness analysis shows that KNN is also a good algorithm even under mislabeling conditions on the dataset. The DT algorithm also has good performance over the other algorithms that have almost 0 in both scores. We can conclude that KNN can be safely used to perform the water-body segmentation even if the training dataset contains mislabeled data.

In the future, we will expand our perspective by including more detailed features and deep representations for low-level and high-level classifications. In addition, we will make a temporal analysis over images of the same place on different dates so that we will be able to predict possible alluviums.

Acknowledgements The present work was supported by the Vice rectorate for Research of the Universidad Nacional de Ingeniería (VRI–UNI). The satellite imagery was provided thankfully by the Head Space Agency of Peru (CONIDA).

References

1. Aksoy, S.: Spatial techniques for image classification. In: Image processing for remote sensing, pp. 241–284. CRC Press (2007)
2. Al-Ahmadi, F., Hames, A., et al.: Comparison of four classification methods to extract land use and land cover from raw satellite images for some remote arid areas, kingdom of Saudi Arabia. Earth **20**(1), 167–191 (2009)
3. Albright, D., Burkhard, S., Lach, A.: Commercial satellite imagery analysis for countering nuclear proliferation. Annual Review of Earth and Planetary Sciences (2018)
4. Introduction neural networks in remote sensing: ATKINSON, Peter M.; TATNALL, A.R. Int. J. Remote Sens. **18**, 699–709 (1986)
5. Breiman, L.: Random forests. Machine learning 45, 5–32 (2001)
6. Burges, C.J.: A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery 2, 121–167 (1998)
7. Chandrakala, M., Amsaveni, R.: Classification of remote sensing image areas using surf features and latent dirichlet allocation. International Journal of Advanced Research in Computer Science and Software Engineering **3**(9), 178–182 (2013)
8. Christidis, N., Betts, R.A., Stott, P.A.: The extremely wet march of 2017 in Peru. Bull. Am. Meteor. Soc. **100**(1), S31–S35 (2019)
9. Duan, L., Hu, X.: Multiscale refinement network for water-body segmentation in high-resolution satellite imagery. IEEE Geoscience and Remote Sensing Letters (2019)
10. Gao, B.C.: Ndwila normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. **58**(3), 257–266 (1996)
11. Goetz, A.F., Vane, G., Solomon, J.E., Rock, B.N.: Imaging spectrometry for earth remote sensing. science 228(4704), 1147–1153 (1985)
12. Hart, A., Hearn, G.: Mapping geohazards in the watersheds above leh, ladakh: The use of publicly-available remote sensing to assist risk management. International Journal of Disaster Risk Reduction **31**, 789–798 (2018)
13. Hunt E B, M.J., Stone: Experiments in induction. Academic Press, New York (1996)
14. Jabari, S., Zhang, Y.: Very high resolution satellite image classification using fuzzy rule-based systems. Algorithms **6**(4), 762–781 (2013)
15. Jiang, Z., Huete, A.R., Chen, J., Chen, Y., Li, J., Yan, G., Zhang, X.: Analysis of ndvi and scaled difference vegetation index retrievals of vegetation fraction. Remote Sens. Environ. **101**(3), 366–378 (2006)
16. KELLER, James M.; GRAY, M.R.: A fuzzy k-nearest neighbor algorithm. IEEE transactions on systems, man, and cybernetics pp. 580–585 (1985)

17. Kleinbaum, D.G., Dietz, K., Gail, M., Klein, M., Klein, M.: Logistic regression. Springer (2002)
18. Längkvist, M., Kiselev, A., Alirezaie, M., Lout_, A.: Classification and segmentation of satellite orthoimagery using convolutional neural networks. Remote Sensing 8(4), 329 (2016)
19. Miao, Z., Fu, K., Sun, H., Sun, X., Yan, M.: Automatic water-body segmentation from high-resolution satellite images via deep networks. IEEE Geosci. Remote Sens. Lett. **15**(4), 602–606 (2018)
20. Mustafa, F.A., Bayat, O.: Water surface area detection using remote sensing temporal data processed using matlab. The International Journal of Electrical Engineering & Education p. 0020720919851500 (2019)
21. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)
22. Shahbaz, M., Guergachi, A., Noreen, A., Shaheen, M.: Classification by object recognition in satellite images by using data mining. Proceedings of the World Congress on Engineering. **1**, 4–6 (2012)
23. Sheng, Y., Song, C., Wang, J., Lyons, E.A., Knox, B.R., Cox, J.S., Gao, F.: Representative lake water extent mapping at continental scales using multi-temporal landsat-8 imagery. Remote Sens. Environ. **185**, 129–141 (2016)
24. Yang, X., Zhao, S., Qin, X., Zhao, N., Liang, L.: Mapping of urban surface water bodies from sentinel-2 msi imagery at 10 m resolution via ndwi-based image sharpening. Remote Sensing **9**(6), 596 (2017)

A Stock Market Forecasting Model in Peru Using Artificial Intelligence and Computational Optimization Tools

Miguel Angel Cano Lengua, Mirko Jerber Rodríguez Mallma, and Erik Alex Papa Quiroz

Abstract It is proposed the development of a forecast model capable of predicting the behavior of the price indices and quotes of the shares traded on the Lima Stock Exchange, based on the use of artificial intelligence techniques such as artificial neural networks and fuzzy logic based on computational optimization methods. The proposed model considers the forecast, in addition to the historical quantitative data of the share price, the inclusion of qualitative macroeconomic factors that significantly influence the behavior of the time series of the stock markets. It is about harnessing the ability of artificial neural networks to work with nonlinear quantitative data and their capacity for learning and also take advantage of the fuzzy logic technique to simulate the way of reasoning of human beings by defining judgment rules or knowledge base and their evaluation through inference mechanisms. The main contribution is to demonstrate that the proposed model is capable of obtaining more optimal approximations in the forecast of the financial time series.

Keywords Artificial neural networks · Stock market forecasting · Computational optimization

1 Introduction

The study of the stock market is important to make an adequate forecast of the behavior of stock prices and other quantitative characteristics. In Peru, this activity began in 1860 with the creation of the Lima mercantile exchange. Currently, because

M. A. Cano Lengua (✉)
Universidad Privada Norbert Wiener, Lima, Perú
e-mail: miguel.cano@uwiener.edu.pe

M. J. Rodríguez Mallma
Universidad Nacional de Ingeniería, Lima Perú and Universidad Ricardo Palma, Lima, Perú
e-mail: mjrodriguezm@uni.pe

E. A. Papa Quiroz
Universidad Privada del Norte, Universidad Nacional Mayor de SanMarcos, Lima, Perú
e-mail: erikpapa@gmail.com

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_7

the market does not behave in a linear manner, and in the prediction, it is necessary to include variables from both the fundamental analysis and the technical analysis, a series of investigations on the application of alternative and multidisciplinary methods have recently been developed in the analysis of financial markets, such as chaos theory, fractals, wavelets, support vector machines, hidden Markov models, artificial intelligence, and combinations of these methods, see Smith and Gupta [1].

Recent researches regarding the prediction of the stock market in Peru are as follows: Cáceres [2] using support vector regression, Mejía [3] using time series, Vegas [4] using data mining for high markets frequency. This work uses a neural network-based methodology adding fuzzy logic reasoning for the preprocessing of input variables from qualitative nature.

This article focuses on the use of artificial neural networks, fuzzy logic, and computational optimization, to develop and build a suitable model that achieves a better approximation in the forecast of the behavior of stock prices, compared to traditional forecasting methods used in the stock market of Peru. This approach seems to be very promissory because contributes to the study of forecasting stock exchange in Peru.

The content of this article is as follows. In Sect. 2, we present the proposed model. Section 3 details the methodology of the selection of macroeconomic variables, the indices, and indicators of the stock market. Section 4 presents the numerical experiments and the main contributions of this study. Finally, in Sect. 5, we give our conclusions.

2 Proposed Model

In this work, a model based on the sequential hybrid system is used and will be carried out in two parts: The first one corresponds to the definition and development of the fuzzy logic system for the processing of qualitative values, and the second one corresponds to the development and definition of the artificial neural networks model that will be in charge of the processing of the set of quantitative variables and the result variables that the fuzzy logic system has previously processed in the first part.

The fuzzy logic system will carry out a treatment of qualitative variables, through a process of defining business rules, to determine in a more optimal way the influence and implication of qualitative variables in the forecast of time series in a market of values, whose output will be used as one more input in the proposed neural networks model.

Figure 1 shows the architecture for the implementation of the hybrid model proposed in this research, which is composed of a model of artificial neural networks and a fuzzy logic system.

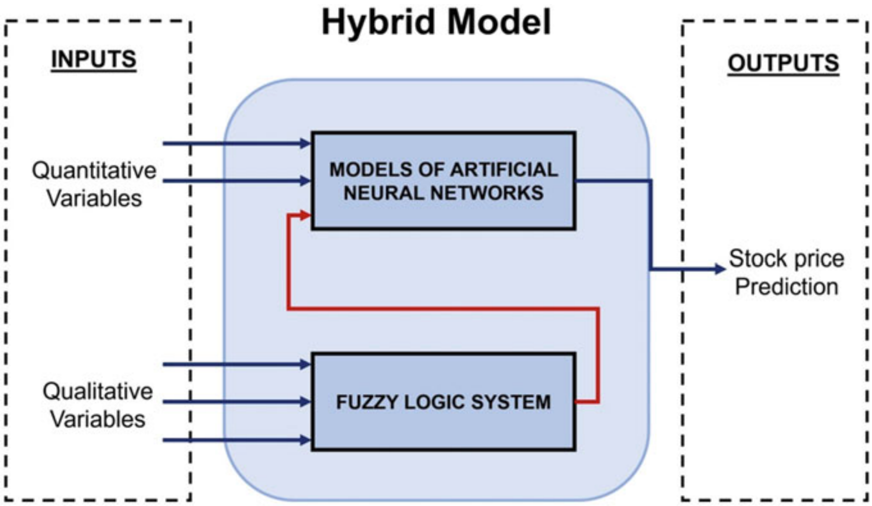


Fig. 1 Proposed hybrid model

2.1 Description of the Proposed Model

The CRISP-DM methodology was used, which is one of the most standardized and accepted when facing data mining projects. The design phases are

Phase 1: Understanding the business or problem. Phase 2: Understanding the data. Phase 3: Data preparation. Phase 4: Modeling. Phase 5: Evaluation. Phase 6: Implementation.

2.2 Definition of the Proposed Model

In Fig. 2, we show the unified data of both the qualitative and quantitative variables from the model.

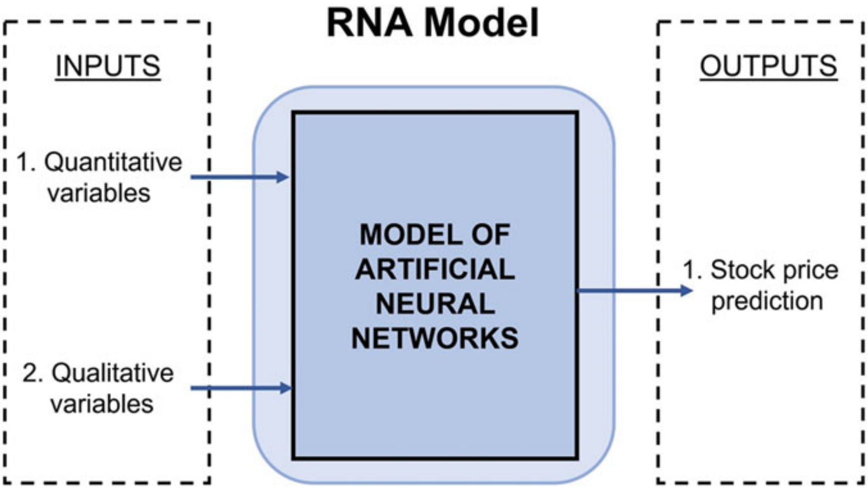


Fig. 2 Input variables of the model (qualitative and quantitative)

2.3 The Architecture of the Artificial Neural Networks Model

The architecture of the artificial neural networks model proposed in this work is multilayer feedforward neural networks (MFNN). This consists of an input layer, an output layer, and at least one layer of hidden neurons between the input and output layers. The information flow ranges from left to right, with x inputs that pass through the network via connections with assigned weights to the hidden layer of neurons and, subsequently, to the output layer. The weights that connect the input elements i with the hidden neuron j are denoted by $w_{i,j}$, while the weights that connect the hidden neuron j with the output neuron k are denoted by $v_{k,j}$ [1].

For the construction of the artificial neural networks model proposed in this research, the methodology proposed by Kaastra and Boyd [5] will be followed for the design of neural network models applied to financial and economic time series predictions, which is composed of eight steps and described below. In the inflation case, which is the rise in the price of goods and services, its impact on the stock market is generally known as negative. Also, the stock price is very sensitive to the financial situation of the company which can be impacted by the GDP index. In the case of interest rate, which is the percentage of principal charged by the lender for the use of its money, its value could be used as a measure of the economy's health which also affects the stock market. Finally, exchange rate fluctuations affect stock prices.

3 Description of Variables

3.1 Macroeconomic Qualitative Variables

Several studies have been carried out to determine the influence of macroeconomic factors on the behavior of stock prices in the stock market. In this study, we consider four qualitative macroeconomic variables: inflation, mining sector Gross Domestic Product (GDP), interest rate and exchange rate [6]. In the inflation case, which is the rise in the price of goods and services, its impact on the stock market is generally known as negative. Also, the stock price is very sensitive to the financial situation of the company which can be impacted by the GDP index. In the case of interest rate, which is the percentage of principal charged by the lender for the use of its money, its value could be used as a measure of the economy's health which also affects the stock market. Finally, exchange rate fluctuations affect stock prices.

3.2 Indicators as Quantitative Variables

In addition to the macroeconomic variables that significantly influence the behavior of stock prices in the stock market, it is important to take into account various indices or indicators that allow measuring the behavior and health of the stock market over time [7]. In this study, four of the most important indicators to know the oscillation of the Peruvian stock market were used as quantitative variables for our proposed model. These indicators are the general index of the Lima Stock Exchange (IGBVL), the selective index of the Lima Stock Exchange (ISBVL), the sectoral index (SI) and the national capitalization index (INCA). Both IGBVL and ISBVL are liquidity indexes that show, as an average, the tendency of the most important stock prices on the Peruvian stock market. The SI is the average index for the whole Peruvian mining sector, and the INCA is a capitalization index that reflects the price behavior of a diversified portfolio of 20 shares of companies whose majority economic activity is in Peru and that are listed on the Lima Stock Exchange.

The dataset has been obtained mainly from two sources. For macroeconomic variables, the Peruvian central bank provides free access to datasets [8].

4 Numerical Experiments

For this study, one of the largest companies in the mining sector in Peru has been studied, the "*Volcan*" enterprise. Historical information on the price of its shares traded on the Lima Stock Exchange in the period between January 2001 and December 2010 is considered. Table 1 presents the basic information on this

Table 1 Basic volcano information extracted from [9]

Enterprise data	
mnemonic	VOLCAN
RUC	20,383,045,267
Code	VOLCABC1
Number of outstanding shares	1,249,669,371
Number of shareholders	8841
Nominal value	S/. 1.00
Date of listing in the BVL	31/03/1998

company. In the period of study (from January 2001 to December 2010), 2411 historical data have been identified for each of the variables used in the study. The models to follow all have the same input, a single output (1) and differ in the processing of the inside of the layer. We denote MSE as the mean square error and iterations are the repetitions that perform to obtain the solution. The mean square error is obtained using the maximum descent optimization method.

Several topologies have been used in order to find the best model for stock price forecasting purposes. As Table 2 shows, the topology is defined as (p–q–1) where "p" indicates the number of inputs (nodes) used in the input layer, "q" indicates the number of nodes used in the hidden layer, and 1 indicates that all models used one node in their output layer. Figure 3 summarizes the performance of the topology model at the training phase.

Table 2 Test models and topologies

Test Model	Topology (p–q–1)	MSE	Iterations
Model 1	7–3–1	0.0740	49
Model 2	**7–4–1**	**0.0441**	**22**
Model 3	**7–5–1**	**0.0544**	**113**
Model4	7–6–1	0.0934	77
Model 5	**7–7–1**	**0.0428**	**179**
Model 6	7–8–1	0.0987	43
Model 7	7–9–1	0.0778	118
Model 8	7–10–1	0.1020	161
Model 9	7–11–1	0.1090	123
Model 10	7–12–1	0.1220	69
Model 11	7–13–1	0.0856	117
Model 12	7–14–1	0.0936	89
Model13	7–15–1	0.1530	61
Model 14	7–18–1	0.2340	56
Model 15	7–20–1	0.7040	95

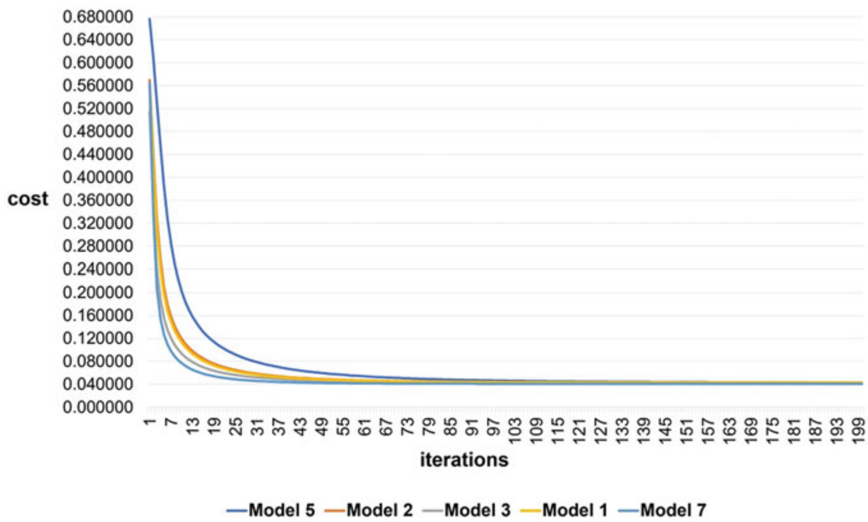


Fig. 3 Error for each iteration of the five best models

5 Results

From the experimentation process, we have identified the best results to forecast the price indices and quotes of the shares traded on the Lima Stock Exchange were obtained with model 5 whose mean squared error has the lowest value 0.0428; however, the number of iterations on this model is also the highest which is an indicator that artificial neural networks need high levels of computational resources and are time-consuming algorithms. Since our results are measured in terms of mean squared error metric, we can use it to compare with any other forecasting method or algorithm and evaluate its accuracy by comparison. The results of this study suggest that artificial neural networks can be successfully applied to process sequenced data type by analyzing time series and make predictions with a good level of confidence to support decision-makers on stock market prediction tasks.

6 Conclusions

In this work, we introduced a hybrid forecasting model applied to the stock market in Peru by using artificial intelligence tools and computational optimization methods. Our work uses both qualitative and quantitative variables that impact on the behavior of stock markets. The relevance of this work is that our proposed hybrid model can significantly improve the stock market prediction accuracy since more information was analyzed. Future work on stock market forecasting by using artificial intelligence tools should be performed at other companies in Peru to confirm that our results can be generalized. Also, we should include new artificial neural networks-based algorithms such as recurrent neural networks, which has demonstrated good performance in analyzing sequenced data type [10]. Finally, a decision support system can be applied to assess our proposed model predictions and then establish a business case in terms of return of investments after a real stock market investment in Peru, taking into account different scenarios and levels of risk involved.

References

1. Smith, K., Gupta, J.: Neural Networks in Business: Techniques and applications, Pennsylvania. Idea Group Publishing, USA (2014)
2. Cáceres V. Stock Exchange Price Prediction using Support Vector Regression. (Pre-degree thesis) University of Lima, Lima (2018)
3. Mejia L. Modeling the General Index of the Lima Stock Exchange with Time Series. (Pre-degree thesis) National University of Trujillo, Trujillo (2017).
4. Vegas, I.: Stock Market Prediction Through Data Mining For High Frequency Market. Final Degree Project)Polytechnic University of Madrid, Madrid (2016)
5. Kaastra, I., Boyd, M.: Designing a neural network for forecasting financial and economic time series. Neurocomputing Elsevier Science **10**, 215–236 (1996)
6. Gay, Jr., Robert D., Nova Southeastern, Effect Of Macroeconomic Variables On Stock Market Returns For Four Emerging Economies, International Business & Economics Research Journal (2016).
7. Pastor, J.: Stock Exchange - Operation and Stock Market Valuation. Lima, Peru (2011)
8. Central Reserve Bank of Peru. https://estadisticas.bcrp.gob.pe/estadisticas/series/. (Consulted in Dezember 16, 2019).
9. Lima Stock Exchange. https://www.bvl.com.pe/inf_corporativa64801_Vk9MQ0FBQzE.html, last accessed 2019/09/06.
10. Oeda, S., Kurimoto, I., & Ichimura, T. Time Series Data Classification Using Recurrent Neural Network with Ensemble Learning. Knowledge-Based Intelligent Information and Engineering Systems, 742–748 (2006).

Extreme Learning Machine for Business Sales Forecasts: A Systematic Review

Edu Saldaña-Olivas and José Roberto Huamán-Tuesta

Abstract Technology in business is vital, in recent decades technology has optimized the way they are managed making operations faster and more efficient, so we can say that companies need technology to stay in the market. This systematic review aims to determine to what extent an Extreme Learning Machine (ELM) system helps sales forecasts (SF) of companies, based on the scientific literature of the last 17 years. For the methodology, the systematic search for keywords began in the repositories of Google Scholar, Scielo, Redalyc, among others. Documents were collected between 2002 and 2019 and organized according to an eligibility protocol defined by the author. As an inclusion criteria, the sources in which their conclusions contributed to deepening the investigation were taken and those that did not contribute were excluded. Each of the results represented in graphs was discussed. The main limitation was the little information on the subject because it is a new topic. In conclusion, an ELM system makes use of both internal and external data to develop a more precise SF, which can be used not only by the sales and finance area but also to coordinate with the production area a more exact batch to be produced; this has a great impact on the communication and dynamism of companies to reduce costs and increase profits.

Keywords Business sales forecast · Extreme learning machine · Systematic review

1 Introduction

Technology has great relevance in commercial operations; its main role in business is to boost growth and improve operations. Without technology, companies will almost certainly not be able to survive in today's competitive and globalized market.

E. Saldaña-Olivas (✉) · J. R. Huamán-Tuesta
Universidad Privada del Norte (UPN), Trujillo, Peru
e-mail: geduso@hotmail.com

J. R. Huamán-Tuesta
e-mail: jose.huaman@upn.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_8

Currently, the great concern of organizations lies in having accurate inventories in their stores. Faced with this problem, there are several techniques that a company can use to acquire the amount of inventory necessary to allow it to reach or exceed the sales target. Here, it is important to note that these techniques can be performed regardless of business turnover, billing size, the nature of the company, or its scope (whether local or international). One of these alternatives, and the most used to present better results, is the projection of sales forecasts, which is basically a forecasting system for a future event that by its nature is uncertain and random, whose responsibility falls on all areas of the company, the same that are fed with information so that the forecasts are more accurate [1]. Forecasting is a multifunctional process and, therefore, its management requires the same type of tools as the management of other multifunctional processes. The forecasting process involves staff from many functions, and not everyone knows the forecasting process in its entirety. Production planning, inventory management, and forecasting should be understood as a whole before it is possible to set justified objectives for an individual part of it [2]. In recent decades, computer science has transformed work in almost every sector of the economy. We are at the beginning of an even bigger and faster transformation due to recent advancements in machine learning (ML), a branch of artificial intelligence capable of accelerating the pace of automation, bringing implications for the economic flow and the workforce in the future of global trade [3]. In an effort to keep improving, extreme learning machines have been developed, an emerging ML technique that has become an area of research in recent years due to the significant contributions made by numerous researchers around the world. The objective of this study is to determine to what extent an extreme learning machine system helps companies' sales forecasts, based on the systematic review of the scientific literature of the last 17 years. Studying these tools has relevance, and influences directly and indirectly, on the profits of a company, since these tools have a great impact on communication and dynamism between areas, helping to reduce expenses, therefore, to increase profits, and to position themselves better against the competition with stronger and more accurate demand estimates. This systematic review is structured as follows: First, we describe the methodology with the inclusion and exclusion criteria; subsequently, we detail the fields of the chosen sources, included and excluded, we present the results and discuss each one of them, with the limitations that we had for the investigation; and we close with the conclusions.

2 Methodology

2.1 Information Search Process

In this paper, a systematic review (SR) of the scientific literature in the last 17 years, between 2002 and 2019, was carried out, taking as a guideline the research and analysis of the theoretical basis of this methodology. SR as scientific research in which the

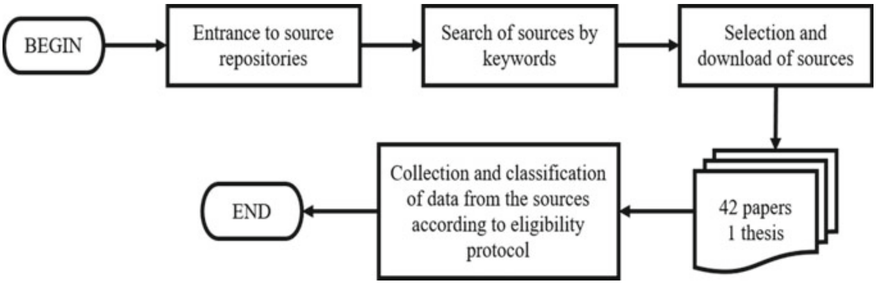


Fig. 1 Flowchart of the information search process

units of analysis are the original primary studies, from which it is intended to answer a research question clearly formulated through a systematic and explicit process [4].

The objective of this investigation was to determine to what extent does an ELM system helps the SFs of companies, described in the scientific literature of the last 17 years. For this, Google Scholar, Scielo, and Redalyc were used as repositories, mainly, in which the following keywords (and their pair in Spanish) were used individually and in combination: sales forecast, machine learning, and extreme learning machine; from the results, theses and papers were selected and downloaded for the elaboration of this review. Figure 1 shows a flowchart of the information search process.

2.2 Inclusion and Exclusion Criteria

Original papers and theses published based on indexed scientific data, in English, Spanish, and Portuguese, between 2002 and 2019, were included, which:

- It defines the theoretical basis of SF, ML, and ELM,
- It describes the importance of an SF in the processes of a company,
- It compares the efficiency between SF models and how to make them,
- It compares ML and ELM system models for forecasting,
- It determines how an ML and ELM system affects the processes of a company,
- And it determines the suitability of using ELM systems for SFs.

Sources whose conclusions did not meet any of the points of the inclusion criteria raised by the author were excluded, and were not included for the deepening of this research. Figure 2 shows a flowchart of the inclusion and exclusion criteria.

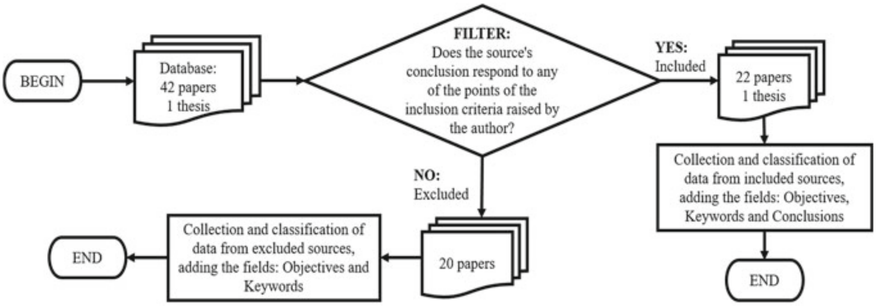


Fig. 2 Flowchart of the inclusion and exclusion criteria

3 Results and Discussion

The searches for sources were carried out based on the variables, for which the keywords were used, this helped to deepen the topic and clarify the objective raised in this research. However, there are few previous research studies on the specific subject, because its inquiry has recently gained momentum in recent years. In addition to this, most of the articles only had paid access, so we had to turn to people who had purchased articles to have full access to them. Most of the sources were in English, which was not inconvenient because the language is mastered; however, with articles in Portuguese, there was a slight difficulty in understanding, despite the use of Google Translator, due to little mastery of the language.

The search for sources in the repositories was virtual and yielded a total of 43 documents extracted from Google Scholar (20), Scielo (14), Redalyc (5), Hindawi (2), Taylor & Francis (1), and World Scientific (1), through the individual and combined use of keywords. These documents are distributed in 42 papers and only (1) thesis. To this total, a filter was applied according to the defined inclusion and exclusion criteria. The sources selected by inclusion criteria, which was determined by the relationship of the sources with the objective of this investigation, were the following: [2, 3, 5–25]. In total, 23 sources were included to deepen the topic according to the period between 2002 and 2018. Therefore, the sources not considered by exclusion criteria were: [26–45]. In total, 20 sources were excluded.

Figure 3 shows the number of sources by periods of years, and most of the publications are concentrated in 2015 with a total of eight (8), followed by the years 2014 and 2017 with five (5), and 2011 and 2019 with four (4). The series of events that ocurred in 2015 prompted the study and use of ML: Amazon launched its own ML platform [46]; Microsoft created the “Distributed Machine Learning Toolkit,” which enabled efficient distribution of ML problems across multiple computers [47]; Elon Musk and Sam Altman, among others, founded the non-profit organization OpenAI, endowing it with US$ 1 billion to ensure that the development of artificial intelligence has a positive impact on humanity [48].

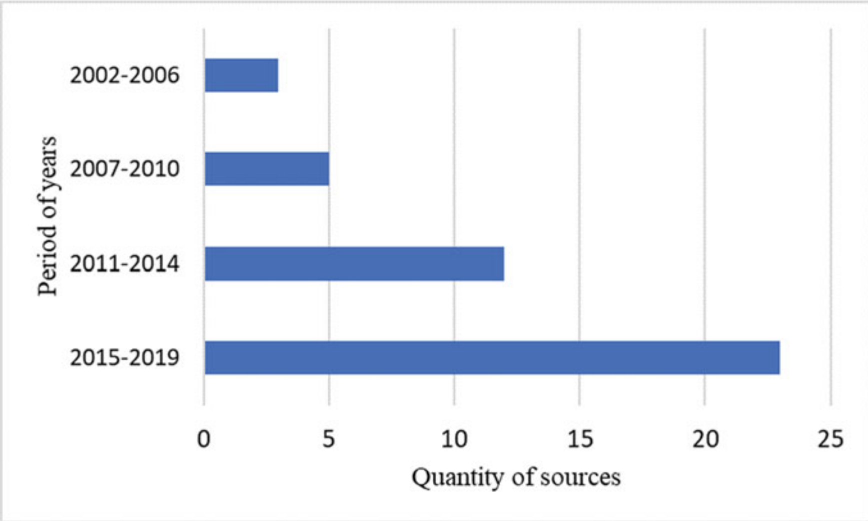


Fig. 3 Number of sources by periods of years

When the years of publication were grouped by time periods, it was clear that the subject of study has gained momentum in recent years. The events that accompanied the momentum of ML in these years were: In 2016, Google DeepMind beat professional gamer Lee Sedol in Go (considered one of the most complicated board games) by 4 games to 1, expert Go players affirm that the algorithm was able to perform "creative" movements that had not been seen so far [49]; in 2017, OpenAI trained chatbots or conversational agents, who invented their own language to cooperate and achieve their goal effectively [50]; soon after, Facebook also successfully trained agents to negotiate and even lie [51]; an algorithm also developed by OpenAI defeated the best players in 1-versus-1 matches of the online game Dota 2 [52]. The percentage of sources found by countries involved is Brazil with 23% of the total (11 papers), China with 18% (nine papers), and the USA with 10% (five papers) as shown in Fig. 4; and this is directly related to the contribution of these countries to the scientific ecosystem. The English language is the one that monopolized the results of the SR with 86% of the total sources (37 papers), followed by sources in Spanish and Portuguese with 7% each (adding six sources between them). To better understand this phenomenon, Carlo Duarte, oceanographer at King Abdullah University of Science and Technology in Saudi Arabia, explained in an interview with "El Mundo": "English is the vehicular language of science, as Latin was before and, for a while, the French. Proficiency in English is essential to access research results, which are mostly published in this language, and share the results of our own research" [53].

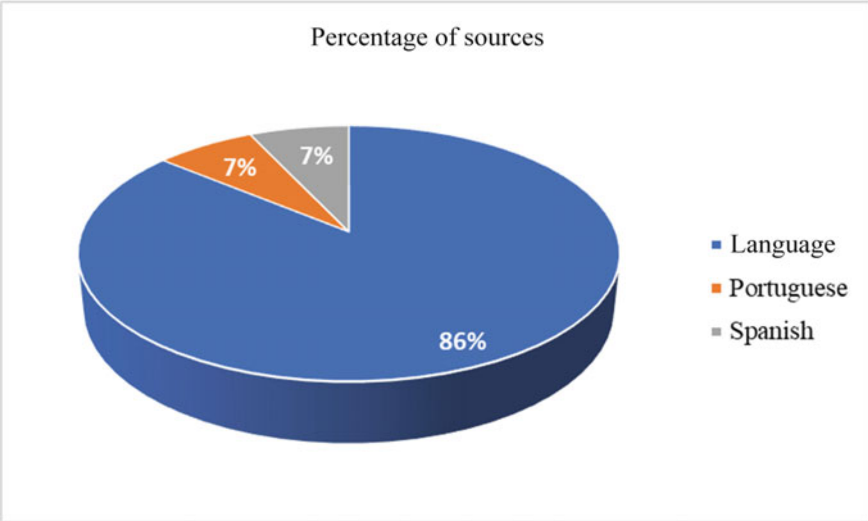


Fig. 4 Sources by language

4 Conclusions

A large-scale SF based on ELM has a reduced error in the square root of the forecast mean and improves its accuracy and speed; furthermore, an hybrid ELM model can effectively extract the underlying information from the sales data and improve SF performance by producing lower prediction error, and improves training speed more than other SF models.

This work aims to help companies understand the benefits of using a SF and implement an ELM system for it, in order to increase their profits and improve their internal processes. SFs are directly related to the increase in sales, because an accurate forecast is a fundamental basis for decision-making in the management of a company, providing relevant information to coordinate with the production area and thus reduce the loss of products without turnover, which translates into more products sold and fewer losses due to in-stock or expired products. It is understood that there is no specific SF model that is more beneficial than another, within the traditional models; but it is concluded that the use of ELM systems greatly improves the accuracy of SF compared to traditional models.

Although there is not much literature that refers explicitly to ELM systems for SFs, more than a limiter it is an opportunity for this paper to be an important contribution for people interested in the subject. That is why an in-depth study with real data from a company or a product is suggested to better understand how an ELM algorithm works in the efficiency of a SF and the cost–benefit of implementing it.

References

1. Vargas, M.: La importancia de implementar el uso de pronósticos en las empresas. Conexión ESAN, Lima, Peru. Available at: https://www.esan.edu.pe/conexion/actualidad/2015/05/11/importancia-implementar-el-uso-de-pronosticos-empresas/ (2015)
2. Kerkkänen, A., Korpela, J., Huiskonen, J.: Demand forecasting errors in industrial context: Measurement and impacts. Int. J. Product. Econ. **118**(1), 43–48. Lappeenranta University of Technology, Lappeenranta, Finland (2009). https://doi.org/10.1016/j.ijpe.2008.08.008
3. Brynjolfsson, E., Mitchell, T.: What can machine learning do? Workforce implications. Science **358**, 1530–1534. Science, USA (2017). https://doi.org/10.1126/science.aap8062
4. Ferreira González, I., Urrútia, G., Alonso-Coello, P.: Revisiones sistemáticas y metaanálisis: bases conceptuales e interpretación. Revista Española de Cardiología, **64**(8), 688–696. Revista Española de Cardiología, Spain (2011). https://doi.org/10.1016/j.recesp.2011.03.029
5. Sokolov-Mladenović, S., Milovančević, M., Mladenović, I., Alizamir, M.: Economic growth forecasting by artificial neural network with extreme learning machine based on trade, import and export parameters. Comput. Human Behav. **65**, 43–45. University of Nis, Serbia & Islamic Azad University, Iran (2016). https://doi.org/10.1016/j.chb.2016.08.014
6. Chen, F.L., Ou, T.Y.: Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. Expert Syst. Appl. **38**(3), 1336–1345. Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan (2011). https://doi.org/10.1016/j.eswa.2010.07.014
7. Gao, M., Xu, W., Fu, H., Wang, M., Liang, X.: A novel forecasting method for large-scale sales prediction using extreme learning machine. In: 7th International Joint Conference on Computational Sciences and Optimization, pp. 602–606. Beijing, China (2014). https://doi.org/10.1109/CSO.2014.116
8. Landa-Torres, I., Ortiz-García, E.G., Salcedo-Sanz, S., Segovia-Vargas, M.J., Gil-López, S., Miranda, M., Del Ser, J.: Evaluating the internationalization success of companies through a hybrid grouping harmony search-extreme learning machine approach. IEEE J. Select. Topics Signal Process. **6**(4), 388–398. IEEE Signal Processing Society (2012). https://doi.org/10.1109/JSTSP.2012.2199463
9. Liakos, K.G., Busato, P., Moshou, D., Pearson, S., Bochtis, D.: Machine learning in agriculture: a review. Sensors **18**(8), 1–29. MDPI Open Access Journals (2018). https://doi.org/10.3390/s18082674
10. Veenadhari, S., Misra, B., Singh, C.D.: Machine learning approach for forecasting crop yield based on climatic parameters. In: 2014 International Conference on Computer Communication and Informatics. International Conference on Computer Communication and Informatics, Coimbatore, India (2014). https://doi.org/10.1109/ICCCI.2014.6921718
11. Huang, G.B.: What are extreme learning machines? Filling the gap between frank Rosenblatt's Dream and John von Neumann's Puzzle. Cognit. Comput. **7**, 263–278. Nanyang Technological University, Singapore, Singapore (2015). https://doi.org/10.1007/s12559-015-9333-0
12. Aucapiña, J.: Modelo de elaboración de pronóstico de ventas mediante el uso de redes neuronales artificiales y SVR. Pregrade thesis. Universidad de Cuenca, Ecuador. https://dspace.ucuenca.edu.ec/handle/123456789/31256 (2018)
13. Corres, G., Passoni, L.I., Zárate, C., Esteban, A.: Estúdio Comparativo de Modelos de Pronóstico de Ventas. Iberoamerican J. Indus. Eng. **6**(11), 113–134. Universidad Nacional Del Centro de La Provincia de Buenos Aires & Universidad Nacional de Mar del Plata, Argentina (2015). https://doi.org/10.13084/2175-8018/ijie.v6n11p113-134
14. Wacker, J.G., Lummus, R.R.: Sales forecasting for strategic resource planning. Int. J. Oper. Product. Manage. **22**(9–10), 1014–1031. International Journal of Operations and Production Management (2002). https://doi.org/10.1108/01443570210440519
15. Sanders, N.R., Graman, G.A.: Quantifying costs of forecast errors: a case study of the warehouse environment. Omega **37**(1), 116–125. Texas Christian University & Michigan Technological University, USA (2009). https://doi.org/10.1016/j.omega.2006.10.004

16. Zhang, Y., Wei, Z., Zhang, L., Lin, Q., Du, J.: Improved online sequential extreme learning machine for simulation of daily reference evapotranspiration. Tecnología y Ciencias del Agua 8(2), 71–76. Instituto Mexicano de Tecnología del Agua, Mexico (2017). https://doi.org/10.24850/j-tyca-2017-02-12
17. Kan, G., He, X., Ding, L., Li, J., Hong, Y., Lei, T., Huang, P.: Daily streamflow simulation based on the improved machine learning method. Tecnología y Ciencias del Agua **8**(2), 51–60. Instituto Mexicano de Tecnología del Agua, Mexico (2017). https://doi.org/10.24850/j-tyca-2017-02-05
18. Dais de FariaI, M., dos Santos de Carvalho, J.L.F.: Planejamento de cenários e amplificação da variabilidade da demanda. Revista Eletrônica de Administração **17**(3). Universidade Federal do Rio de Janeiro, Brazil (2011). https://doi.org/10.1590/S1413-23112011000300010
19. Scarpin, C.T., Steiner, M.T.A.: Proposal for a strategic planning for the replacement of products in stores based on sales forecast. Pesquisa Operacional **31**(2), 351–571. Universidade Federal do Paraná, Brazil (2011). https://doi.org/10.1590/s0101-74382011000200008
20. de Melo, D.C., Alcântara, R.L.C.: A gestão da demanda em cadeias de suprimentos: uma abordagem além da previsão de vendas. Gestão & Produção **18**(4), 809–824. Universidade Federal de Uberlândia & Universidade Federal de São Carlos (2012). https://doi.org/10.1590/s0104-530x2011000400009
21. Claveria, O., Torra, S., Monte, E.: Modelling tourism demand to Spain with machine learning techniques. The impact of forecast horizon on model selection. Revista de Economía Aplicada 24(72), 109–132. Universidad de Zaragoza, Spain (2016). https://redalyc.org/articulo.oa?id=96949059006
22. Lu, C.-J., Shao, Y.E.: Forecasting computer products sales by integrating ensemble empirical mode decomposition and extreme learning machine. Math. Prob. Eng. **2012**, 1–15. University of Science and Technology & Fu Jen Catholic University, Taiwan (2012). https://doi.org/10.1155/2012/831201
23. Hu, K., Zhou, Z., Weng, L., Liu, J., Wang, L., Su, Y., Yang, Y.: An optimization strategy for weighted extreme learning machine based on PSO. Int. J. Patt. Recognit. Artif. Intell. **31**(1). World Scientific (2017). https://doi.org/10.1142/s0218001417510016
24. Tran, T. Van, Wang, Y.N.: An evolutionary extreme learning machine based on chemical reaction optimization. J. Inf. Optimiz. Sci. **38**(8), 1265–1290. Hunan University, China & Hanoi University of Industry, Vietnam (2017). https://doi.org/10.1080/02522667.2016.1220094
25. Cao, J., Lin, Z.: Extreme learning machines on high dimensional and large data applications: a survey. Math. Prob. Eng. **2015**, 1–13. Hangzhou Dianzi University, China & Nanyang Technological University, Singapore (2015). https://doi.org/10.1155/2015/103796
26. Tsai, C.-F., Wang, S.-P.: Stock price forecasting by hybrid machine learning. In: Proceedings of the International Multi Conference of Engineers and Computer Scientists, vol. 1. International Multi Conference of Engineers and Computer Scientists (2009)
27. Choudhry, R., Garg, K.: A hybrid machine learning system for stock market forecasting. Eng. Technol. Int. J. Comput. Inf. Eng. **2**(3), 689–692. World Academy of Science (2008)
28. Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F., Min, H., Deng, X.: Empirical analysis: stock market prediction via extreme learning machine. Neural Comput Appl **27**(1), 67–78. Neural Computing and Applications (2014). https://doi.org/10.1007/s00521-014-1550-z
29. Yoo, P.D., Kim, M.H., Jan, T.: machine learning techniques and use of event information for stock market prediction: a survey and evaluation. In: International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce. CIMCA-IAWTIC'06, Vienna, Austria (2005). https://doi.org/10.1109/cimca.2005.1631572
30. Huang, G.B, Ding, X., Zhou, H.: Optimization method based extreme learning machine for classification. Neurocomputing **74**(1–3), 155–163. Nanyang Technological University, Singapore & Xi'an Jiaotong University, China (2010). https://doi.org/10.1016/j.neucom.2010.02.019
31. Patel, J., Shah, S., Thakkar, P., Kotecha, K.: Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Syst. Appl. **42**(1), 259–268. Nirma University, India (2015). https://doi.org/10.1016/j.eswa.2014.07.040

32. Ramalho, G.L.B., Filho, P.P.R., De Medeiros, N.S.F., Cortez, P.C.: Lung disease detection using feature extraction and extreme learning machine. Revista Brasileira de Engenharia Biomedica **30**(3), 207–214. Federal Institute of Education, Science and Technology of Ceará & Federal University of Ceará, Brazil (2014). https://doi.org/10.1590/rbeb.2014.019
33. Kumar, R., Singh, M.P., Kumar, P., Singh, J.P.: Crop selection method to maximize crop yield rate using machine learning technique. In: 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials. ICSTM, Chennai, India (2015). https://doi.org/10.1109/icstm.2015.7225403
34. Reena Rose, R., Meena, K., Suruliandi, A.: An empirical evaluation of the local texture description framework-based modified local directional number pattern with various classifiers for face recognition. Braz. Arch. Biol. Technol. **59**(2), 1–17. Brazilian Archives of Biology and Technology, Brazil (2016). https://doi.org/10.1590/1678-4324-2016161057
35. Cao B., Xie Y., Yang C.: Online hybrid modeling method with application for predicting Bauxite production indicators. Revista Escola da Minas **68**(4), 417–426. Central South University, China (2015). https://doi.org/10.1590/0370-44672014680245
36. Finotti, R.P., Cury, A.A., de Barbosa, F.S.: An SHM approach using machine learning and statistical indicators extracted from raw dynamic measurements. Latin Am. J. Solids Struct. **16**(2), 1–17. Universidade Federal de Juiz de For a, Brazil (2019). https://doi.org/10.1590/1679-78254942
37. Okada, R., Costa, J.F.C.L., Rodrigues, Á.L., Kuckartz, B.T., Marques, D.M.: Scenario reduction using machine learning techniques applied to conditional geostatistical simulation. REM Int. Eng. J. **72**(1), 63–68. Universidade Federal do Rio Grande do Sul, Brazil (2019). https://doi.org/10.1590/0370-44672018720135
38. Ávalos, H., Gómez, E., Guzmán, D., Ordóñez-Camacho, D., Román, J., Taipe, O.: Where to park? Architecture and implementation of an empty parking lot, automatic recognition system. Enfoque UTE **10**(1), 54–64. Universidad UTE, Ecuador (2019). https://doi.org/10.29019/enfoqueute.v10n1.445
39. De-La-Hoz, E.J., De-La-Hoz, E.J., Fontalvo, T.J.: Metodología de Aprendizaje Automático para la Clasificación y Predicción de Usuarios en Ambientes Virtuales de Educación. In: Información Tecnológica, **30**(1), 247–254. Información Tecnológica (2019). https://doi.org/10.4067/S0718-07642019000100247
40. Souza, C.M.P. de, Thomazini, A., Schaefer, C.E.G.R., Veloso, G.V., Moreira, G.M., Fernandes Filho, E.I.: Multivariate analysis and machine learning in properties of Ultisols (Argissolos) of Brazilian amazon. In: Revista Brasileira de Ciência Do Solo, **42**, 1–20. Revista Brasileira de Ciência Do Solo, Brazil (2018). https://doi.org/10.1590/18069657rbcs20170419
41. Peixoto, E.C., Pinto, L.R.: Gerenciamento de estoques via previsão de vendas agregadas utilizando simulação. Production **16**(3), 569–581. Production, Brazil (2006). https://doi.org/10.1590/s0103-65132006000300016
42. León-Borges, J.A., Noh-Balam, R.-I., Rangel Gómez, L., Philip Strand, M.: The machine learning in the prediction of elections. In: Revista Electrónica de Computación, Informática Biomédica y Electrónica, **4**(2). Universidad de Guadalajara, Mexico (2015). https://www.redalyc.org/articulo.oa?id=512251502001
43. Haghighi, M.: Data mining and machine learning: an overview of classifiers. In: Ciência e Natura, **37**(2), 76–86. Payam Noor University, Sosangerd, Iran (2015). https://doi.org/10.5902/2179460x20756
44. Mancera Valetts, L.P., Baldiris Navarro, S.M., Betancur Chicué, V.: Indicators of ADHD symptoms in virtual learning context using machine learning technics. In: Revista Escuela de Administración de Negocios, **79**, 22–37. Universidad EAN, Colombia. https://www.redalyc.org/articulo.oa?id=20643042002 (2015)
45. Montero-Romero, T., López-Martin, M. del C., Becerra-Alonso, D., Martinez-Estudillo, F.J.: Extreme learning machine to analyze the level of default in Spanish deposit institutions. In: Revista de Métodos Cuantitativos para la Economía y la Empresa **13**, 3–23. Universidad Pablo de Olavide, Spain. https://www.redalyc.org/articulo.oa?id=233124421001 (2012)

46. Amazon se apunta al Machine Learning "low-cost". Cleverdata, Spain. Available at: https://cleverdata.io/amazon-machine-learning-low-cost/ (2015)
47. Microsoft open sources distributed machine learning toolkit for more efficient big data research. Microsoft Research Blog, United States of America. Available at: https://www.microsoft.com/en-us/research/blog/microsoft-open-sources-distributed-machine-learning-toolkit-for-more-efficient-big-data-research/ (2015)
48. Elon Musk's billion-dollar crusade to stop the A.I. apocalypse. Vanity Fair, United States of America. Available at: https://www.vanityfair.com/news/2017/03/elon-musk-billion-dollar-crusade-to-stop-ai-space-x (2017)
49. Google AI beats Go world champion again to complete historic 4–1 series victory. Tech Crunch, United States of America. Available at: https://techcrunch.com/2016/03/15/google-ai-beats-go-world-champion-again-to-complete-historic-4-1-series-victory/ (2016)
50. These AI bots created their own language to talk to each other. Vox, United States of America. Available at: https://www.vox.com/2017/3/23/14962182/ai-learning-language-open-ai-research (2017)
51. Facebook teaches bots how to negotiate. They learn to lie instead. Wired, United Kingdom. Available at: https://www.wired.co.uk/article/facebook-teaches-bots-how-to-negotiate-and-lie (2017)
52. For the first time, AI crushed a professional Dota 2 player at the game's biggest tournament. Quartz, United States of America. Available at: https://qz.com/1052409/openai-just-beat-a-professional-dota-2-player-at-the-international-2017/ (2017)
53. La ciencia sólo habla en inglés. Diario El Mundo, Spain. Available at: https://www.elmundo.es/ciencia/2015/09/17/55f9b8ef22601d962a8b45b2.html (2015)

Automatic Balancing System of Single-Phase Consumer Units Connected to the Low-Voltage Distribution Network

Alex Sander Leocádio Dias and Israel Gondres Torné

Abstract Charge imbalance is common events in electrical distribution systems, which are introduced by the irregular allocation of single-phase loads by the phases that make up the three-phase power distribution. The direct consequence of this imbalance is the drop in operating performance and the reduction in the useful life of network assets (transformers, cables, etc.), as well as affecting the quality of energy distributed from a product standpoint over distances from active sources of the distribution circuits, therefore voltage drops are present. The technique used to correct voltage drops is to redistribute the loads more evenly, but this activity is not an easy task and almost always the effectiveness of the way it is performed in most utilities is questionable, including incurring infringement claims. Quality indicators are inadequate supply voltage level at delivery points and customer connection. The model proposed in this work has the aid of the load-balancing task, using a fuzzy logic tool for the interpretation of the dynamic behavior of the loads connected in the system and the optimization of the best allocation of single-phase consumer units with an algorithm based on the simplex model. Both decisions making with fuzzy logic and distribution were simulated with data provided by the local concessionaire Eletrobrás Amazonas through the grid management system.

Keywords Computational intelligence · Fuzzy logic · Load balancing · Secondary distribution network

A. S. L. Dias · I. G. Torné (✉)
Higher School of Technology, State University of Amazonas, Manaus, Brazil
e-mail: igondrest@gmail.com

A. S. L. Dias
e-mail: alex.sander06@gmail.com

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_9

1 Introduction

In Brazil, the growth of electricity consumption has always been the subject of speculation, so much so that in the 2008 scenario of Brazil, whose economy was accelerating, there was some concern about the absorption capacity of energy demand by the industrial and commercial sector [1].

Considering the complexity of operation and maintenance, the distribution subsystem is the most complex of the three subsystems, since it is at this stage that the interface between the distributors (private, mixed or state utilities) and the consumer units (UC's) occurs [2]. In line with [3], the distribution can still be subdivided into primary and secondary distribution and this was largely impacted by the issuance of normative resolution—RN 223 of the National Electricity Agency—ANEEL, which according to its wording obliges the concessionaires to meet free of charge requests for new connections as long as the applicant's installed load does not exceed 50 kW. Because of this RN, there was a difficulty in answering new calls at the local concessionaire, that is, it overloaded the sector responsible for this service, given the significant growth of both regular and irregular calls (clandestine calls). All these factors negatively affect the power distribution of this concessionaire, leading to complaints and fines for breach of collective indications of both product and service quality. However, of all the problems of the distribution subsystem, the burning of the transformer asset is the most impacting for both the concessionaire and the consumers served by it. This work proposes a method to mitigate one of the main causes of the burning of this asset, which is the severe imbalance in its secondary, which is closely linked to regions with large non-technical losses, with the predominance of energy deviations and clandestine connections.

Concerning clandestine connections, the cause of which is due to social problems, and the unplanned and unplanned expansions of the city of Manaus can be cited, it occurs that such connections generate harmful impacts to distributed energy, since most of them are carried out criminally, without the use appropriate connections between network installations and the customer and without measurement approved by the concessionaire, in addition to non-compliance with the technical criteria provided for in the concessionaire standard. All these factors together and added to the advanced age of the networks, in addition to the contingency faced in the electrical system as a whole, require robust asset optimizations, aiming at preservation and continuity of supply, in addition to improving operational quality and performance indicators.

In order to optimize the physical resources in the face of the sudden demand of the loads (consumer units) connected to the secondary distribution network, the present work discusses a low-cost option, but using artificial intelligence resources and the use of electronic devices, power for optimal or suboptimal allocation of connected loads along the distribution network. Aiming at improving the concessionaire's performance against the indicators, increasing the useful life of the assets and postponing the reinforcement and expansion of the networks, in line with what is established in [4].

The balancing of UCs is one of the first procedures to be adopted as a prophylactic measure in problems of violation of energy quality indicators from the product point of view (poor or critical voltage), but its need is not so evident until they occur. Complaints by UCs are evidenced in the RN 398 sample set (PRODIST-ANEEL), and however, the method applied in the concession area of Eletrobrás Amazonas Distribuidora de Energia SA is still rudimentary, without the aid of continuous measurements, without the support of automation or with a history of load demand requests. Sometimes the process is done based on the knowledge acquired by the performing teams, and its result is not assertive and passive to question, besides imposing the need to expose people to electrical risk during system interventions, perform shutdowns of transformation units, which affects collective quality indicators, such as those that measure service continuity. The whole scenario described shows the lack of a more effective, intelligent action that optimizes the available resources. In this sense, this work proposes the adoption of a device capable of balancing loads (UCs) connected to the distribution network in a way guided by a simplex model algorithm, adopting artificial intelligence for the decision-making stage using fuzzy logic and without the presence of human intervention in the switching process, which is now performed by switching circuits as a final control element, using the power electronics.

2 State of the Art

2.1 Balancing of Consumer Units—UCs

The optimization of the electric power distribution networks is the subject of many works, almost all concentrated in the medium voltage networks, being scarce works with the low network, however, in [5] a completely combined programming model was presented, which addressed the improvement of load redistribution losses in distribution circuits.

The allocation of loads of more equitable manner throughout the three phases of the distribution system is configured as a combinational problem in this respect [6, 7] has an integer linear programming model with the use of solver open, supporting the design of distribution networks projects. Sidewalk in these works [8] presents a computational system in which it is possible to simulate the load distribution, resulting in the reduction of neutral current, very present in severe imbalances and preserving the integrity of conductors and transformers in distribution networks.

However, it was in [8, 9] that mathematical modeling was presented for various maintenance and optimization actions in distribution networks, focusing on load balancing in this work, as it served as the foundation for the development of this work.

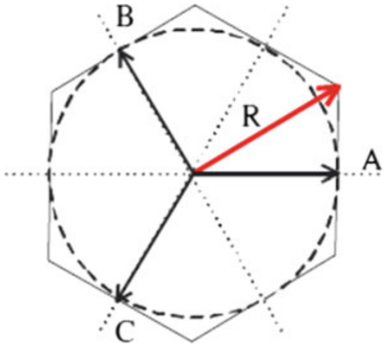


Fig. 1 Graphical representation of the approach constraints of a load balancing. *Source* [9]

As Fig. 1 shows the graphical representation of load balancing, with the polygon formed by six edges representing the approximations of balancing, i.e., the greater the number of edges, the closer the ideal balance would be, however, the fact that it increases the amount of edges that requires greater mathematical effort and when implemented in algorithm, results in greater computational effort.

With (1) summarizes the circuit balance determination.

$$\mathrm{bal} = \left(1 - \frac{R}{A + B + C}\right) \tag{1}$$

At where:

bal = Load balancing in the circuit;

A, B and C = Total loads in each phase;

R = Total circuit unbalance.

2.2 Fuzzy Logic

Proposed by L. A. Zadeh (1965), it is a promising branch of artificial intelligence that is based on the issue of pertinence, allowing it to work with vague or difficult to interpret information. The difference compared to Boolean logic (true and false) is that in fuzzy logic propositions can present degrees between the two extremes. The fuzzy logic is applied in control systems, system modeling, economical engineering, and in this work in the decision-making step in order to optimize the distribution of single-phase PAs, thus translating and seeking to minimize the uncertainties inherent to the dynamic behavior of loads connected to the distribution subsystem.

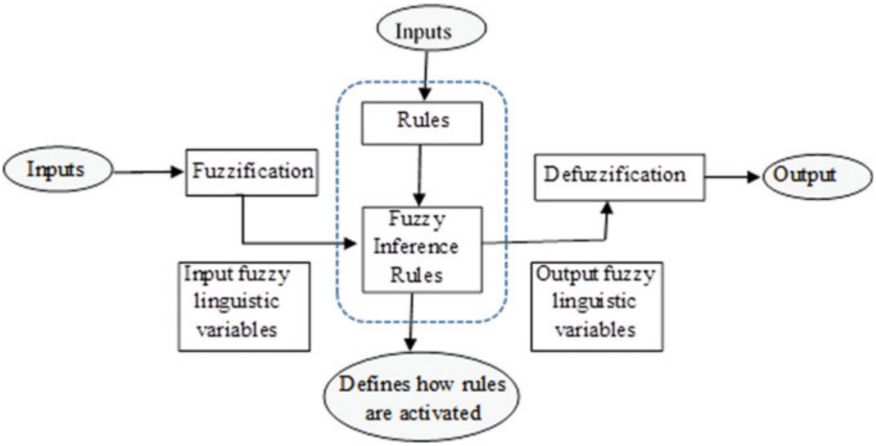


Fig. 2 Structure of the fuzzy logic controller. *Source* Adapted from [11]

2.3 Mamdani Model

The main idea of Mamdani's method is to describe process states through linguistic variables and use these variables as inputs to control rules; the rules connect input variables with output variables and are based on the description of the fuzzy state that is obtained by defining the linguistic variables. Each crisp input (either a real number or a real number of real numbers) is expected to match a crisp output and, in general, a fuzzy system matches each input with output. In this case, [10] states that a fuzzy system is a function of *Rn* in *R*, built by a specific methodology according to 3 modules. Figure 3 shows the general structure of a fuzzy logic-based controller (Fig. 2).

2.4 Mathematical Optimization

In general, concepts and optimization problems seek to maximize or minimize a real function by choosing input values from a specific set and evaluating the assumed value for a given function, which is, finding the best results for an objective function in a given domain. In [12], mathematical programming studies problems, properties, solution algorithm creation, and real-world applications.

Also, according to Boyd and Vanderberghe [12], an optimization is composed of variables, objective function, and constraints, and the objective function is the mathematical expression that one wishes to optimize, whereas the constraints are the limits for the variables such that the solution is feasible and the variables are the values for the objective function. (2) presented the generic form of problems associated with optimization.

$$\min_{sujeito\, f_i(x) \leq b_i;\, i=1,\ldots,n} f_0 \quad (\boldsymbol{x}) \tag{2}$$

At where:

Min f_0 (x) = objective function;

B_i = Restrictions.

2.5 Simplex Method

Still, according to [12], the simplex method walks in the vertex region until obtaining a solution that does not have better than it in the neighborhood, and therefore, this solution found is the optimal one. However, there are scenarios where the optimal solution cannot be reached, in which case there is no viable optimal solution to the problem, because of incompatible constraints or the variables tend to infinity, causing the objective function to find no limits. Simplex is commonly used in optimal linear-PL programming solutions when it meets the following criteria:

1. All variables are no negative;
2. All bi is nonnegative;
3. All initial equations of the systems are of the same or smaller type (≤), such that only the gap variables are found.

It happens that sometimes some of the criteria cannot be met, and for that, the technique of the two-phase simplex is employed.

2.6 Usual Allocation of UCs in Distribution Circuits

According to the procedures in [13], low-voltage consumer units, when connected to bare cable distribution networks, are usually anchored close to the supporting structures of the networks near the posts and due to the incompatibility of the metal alloys that make up the network (usually aluminum) and service branches (usually copper insulated cables), wedge-type or parallel-type connectors compatible with the gauges in the circuit sections are employed, thus preventing oxidation at the connection points. Figure 3 generally demonstrates the allocation of consumer units along with the n nodes that make up the low-voltage distribution network.

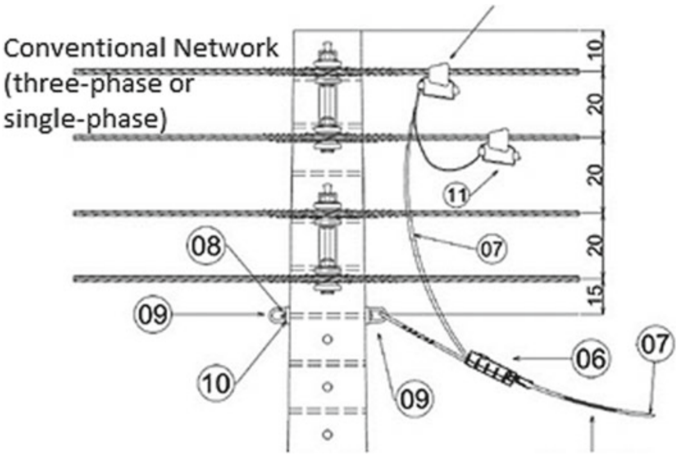


Fig. 3 UC extension connected to the distribution network. *Source* [13]

3 Materials and Methods

3.1 Data Collection and Work Design

The survey of the data applied in this work was carried out with the local concessionaire, Eletrobrás Amazonas Distribuidora de Energia S.A., directly from its SGDREDE database, whose purpose is the remote monitoring of transformation units scattered in various neighborhoods of Manaus city.

The adoption of the limits for the linguistic variables of the fuzzy sets employed in the decision making of the balanced system proposed in this paper obeyed the expert knowledge and experience of the concessionaire's design body, but was also based on the two-year histograms 2017 and 2016, aiming at a better approach to the objective of this theme.

The methodology applied in this proposal of the automatic balancing system of UCs was carried out in first-moment searches in collections of similar themes, works published in journals and periodicals, in order to better base the foundation of the theme. In a second stage, the hardware schematic was developed, using the CAD PROTEUS 8 Tool. The development of the control system, i.e., the third stage of the work, was performed in a computer environment simulation, the MATLAB R2014a®, which the Fuzzy Logic Toolbox allows the creation of Mamdani models, as well as an open solver available in the Excel 2010 tool for validation of the simplex method algorithm model.

3.2 Fuzzy Block Determination for Decision Making

Figure 4 shows the Mamdani model proposed in this paper, whose purpose is to assist the daily control strategy to be performed by the proposed system in the interpretation of the inputs, which are presented as the demand requests of the three system phases and the time of day, so depending on the interpretation by the model, the UC's balance enable block can be activated.

Mamdani model variables. The construction of the proposed Mamdani method uses four input impacts: Phase A Demand; Phase B Demand; Phase C Demand; and Schedules, for output only one output: Balance Ckt. The division of impacts into language levels was defined according to Table 1.

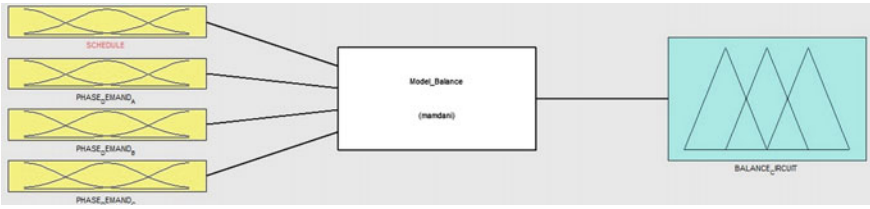


Fig. 4 Proposed Mamdani model

Table 1 Description of linguistic variables and result

Linguistic variables		Fuzzy set
Type	Description	
INPUT	Schedule	Light (HL)
		Moderate (HM)
		Critical (HC)
	Phase A Demand	Light (DFLA)
		Moderate (DFMA)
		Critical (DFCA)
	Phase B Demand	Light (DFLB)
		Moderate (DFMB)
		Critical (DFCB)
	Phase C Demand	Light (DFLC)
		Moderate (DFMC)
		Critical (DFCC)
OUTPUT	Balance circuit	*ON* (ECL)
		OFF (ECD)

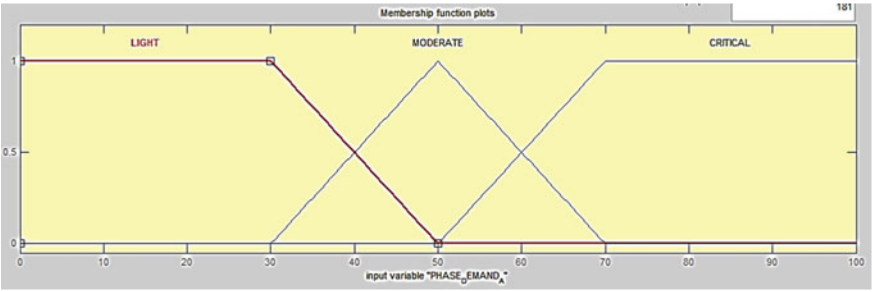


Fig. 5 Variable Phase A Demand

Input variables. Demand from the connected consumer units in the transformation unit phases varies throughout the day, but the winding of each transformation unit can only provide a power that is compatible with its transformation capacity and when overdemand occurs, the entire system is put at risk of overload shutdown or untimely actuation of the protections. Figure 6 shows the parameterization in MATLAB's Fuzzy Logic Toolbox for the input variable called Phase A Demand, and the same is true for the other phases that make up the system (Fig. 5).

The circuit consumption profile in home feeders shows that in the early hours of the day, many loads are less likely to be connected and requesting large power blocks, but at system peak time (most critical time defined by RN between 18:30 and 21:30), any sudden increase may incur a severe risk to customer service. The parameterization of the variable named time is shown in Fig. 6.

Rule base. The determination of the linguistic rules, which totaled for this application 81, followed the expert advice of the concessionaire that was the target of the study, based on the occurrences recorded in the Administration and Management System—SIAGE and SGDREDE, in addition to observing the cost ratio unit of operation and quality of supply. Figure 7 demonstrates a part of the 81 rules that make up the model.

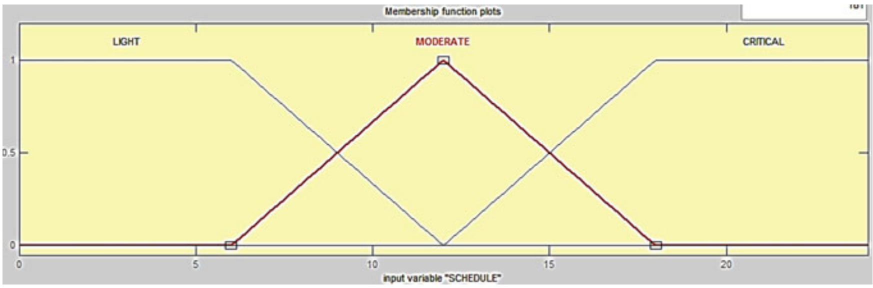


Fig. 6 Time variable

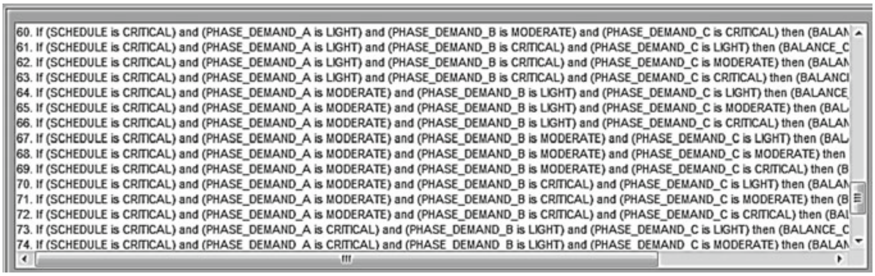


Fig. 7 Rules of inference of linguistic variables

3.3 Reconfiguration of Simplex-Based Solver Assist UCs

In order to perform the distribution of the UCs by the three phases of the secondary low-voltage distribution system—SDBT in order to obtain the balance in load allocation (best distribution or optimal distribution), the present work adopts the distribution method based on simplex, which is briefly based on offerors and plaintiffs, subject to restrictions inherent in the process.

(3) and (4) present the restrictions imposed on the algorithm in the search for the best load balancing.

$$A + \frac{2 \times \text{Demtotal} \times \text{bal}}{3} \leq \text{Dem total} \tag{3}$$

$$\frac{2 \times \text{Demtotal} \times \text{bal}}{3} - A \leq \frac{\text{Dem total}}{3} \tag{4}$$

At where:

Dem total = total circuit demand;

A = phase demand any phase.

The offerors would be the three phases of the transformation unit with their maximum and finite capacity to supply power and the demanders would be all the loads connected to the secondary of the transformation unit.

3.4 UC's Switching Interface

The switching process of UCs is performed by electronic keys, in which case the SCR thyristors are used in the anti-parallel configuration, as shown in the schematic of Fig. 8. To avoid switching by electrical noise and out of time determined by the control block, the main thyristors are protected with a snubber circuit-type relief filter.

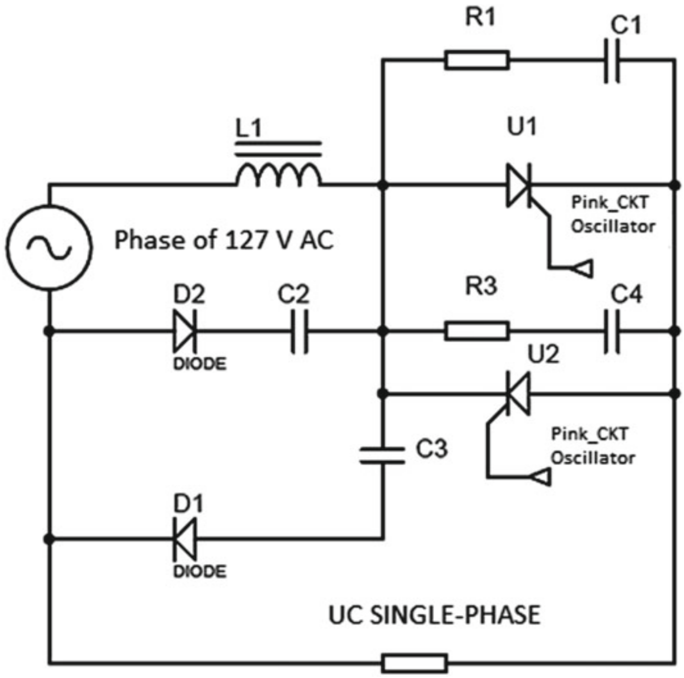


Fig. 8 Simplified schematic of the consumer unit switching circuit

4 Results and Discussion

To simulate the system proposed in this work, a circuit selection was performed based on the measurements indicated by the SGDREDE, and the target circuit of this study presented high unbalance and served a zone with a high rate of technical and non-technical loss. Figure 9 shows the sketch of the circuit, whose data were simulated in this work.

Table 2 presents the circuit characteristics in the elaboration stage of this work.

Figure 10 shows the flow diagram of the proposal's internal system operations, where all the sub-blocks that make up the system as a whole are arranged.

Based on the information in Table 2, the electronic input sensing circuits of the balancing system in Fig. 10 send data to the microcontroller input, sensitizing it to initiate an analysis of the demand per phase and the amount of UCs per phase. circuit, based on the decision making of the fuzzy block, obeying the fuzzy logic and parameterization of Table 1 of the fuzzy sets, the fuzzification of the inputs occurs, provided that the set of rules that best expresses the interaction is observed. The output of the fuzzy block enables the optimization block, which with a simplex model-based algorithm seeks load redistribution and the microcontroller enables the switching block of consumer units connected to the network.

Table 3 shows the result of the simulation of the single-phase PA redistribution

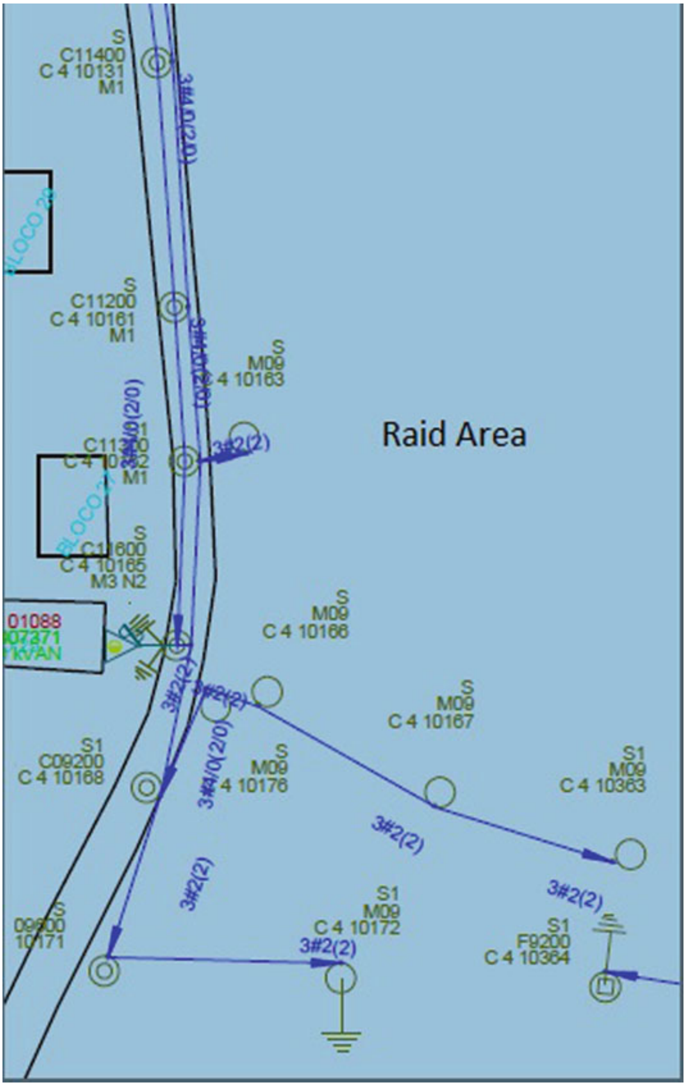


Fig. 9 Sketch of circuit CC-03-0188. *Source* SGDREDE, 2018

by the three phases of the distribution circuit, that is, the distribution performed by the solver (optimization block) of the system proposed in this work phase capacity constraints and maximize load balancing.

Figure 11 shows the behavior of the fuzzy decision-making block, which responses by enabling the optimization block in the face of demand greater than 80% in Phase A and time greater than 18 h, therefore at the tip.

Table 2 CC030188 circuit distribution and circuit technical data

Code	Nº Alm	Pot. Transf	SE	No. of UCs	1φ	2φ	3φ
0188	3	150	CC	71	34	23	14
Phase	Current (A)	Voltage F-N	Voltage F-F	Qty UC1φ	Qty UC by Potência (kVA)		
					0.5	1.5	3.0
A	409.4	140	230	16	14	12	8
B	472	139.8	228	10			
C	557.8	139.5	227	8			

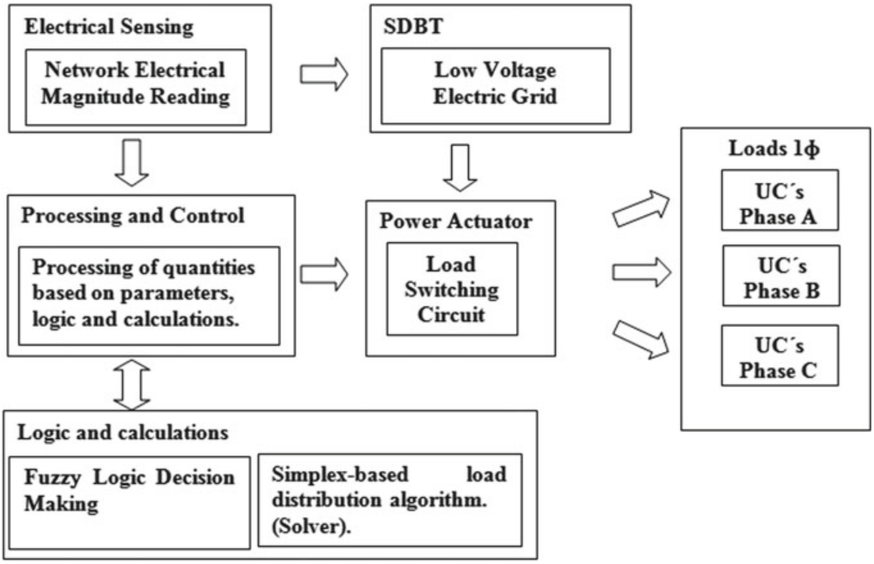


Fig. 10 Simplified diagram of action flow in the balancing system

Table 3 Distribution of UCs for the CC030188

Load distribution			
UC	System phases		
	A	B	C
TYPE 1	6	4	4
TYPE 2	4	4	4
TYPE 3	2	3	3

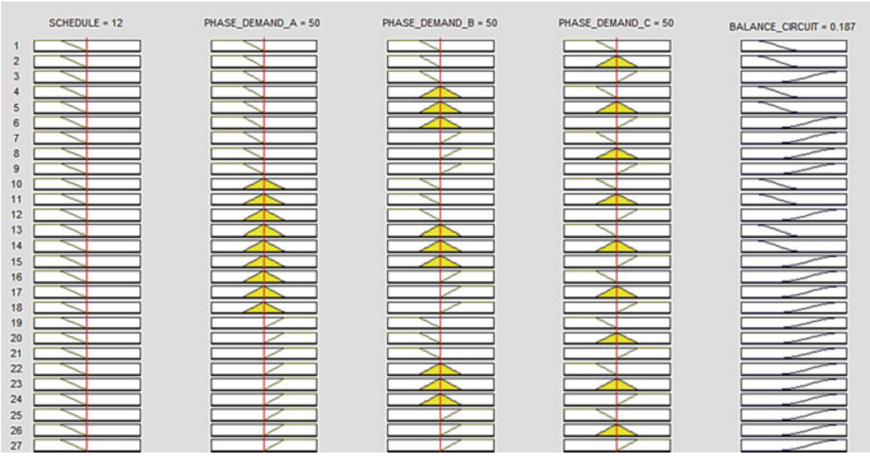


Fig. 11 Fuzzy block response graph

5 Conclusion

Using fuzzy sets artificial intelligence methods for better decision making in the face of increased energy demand in a power distribution system and using a combined optimization algorithm inspired by the simplex model, with respect to each phase of the transformation unit, open the possibility of more effective reconfiguration from the point of view of reducing the imbalance in three-phase distribution systems.

The use of embedded and microcontroller systems, allied with power switches (SCR) in the anti-parallel configuration and playing the role of actuators in the handling of the UCs, proves to be a viable and economically more favorable option than the traditional team shift method.

The simulated results showed that, in general, the balancing model proposed here made it possible, from the information collected from the network, to determine the necessary redistribution to balance the loads and at the same time to reduce the need for interventions to improve the distribution of loads energy, i.e., reducing the cost of commuting, optimizing the resources of available assets, improving the performance of imbalance losses, and stress levels associated with severe imbalances.

References

1. Rocha, O., Dos Anjos, J.P., De Andrade, J.B.: Energy trends and the water-energy binomium for Brazil. Anais da Academia Brasileira de Ciências **87**(2), 569–594 (2015)
2. Creder, H.: Instalações Elétricas, 16th edn. LTC, Rio de Janeiro (2016)
3. ANEEL: Resolução 223, Condições gerais para elaboração dos planos de universalização de energia elétrica. 2003. https://www.aneel.gov.br. Last Accessed 2017/10/27

4. Rajalakshmi, K., Kumar, K.S., Venkatesh, S., Belwin, J.: Reconfiguration of distribution system for loss reduction using improved harmony search algorithm. In: 2017 International Conference on High Voltage Engineering and Power Systems (ICHVEPS), on Proceedings IEEE, pp. 377–378. Sanur (2017)
5. Hasanpour, R., Kalesar, B.M., Noshahr, J.B., Farhadi, P.: Reconfiguration of smart distribution network considering variation of load and local renewable generation. In: 2017 IEEE International Conference on Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), on Proceedings IEEE, pp. 1–5. Milan (2017)
6. Radwan, A.A., Foda, M.O., Elsayed, A.M., Mohamed, Y.S.: Modeling and reconfiguration of middle Egypt distribution network. In: 2017 Nineteenth International Middle East Power Systems Conference (MEPCON), on Proceedings IEEE, pp. 1258–1264. Cairo (2017)
7. Oliveira, R.: Ferramenta de otimização para projetos de redes de distribuição secundárias de energia elétrica. Dissertação de mestrado. Universidade Tecnológica do Paraná, Curitiba (2008)
8. Yin, N.N., Thuzar, M., Thwe, E.P.: Analysis of loss reconfiguration for distribution network system. In: 21st International Computer Science and Engineering Conference (ICSEC), on Proceedings IEEE, pp. 1–5. Bangkok (2017)
9. Bettoni, L.M.: Modelos PLIM para otimização de circuito da rede de distribuição secundária de energia elétrica. Dissertação de Mestrado – Universidade Tecnológica Federal do Paraná. Curitiba (2011)
10. Muñoz, M., Miranda, E.: A fuzzy system for estimating premium cost of option exchange using Mamdani Inference: Derivates Market of Mexico. In: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), on Proceedings IEEE, pp. 888–895. Vancouver, (2016)
11. Nogueira, E.L., Nascimento, M.H.R.: Inventory control applying sales demand prevision based on fuzzy inference system. J. Eng. Technol. Ind. Appl. **03** (2017)
12. Boyd, S., Vanderberghe, L.: Convex optimization, 7ª edn. Cambridge University Press (2009)
13. ELETROBRAS Amazonas Distribuidora de Energia S.A. Norma de Distribuição de Energia Elétrica – NDEE 2. www.eletrobrasamazonas.com/normastecnicas.html. Last accessed 2017/10/27

Model Predictive Controllers: Development of a Tuning Methodology Using Sensitivity Functions

André Luiz P. de Lima and Luz Adriana Alvarez

Abstract In this work, a practical tuning methodology for model predictive controllers (MPC) is introduced, in which the control parameters are determined using sensitivity functions. The controller applied in this project was the generalized predictive controller (GPC), which possesses a considerable presence in the industry, relative to other types. The closed-loop system with the GPC controller can be represented through sensitivity functions. Here, we define indexes based on those sensitivity functions that measure the impact of each tuning parameter on the closed-loop performance. The sensitivity functions can be associated with a specific control objective, such as error minimization, disturbance rejection and reduction of control effort. Then, through an optimization problem, we propose to find the control parameters that let the processing system to reach the desired control objectives. The importance of each objective can be defined through a numerical weight. The proposed approach is tested through simulation in a typical chemical engineering system: the level control of a tank reactor.

Keywords Model predictive control · Tuning · Sensitivity functions

1 Introduction

Advanced controllers have been producing a positive impact in improving industrial processes operation, and inside this group, it is worth mentioning model predictive controllers (MPC). They are able to promote significant process changes, thanks to the capability to work with multivariable problems close to the constraints. Besides,

A. L. P. de Lima (✉) · L. A. Alvarez
Department of Chemical Systems Engineering, State University of Campinas, Campinas, Brazil
e-mail: a134946@dac.unicamp.br

L. A. Alvarez
e-mail: luzadri@unicamp.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_10

MPC's are intuitive, a process model allows the output error estimation based on internally calculated inputs, and an optimizer that minimizes the output error by determining the best input sequence. However, despite its advantages, MPC still lacks a well-determined tuning methodology [1].

When discussing the tuning of an MPC, it should be considered that there are many different parameters to be adjusted that can overlap the effects of each other, creating an issue when dealing with this problem [2]. Besides that there are many different types of model predictive controllers, and each one presents a specific set of parameters to be tuned, which will affect the optimization problem to be solved [3]. Many articles published in the past few years have been focusing on the universalization of tuning methods, in order to improve robustness and standardization [4–6], showing how this topic is still relevant even in recent times.

Among the many parameters that require adjustments in an MPC, we may highlight the weight matrices contained in its cost function, which can be classified by their characteristics: while one is responsible for minimizing the output error (Q), the other one reduces the input effort in controlling the system (R). These parameters are usually tuned by trial-and-error, damaging its general attractiveness.

In this work, a novel methodology for tuning the Q and R parameters is proposed. The approach is based on the closed-loop sensitivity functions, which are treated as indexes in a global equation that represents the cost function of an optimization problem. Between the many types of controllers that could have been used for this purpose, it has been chosen to apply the generalized predictive controller (GPC) because of its presence both in the industrial and scientific environments [7]. There are three different objectives that are contained in the tuning function: the output error minimization, the reduction of control effort and the disturbance rejection. The method was applied to a single-input, single-output (SISO) problem.

2 Sensitivity Functions

In the MPC tuning field, the concept of sensitivity function is well known, with other works applying them in order to obtain adequate methods [8, 9]. Sensitivity functions can be defined as expressions that represent the relation between system input data (such as disturbance, reference and noise) and variables of interest for the system (such as inputs and outputs). In order to determine the closed-loop influence of the disturbance on the output, it is quite frequent the use of this relation, which is desired to be as small as possible [1].

As an example, let us consider Eq. 1, which details the closed-loop dynamics of the system in the presence of disturbance. Here, y represents the system output, while d and w represent the disturbance and the reference, respectively. This allows us to define the term that relates the output to the input data (in this case, d and w). We may analyze the physical meaning of Eqs. 2 and 3 and specify rules about the desired closed-loop behavior. Equation 3, for example, represents the effect of a reference change on the output. For such variables, the ideal situation is to obtain output values

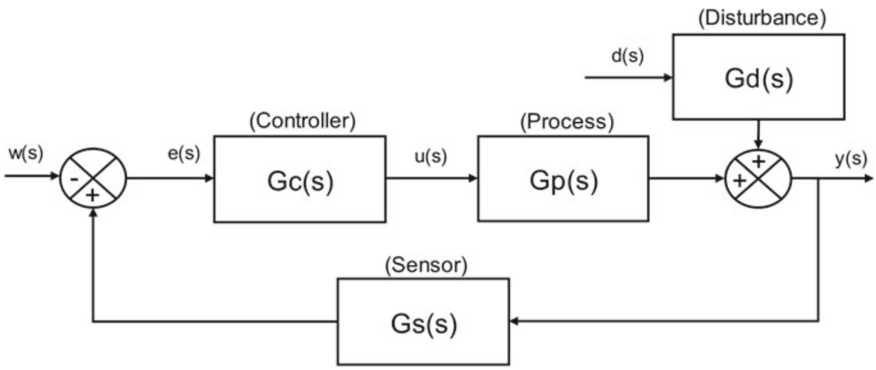


Fig. 1 Block diagram of a closed-loop system

as close as possible to reference values, which means $E(s)$ close to one. Equation 2, on the other hand, can be interpreted as the influence that disturbance has on the output, and it is usual to dictate that this should be null.

$$y(s) = D(s) * d(s) + E(s) * w(s) \tag{1}$$

$$D(s) = \frac{y(s)}{d(s)} \tag{2}$$

$$E(s) = \frac{y(s)}{w(s)} \tag{3}$$

We may apply this concept to a closed-loop system containing a controller, process and sensor, as shown in Fig. 1. For this system, we may define a set of sensitivity functions, they are summarized in Table 1 with a brief description of their physical meaning.

3 Tuning Methodology Using Sensitivity Functions

As mentioned before, the methodology is focused on the tuning of the weighting matrices Q and R. Here, we propose to minimize a function FT according to Eq. 4. In this, I represents the set of indexes considered in the proposed method, which depends on the weighting parameters (i.e., R and Q).

$$\min_{R,Q} FT(I(R, Q)) \tag{4}$$

Three different indexes related to different objectives:

Table 1 Sensitivity functions in a closed-loop system

Sensitivity functions	Description
$\frac{y(s)}{d(s)} = \frac{G_d}{G_c G_p G_s + 1}$	The relation between system output and disturbance. Represents the influence of the disturbance on the output. It is usually desired for its value to be near zero
$\frac{y(s)}{w(s)} = \frac{G_c G_p}{G_c G_p G_s + 1}$	The relation between system output and reference. Represents the proximity between those variables. It is desired for them to be equal (value of 1)
$\frac{u(s)}{d(s)} = \frac{G_d}{G_p (G_c G_p G_s + 1)}$	The relation between system input and disturbance. Represents the disturbance effect on the input. It is usually desired for its value to be small to avoid an excessive control effort
$\frac{\Delta u(s)}{w(s)} = \frac{\Delta G_c}{G_c G_p G_s + 1}$	The relation between the control effort and reference. Represents the effect of a reference change on the control effort. It is usually desired for its value to be near zero

- **Minimization of error (I_{St}).** This index aims the approximation between the system output and reference.
- **Reduction of effort (I_{dMt}).** This index will focus on reducing the control effort.
- **Rejection of disturbance (I_{S0}).** It aims to minimize the effect promoted by the disturbance in the system.

The methodology consists of the minimization of the difference between each index and its desired value I^*, as described in Eq. 5.

$$FT = w_{\mathrm{ST}} * \left(I_{\mathrm{St}} - I^*_{\mathrm{St}}\right)^2 + w_{dMT} * \left(I_{\mathrm{dMt}} - I^*_{\mathrm{dMt}}\right)^2 + w_{\mathrm{S0}} * \left(I_{\mathrm{S0}} - I^*_{S0}\right)^2 \quad (5)$$

Here, each difference is squared and weighted. Note that in this formulation, the optimization problem of the MPC tuning procedure is more intuitive than the MPC problem itself, since each weight is related to a control objective. Additionally, the system is previously scaled in order to ease the choice of the weighting values w_{ST}, w_{dMT} and w_{S0}. The proposed indexes and their respective ideal value $I*$ are presented in Table 2.

The indexes are obtained through their frequency response to the disturbance. Here, we consider the frequency range where the system can operate. The frequencies may have values between 0 and f_{crit}, which is the maximum frequency to operate the system. Frequency values above the maximum frequency promote the "aliasing"

Table 2 Typical sensitivity functions in a closed-loop system

Type	Symbol	Indexes (I)	Ideal (I*)
Error minimization	I_{St}	$\frac{y(s)}{w(s)} = \frac{G_c G_p}{G_c G_p G_s + 1}$	1
Effort reduction	I_{dMt}	$\frac{\Delta u(s)}{w(s)} = \frac{\Delta G_c}{G_c G_p G_s + 1}$	0
Disturbance rejection	I_{S0}	$\frac{y(s)}{d(s)} = \frac{G_d}{G_c G_p G_s + 1}$	0

phenomenon, in which it is impossible for the system to distinguish the oscillatory signal due to the limitation in the sampling time T_S. Equation 6 describes the value of f_{crit}.

$$f_{\text{crit}} = \frac{\pi}{T_S} \tag{6}$$

Each index $\mathbf{I}_x$ is calculated at the system operating range. It considers the integral of the amplitude ratio of the frequency response of each sensitivity function, as described by Eq. 7. For the case where the disturbance is known, the amplitude ratio can be multiplied by a Gaussian function [represented through function $g(\cdot)$] with a center located in the disturbance frequency, as described in Eq. 8. Both approaches will be compared in order to check their performance in tuning a SISO system with known disturbance.

$$\mathbf{I}_x = \frac{\int_0^{f_{\text{crit}}} |I_x(j\omega)| \mathrm{d}\omega}{\int_0^{f_{\text{crit}}} \mathrm{d}\omega} = \frac{\int_0^{f_{\text{crit}}} |I_x(j\omega)| \mathrm{d}\omega}{f_{\text{crit}}} \tag{7}$$

$$\mathbf{I}_x = \frac{\int_0^{f_{\text{crit}}} g(\omega) |I_x(j\omega)| \mathrm{d}\omega}{\int_0^{f_{\text{crit}}} \mathrm{d}\omega} = \frac{\int_0^{f_{\text{crit}}} g(\omega) |I_x(j\omega)| \mathrm{d}\omega}{f_{\text{crit}}} \tag{8}$$

4 Process Simulation

The methodology is applied in a typical example in the chemical engineering field. Consider the level control inside a tank reactor, in which the input stream is fixed, and the output stream is located at the bottom of the tank. The reactor level is controlled via the use of a valve in the output stream, then the valve opening is considered as the manipulated variable. The process model is presented in Eq. 9. Here, L is the liquid level (in m), F_o is the input flow rate (in m^3/min), A is the transversal area of the tank (in m^2), g is gravity (in m/min^2) and C_V is the valve opening (in %).

$$\frac{\mathrm{d}L}{\mathrm{d}t} = \frac{1}{A} F_o - \frac{C_v \sqrt{g\mathrm{L}}}{A} \tag{9}$$

Process Operation Scenario: During this simulation, we have that F_o has a value of 0.06 m^3/min, the diameter of the tank is 0.5 m and the nominal value for C_V equals to 0.0002 m^2. The maximum level capacity of the tank is 1.4 m.

Process Transfer Function: The model is simulated in order to obtain a transfer function that represents this one. The variables have different ranges: The manipulated variable operates between 0 and 100%, while the tank level is between 0 and 1.4 m. The obtained transfer function was scaled to improve the numerical

Table 3 Simulation scenario for the tank system

Reference parameters change with time					
Time (min)	0	24	136	200	400
Reference (m)	1.1	0.9	1.1	1.1	1.1
d_{SP}	0	0	0	15	−10
Reference and disturbance parameters	$\omega = 0.05$ (min^{-1}); $t_{est} = 24$ min; $t_{sim} = 504$ min;				
Controller parameters	$H_C = 15$; $H_P = 10$; $T_S = 8$ min; $\tau_{cur} = 8$ min;				

performance, it is presented in Eq. 10, considering a time scale in minutes.

$$G_P(s) = \frac{-2.586}{8.69s + 1}[\%/\%] \tag{10}$$

Reference Curve: Here, we defined a reference curve that behaves exponentially, like a first-order function. This concept has been applied in other works and has shown an increase in efficiency and process performance [10]. In this simulation, the time constant for the reference curve (τ_{cur}) equals to 8 min.

Disturbance: In Eq. 11, the transfer function of the disturbance is presented. The Eq. 12 describes the disturbance behavior over time, showing the presence of oscillation. The variable d_{SP} is a scalar value.

$$G_d(s) = \frac{1.5}{10s + 1}[\%/\%] \tag{11}$$

$$\mathrm{d}(t) = d_{SP} + 10 * \sin(\omega t) \tag{12}$$

Simulation Scenario: Table 3 describes the simulation parameters along the time. We have in this case that t_{est} and t_{sim} are, respectively, the reference stabilization time (approximated time for the curve to reach 95% of its set point) and the total simulation time, while H_C and H_P are the GPC control and prediction horizon, respectively.

5 Results and Discussion

5.1 Error Minimization (I_{St})

To analyze the first index, we performed a tuning procedure with $w_{St}= 1$ and $w_{dMt}= w_{S0}= 0$ for both simulations utilizing the different frequency response calculations methods (Eqs. 7 and 8). The results are presented in Table 4, Here, the methods are compared in terms of integrated square error (*ISE*). We observed that the method of applying Gaussian produced an error-index approximately 20% smaller than the method proposed in Eq. 7. The simulation of the closed-loop system with the tuning results obtained for the second method is shown in Fig. 2.

Table 4 Results for simulation with w_{St}=1 and w_{dMt}= $w_{S0} = 0$

Parameters	Method without Gaussian	Method with Gaussian
R	0.1835	0.1528
Q	72.0026	68.2812
ISE	0.0084	0.0065

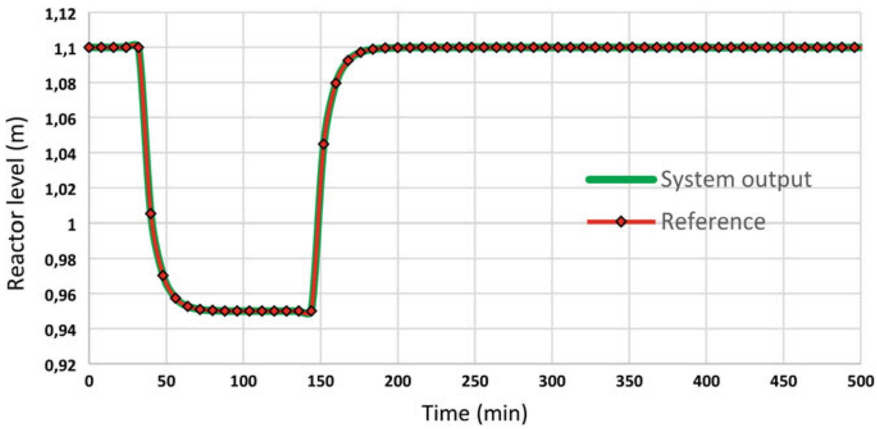


Fig. 2 System response after tuning with $w_{St} = 1$ and $w_{dMt} = w_{S0} = 0$ (Gaussian method)

5.2 *Reduction of Effort*

In order to quantify the reduction of control effort, we propose to apply an index similar to *ISE* for the input variation, the integrated square input variation (*ISV*). The tuning method was applied to the closed-loop system with different weights. Table 5 summarizes the results of the simulations. Notice that the increase in the weight resulted in a decrease of the *ISV*, consequently, the control effort. Specially, the reduction of the effort seems to intensify since both w_{St} and w_{dMt} have similar values, reinforcing the intuitiveness of the method.

Table 5 Results for simulations changing the weight for reduction of effort

Simulation	*R*	*Q*	ISE	ISV
$[w_{ST}, w_{S0}, w_{dMt}] = [1; 0; 0]$	0.1528	68.2812	0.0065	2766.9141
$[w_{ST}, w_{S0}, w_{dMt}] = [1; 0; 0.5]$	0.0856	18.1585	0.0288	2761.5660
$[w_{ST}, w_{S0}, w_{dMt}] = [1; 0; 1]$	244.5960	366.0569	208.9728	2146.1654
$[w_{ST}, w_{S0}, w_{dMt}] = [1; 0; 3]$	199.9999	100.0010	865.1708	1776.2346

Table 6 Results for simulation with different weights for rejection of disturbance

Simulation	R	Q	ISE
$[w_{\mathrm{ST}}, w_{\mathrm{S}0}, w_{\mathrm{dMt}}] = [1; 0; 1]$	244.5960	366.0569	208.9728
$[w_{\mathrm{ST}}, w_{\mathrm{S}0}, w_{\mathrm{dMt}}] = [1; 1; 1]$	50.5989	136.0242	89.2377

5.3 Rejection of Disturbance

In this case, a new scenario was calculated, where the tuning weights are equal. Table 6 shows that in comparison with the simulation without the action of the disturbance rejection, the ISE is smaller.

6 Conclusions

In this work, a methodology for tuning the weighting factors of an MPC controller was proposed, focusing on three different actions: minimization of output error, reduction of control effort and rejection of disturbance. The method was simulated in the level control of a tank, where we evaluated the effectiveness of the indexes related to each objective. The first index minimized the output error, resulting in a response behavior close to the reference. The results for the control effort were measured through the ISV. The controller performance was improved using a Gaussian function. It is important to highlight that the method is intuitive, practical and adaptable since it is simple to define the importance of each objective. The results showed that the desired behavior was achieved, resulting in a simple method for MPC tuning. In future work, this method will be extended to multivariable systems.

References

1. Maciejowski, J.: Predictive Control: With Constraints, 1st edn. Pearson Education Limited, Harlow (2002)
2. Francisco, M., Álvarez, H., Vega, P., Revollar, S.: Stable model predictive control tuning considering asymmetric bounded signals. Comput. Aided Chem. Eng. **28**, 493–498 (2010). https://doi.org/10.1016/S1570-7946(10)28083-5
3. Holkar, K.S., Waghmare, L.M.: An overview of model predictive control. Int. J. Control. Autom. **3**, 47–63 (2010)
4. Shadmand, M.B., Jain, S., Balog, R.S.: Autotuning technique for the cost function weight factors in model predictive control for power electronic interfaces. IEEE J. Emerg. Sel. Top Power Electr. **7**, 1408–1420 (2018). https://doi.org/10.1109/JESTPE.2018.2849738
5. Santos, J.E.W., Trierweiler, J.O., Farenzena, M.: Robust tuning for classical MPC through the multi-scenarios approach. Ind. Eng. Chem. Res. **58**, 3146–3158 (2019). https://doi.org/10.1021/acs.iecr.8b05485

6. Luzi, M., Vaccarini, M., Lemma, M.: A tuning methodology of model predictive control design for energy efficient building thermal control. J. Build. Eng. **21**, 28–36 (2019). https://doi.org/10.1016/j.jobe.2018.09.022
7. Garriga, J.L., Soroush, M.: Model predictive control tuning methods: a review. Ind. Eng. Chem. Res. **49**, 3505–3515 (2010). https://doi.org/10.1021/ie900323c
8. Hosseinnia, S.H., Lundh, M.: A general robust MPC design for the state-space model: application to paper machine process. Asian J. Control **18**, 1891–1907 (2016). https://doi.org/10.1002/asjc.1252
9. Francisco, M., Vega, P., Álvarez, H.: Robust integrated design of processes with terminal penalty model predictive controllers. Chem. Eng. Res. Des. **89**, 1011–1024 (2011). https://doi.org/10.1016/j.cherd.2010.11.023
10. Yamashita, A.S., Alexandre, P.M., Zanin, A.C., Odloak, D.: Reference trajectory tuning of model predictive control. Control Eng. Pract. **50**, 1–11 (2016). https://doi.org/10.1016/j.conengprac.2016.02.003

Development of an Intelligent Parking Aid System

Lucas Amaral Costa, Alessandro Corrêa Victorino, and Pedro Antônio Diniz Chaves

Abstract The increase in the number of cars in urban areas is causing an urban problem regarding the occupation of large areas for parking space. One possible solution for this problem is within the vertical parking. Thus, in a previous paper published by this research group, an innovative device that parks two vehicles in the same space required by only one was created. The patent request for this device has the registration number BR1020140118142. In the present discussion, technologies that communicate with the driver and help him park easily and safely have been incorporated into this device. This technology implanted in the device is composed of a position sensor and two cameras. In the first step, the proper position sensor and cameras for the problem were selected. On a second step, the optimal positions for the chosen laser sensor and cameras were defined. On a third and last step, the maneuvering of the vehicles during the parking task was simulated in the software Blensor®. The obtained results were effectively used to determine the vehicle's position precisely, assisting the driver in the task of parking and avoiding collisions.

Keywords Algorithm · Laser sensor · Parking aid system · Simulation · Vertical parking

L. A. Costa (✉) · A. C. Victorino · P. A. D. Chaves
Federal University of Minas Gerais, Belo Horizonte, MG 31270-901, Brazil
e-mail: llucasac@ufmg.br

P. A. D. Chaves
e-mail: pedrodinizch@gmail.com

A. C. Victorino
Universitè de Technologie de Compiègne, Rue Roger Coutolenc 60200, Compiègne, France
e-mail: alessvict@gmail.com

© Springer Nature Switzerland AG 2021
Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_11

1 Introduction

The Engineering School of the University of São Paulo presented the study [1] showing that about one quarter (25%) of the built area of São Paulo capital is occupied by parking space, based on historical records evaluated from 1930 until 2010. The parking problem [2] requires a solution, once this parking space could be much more efficiently used, as for the housing area. A possible solution for the mentioned problem is verticalizing the parking lot, efficiently increasing the number of parked cars in a reduced area. Although this idea already exists, a vertical parking device was proposed in a previous discussion [3, 4] as an efficient innovation [5] for the suchlike devices currently on the market. The proposed device in Fig. 1 has a combination of simplicity in the assembly and state-of-the-art attributes. The device is composed of a lift platform, an electric hoist, and a movable pulley. It has a lifting platform that allows parking two automobiles in the same area required by one automobile. Also, intending to facilitate the parking by freeing up the maneuver area, its lift platform has only 3 fixities (instead of 4), i.e., one corner is hanging. Hence, the lift platform is designed with a structural carbon steel grid dimensioned to sustain all the strain on the platform as in [6, 7].

The operation of the parking device is simple. The first car (the red one in Fig. 1) parks on the lift platform. Then, the electric hoist raises the lift platform through a movable pulley, that divides the applied force by two. Finally, the second vehicle (the black one in Fig. 1) parks below the first car. The lift platform features a geometry that prevents fluids such as oil and water from falling over the bottom car. Also, the platform works as a second-class lever, with the resistance force acting between the electric hoist and the fulcrum.

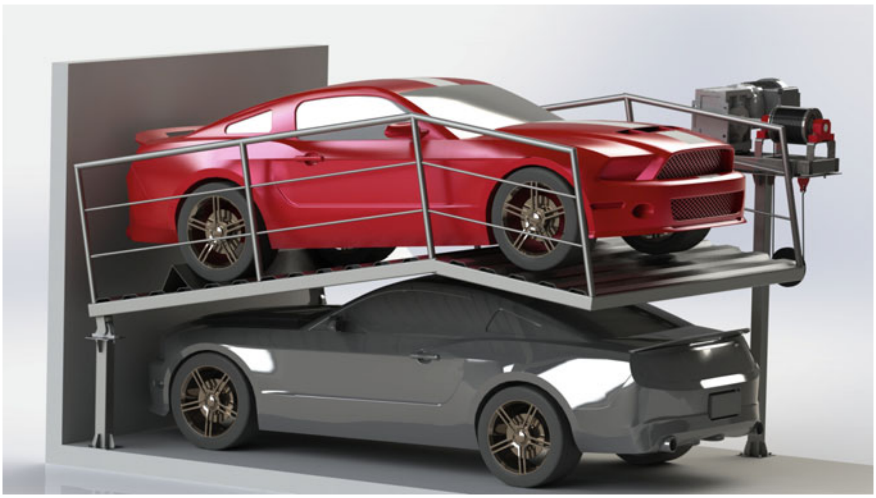


Fig. 1 Parking duplicator device

For the purpose of making the parking task easier, this modern vertical parking system is also equipped with a position sensor and two cameras that communicate with the driver to help his parking. The outputs from the sensor and the cameras are merged, and the pose (position and orientation) of the vehicle relative to the free parking space is tracked while the car is parking. This is accomplished through an algorithm that also predicts the next pose of the vehicle based on its prior state. A small display shows the distance of the car to any object closer than 50 cm and a beep warns the driver if a collision is imminent, or if the car is oriented in a position that probably will not lead to properly parking. This technology aims the proper positioning of the vehicles while parking and the driver's safety. The present paper is centered on the development of this assistive technology. Except for the writing of the algorithm, it comprises all the steps in the conception of the aid system.

In the first stage, this paper approaches the selection of the specific position sensor and cameras to be used on the parking device. The position sensor and the cameras detect with precision the position of the car and other obstacles in relation to the parking space as a function of time as the car is parking. In the second stage, the optimal positions for the sensor and cameras were defined. On a third step, the parking task was simulated on the software Blensor®. The obtained results were effective when applied in precisely determining the vehicle's pose, assisting the driver in the task of parking and avoiding collisions.

2 Materials and Methods

The methodology adopted in the development of the sensor-based system aiming to facilitate the maneuvering of the car was divided into 3 steps: Sect. 2.1 Selection of the Proper Sensor and Cameras; Sect. 2.2 Definition of the Optimal Location for the Sensor and Cameras; and Sect. 2.3 Simulations on the Software Blensor®. It is important to notice that these steps did not happen linearly, they happened simultaneously or with intersection points.

2.1 Selection of the Proper Sensor and Cameras

In the first step, the appropriate sensor and cameras for the presented problem were selected. The fusion of the outputs from the position sensor and the cameras must show the precise pose of the vehicle as a function of the time. Thus, the vertical parking device needs a sensor and cameras that can precisely determine the vehicles' pose throughout the entire maneuvering. At the same time, they cannot be financially expensive, as this would impair the commercialization of the device. So, in this step, the position sensors and cameras available on the market were technically and financially compared. Besides that, suchlike parking aid systems such [8] were studied helping in the selection of the proper sensor. Simulations of various sensors and cameras took place on the software Blensor®. Thereby, the sensor and camera that offered the best benefit–cost ratio were selected.

2.2 Definition of the Optimal Location for the Sensor and Cameras

In the second step, the position in which the sensor and the cameras would be installed was defined. The sensor and cameras need to detect the vehicle from the start to the finish of the parking task, beginning at a distance of at least 7 m from de device. This minimum distance is necessary so that the driver can maneuver the vehicle and get to the park. The locations of the sensor and cameras were evaluated so that there were few obstacles between them and the vehicle. Suchlike analyzes were made through the prototype in Fig. 2, which was modeled in the software SolidWorks® and 3D printed. It helped with the definition of the optimal location for the sensor and cameras. The maneuvering of the two vehicles was simulated and studied on the 3D print aiming to understand the location that would make both cars (superior and inferior) be in the range of the position sensor and the cameras during their whole parking task. Situations such as possible blind spots and interferences were also

Fig. 2 Small-scale 3D printed prototype

evaluated through this same method in the present step. Thus, the optimal locations for the sensor and cameras were defined.

2.3 Simulations on the Software Blensor®

In the third step, the sensor and cameras on de parking device were simulated in the software Blensor® [9]. First, the vertical parking device was modeled in the software SolidWorks®, then it was exported to the software Blensor®. On Blensor®, the sensor and the cameras were installed in the parking device according to the location defined in step "Sect. 2.2. Definition of the Optimal Location for the Sensor and Cameras." Then, the sensor and the cameras were simulated during the maneuvering of the vehicles in the parking task. The obtained results are exhibited in Sect. 3. Results.

2.4 Development of the Algorithm

This step is not concluded yet. It will be presented in future work. It consists of the development of an algorithm and a control system [10] that will merge the data from the cameras with the information from the sensor, detecting the pose of the vehicle through corner recognition, measuring the distance from the vehicle to obstacles, interacting with the driver, avoiding collisions, and assisting the driver throughout the entire task of parking. By means of Kalman Filtering [11], this algorithm predicts the next pose of the vehicle based on its prior state. Also, if a person or an animal enters the parking device operational area, presenting a possible injury condition, this algorithm will stop the parking device and alert the driver with a beep.

3 Results

The selection of the proper sensor and cameras and the definition of its optimal location in the vertical parking have a strong influence on the efficiency of the parking task. Therefore, the results present the selection and the location of the cameras and the sensor which explore the maximum efficiency of the devices.

3.1 Selection of the Proper Sensor and Cameras

Several lidar laser sensors were simulated on the software Blensor® to define the best one for the vertical parking. The generic lidar laser sensor with one ray that reaches 10 m proved to be efficient in the action of detecting the position of the vehicle near the vertical parking. Besides that, it is cheap in the market of laser sensors which makes it a good option considering the benefit–cost ratio.

Also, in the software Blensor®, many types of cameras were simulated to define the best one for the vertical parking. The type of camera Canon APS-C proved to be efficient in the action of detecting the position of the vehicle in relation to the time while it parks. The type Canon APS-C proved to be a good option for the device because it can see very well the space necessary for the vehicle to make maneuvers to park and it can follow the vehicle during the entire parking task. Besides that, this type presents a better benefit–cost ratio than other types of cameras for the needs of the device.

3.2 Definition of the Optimal Location for the Sensor and Cameras

The definition of the best location for the sensor is very important to improve sensor efficiency. After testing many places in the software Blensor®, the best location was defined. As presented in Fig. 3, it is located in the same column where the electric hoist is, and at 0.73 m above the floor. This height is ideal because the generic lidar sensor can detect the position of the vehicle reaching more areas of the vehicle than other heights. Also, the lidar sensor is installed at an angle of $-35°$ in relation to the red cartesian axis shown in Fig. 3. This angle is important because the sensor angle together with the sensor height can allow the parking aid system to properly see obstacles and detect the vehicle during the entire task of parking, including within the platform.

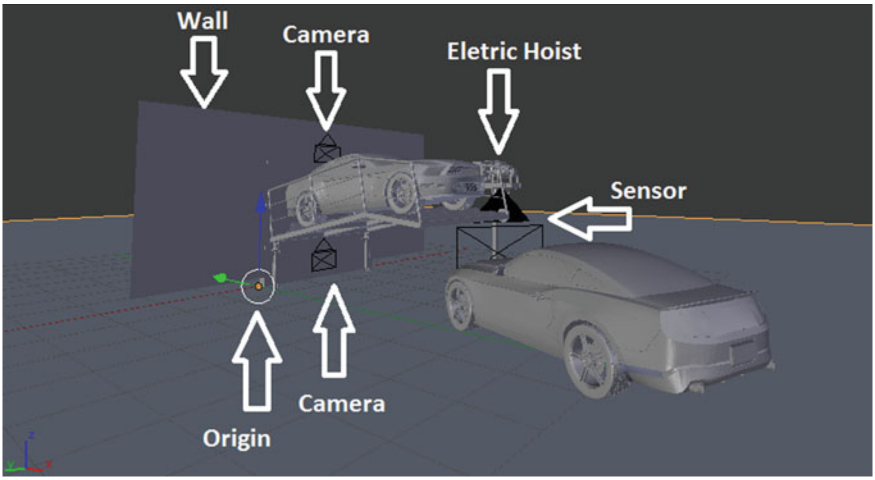


Fig. 3 Location of cameras and sensors

The position of the camera in the vertical parking is essential so that its maximum potential is used in the device. The camera was simulated in different places on the software Blensor®, and it was concluded that the device needs two cameras. Figure 3 shows the locations of the installation of the sensor and the cameras. One camera is located at the wall at a height of 2.7 m above the floor. This height is necessary so that the camera can properly see the space available for parking when the platform is not suspended. After the platform with a vehicle is suspended, the camera's vision becomes poor, so another camera is needed to see the other car that will park underneath. The other camera is located below the platform at a height of 0.45 m above the floor. Now, this height is necessary so that the camera can properly see the free space for parking when the platform is suspended, i.e., when the vehicle is parking in the bottom parking slot. Therefore, the two cameras are essential to the operation of the device when the platform is suspended, or when it is not.

It is possible to see that the sensor is big in Fig. 3 because in the software Blensor® it is unfortunately not possible to let the sensor in its original size, which would be smaller than presented in Fig. 3.

3.3 Simulation on the Software Blensor®

After proper selection of the sensor and cameras and definition of the optimal location for the sensor and the cameras, the whole system was simulated in the software Blensor®. Therefore, four images were generated as shown in Fig. 4, two of them show the camera and sensor views. The camera view in Fig. 4 shows the simulation of the camera with excellent results because the chosen camera can follow the vehicle during the entire maneuver with an accurate definition of the pose of the vehicle in the space of the parking lot. Also, the scan view in Fig. 4 can detect the position of the vehicle with precision in the parking area. In the image Scan View in Fig. 4, the orange line is the back of the car seen by the laser. Also, it is possible to realize that a piece of the platform was detected, which was expected as the laser can reach the vehicle with precision until the very end of the parking task when the vehicle is finally on the platform. The coordinates of the sensor view are presented on the image with the description "Location" on Fig. 4, in relation to the origin presented on "General View" in Fig. 4.

The driver will receive the information from the parking aid system, which will help him to avoid collisions and park smoothly. A small display will show the distance from the vehicle to any obstacle closer than 50 cm in relation to any exterior part of the vehicle. A beep will alert the driver if a collision is imminent. The frequency of this beep increases as the vehicle gets closer to any obstacle. Thus, the driver will get assistance throughout the maneuver and collisions will be avoided.

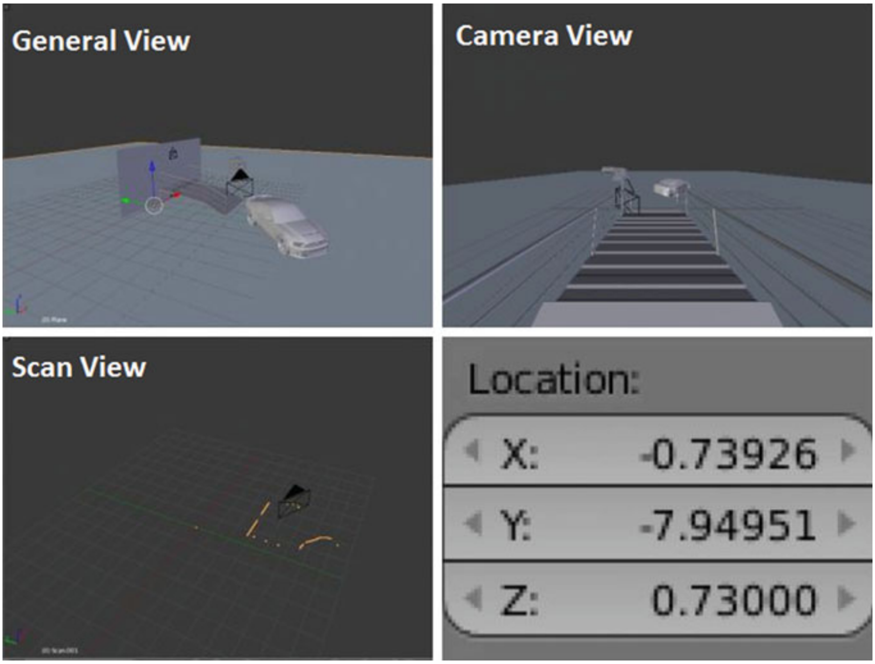


Fig. 4 Parking maneuver

4 Conclusion

The present paper proposed a parking aid system able to assist the driver throughout the entire parking task in a specific parking duplicator device presented on a previous discussion [3]. This parking aid system is composed of cameras and a laser sensor. After its development according to Sect. 2 of this paper, its operation was simulated on the software Blensor®, where the cameras and sensors were evaluated. The obtained results showed that the developed system is robust. The selected sensor/cameras and their locations were able to precisely detect the vehicle throughout the entire maneuver on both parking spaces. The system can effectively warn the driver to avoid collisions and assist him through the entire task of parking.

The next step in this research is the development of the algorithm that will make the fusion of the data from the sensor and the cameras, localizing the car through corner recognition, evaluating the distance of the vehicle to obstacles, and interacting with the driver while assisting him.

Acknowledgements The authors are grateful for the support of the Graduate Program of the Federal University of Minas Gerais, Coordination for the Improvement of Higher Education Personnel (CAPES), Research Support Foundation of the Minas Gerais State (FAPEMIG), National Council for Scientific and Technological Development (CNPq) and Research Development Foundation (Fundep) in the development of the present academic work which is part of the master's thesis [12].

References

1. Leite Júnior, H.F., Alencar C.T., John V.M.: Evolution of Space for the Automobiles in Relation to the Total Built Area of the Buildings from São Paulo. 11ª Conferência Internacional da LARES, São Paulo (2011) (in press)
2. A proporção de habitantes por carro nas capitais, NEXO Journal, https://www.nexojornal.com.br/grafico/2016/01/29/A-propor%C3%A7%C3%A3o-de-habitantes-por-carro-nas-capitais. Last accessed 8 Aug 2019
3. Costa, L.A., Ferreira, M.F.S., Ferreira, A.G., Teixeira, L. F., Costa, J.M.: Development of a parking duplicator mechanical device driven with pulley. In: BTSym 2017 Proceedings. ISSN 2447-8326. V.1 (in press)
4. Costa, L.A.: Dispositivo Duplicador de Vagas de Garagem. BR 10 2017 011814 2, 6 Feb 2017
5. Tidd, J., Bessant, J., Pavitti, K.: Gestão da Inovação, 5th edn. Bookman, Porto Alegre (2015)
6. Costa, L.A., Victorino, A.C.: Stress-strain analysis of a parking system's grid. In: BTSym 2018, Proceedings. ISSN 2447-8326. V.1 (in press)
7. Costa, L.A, Victorino, A.C.: Stress-strain analysis and optimization of a parking duplicator hanging grid. In: Iano, Y., Arthur, R., Saotome, O., Vieira Estrela, V., Loschi, H. (eds.) Proceedings of the 4th Brazilian Technology Symposium (BTSym'18). Smart Innovation, Systems and Technologies, vol. 140. Springer (2019) (in press). https://doi.org/10.1007/978-3-030-16053-1_31
8. Jung, H.G., Cho, Y.H., Yoon, P.J., Kim, J.: Scanning laser radar-based target position Designation for parking aid system. IEEE Trans. Intell. Transport. Syst., 406–424. ISSN 1558-0016 (2008) (in press)
9. Blensor, Blender Sensor Simulation, Universität Salzburg. https://www.blensor.org/. Last accessed 16 June 2019
10. Ogata, K.: Modern Control Engineering, 5th edn. Pearson, Upper Saddle River (2010)
11. Grewal, M.S., Andrews, A.P.: Kalman Filtering: Theory and Practice Using MATLAB, 2nd edn. Wiley (2001)
12. Costa, L.A.: Development of an Intelligent Parking Duplicator Mechanical Device. Master's thesis. Universidade Federal de Minas Gerais, Belo Horizonte. https://hdl.handle.net/1843/31612 (2019).

High-Dynamic-Range Imaging Using Nanowires

Juan C. Barbaran, Victor Murray, Munib Wober, and Fawwaz Habbal

Abstract High-dynamic-range (HDR) imaging applications are increasing rapidly due to their advantages to increase the color quality of the pictures that can be accessed now even from the cellphones. Combining HDR with multispectral images could give us the opportunity to not only have great color information but also to analyze different compositions of materials. However, the use of these techniques is still limited due to portability and costs. In this work, we present linear and nonlinear methods to generate HDR images using a novel single capture multispectral camera based on nanowires sensors. The results presented here show an error reduction in color estimation compared to previous methods.

Keywords High-dynamic-range imaging · Multispectral imaging · Nanowires

1 Introduction

Several methods to capture multispectral images have been developed over the years due to the different advantages of these images in different applications like precision agriculture [1], cancer detection [2], or food quality [3]. Some of the techniques

J. C. Barbaran · V. Murray (✉)
Department of Electrical Engineering, Universidad de Ingenieria y Tecnologia, Lima, Peru
e-mail: vmurray@ieee.org

J. C. Barbaran
e-mail: juan.barbaran@utec.edu.pe

V. Murray
Deparment of Electrical and Computer Engineering, University of New Mexico, Albuquerque, NM, USA

V. Murray · M. Wober · F. Habbal
School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
e-mail: mwober@seas.harvard.edu

F. Habbal
e-mail: habbalf@seas.harvard.edu

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_12

use a scanning process, which means that the filters are tuned or changed every time that a new spectral frequency is needed to be captured, or use a line-scan hyperspectral imager, while other sensors use snapshot methods to capture all the spectral frequencies at the time. One disadvantage of using the first technique is that motion artifacts are generated into the generated images. For this reason, in recent years, there are different efforts to build a sensor that can capture multispectral images in a single capture which could imply reducing the motion artifacts.

Likewise, non-scanning techniques do not use moving parts, thereby, in several cases, the manufacturing cost is low, the design is simple, and the sensors become portable [4]. In the industry, the most used technique to capture multiband images in a single capture is the color filter array (CFA), which uses a Bayer array to obtain three bands: red (R), green (G), and blue (B). The extension of this technique to images with more than three bands is called a multispectral filter array (MSFA). In MSFA, like CFA, there is a pattern that is repeated throughout the sensor called a macropixel, which corresponds to a set of geometrically distributed filters. On the other hand, the use of high-dynamic-range (HDR) imaging is increasing basically thanks to the new cellphones [5]. The combination of HDR with multispectral images could give us the opportunity to both analyze different compositions of materials and to obtain great color information. However, the use of these techniques is still limited for most people due to the high costs and the dimensions of these cameras.

The main objective of this work is to show a novel method to recover color information using a novel prototype of a multispectral camera that is based on nanowires sensors. Color information is always a must in any single camera since it allows us to relate what we are looking with that we are analyzing in different parts of the spectrum. In the next subsections, we describe the novel multispectral camera prototype and how it obtains the images. Then, in the next sections, we describe a new technique to obtain multispectral information. Later, the results and discussions are presented.

1.1 Multispectral Filter Array-Based in Nanowires

A prototype of an MSFA camera has been developed, in the Harvard John A. Paulson School of Engineering and Applied Sciences, using nanowires in the manufacture of filters. Since the nanowires have the characteristic of emitting colors [6] and the fact that they could also absorb colors outside the visible spectrum, for example, in the near-infrared (NIR) [4], they also have the property of refracting light at different wavelengths. This characteristic is due to the dependence of the wavelength to pass through certain spaces, in this case, the space between each nanowire [7], thus it can be considered that the refraction depends on the radius of the same. Grouping nanowires of the same radius, in a determined area, can generate a filter capable of absorbing a specific wavelength. For the construction of these filters, first, a group of nanowires is fabricated in a silicon wafer and covered with a transparent polymer, after that, the encapsulated nanowires in the polymer are separated from the wafer using a knife. Finally, the entire set of filters is superimposed on monochromatic light sensors, allowing them to filter the designed frequency. This technique allows us to manufacture multispectral cameras in a size equal to the conventional digital cameras, that is, much smaller than common multispectral cameras in the market.

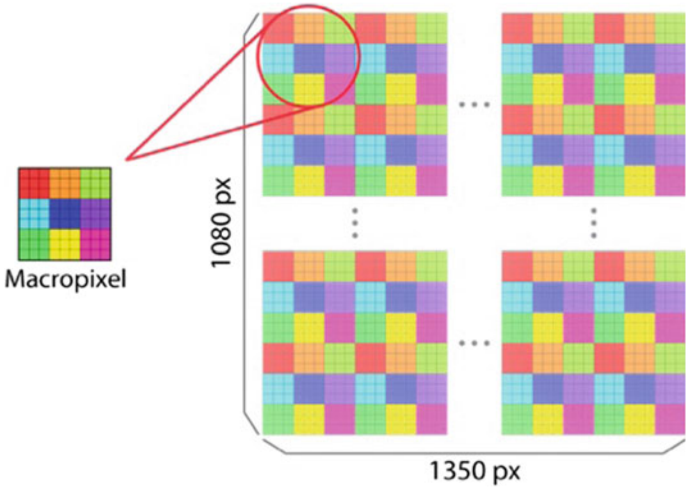


Fig. 1 Distribution of the nanowires filters in the sensor. Each of the colors represents a type of filter grouped into sections of 3 × 3 pixels called the spectral band. Likewise, a group of 9 × 9 pixels containing the 9 filter types is defined as a macropixel

The used prototype presents nine multispectral bands using an array of macropixels shown in Fig. 1. Each filter is grouped in squares of 3 × 3 pixels and the total set of filters has a size of 1080 × 1350 pixels. These filters were superimposed in a monochromatic DMK 27AUP031[1] camera, which has a CMOS sensor MT9P031 with a resolution of 2592 × 1944 pixels (5 megapixels), where each of them measures 2.2 × 2.2um. The data stored by the camera has an 8-bit resolution.

1.2 Image Dataset

The dataset created for the current status of the research can be divided into three groups: (i) images of the Macbeth grid, (ii) images of bottles, and (iii) white background images. In addition, for each group, three cameras were used, where each camera corresponds to a camera similar to the prototype but with no alteration. In

[1]DMK 27AUP031 Monochrome Camera. Information available at www.theimagingsource.com.

this sense, we have three types of images for each group: (i) color images, (ii) gray images, and (iii) images captured with the prototype (multispectral). All captures have a set of images of the same scene with 12 different exposure times ranging from 1 to 1/600 s. In Fig. 2, we show some examples of the images in the database. On the left, we show the color images obtained with a color camera with the same characteristics in the sensor as the prototype. On the right, we show examples of images obtained with the prototype.

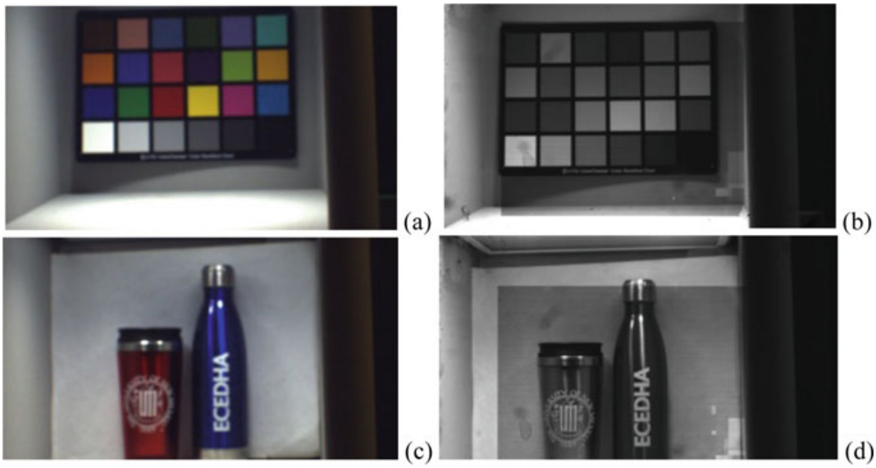


Fig. 2 Some images used as the dataset. **a** Macbeth grid from a color camera. **b** Macbeth grid with the prototype. **c** Bottles with the color camera. **d** Bottles with the prototype

The most important images are the pictures of the Macbeth grid shown in Fig. 2a, b, which is an instrument used to calibrate all types of color cameras. It has twenty-four colors distributed in 4 rows and 6 columns. The main characteristic of this color checker is that the RGB values of each color are known, and for that reason, images of Macbeth grid captured with the prototype will be used to train and validate the model obtained from the color reconstruction.

2 Methodology

In the previous work, [8], to estimate color, information using multispectral images from the prototype was carried out in three stages: (i) demosaicing or separation of samples, (ii) model acquisition, and (iii) post-processing. In a second work, [9], high-dynamic-range imaging was introduced before the first stage to improve the results.

In this work, a new stage is introduced before the separation of samples that consists of performing a saturation correction in each of the pixels; likewise, the use of multivariable nonlinear regression is introduced to obtain the color model and is compared with previous results. The following section will detail the methodology performed and each of its parts.

2.1 Reducing Image Degradation

Assuming that pixel values should be fully saturated when a light source covering the full spectrum is applied, a correction to the degradation is applied. For this, a white background image is used under different exposure times. Thus, the image with all the filters was multiplied by a correction matrix, which was obtained using the white background images. The median of this was calculated and divided between each pixel of the image using:

$$A_{ij} = \frac{\mathrm{median}\{B\}}{B_{ij}}, \tag{1}$$

where A is the correction matrix, B the matrix the white background images, and (i, j) each position of the pixels in the image.

2.2 High-Dynamic-Range Imaging

High-dynamic-range (HDR) images are images generated from two or more pictures with different exposure times. Thus, the dynamic range increases in the image, i.e., the darker and lighter parts will be highlighted correctly. Due to that characteristic, the second step for color reconstruction is to obtain an HDR image from the images obtained in the previous step under different exposure times.

2.3 Multispectral Demosaicing

The next step is to separate the data of each of the nine spectral frequencies. In the beginning, we have a 1080×1350 image where the nine types of filters are found. First, we separate that image into nine different ones, in which each one only contains the information of a single multispectral filter. Since the prototype has an orientation error between the pixels and the filters that originated when joining the two parts (nanowires on to the image sensor), we only take the central pixel of each group of 3×3, since that pixel has the highest probability of not presenting the orientation error.

With this procedure, the nine images are reduced to 360 × 450 pixels. These new images have missing information in some of the pixels, and for that reason, the final step is performing a linear interpolation to complete the data.

2.4 Model Acquisition

Using the nine images belonging to the Macbeth grid, obtained in the previous step, a model for the reconstruction of color was obtained. First, we need to put the matrix form the information of each pixel. Thus, rectangular portions of the image that contain a color of the twenty-four available from the Macbeth grid were selected. In the obtained filter matrix, each row would correspond to one pixel and each column to a different filter. The size of this matrix is $n \times 9$, where n is the total number of pixels.

Then, a second color matrix was created that will contain the ideal color values that each pixel should have. These values were obtained from the datasheet of the Macbeth grid. In this second matrix, each row corresponds to one pixel and each column corresponds to a color value (R, G, or B). These two matrices were used to obtain coefficients that represent the color model. These coefficients were obtained using three types of multispectral regressions: (i) linear regression, (ii) nonlinear regression of second order, and (iii) nonlinear regression of third order. These models can be defined as $Y = bX + E$ or

$$\begin{bmatrix} y_1 \\ y_3 \\ \vdots \\ yn \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{np} \\ 1 & x_{31} & x_{32} & \dots & x_{np} \\ \vdots & \vdots & & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} b_1 \\ b_3 \\ \vdots \\ b_n \end{bmatrix} + \begin{bmatrix} e_1 \\ e_3 \\ \vdots \\ e_n \end{bmatrix}, \tag{2}$$

where y represents the target value, x_{ij} the sampled value, b the coefficients, and e the error. Then, we can use linear and nonlinear methods of the form of:

$$y_i = b_0 + \sum_{j=1}^{n} b_j x_{ij} + e_i, \tag{3}$$

$$y_i = b_0 + \sum_{j=1}^{n} b_j x_{ij} + \sum_{k=1}^{n} b_k {x_{ik}}^2 + e_i, \tag{4}$$

$$y_i = b_0 + \sum_{j=1}^{n} b_j {x_{ij}}^2 + \sum_{k=1}^{n} b_k {x_{ik}}^3 + e_i. \tag{5}$$

The first model is a linear polynomial, while the second model is the combination of a linear and a quadratic polynomial, and, the latter, combines a quadratic and a cubic polynomial. The objective of this method is to obtain the coefficients b_0, b_j, b_k and e_i for every color band (R, G, B). Using these coefficients, the R, G, and B values of each pixel are obtained.

2.5 Post-processing

The last step in color reconstruction is post-processing. First, the outliers of each color band (RGB) are removed. Recalling that we are using an 8-bit camera, the values are in a range between 0 and 255. For this reason, negative values or greater than 255 are considered atypical or saturated. To eliminate these values, any negative pixel was replaced by 0 and any value greater than 255 by this same value. The second procedure performed is a 3×3 median filter. This filter aims to blur the image to avoid sudden changes of color between adjacent pixels.

3 Results and Discussions

For the tests, the images of the Macbeth grid were used as ground truth. We use two different methods to measure the error: the mean squared error (MSE) and the structural similarity index (SSIM) [10]. The MSE measures the average squared of the numerical difference that exists between the reconstructed value and the ideal value that should be obtained for each pixel. The SSIM measures how similar are two images in structure and form regardless of the numerical values of the pixels. The value of SSIM varies between 0 and 1, being 0 totally different images and 1, identical.

In Table 1, we show the results of the three regression methods (3), (4), and (5), for six combinations of exposures times in terms of the peak signal-to-noise ratio (PSNR), derived from the MSE and SSIM. Both nonlinear regression methods have better results compared with linear regression. However nonlinear regression of third grade has the best results in PSNR and SSIM. Also, it can be noticed that the best combination of exposures times is [1/10, 1/20] seconds.

Some of the color images obtained are shown in Fig. 3. The first set of images, Fig. 3a–d, was obtained using linear regression and the second one, Fig. 3e–f, using nonlinear regression of third order. The most important characteristic that has to be highlighted from those images is that there is a big difference between the images obtained with and without saturation correction. In the pictures with the correction, some defects of the prototype originated in the building process were eliminated and also the background is smooth and has homogeneous tonalities.

Table 1 Results of color reconstruction using correction in terms of SSIM and PSNR for six different combinations of exposure times in each of the regression methods: linear and nonlinear (NL)

Exposure times (s)	Error type	Linear	NL 2nd Order	NL 3rd order
[1/5, 1/10]	PSNR	21.8044 dB	23.2388 dB	23.8695 dB
[1/10, 1/20]	PSNR	21.5604 dB	23.0131 dB	23.5959 dB
[1/20, 1/40]	PSNR	20.7309 dB	21.9950 dB	22.8471 dB
[1/5, 1/10, 1/20]	PSNR	21.6298 dB	23.0740 dB	23.7162 dB
[1/10, 1/20, 1/40]	PSNR	21.2076 dB	22.6478 dB	23.1363 dB
[1/5, 1/10]	SSIM	0.4580	0.5541	0.5248
[1/10, 1/20]	SSIM	0.4525	0.5392	0.5474
[1/20, 1/40]	SSIM	0.4510	0.5193	0.5589
[1/5, 1/10, 1/20]	SSIM	0.4529	0.4856	0.5532
[1/10, 1/20, 1/40]	SSIM	0.4509	0.4633	0.5387

4 Conclusions

We have presented a new technique to estimate high-dynamic-range color information using linear and nonlinear methods from data obtained using a novel single-shot multispectral camera. This novel prototype is based on nanowires sensors located on the top of a regular image sensor. Each group of nanowires sensors can select a specific frequency based on their dimensions. The results presented here reduce the error, in terms of both PSNR and SSIM, compared to previous methods. The nonlinear methods are about 22% better than the linear systems. The matrix correction used as part of the process is key to reduce the errors due to the degradation that appears during the implementation of the nanowires sensors. As future work, the goal is to reach a higher SSIM value that could imply that the HDR images are getting closer to commercial cameras but with these novel nanowires sensors.

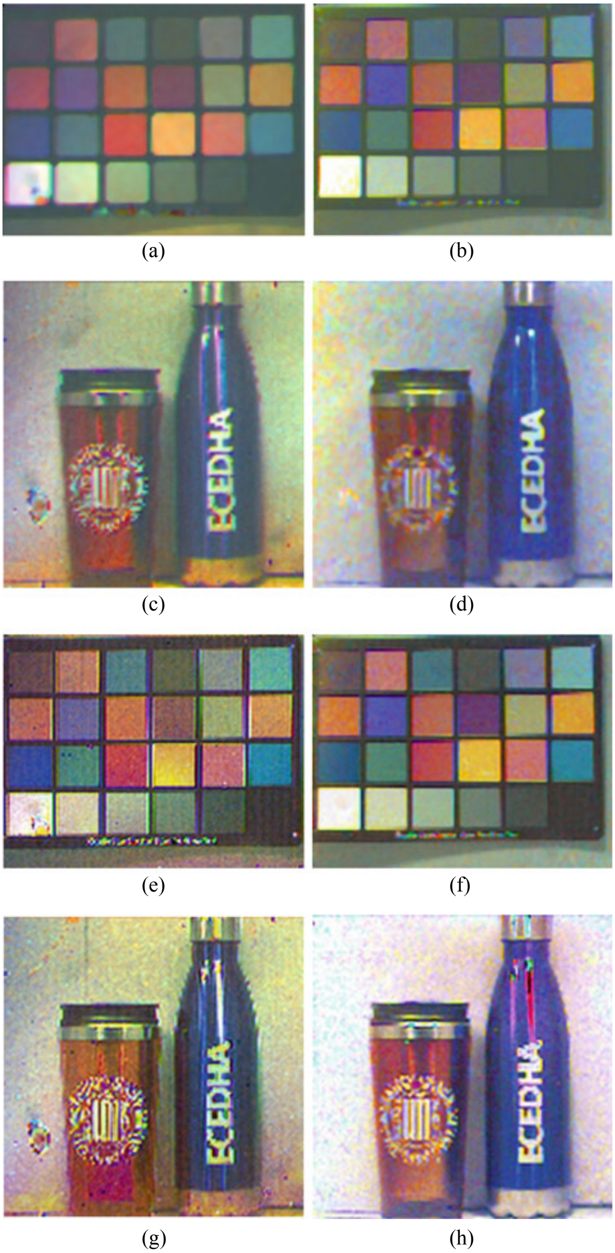


Fig. 3 Color image reconstruction for different methods using high-dynamic-range imaging. **a**–**d** Linear regression. **e**–**h** Nonlinear regression of third order. The left column images were obtained without using the matrix correction and the right column images were obtained using the matrix correction. (**a**), (**b**), (**e**) and (**f**) Macbeth grid. (**c**), (**d**), (**g**), and (**h**) test bottles (see Fig. 2 as reference)

References

1. Barbaran, J.C., Murray V., Wober, M., Habbal, F.: Color estimation using nanowires sensors. In: IEEE International Conference on Electronics, Electrical Engineering and Computing, Lima (2019)
2. Barbaran, J.C., Murray V., Wober, M., Habbal, F.: A first approach to multispectral high-dynamic-range imaging using nanowires sensors. In: IEEE International Conference on Electronics, Electrical Engineering and Computing, Lima (2019)
3. Kriegmair, M.C., et al.: Multiparametric cystoscopy for detection of bladder cancer using real-time multispectral imaging. European Urology (2019)
4. Liu, J., et al.: Rapid and non-destructive identification of water-injected beef samples using multispectral imaging analysis. Food Chem. **190**, 938–943 (2016)
5. Lum, C., et al.: Multispectral imaging and elevation mapping from an unmanned aerial system for precision agriculture applications. In: Proceedings of the 13th International Conference on Precision Agriculture (2016)
6. Park, H., et al.: Filter-free image sensor pixels comprising silicon nanowires with selective color absorption. Nano Lett. **14**(4), 1804–1809 (2014)
7. Park, H., Seo, K., Crozier, K.B.: Adding colors to polydimethylsiloxane by embedding vertical silicon nanowires. Appl. Phys. Lett. **101**(19), 193107 (2012)
8. Seo, K., et al.: Multicolored vertical silicon nanowires. Nano Lett. **11**(4), 1851–1856 (2011) (PMID: 21413684)
9. Sharma, A.: Exploring the basic concepts of HDR: dynamic range, gamma curves, and wide color gamut. Inf. Display **35**(5), 14–18 (2019)
10. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

An Insight into Applications of Internet of Things Security from a Blockchain Perspective

Lucas Alves Rodrigues de Sá, Yuzo Iano, Gabriel Gomes de Oliveira, Diego Pajuelo, Ana Carolina Borges Monteiro, and Reinaldo Padilha França

Abstract The Internet of Things (IoT) has grown wildly in the last years but the security of the generated data has not improved at the same proportion as the number of devices connected to the Internet has multiplied, and this is very worrying because privacy is one of the most important topics when a private IoT application is developed. As the growing number of IoT application is taking place more strategies for data security is required to deal with different scenarios of the wide and vast type of IoT applications. An investigation in order to find proper solutions to common IoT applications that have been challenged for improvements in data security is the goal of this paper. Through blockchain technology that offers security naturally without any expensive investment, the data security that is required by IoT devices become reachable. This paper presents a discussion in order to clarify how the security on Internet of Things applications can be enhanced using blockchain providing a correlation between the capabilities of both technologies.

Keywords IoT · Security · Blockchain · Data privacy · IoT decentralization

L. A. R. de Sá (✉) · Y. Iano · G. Gomes de Oliveira · D. Pajuelo · A. C. Borges Monteiro · R. Padilha França
School of Electrical and Computer Engineering (FEEC), University of Campinas—UNICAMP, Av. Albert Einstein, 400, Barão Geraldo, Campinas, SP, Brazil
e-mail: l226526@g.unicamp.br

Y. Iano
e-mail: yuzo@decom.fee.unicamp.com

G. Gomes de Oliveira
e-mail: oliveiragomesgabriel@ieee.org

D. Pajuelo
e-mail: diego.pajuelo.castro@gmail.com

A. C. Borges Monteiro
e-mail: monteiro@decom.fee.unicamp.com

R. Padilha França
e-mail: padilha@decom.fee.unicamp.com

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_13

1 Introduction

Internet of Things (IoT) brings a current and transformative concept about connecting physical objects using sensors, chips, and software to connect and exchange data with other devices and systems over the Internet, where these devices vary from ordinary household objects. Even sophisticated industrial tools, these IoT "things" are related to any object that has had the implementation of sensors and other digital systems to work smarter by exchanging information with people and other objects, not necessarily using Internet connections [1].

Internet of Things devices are becoming more common and increasingly revolutionizing the ways of thinking and building technologies, and they can be any hardware that can collect data from the physical world, such as temperature, size, speed, and share this data with the Internet, through of virtual world [2].

These exchanges of information can occur via Radio Frequency (RFID), Wi-Fi, Ethernet, Bluetooth, among other forms of connection currently existing, and communication network systems exist in different proportions and can be connected to the world wide web or the user's car or just to the home network, for example. In this sense, "things" through digital systems added to them are able to connect to other objects and people, allowing the sending of commands, the return of usage data, the identification of each other's presence, among others. Other applications [1, 3].

IoT refers to a current trend of connecting all kinds of physical objects to the Internet, especially the most unusual ones, it is possible to connect everyday objects like cars, thermostats, appliances, baby monitors to the Internet through embedded devices, perfect communication between people, processes, and things is possible. Through growing computing technologies over the years like cloud, big data, big data analytics, and mobile technologies, physical things can share and collect data with minimal human intervention, making the modern world hyperconnected, digital systems can monitor, record, and adjust every interaction between connected things. The physical world meets the digital world, and they cooperate [4].

Due to the growing usage of devices connected to the Internet, the integrity and access to the dense cloud of data that have been stored on the Internet have worried many investors. Similar to a personal computer and mobile phones, which are likely to be hacked, when using an IoT device is identical, since security and privacy are the biggest challenges facing today as all devices collect user's personal information, share with other devices, and consequently, the data is stored in the manufacturer's databases which can be hacked [1, 2].

However, the more devices connected to the network, the greater the exposure to risks such as data hijacking or theft, alteration of information or even the connected environment, due to security breaches that can occur on both sides, manufacture, and user, from an information security standpoint, we also need to consider the possible security breaches that industries may overlook, as well as from the standpoint of using as an insecure non-encrypted network and insecure web interfaces with authentication. Weak or insufficient, unsafe mobile and cloud interface, and security settings that make software unsafe are examples of failures [5].

Thus, the main challenge of ensuring the security of IoT systems is that some devices have limited capabilities, and therefore cannot perform traditional security functions, where many of these IoT devices were not designed with security in mind because their main objective was to add functionality at a low cost. One answer to these two questions lies in a technology known as the blockchain, technology incorporated in bitcoin that is a cryptocurrency introduced in 2009 by Satoshi Nakamoto, where blockchain can be conceptualized as a decentralized database [6, 7].

Blockchain technology is a public ledger (or ledger) that records a virtual currency transaction, the most popular of which is bitcoin, so this record is reliable and unchanging, it records information in a group of block transactions, marking each block with a timestamp and date, and at each time period (about minutes on the blockchain), a new transaction block is formed which binds to the previous block; as the amount of traded bitcoins, who sent/received, and when this transaction was made, and wherein the book it is recorded, in the virtual currency transaction scenario. This shows that transparency is one of the key attributes of blockchain. In short, blockchain technology is a public, distributed ledger that records all virtual transactions in a blockchain that anyone can participate in [8, 9].

In this sense, this paper aims to clarify how the security on Internet of Things applications can be enhanced using blockchain providing a correlation between the capabilities of both technologies.

2 Methodology

This survey carried out a bibliographic review focusing on the rationale for linking security on the Internet of Things applications with the use of blockchain technology developing a correlation between both technologies.

3 Background Knowledge

The concept of Internet of Things (IoT) refers to a wide variety of smart devices that connect to the Internet, interacting with each other to automate processes, is based on the concepts of connectivity and intelligence, for better viewing in an industrial plant scenario, containing thousands of sensors in a factory can give the impression that everything is under control, however, if the collected data is not used for achieving operational excellence, so much information will only generate waste [1].

IoT offers a compelling insight into applications that can turn raw data from hundreds of thousands of sensors into meaningful information, as critical issues such as performance, availability, and security take precedence over interoperability. Interoperability in IoT is related to the software that must have the ability to transmit and receive information correctly, i.e., senders and receivers must understand data from the same perspective, creating the complexity of IoT, which has been developed in different segments, making it difficult to standardize protocols as well [10].

Thus, the challenge related to the integration of layers in an IoT application is still just in its infancy, and considering that the total potential value of an IoT application depends on the ability to achieve interoperability between different IoT systems. An IoT system can be the very definition of a heterogeneous system, comprising a widely diverse set of devices and resources collected in its multiple layers, which to work together, and these parts need a common structure of protocols and higher-level models to safely recognize each other and exchange information about data, resources, and status. Thus, this diversity of devices and resources, even in a single IoT application, requires a more extensive set of standard methods to deal with possible interactions between the various parts of an IoT system [5].

IoT device interoperability is nowadays based on the client-server model (Fig. 1) where the clients are the IoT devices, and the servers are databases present in the Internet; this database normally are characterized as centralized databases; in other words, the management and data storage of all IoT devices data is centralized in a single database. The client-server model does not offer data security in its nature, then is necessary for the implementation of security techniques like data encryption to keep the data secured on the Internet. Most IoT devices are constrained having no great power to run the encryption algorithms needed to bring security to the data generated by the IoT devices [11].

According to [12], nowadays, in order to have safety automation in an IoT process, the involved IoT devices need to accomplish some requirements like:

- Device and data security, including the authentication of devices and confidentiality and integrity of data;
- Implementing and running security operations at IoT scale;
- Meeting compliance requirements and requests;
- Meeting performance requirements as per the use case.

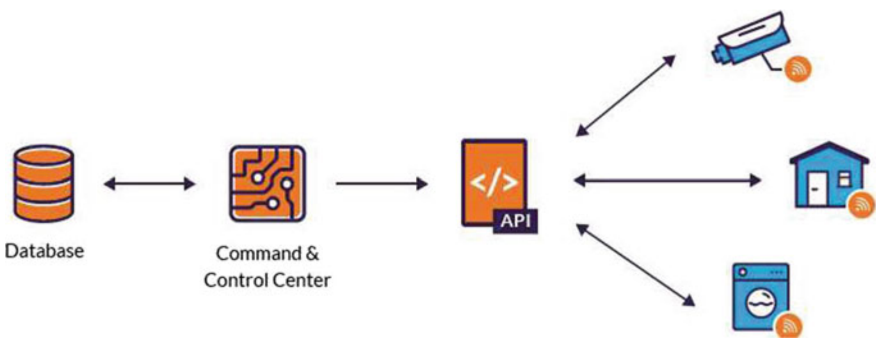


Fig. 1 Current manner management [10]

The same author also presents some IoT devices vulnerabilities and security issues as follow:

- Unpatched vulnerabilities, regarding connectivity issues or the need for end users to manually download updates directly from a C&C center, often result in devices running on outdated software, leaving them open to newly discovered security vulnerabilities.
- Weak authentication, with respect to manufacturers, often release IoT devices (e.g., home routers) containing easily decipherable passwords, which might be left in place by vendors and end users. When left open to remote access, these devices become easy prey for attackers running automated scripts for bulk exploitation.
- Vulnerable APIs, concerning as a gateway to a C&C center, APIs are commonly targeted by a variety of threats, including man in the middle (MITM), code injections (e.g., SQLI), and distributed denial of service (DDoS) assaults.

The author of [13] concludes that security is the key issue affecting IoT development; only by ensuring security can IoT be more universal.

In order to guarantee the implementation of an IoT application in a more secure way, a technology known as blockchain can be used, it was first introduced in 2009 by Satoshi Nakamoto with the advent of the cryptocurrency known as bitcoin [14]. From a generic perspective, a blockchain is defined as an immutable ledger [15]; in this way, everything that happens within the blockchain network cannot be changed.

Technically speaking the blockchain is a chain of blocks where each block interacts with the previous block (Fig. 2); in other words, each block is a peer of a P2P (peer-to-peer) network; in this way, each peer or block has a digital identifier known as a hash. The hash function is an algorithm used in cryptocurrencies that transform large numbers of information into a fixed-length hexadecimal number sequence, each hash is created with the aid of a double-SHA-256 algorithm, which creates a random 512-bit number or 64 bytes [7, 16, 17].

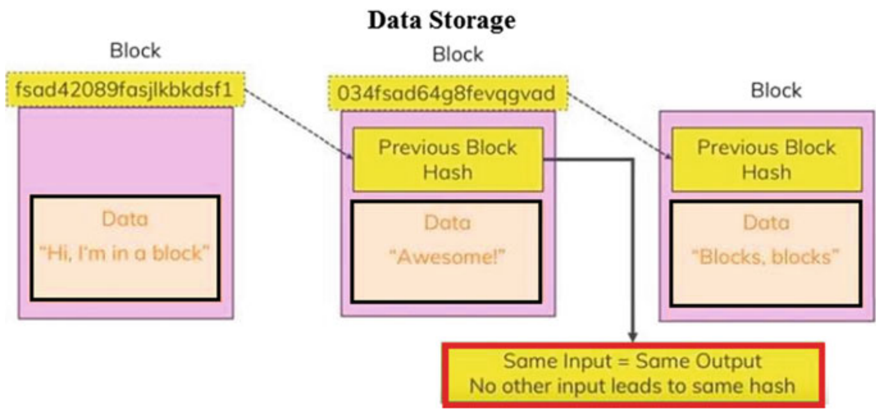


Fig. 2 Blockchain organization

So, the hashes are generated by an algorithm SHA-256, such as each block has a respective hash, the subsequent block points to the hash of the previous block, wherein this interaction this link between blocks continue until the first block of the entire blockchain can be reached, which block is known as the genesis block. This connection between block forms a chain of the block that means a blockchain [16, 17].

This is also the point that makes cryptocurrencies so secure, due to blockchain, since each block has a hash based on the previous block. Therefore, if someone modifies a block already added to the blockchain to create a fake transaction, then this hash will be modified; and since each block is built on information from the previous block, a modification to the blockchain would affect the entire chain, resulting in users' ability to identify fraud and keep the system free from failure [17].

That way, since IoT, looks for methods that are secure enough to automate processes and exchange data in real-time; blockchain could be the perfect fit, whereas the number of sensors in vehicles, factories, buildings, and urban infrastructure grows, companies are looking for safe and automated ways to enable a process network. In this scenario, blockchain could offer a standardized method for accelerating data exchange and enabling processes between IoT devices without an intermediary, commonly used a server that acts as the central communication according to requests [18].

IoT devices are constrained, this means that they are very limited devices in terms of memory, power of processing and requires security to be lightweight, scalable, distributed, and private. Blockchain has the potential to overcome these IoT device requirements because it is distributed, secure, and private in nature [19] (Fig. 3).

Or in a possible blockchain-based network, but in an interconnected transaction flow, where transactions can be verified directly by the users requesting them, as long as they perform two other transactions previously, and can have a limit of transactions per second that can be processed is directly related to the number of users on the network [17, 19].

The major benefit of using blockchain in an IoT application is because the security of a blockchain increases with the number of devices and is also more difficult to carry out the dreaded 51% attacks [20]. In this sense, blockchain will allow IoT ecosystems to break with the traditional connector-based network paradigm, where devices rely on a central cloud server to identify and authenticate individual devices [7].

In the last few decades, the centralized model has been the most used; however, it can be problematic since the number of network nodes increases in the order of millions and consequently results in billions of transactions. Consequently, there is an exponential increase in computational requirements, which culminates in costs. In this context, servers can be seen as bottlenecks, creating vulnerability for IoT networks to be attacked by denial of service (DoS/DDoS). In view of this, it will be a laborious task to establish centralized networks in several industrial sectors, mainly in large farms, since the IoT nodes will be able to expand in large areas with a lack of connection possibilities [7, 16].

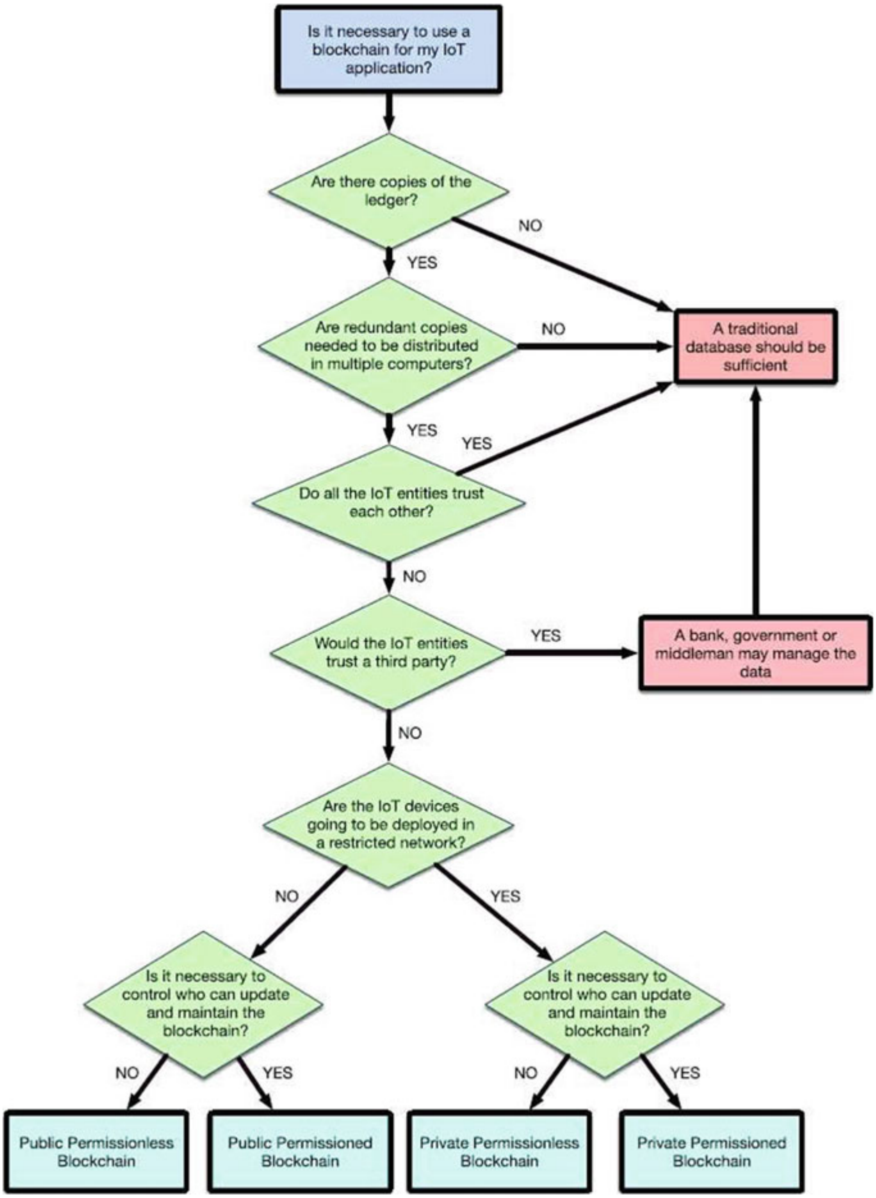


Fig. 3 Flow diagram for deciding when to use blockchain in an IoT application [19]

By contrast, using blockchain will allow for safer network meshing where IoT devices reliably interconnect, avoiding threats such as device spoofing and impersonation as long as each legitimate node can be registered. In the lock chain, devices may be able to authenticate and identify each other without requiring central servers or certification authorities, where the network will scale to support billions of devices without the need for additional resources [7, 16].

4 Results and Discussion

In [19], with blockchain implementation was created a chain of smart home devices that improved security naturally. In a smart home blockchain network, the IoT devices share their data in the blockchain network that protect them against hackers, all transactions performed by the IoT devices have to be validated by the miners. The validation procedure is an inherent feature of blockchain.

Although still in its early stages of development, IoT is primarily comprised of technologies that enable data collection, remote monitoring, and device control due to hyper connection of users when using IoT devices, the need to invest in security solutions is increasing. As IoT progresses, it will become a network of stand-alone devices that can interact with each other and their environment and make smart decisions without human intervention.

The Internet of Things has enabled the creation of a hyperconnected world, where the number of sensors in vehicles, houses, buildings, and urban infrastructures grows, but there is also a growing need for safe and automated measures to enable these devices for networks and networks; since most devices and objects used in everyday life by society at large are gaining the ability to connect to the Internet or device networks that share information with each other, and these IoT applications now require enhanced security that sensors and integrated devices will transmit information to each other and over the Internet.

Blockchain enters this scenario to ensure the protection of these communications by preventing devices from being compromised by cyberattacks and user behavioral patterns from being unraveled, where blockchain technology has become a great ally for companies by providing a standardized method for accelerating data exchange. Enabling processes to be executed between IoT devices without intermediaries and securely.

Blockchain will also enable data monetization, where owners of IoT devices and sensors can share IoT-generated data in exchange for real-time micropayments. It can help companies improve security through decentralized interaction and data exchange. Optimize the reliability of devices, applications, and platforms. By uniting blockchain + IoT, companies are able to reduce challenges through innovative processes. The encryption and decentralized control offered by blockchain are highly effective alternatives to replace traditional security mechanisms.

This technology can easily deal with IoT security issues ensuring that the data shared in the blockchain network will not be changed or deleted, can be guaranteed due to the fact that, in a blockchain network to an intruder modify or delete a date it has to have control of most part of the devices that are members of the network. Still considering that this attacker has to rewrite the modification or deletion on the consecutive blocks from the block where the data was modified or deleted, depending on how big is the blockchain network, the effort to be practically impossible because it requires much power of processing that can't be offered by a conventional computer.

Blockchain, therefore, presents many promising opportunities for the future of IoT, but the challenges nonetheless remain, such as consensus models and computational transaction verification costs, still in the early stages of blockchain development, with these obstacles eventually overcome, further paving the way for many application possibilities.

5 Conclusions

Despite the ease and benefits of connecting objects around us today, IoT is still a recent technology that is constantly improving and growing, but a downside to being fully connected all the time is data leakage, intrusion, privacy, and any viruses that may cause problems with digital systems. However, the focus of technology and endeavors with innovative ideas is to provide users with greater peace of mind and comfort, since data obtained from physical objects connected to networks only have to contribute positively to our future.

Equipping Industry 4.0 smart devices, connected cars and appliances, businesses, and factories with the right security solutions from the start is critical to helping prevent attacks, even though gateways that connect IoT devices to enterprise and manufacturer networks need to be protected. However, in contrast to human-controlled devices, gateways that go through a single authentication process can make perfect sources of infiltration on corporate networks; in this context, it is clear that the more security implemented the better for overall system protection. Still considering that having an effective risk management mechanism is an important part of this process.

So, in the context of blockchain for that kind of IoT application that supports the implementation of blockchain, the integrity of the IoT devices would naturally take place because the transactions are recorded and grouped into blocks, the blocks are protected because they use the hashing encryption, due to decentralization and data security provided by this encryption, theoretically, it is impossible to make changes when something is written to the blockchain [19].

Whereas applying the right security strategy in addition to helping to protect a company's data also maintains customer trust. Finally, the right security solutions for IoT help both an enterprise activate new business models and save costs.

References

1. Wortmann, F., Flüchter, K.: Internet of things. Bus. Inf. Syst. Eng. **57**(3), 221–224 (2015)
2. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
3. Padilha, R., Iano, Y., Monteiro, A.C. B., Arthur, R., Estrela, V.V.: Betterment proposal to multipath fading channels potential to MIMO systems. In: Brazilian Technology Symposium, pp. 115–130. Springer, Cham (2018)
4. Fleisch, E.: What is the internet of things? An economic perspective. Econ. Manage. Finan Markets **5**(2), 125–157 (2010)
5. Gilchrist, A.: Industry 4.0: the industrial internet of things. Apress (2016)
6. Zhou, J., Cao, Z., Dong, X., Vasilakos, A.V.: Security and privacy for cloud-based IoT: challenges. IEEE Commun. Mag. **55**(1), 26–33 (2017)
7. Tasatanattakool, P., Techapanupreeda, C.: Blockchain: challenges and applications. In: 2018 International Conference on Information Networking (ICOIN), pp. 473–475. IEEE (2018)
8. Mettler, M.: Blockchain technology in healthcare: the revolution starts here. In: 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), pp. 1–3. IEEE (2018)
9. Bagchi, R.: Using Blockchain technology and smart contracts for access [Art] (2017)
10. Olson, N.: The internet of things (2016)
11. Aloi, G., Caliciuri, G., Fortino, G., Gravina, R., Pace, P., Russo, W., Savaglio, C.: Enabling IoT interoperability through opportunistic smartphone-based mobile gateways. J. Network Comput. Appl. **81**, 74–84 (2017)
12. Jin-Cui, Y., Bin-xing, F.: Security model and key technologies for the Internet of things [Art]. J. China Univ. Posts Telecommun. (2011)
13. Thales: what are the key requirements of IoT security? Thales, 31 08 2019. [Online]. Available: https://www.thalesesecurity.com/faq/internet-things-iot/what-are-key-requirements-iot-security. Accessed 31 08 2019
14. Bagchi, R.: Using blockchain technology and smart contracts for access [Art]. University of Helsinki (2017)
15. Dorri, A., Kanhere, S.S., Jurdak, R., Gauravaram, P.: Blockchain for IoT security and privacy: the case study of a smart home [Art]. In: The University of New South Wales (2017)
16. Carlozo, L.: What is blockchain? J. Accountancy **224**(1), 29 (2017)
17. Mauri, L., Cimato, S., Damiani, E.: A comparative analysis of current cryptocurrencies. In: ICISSP, pp. 127–138 (2018)
18. Kshetri, N.: Can blockchain strengthen the internet of things? IT Prof. **19**(4), 68–72 (2017)
19. Fernández-Caramés, T.M., Fraga-Lamas, P.: A review on the use of blockchain for the internet of things. IEEE Access 6:32979–33001 (2018)
20. Mustard, S.: Will blockchain technology disrupt the ICS world?. ISA (2019). [Online]. Available: https://www.isa.org/intech-plus/2018/january/will-blockchain-technology-disrupt-the-ics-world/ Accessed 23 July 2019

IP Network Management from the Perspective of Distributed Artificial Intelligence: State of the Art

Juan Carlos Zegarra-Vásquez and Nestor Mamani-Macedo

Abstract An approach to study Distributed artificial intelligence is through software agent-based intelligent systems. Instead of having a large centralized system that collects complete system information, there is a small number of subsystems or agents, which are focused on the effort to cooperate to solve a given problem. Agents can communicate with each other, exchange views and find various strategies to obtain a specific solution. Due to the exponential growth of network devices and the massive use of broadband networks, current centralized management systems are reaching limits in terms of managing the amount of information to be managed, which is why distributed management based on the paradigm of intelligent agents can be applied in this domain. In the literature, there are several experimental and academic prototypes that attempt to demonstrate how the use of distributed artificial intelligence can be an interesting alternative to the classic centralized network management systems. This article benchmarks these prototypes, emphasizing two aspects (i) the strategy and (ii) the inherent functional blocks of each proposal.

Keywords Multi-agent systems · Agent-oriented software engineering · Mobile agents · SNMP protocol

J. C. Zegarra-Vásquez
Facultad de Ingeniería Electrónica y Eléctrica, Universidad Nacional Mayor de San Marcos, Lima 01, Peru
e-mail: juan.zegarra@perunetworks.net

N. Mamani-Macedo (✉)
Facultad de Ciencias Matemáticas, Universidad Nacional Mayor de San Marcos, Lima 01, Peru
e-mail: nmamanim@unmsm.edu.pe

Facultad de Ingeniería, Universidad Peruana de Ciencias Aplciadas, Lima 01, Peru

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_14

1 Introduction

It can be affirmed that network management is the set of activities carried out so that the network provides an adequate level of service to the users. The main objectives of network management are to improve the availability and performance of system elements, as well as to increase their effectiveness [1]. The growth of networks meant serious changes in network management, it making necessary to control an increasing number of equipment and technologies from different suppliers, and gradually introduce new services [2]. The efficiency of network administrators to manage networks decreases as they become more complex and heterogeneous. Managing large networks created a crisis for many organizations [3]. That is why it is not enough to know the faults that may be occurring on the network at a given time, but it is essential to be able to anticipate them and know how to react in case they occur. In that sense, intelligent agents are a suitable tool for self-management in the area of telecommunications, specifically autonomous networks [4].

2 Theoretical Framework

Distributed artificial intelligence (DAI) has emerged as a new branch of artificial intelligence that aims to study intelligent systems formed by a set of agents [5], with the aim of solving problems where collective behavior is more efficient than individual behavior. One of the branches of the DAI is systems with multiple agents or multi-agent systems (MAS), which studies the behavior of a collection of autonomous agents with the aim of solving a problem [6]. In generic form, we can say that an agent is any entity with the ability to perceive the environment in which it is located, and that has the ability to act on the same environment [7]. Software engineering is the application of a systematic, disciplined and quantifiable approach to software development, operation and maintenance [8]. Object-oriented methodologies (OO) fail to meet the development needs of these new systems [9] because an object is not autonomous, nor proactive. As for the agent-oriented software engineering, there is no uniform consensus regarding the software engineering methodology to be used, so, several proposals for agent-oriented methodologies have been made: TROPOS, MaSE, Message, Ingenias, Zeus, Prometheus, GAIA, MAS-CommonsKADS, Vowel Engineering, among others [10].

The ISO management model classifies the tasks of network management systems into five functional areas. The task of the person in charge of managing a business network will be to evaluate the management platform to be used in terms of the extent to which said platform forms the management problem of these areas [11]:

- Configuration Management. It obtains data from the network and uses it to add, maintain and remove components and resources to be monitored.
- Fault Management. Location and recovery of network problems.

- Security Management. Offer mechanisms that facilitate the confidentiality, availability and integrity of communications.
- Accounting Management. It measures network usage parameters that allow decisions to be made to optimize the network infrastructure.
- Performance or Performance Management. It maintains the level of service that the network offers to its users.

3 Intelligent Agent Integration and Network Management Solutions

In recent years, several researchers have proposed work that integrates network management with intelligent agents in different fields. Each work addresses a particular problem of a part of the reality of management, so it is important to take into account research advances in this field.

3.1 An Improved Network Monitoring System that Uses Multi-agent-Based Technology

The authors [12] present a system based on multiple agents to monitor nodes in a network environment. The models developed for the proposed system define certain deliberative agents that interact with each other to achieve the objectives and requirements of the multi-agent organization and the integration of fault services, security services and configuration services into a single platform. The automated network monitoring system integrates configuration, fault and security management services into a single software platform to monitor nodes in a network environment effectively using a multi-agent-based approach as recommended by Akinyokun et al. [13]. Figure 1 shows the system architecture based on eight agents that cooperate in the same domain to perform network monitoring services to meet the main objective.

The functions of the agents are

- **Server Agent**: Performs the initial configuration, network testing, data collection, analysis and mining functions programmed by the network administrator. You can accept and send requests to and from other agents. Send messages to the report manager for later transmission to the network administrator's screen notification system and also integrate with Web services for SMS alerts via mobile phone.
- **Client Agent**: Searches for information on each computer on which it resides and also gives instructions to the client's operating system to perform specific tasks such as the shutdown operation.
- **Failure Detection Agent**: The fault detection agent is designed as a reactive agent that monitors all nodes in the local area network to detect nodes in an active or inactive state.

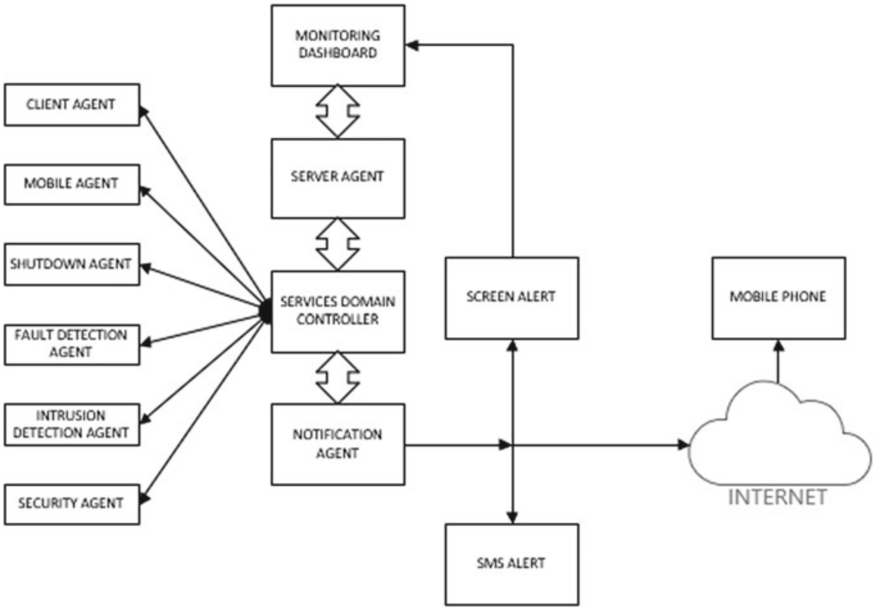


Fig. 1 Proposed system architecture based on multiple agents

- **Intrusion Detection Agent**: The intrusion detection agent uses an anomaly detection mechanism in a node (s) of the network environment. The agent is trained for the task of intrusion detection using Fuzzy ART techniques.
- **Notification Agent**: Performs the function of sending an almost real-time alert message of all network events as alert messages on the network monitor screen and short message service (SMS).
- **Shutdown Agent**: It is an autonomous and reactive agent designed to shut down all profiled nodes with the active state in the network, using a pre-programmed time.
- **Security Agent**: Creates a secure end-to-end communication between the server agent and the client agent residing on each node of the network, encrypting and decrypting messages.
- **Service domain controller**: Directly oversees and controls all agent services in the domain. It must first be started before activating any other agent.
- **Mobile Agent**: It is the messaging agent that can be sent from the service domain controller to other nodes in the network environment.

According to the authors, the tests of the proposed network monitoring application using the multi-agent model proved highly efficient. The integration of the services of the ontology of failures, detection of intruders and configuration in a single platform that uses multiple agents for the network environment is an important advance to solve the problems of the administrator of the network.

3.2 Network Management Based on Multi-agent Architecture

The authors of the document [4] have the objective of proposing a network management architecture based on intelligent agents. The focus of this architecture is the autonomous management of networks by designing a collaboration method based on the task-oriented domain. This method implements a negotiation mechanism for agents to improve the performance of network management through a multi-agent hierarchical architecture.

The specific network administrator maintains high-level policies. At the lower level, there is a set of agents that collect data related to the QoS parameters, which when detecting the current state of the network, these agents transfer the data to the level that contains the decision-making module, through a negotiation mechanism between a set of collaborating agents that are governed by network policies. The proposed multi-agent architecture specifies two types of agents:

- **Sensing Agents (SA)**: They are reflex agents based on models that store information about the current state of the world using an internal model; the action performed by an agent is based on their perception and rules.
- **Collaborating Agents (CA)**: They are utility-based agents that use a world model and a utility function that will allow rational decision making by each agent.

In this way, the authors use the task-oriented domain as a way to use artificial intelligence methods to implement a negotiation protocol because the agents have a function that qualifies the usefulness of each agreement so that they can make a decision and optimize its current utility.

3.3 Network Management Using Multi-agent Systems

The authors [14] present a multi-agent network management system that defines a set of intelligent agents that interact to achieve the objectives and requirements of the multi-agent organization. These agents have the property of being adaptive, acquiring knowledge and skills to make decisions according to the real state of the network that is represented in the information base (MIB) of devices that use the SNMP protocol. The ideal state of the network policy is defined by the end-user, who configures the values that the performance variables and other parameters must have, such as the frequency with which these variables must be monitored.

Architecture and Construction of a Multi-agent System: The multi-agent system is composed of the “Agent Interface”, “Managing Agent”, “Information Agent” and “Trap Agent”. Figure 2 shows the system collaboration diagram where agents and messages are exchanged with their own performatives, and the action implemented in the ontology is represented.

Each of the agents in the system is described below:

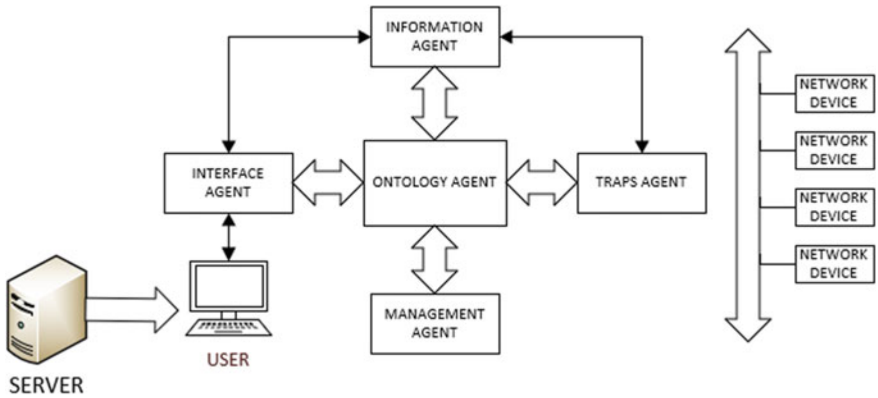


Fig. 2 Agent collaboration diagram

Interface Agent: This agent is dedicated to receiving the IP and OID addresses specified by the user, this information is necessary to know the device, in addition, this agent will present the information related to abnormal actions during the operation of the network, and this information is displayed as notifications and alarms.

- **Information Agent**: This agent is dedicated to interacting with the "Trap Agent", with the "Managing Agent" and with the SNMP API used for communication with the network devices. This agent acts as an intermediary between the multi-agent system and the monitored network.
- **Traps Agent**: This agent is dedicated to receiving SNMP traps that are sent from the "Information Agent" and stores the data (date, source IP address and OID-SNMP) in a database. Notifies the "Managing Agent" about the notification received.
- **Management Agent**: It is the intermediary between the "Interface Agent" and the "Information Agent". Receive the OID request and send it to the "Information Agent"; in addition, it receives the results and responds to the originating agent ("Agent Interface").
- **Ontology**: The ontology developed consists of concepts such as IP address, OID, value, description and the actions search, update and alarm.

The authors indicate that the execution of the system in a controlled environment gave the expected results in accordance with the limitations of the established policies and the defined thresholds. The effect of the execution of the multi-agent system is to monitor the network and be able to detect if the policy associated with the device is being violated.

Table 1 Comparison of related works from the perspective of network management

Research work	Area of application	ISO management model	Network management protocol	Agent architecture
Network management based on the multi-agent architecture [4]	Generic architecture	Configuration management–performance management	Communicative acts of the agents	Layer model (collaborative and reactive agents)
An improved network monitoring system that uses multi-agent-based technology [12]	Local area network	Fault management–security management–configuration management	Mobile agents and communicative acts of the agents	Hybrid (deliberative and reagents)
Network management using multi-agent systems [14]	Local area network	Fault management–performance management	Simple network management protocol	Collaborative agents

4 Discussion

Research projects related to network management using software agent engineering are currently in the early stages of research, as well as tools, methodologies and technical specifications. As observed in Table 1, most of these works focus on local area networks (LANs), so research projects that cover telecommunications networks, WAN networks or that meet all areas are almost nonexistent of application of the ISO management model. It is not expected that this issue will be developed widely until it is mature enough for the industry to market products based on this paradigm, which in theory has important advantages over traditional centralized management models.

5 Conclusions

Academic studies that relate network management and multi-agent systems cover only some specific aspects of management in small local area networks. Although the academic prototypes shown in this paper show that the implementation of network management systems using software agents is feasible, it is expected that a future will appear new innovative proposals in this regard, which provide new features for the solution of the problems that current systems are presenting due to the growth of the networks and consequently the complexity of their management.

There is no consensus and even less a standard, regarding the use of agent-oriented software development methodologies. Because of this, the development of agent-oriented systems has not been massively developed and driven in the creation of commercial solutions, as it does in the development of object-oriented software. We must bear in mind that the characteristics of the development of multi-agent systems are much more complex than the development of object-oriented since the agent must have the autonomy, social ability, communication with his environment, etc., in comparison with the object that it is simply reactive. There is still much to investigate on the multi-agent systems and network management, it is expected that new innovative proposals will appear in the future, which will provide new features for the solution of the problems that current systems are presenting due to the growth of the networks and consequently the complexity of its management.

References

1. Barba, A.: Gestión de Red. Alfaomega (2001)
2. Millán, R.: Mag Commun World **189**, 54 (2004)
3. Hodeghatta, U.: Challenges of implementing network management solution. Int. J. Distrib. Parallel Syst. (IJDPS) **2**(5), 67 (2011) (India)
4. Raya, K., Gaxiola, C., Castañón-Puga, M.: Network management based on multi-agent architecture (2014)
5. Arias, F., Zapata, G., Vásquez, L., Ovalle, D.: Homepage http://grupotyt.medellin.unal.edu.co/grupo-tyt/produccion-academica. Last accessed 01 Oct 2019
6. O'Hare, G., Jennings, N.: Foundations of Distributed Artificial Intelligence. Wiley (1996)
7. Russell, N.: Inteligencia Artificial. Pearson (2008)
8. IEEE Standard Glossary of Software Engineering Terminology. In: IEEE std 6. ISBN 155937067X. (1990)
9. Zambonelli, F., Omicini A.: Challenges and research directions in agent-oriented software engineering. Auton. Agents Multi-Agent Syst. 9 (2004)
10. Moreno, M., Rosete-Suárez, A., Simón-Cuevas, A.: Ingeniería de software orientada a agentes: roles y metodologías. Ingeniería Industrial. **XXVII**, 25–28 (2006)
11. Millán, R.: Magazine Windows NT 2000 Actual Nº 12, Prensa Técnica S.A. (1999)
12. Akinyokun, O.C., Ekuewa, J.B., Arekete, S.A.: Development of an agent-based system for monitoring software resources in a network environment: In: Artificial Intelligence Research, Sciedu Press, vol. 3, No. 3 (2017)
13. Akinyokun, O.C., Ekuewa, J.B., Arekete, S.A.: Development of agent-based system for monitoring software resources in a network environment. Artif. Intell. Res. **3**(3) (2014) (Sciedu Press)
14. Isaza, G., Mejía, M.H., Castillo, L., Morales, A., Duque, N.: Network management using multi-agents system. Distrib. Comput. Artif. Intell. J. (2012)

Reference Model to Identify the Maturity Level of Cyber Threat Intelligence on the Dark Web

Ricardo Meléndez Santos, Anthony Aguilar Gallardo, and Jimmy Armas Aguirre

Abstract In this article, we propose a reference model to identify the maturity level of the cyber intelligence threat process. This proposal considers the dark web as an important source of cyber threats causing a latent risk that organizations do not consider in their cybersecurity strategies. The proposed model aims to increase the maturity level of the process through a set of proposed controls according to the information found on the dark web. The model consists of three phases: (1) Identification of information assets using cyber threat intelligence tools. (2) Diagnosis of the exposure of information assets. (3) Proposal of controls according to the proposed categories and criteria. The validation of the proposal was carried out in an insurance institution in Lima, Peru, with data obtained by the institution. The measurement was made with artifacts that allowed to obtain an initial value of the current panorama of the company. Preliminary results showed 196 emails and passwords exposed on the dark web of which one corresponded to the technology manager of the company under evaluation. With this identification, it was diagnosed that the institution was at a "Normal" maturity level, and from the implementation of the proposed controls, the "Advanced" level was reached.

Keywords Cybersecurity · Cyber threat intelligence · Maturity model · Dark web

R. M. Santos · A. A. Gallardo (✉) · J. A. Aguirre
Universidad Peruana de Ciencias Aplicadas (UPC), Lima 15023, Perú
e-mail: u201210536@upc.edu.pe

R. M. Santos
e-mail: u201418245@upc.edu.pe

J. A. Aguirre
e-mail: jimmy.armas@upc.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_15

1 Introduction

The dark web is the hidden network where most stolen data is sold. There is a lot of information hosted here, but few companies know how to access this data, evaluate it and minimize the damage they might cause [1]. Cybercriminals are changing their approach to information trafficking because the risks are much lower compared to other crimes. When data is stolen, cybercriminals turn to the dark web and black markets to make a profit from "digital products," such as compromised login credentials of employees, stolen data, source code, credit cards, among others [2]. According to a recent report on April 2019 made by the renowned threat intelligence company Insights, there is a 129% increase in the first quartile of 2019 compared to the first quartile of 2018 in filtered credentials, likewise, an increase of 212% in credit cards on the dark web [2].

In 2018, about 547,000 customers' data from the British company Bupa were put up for sale on the black market, putting personal information at risk [3]. Failure to have adequate controls, this leak could lead to significant economic losses and possible reputational damage. Open-source intelligence (OSINT), or intelligence collected from publicly available sources, can offer significant value to the proactive cyber intelligence by alerting organizations about threats they were not previously aware of [4]. This document evaluates and proposes a reference model that identifies the level of maturity in the cyber threat intelligence process when considering the dark web as an important source of cyber threats.

2 State of the Art: Maturity Models

We review the previous literature to better understand the existing work and lay the foundations for our research. First, we review OSINT tools for the identification of information assets filtered on the dark web, and then, we review cybersecurity maturity models.

2.1 Tools for Identifying Filtered Assets

In this section, we review cyber intelligence threat tools to identify compromised information assets on the dark web. We take as main criteria: open-source, onion site data volume level, tool installation complexity level, and the available documentation. Table 1 outlines the tools needed for finding network accounts, private accounts, and email accounts filtered on dark web.

Table 2 outlines the tools that were reviewed to find vulnerabilities in information assets, both infrastructure, and web.

Table 3 outlines the tools needed for finding cards or bank accounts.

Table 1 Comparison between tools to find filtered accounts

Criteria	Tools to find filtered accounts		
	Pwndb [5]	Karma [6]	Torcrawl [7]
Open source	Yes	Yes	Yes
Onion site data volume level	High	Medium	High
Installation complexity level	Low	Medium	Low
Documentation	Yes	Yes	No

Table 2 Comparison between tools to find filtered vulnerabilities

Criteria	Tools to find filtered vulnerabilities		
	Photon [8]	Onionscan [9]	Freshonions [10]
Open source	Yes	Yes	No
Onion site data volume level	High	High	Medium
Installation complexity level	Low	Medium	Medium
Documentation	Yes	No	No

Table 3 Comparison between cybersecurity capacity maturity models

Criteria	Tools to find filtered vulnerabilities		
	The inspector [13]	Torbot [14]	Onionoff [15]
Open source	Yes	Yes	No
Onion site data volume level	High	Low	Medium
Installation complexity level	Low	Medium	Low
Documentation	Yes	No	Yes

2.2 Maturity Cybersecurity Models

The main components and structure of the maturity models that focus on cybersecurity were evaluated. According to this review, the maturity models considered are C2M2, SSE-CMM, CCSMM, NICE [11]. The FFIEC Cyber Assessment Tool (CAT) was also reviewed, which, despite not finding it in the review of the preceding literature, is consistent with the cybersecurity framework of the National Institute of Standards and Technology (NIST), as well as with cybersecurity practices accepted by industry [2]. Finally, we take into account COBIT 2019 because it considers cybersecurity as a priority area [12]. Table 4 shows a comparison between them.

Table 4 Comparison between cybersecurity capacity maturity models

Criteria	Maturity cybersecurity models					
	C2M2	NICE	CCSMM	SSE-CMM	CAT-FFIEC	COBIT 2019
Oriented to cybersecurity	Yes	Yes	Yes	No	Yes	No
Application sector	Energy	Labor force	Any	Communities	Finance	Any
Includes threats on the dark web	No	No	No	No	No	No
Documentation level for implementation	Medium	Medium	High	Medium	High	High
Level of complexity of the documentation	High	High	Medium	Medium	Low	Medium

3 Asset Discovery Information: Model Proposed Reference

3.1 Model Description

Following the frameworks with their maturity levels found in the literature review, we propose a reference model to identify the maturity level of cyber intelligence of threats on the dark web. The following stages describe the process by which the model operates. First, the OSINT tools identify if it is compromised or filtered information on the dark web. Then, the level of exposure of the filtered data is measured with tables proposed by the model. With this information, the maturity level is calculated using a formula that obtains measured data. Finally, security controls are proposed so that the institution can increase the level of maturity in the threat identification process, this time considering the dark web scenario. Figure 1 shows the stages of the proposed model.

3.2 Model Phases

Identification. In this stage, each of the assets which are going to be tracked and monitored on the dark web is identified. For the purpose of ordering the information with the artifacts presented by the model, three categories have been considered: carding, accounts, and vulnerability. The first category contains user bank information, for example, bank accounts and cards.

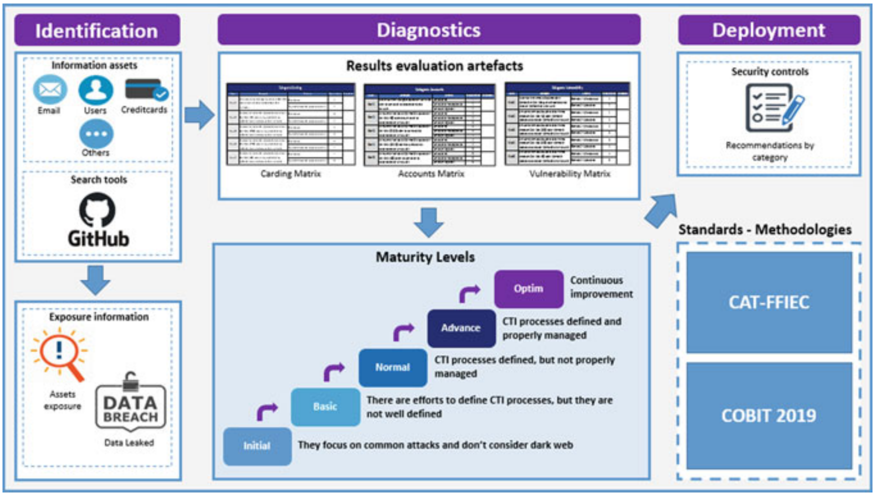


Fig. 1 Referential model design

The second one contains information that allows the identification of a person from the institution, whether internal or external, that is network accounts, administrative or privileged accounts, email accounts. Finally, the last category consists of the vulnerabilities found in the dark web with respect to the information assets of the institution.

Diagnostics. In this process, a measurement of the exposure level of the information assets committed on the dark web is made. To carry out this measurement, tables are used to allow obtaining a numerical value, to later calculate the maturity level.

Finally, by crossing the information obtained previously and with the support of a formula of the proposed model, the level of cyber intelligence maturity of the institution for the dark web scenario can be obtained. Figure 2 shows the maturity model that has been considered, and these are initial, basic, normal, advanced, and optimal.

- *Initial level*. At the initial maturity level, CTI processes are characterized by a tendency to focus only on common attacks and disregard the dark web. The sum of the numerical results that are obtained from the category tables is greater than 200.
- *Basic level*. At the level of basic maturity, there are efforts to define CTI processes of threats; however, these are not well defined in their entirety. The sum of the numerical results that are obtained from the category tables is in the range of 101–200.
- *Normal level*. At the normal maturity level, the CTI processes are already defined; however, they are not properly managed. The sum of the numerical results that are obtained from the category tables is in the range of 51–100.

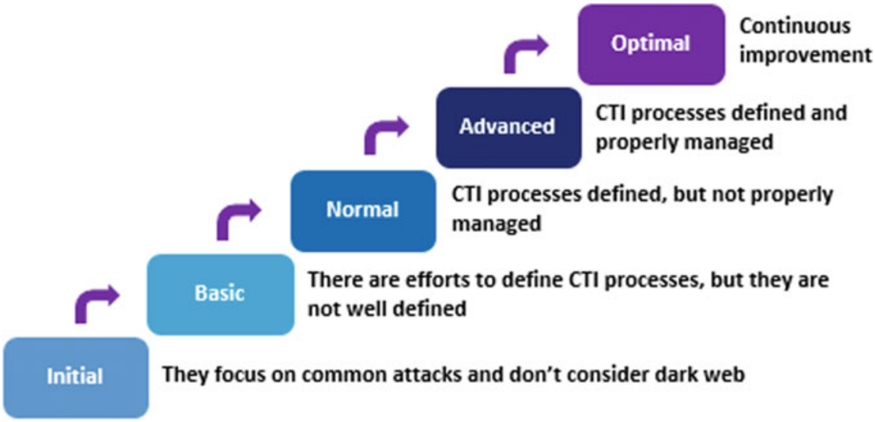


Fig. 2 Maturity levels of the proposed model

- *Advanced level*. At the advanced maturity level, CTI processes are not only defined but also maintain proper management. The sum of the numerical results that are obtained from the category tables is in the range of 10 to 50.
- *Optimal level*. At the optimum maturity level, CTI processes are very well defined and managed. In addition, they are in constant continuous improvement, and activities are carried out that are no longer manual, but almost automatic. The sum of the numerical results that are obtained from the category tables is less than 10.

Deployment. Finally, after obtaining the quantitative results of the diagnosis phase, controls are proposed that allow the institution to increase the current level of maturity. For this, Table 5 of security controls was developed in which the institution must implement according to the level at which it is located.

4 Case Study

4.1 Organization

This investigation was validated in a Peruvian insurance company. For reasons of confidentiality and security of company information, the name will not be disclosed in this document. The organization has approximately 3500 workers within all its areas. This Peruvian company indicated that it was at level 2 according to the FFIEC standard with respect to the security it presents in its operations and systems. The organization has an information security area that is working to raise one more level based on the controls provided by the regulations. They were consulted if within its controls the dark web was being considered, and the answer was that it was not considered.

Table 5 Security controls according to the maturity level

Security controls according to the maturity level		
Level	Category	Control
Initial	Carding	Set encryption method for fields in tables that have card information (scrumble)
		Sending timely emails to users/customers for the cancelation of their cards
	Accounts	Sending mail to the institution about the constant change of its password
		Have an information repository (at least an encrypted excel) for password management
		Establish password expiration periods for users in the institution
	Vulnerability	Perform a vulnerability scan using an annual ethical hacking
Basic	Carding	Set encryption methods for fields in tables that have card information (scrumble)
		Automation for sending timely emails to users/customers for the cancelation of their cards
	Accounts	Establish an awareness program for the institution regarding the correct use of their passwords
		Only the information security team should store privileged account credentials
		Establish password expiration periods for users in the institution
	Vulnerability	Perform a vulnerability scan using a semi-annual ethical hacking
Normal	Carding	Set encryption method for fields in tables that have card information (scrumble)
		The DBA must monitor the productive databases of the organization where card information is located
		Implement an array of roles at the level of roles and profiles in the applications where card information is displayed
	Accounts	Establish an exercise of social engineering on the proper management of passwords in the institution
		Store privileged account credentials in a secure environment
		Implement guidelines (policies and procedures) for password management in the institution
	Vulnerability	Establish an ethical hacking of the institution's SOX applications by the information security team
Advanced	Carding	Set at least MD5 encryption method for fields in tables that contains card information
		Establish a database monitoring for the audit of logs and actions in the database
		Implement an array of roles at the level of roles, profiles, and permission options in applications where card information is displayed

(continued)

Table 5 (continued)

Security controls according to the maturity level		
Level	Category	Control
	Accounts	Establish a data dictionary for the content of passwords, they should not have words that refer to the institution
		Implement monitoring of privileged accounts on all platforms on a monthly basis
	Vulnerability	Establish an ethical hacking of the institution's resources by the information security team
		Implement a vulnerability identification tool in dark web
		Implement alert management for cyber threats on dark web, identification of IOCs, etc.
Optim	Carding	Set AES-256 encryption method for fields in tables that have card information
		Establish automatic alerts immediately to the banking institutions for the cancelation of cards of the users/clients of the institution
		Implement a database monitoring tool for the audit of logs and actions in the database
		Compliance with the requirements of the normal PCI DSS
		Implement an identity access management (IAM) for the capture, registration, and management of user identities with their respective access and permissions on the institution's platforms
	Accounts	Establish a data dictionary for the content of passwords, they should not have words that refer to the institution
		Implement a privileged access management (PAM) system for logging the activity of privileged accounts
	Vulnerability	Establish an ethical hacking of the institution's resources with a specialized provider
		Implement a licensed tool for identifying vulnerabilities in dark web
		Implement a specialized SOC for timely warning and constant monitoring of cyber threats on dark web, identification of IOCs, etc.

A test was carried out on the exposure of the institution's assets on the dark web, and the results were corporate credentials in plain text. With this demonstration, they were receptive to evaluate our reference model in their threat identification process.

5 Validation

Gathering information. The institution provides us the information about the endpoint servers regards assets so that they can complete the information that serves as input in the search.

Search for assets on the dark web. With the information assets collected, a search is performed with the support of OSINT tools. This allows getting results on the level of exposure they present on the dark web. This information will subsequently serve as input for the information security team to take corrective actions regarding the results.

Tables 6 and 7 present the carding and vulnerability results, respectively. Hence, after analysis, no exposed information was found. Also, in Table 8, for the category of the account, 196 emails were found with their passwords exposed. The most critical issue of this finding was that the credential of the IT manager was encountered.

With this information, the maturity level of cyber intelligence of threats of the institution was calculated, and the results indicated the value 67 which refers to the normal level. The proposed controls of this level caused them to consider first the category of the account as the most critical one.

Table 6 Carding results

Carding category				
Level	Concept	Criteria	Weight of criteria	Total weight
1	No findings found on dark web about the accounts of the institution's collaborators	Other banks	1	1
		Corporate group bank cards	1	
2	Information assets exposed on dark web (<10) about the debit or credit cards of the clients/users of the institution are found	Other banks	4	
		Corporate group bank cards	5	
3	Information assets exposed on dark web (10–30) about the debit or credit cards of the clients/users of the institution are found	Other banks	5	
		Corporate group bank cards	6	
4	Information assets exposed on dark web (31–50) about the debit or credit cards of the clients/users of the institution are found	Other banks	6	
		Corporate group bank cards	7	
5	Information assets exposed on dark web (>50) about the debit or credit cards of the clients/users of the institution are found	Other banks	7	
		Corporate group bank cards	8	

Table 7 Vulnerability results

Vulnerability category				
Level	Concept	Criteria	Weight of criteria	Total weight
1	There are no topics or discussions about the information on dark web that are associated with the institution's information assets	Scenario 1—infrastructure	1	1
		Scenario 2—application	1	
2	Topics or discussions from the same people are found on dark web (<10) about the information associated with the institution's information assets	Scenario 1—Infrastructure	1	
		Scenario 2—application	2	
3	Topics or discussions from the same people are found on dark web (10–30) about the information associated with the institution's information assets	Scenario 1—infrastructure	2	
		Scenario 2—application	3	
4	Topics or discussions from the same people are found on dark web (31–50) about the information associated with the information assets of the institution	Scenario 1—Infrastructure	3	
		Scenario 2—Application	4	
5	Topics or discussions from the same people are found on dark web (>50) about the information associated with the institution's information assets	Scenario 1—Infrastructure	4	
		Scenario 2—Application	5	

6 Conclusions

We proposed a reference model to identify the maturity level of cyber threat intelligence. The information assets filtered on the dark web are identified to then diagnose the cyber intelligence threat process based on the exposure level of the committed assets. The model was tested in an institution of the insurance sector in Lima, Peru. The maturity level of cyber intelligence threats of the institution increased from the "Normal" level to the "Advanced" level. These results were obtained through the information provided by the categories tables of the proposed model. The information security area of the institution expressed its satisfaction by having as an additional scenario the dark web within its cybersecurity plan.

Table 8 Accounts results

Accounts category				
Level	Concept	Criteria	Weight of criteria	Total weight
1	No findings found on dark web about the accounts of the institution's collaborators	Network accounts	1	
		Corporate email accounts	1	
		Privileged accounts	1	
2	Information assets exposed on dark web (< 50) about the accounts of the institution's collaborators are found	Network accounts	2	
		Corporate email accounts	2	
		Privileged accounts	3	
3	Information assets exposed on dark web (51–100) about the accounts of the institution's collaborators are found	Network accounts	3	
		Corporate email accounts	3	
		Privileged accounts	4	
4	Information assets exposed on dark web (101–200) about the accounts of the institution's collaborators are found	Network accounts	4	4
		Corporate email accounts	4	
		Privileged accounts	5	
5	Information assets exposed on dark web (>200) about the accounts of the institution's collaborators are found	Network accounts	5	
		Corporate email accounts	5	
		Privileged accounts	6	

As future work, analytical techniques such as sentiment analysis and social network analysis can be used to identify key hackers on the dark web as well as the use of machine learning techniques to identify key features of these hackers. All of the above can greatly increase the maturity level of any organization.

References

1. Samtani, S., Chinn, R., Chen, H., Nunamaker Jr., J.: Exploring emerging hacker assets and key hackers for proactive cyber threat intelligence. J. Manage. Inf. Syst. **34**(4), 1023–1053 (2017)
2. Rosenberg, H.: Banking & Financial Services: Cyber Threat Landscape Report (2019)
3. Otte, J.: Bupa fined after personal data of 500,000 customers offered for sale on dark web The Independent (2018). https://www.independent.co.uk/news/business/news/bupa-data-breach-dark-web-fine-customers-insurance-a8560071.html. Last accessed 28 Oct 2019
4. Bromiley, M.: Threat intelligence: what it is, and how to use it effectively. SANS Institute InfoSec Reading Room, **15**

5. Davidtavarez/Pwndb. (2019). https://github.com/davidtavarez/pwndb. Last accessed 28 Oct 2019
6. Limkokhole/Karma: (2019). https://github.com/limkokhole/karma. Last accessed 28 Oct 2019
7. MikeMeliz/TorCrawl.py. (2019). https://github.com/MikeMeliz/TorCrawl.py. Last accessed 28 Oct 2019
8. s0md3v/Photon: (2019). https://github.com/s0md3v/Photon. Last accessed 28 Oct 2019
9. s-rah/onionscan.: (2019). From https://github.com/s-rah/onionscan. Last accessed 28 Oct 2019
10. Dirtyfilthy/Freshonions-torscraper.: (2019). https://github.com/dirtyfilthy/freshonions-torscraper. Last accessed 28 Oct 2019
11. Rea-Gauman, A.M., San Feliu, T., Calvo-Manzano, J.A. Sanchez-garcia, I.D.: Comparative study of cybersecurity capability maturity models. In: International Conference on Software Process Improvement and Capability Determination (2017)
12. Information Systems Audit and Control Association (U.S.). (2019). Cobit 2019: Introduction and Methodology
13. Moham3dRiahi/Th3inspector. (2019). https://github.com/Moham3dRiahi/Th3inspector. Last accessed 28 Oct 2019
14. DedSecInside/TorBot (2019). https://github.com/DedSecInside/TorBot. Last accessed 28 Oct 2019
15. k4m4/onioff. (2019). https://github.com/k4m4/onioff. Last accessed 28 Oct 2019

Critical Event Manager Module for IP Networks Using Software Agents

Juan Carlos Zegarra-Vásquez and Nestor Mamani-Macedo

Abstract In the management of IP networks, the collection and classification of notifications is a critical task. In addition to unexpected events, such as hardware and software failures, the collected notifications are important to make the proper diagnosis of the failure that occurred and based on that, propose and apply efficient solutions to the problems presented. This work develops a critical event manager using the "Ingenias" methodology for the analysis and design of the prototype. The developed software agent has the characteristic of being proactive, for which it permanently listens to notifications from a network of IP routers and interacting with it to mitigate problems and propose solutions to network failures.

Keywords Software agents · Multi-agent systems · IP network · Management of IP networks · IP network failures

1 Introduction

Network management deals with aspects of planning, organization, supervision, and control of communications elements to guarantee an appropriate level of service according to a specific cost [1]. The liberalization of the old telecommunications monopolies, at the end of the 80s, supposed serious changes in network management, making it necessary, among other things: to control an increasing number of equipment and technologies from different providers and introduce new services gradually [2]. The efficiency of network administrators to manage networks decreases

J. C. Zegarra-Vásquez
Facultad de Ingeniería Electrónica y Eléctrica, Universidad Nacional Mayor de San Marcos, Lima 01, Peru
e-mail: juan.zegarra@ictiperu.com

N. Mamani-Macedo (✉)
Facultad de Ciencias Matemáticas, Universidad Nacional Mayor de San Marcos, Lima 01, Peru
e-mail: nmamanim@unmsm.edu.pe

Facultad de Ingeniería, Universidad Peruana de Ciencias Aplciadas, Lima 01, Peru

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_16

as they become more complex and heterogeneous. Managing large networks created a crisis for many organizations [3]. So, it is not enough to know the faults that may be occurring in the network at a given time, but it is essential to be able to anticipate them and know how to react in case they occur. In that sense, intelligent agents are an adequate tool for self-management in the telecommunications area, specifically autonomous networks. The problem observed in some Network Operation Center (NOC) is the large volume of notifications from the growth of communications networks. In general, network systems and devices generate many notifications. Being able to properly identify and prioritize notifications is a laborious task because the review must be carried out manually [4] by the network operator, which results in the work identifying the source of the fault being slow, unreliable, and inaccurate. How to deal with the problem of unavailability of communications network services intelligently? We consider that among the artificial intelligence techniques, the approach of software agents would be appropriate to develop an intelligent event management system with the capacity to be proactive and effective; to detect and diagnose network problems [5] to support the tasks of the NOC operators.

2 Theoretical Framework

2.1 Agent Concept

In the literature that investigates the field of agents, a weak and not very controversial notion and a second strong controversial notion are distinguished:

The weak notion of an agent: It is a hardware- or computer-based software system that includes the following properties [6]:

- **Autonomy**: Agents operate without the direct intervention of humans or others and have some kind of control over their actions and their internal state.
- **Social ability**: Agents interact with other agents through some kind of agent communication language.
- **Reactivity**: Agents perceive their environment and respond in a timely manner to the changes that occur in it.
- **Proactivity**: Agents do not simply act in response to their environment; they exhibit behavior directed toward goals taking the initiative.

A strong notion of an agent: An agent has the properties of a weak agent, but it is also conceptualized and implemented using concepts that are most commonly applied to humans. For example, it is very common in artificial intelligence to characterize an agent using mental notions, such as knowledge, desires, even some have gone further and considered emotional agents. The basis of the strong notion of an agent is intentionality [7].

2.2 The Social Organization of the Agents

It is the way agents are organized at a given time. Agents make use of social skills and interactive behavior to communicate, cooperate, coordinate, and negotiate with each other. In that sense, agents can be organized in the following ways:

- **Coordination**: It refers to the distribution of tasks in order to achieve common objectives [8].
- **Cooperation**: Through mutual interaction, agents solve tasks jointly or maximize their usefulness.
- **Communication**: Allows synchronizing actions, sending, and receiving knowledge, solving conflicts in order to overcome a problem.

2.3 Tools, Protocols, and Methodologies

The development of this work involved the use of:

- **Ingenias**: It is a methodology and a set of tools for the development of multi-agent systems (MAS) [9], considering this methodology five complementary points of view: organization, agent, tasks and goals, interactions and environment. "Ingenias Development Kit" [10] is a development tool written in Java language that supports the "Ingenias" methodology.
- **JADE (Java Agent Development Framework)**: [11] Is a framework that simplifies the implementation of a MAS through middleware and a graphical interface and a set of tools that support the debugging and deployment phases.
- **The Syslog protocol**: Initially defined in RFC 3164 [12] and later in RFC 5424 [13], Syslog is a service used to send event notifications on IP networks.
- **Esper**: It is an engine written in Java language that allows us to process events or messages in real time and obtain patterns from the events [14]. "Esper" allows us to process complex events and analyze the flow of these events.

3 Background

In recent years, works that integrate network management with intelligent agents in different fields have been proposed. Each work addresses a particular problem of a part of the reality of management, among the main works we have:

- **An improved network monitoring system that uses multi-agent-based technology**: [15] It is based on multiple agents to monitor nodes in a network environment. The models developed for the proposed system define certain deliberative agents that interact with each other to achieve the objectives and requirements of the multi-agent organization.

- **Network management based on multi-agent architecture**: [16] It defines an architecture for autonomous network management by designing a collaborative method based on the task-oriented domain, implementing a negotiation mechanism between agents to improve the performance of network management through a multi-agent hierarchical architecture.
- **Network management using multi-agent systems**: [17] A multi-agent system has the property of being adaptive and has knowledge and skills to make decisions according to the real network state that is represented in the information base for management (MIB: Management Information Base) of devices that use the SNMP protocol.

4 Methodology and Solution Developed

The occurrence of one or more failures in communications protocols, links of the network or hardware of the equipment will cause the routers to notify the critical event manager. Depending on the nature of the failure, the manager will take a passive (inform and indicate the possible resolution of the fault) or active (mitigate the impact of the network failure). The work carried out by the NOC operators and the critical event manager for network fault management can be organized through processes and activities, which are explained in Table 1.

Only some meta-models of the collecting and reasoning agents that implement the proposed characteristics for the system will be described.

Table 1 Management processes and activities of the critical event management module

Agent	Management process	Activity
Collector	Network event collection	Collection of events through notifications from the managed network
Collector	Network event collection	Filtering events. Only relevant events are analyzed, leaving aside those events that are not important
Collector	Network event collection	They are classified as events according to their nature. Classification based on protocol failure, physical link failure, hardware, etc.
Expert	Inference engine	Application of logical rules in a knowledge base to deduce new information
Expert	Inference engine	Detection of causality relationships between events (cause and effect)
Expert	Mitigate negative impact events in the network	Execution of actions on the managed network, in order to mitigate that event that has a negative impact on the network

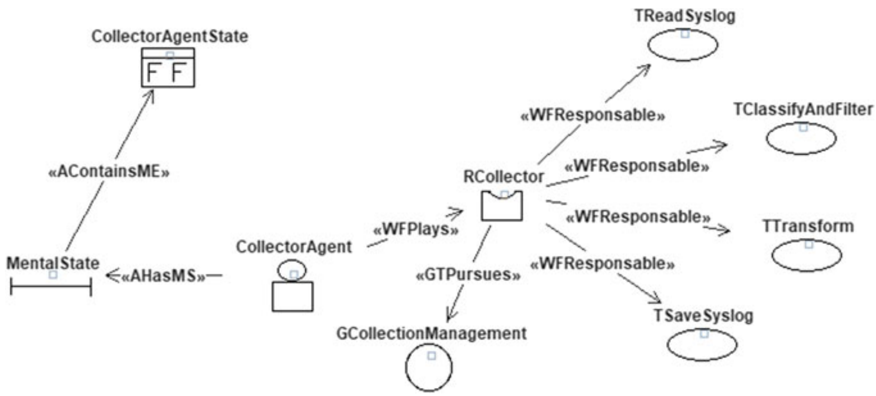


Fig. 1 Agent meta-model: event collector

Meta-model agent of the "Collecting Agent": Manages the collection of notifications obtained from the routers (Syslog). The agent verifies the format of the notifications, filters them, transforms them, and sends them to the "ExpertAgent." Figure 1 shows the "CollectorAgent" model.

The "CollectorAgent" aims to manage event collection or "GCollectionManagement." For this, it plays the role of collector or "RCollector," and in that role it executes the tasks of reading, filtering, classifying, transforming, and storing notifications from the network (TReadSyslog, TClassifyAndFilter, TTransform, TSaveSyslog).

Meta-model goal/task of the "CollectorAgent": The "TReadSyslog" task is one of the tasks that allow fulfilling the objective "GCollectionManagement." "TReadSyslog" reads an event or set of notifications from the network (AEventLogSlot). The data read is made up of Syslog notifications. The data obtained is stored temporarily in a slot (CollectedSyslog) of the "Collector Agent." The task "TClassifyAndFilter" is one of the tasks that also allow fulfilling the objective "GCollectionManagement." "TClassifyAndFilter" reads an event (CollectedSyslog) of the "TReadSyslog" task. The data is classified and filtered as specified in the detailed requirements, then stored temporarily in a slot (SyslogFiltered) of the "Collector Agent." The "TTransform" task is one of the tasks that allow the objective "GCollectionManagement" to be fulfilled. "TTransform" reads an event (SyslogFiltered) of the "TClassifyAndFilter" task. The data is formatted and adequate and then stored temporarily in a slot (SyslogTransformed) of the "Collector Agent." The task "TSaveSyslog" is one of the tasks that also allow fulfilling the objective "GCollectionManagement." "TSaveSyslog" reads an event (SyslogTransformed) from the "TTransform" task. Finally, the data is stored in a database.

Agent meta-model of the "ExpertAgent": The "ExpertAgent" manages the reasoning of the event management module based on the filtered and classified events received from the collection management. Based on previously established patterns, the agent performs reasoning to obtain knowledge of what is happening

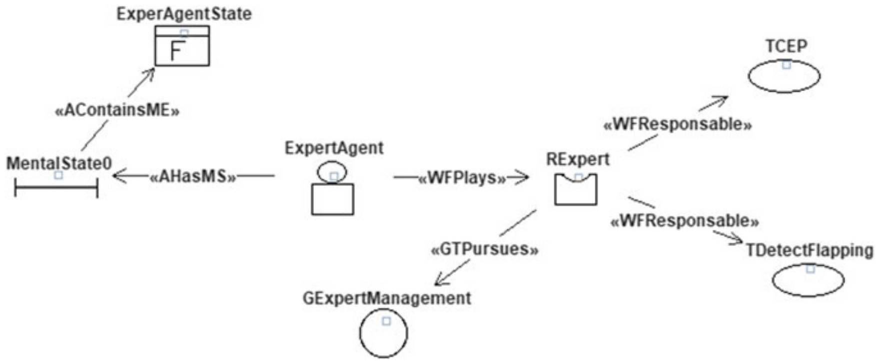


Fig. 2 Agent meta-model: expert

on the network and acts on the network by switching off the network interface that causes flapping. In Fig. 2, the reasoning agent model is shown.

Meta-model goal/task of the "ExpertAgent": The task "TCEP" is one of the tasks that allow fulfilling the objective "GExpertManagement." "TCEP" reads a filtered event (SyslogTransformed) of the "CollectorAgent." The data read is made up of critical event logs, which are processed by the "ComplexEventProcessing" engine.

The task "TDetectFlapping" is one of the tasks that allow fulfilling the objective "GExpertManagement." "TDetectFlapping" reads a fact (SyslogTransformed) from the "CollectorAgent" and as a protection measure, the event manager module executes the shutdown action of the network interface that is causing the failure. The task produces action on an application script called "expect," giving it the data so that the application itself turns off the intermittent network interface. The result of the event correlation work and the action is stored in a register called "PerformedActions."

5 Results and Discussion

The tests of the critical event manager module are divided into two types:

1. **Detection of a relevant event among many unimportant events**: They consist of sending a hundred events (noise) and an important event inserted randomly. The tests were performed with each of the following events:

    a. LINK-3-UPDOWN,
    b. LINEPROTO-5-UPDOWN,
    c. LINK-5-CHANGED,
    d. OSPF-5-ADJCHG,
    e. OSPF-4-CONFLICTING_LSAID,

f. SFF8472-5-THRESHOLD_VIOLATION,
g. OSPF-4-ERRRCV,
h. PLATFORM-ASM-0-FAULT_RECOVERY,
i. PLATFORM-ASM-3- MALLOC_FAILED,
j. MALLOC_FAILED,
k. PLATFORM-ENVMON-0-CARD,
l. PLATFORM-ENVMON-2-CHASSIS_OVERTEMP,
m. PLATFORM-ENVMON-2-FAN,
n. PLATFORM-ENVMON-2-FAN,
o. LINK-3-LINK_FAULT,
p. LINK-3-FCS_ERROR,
q. LINK-3-TOOSMALL,
r. CDP-4-NATIVE_VLAN_MISMATCH,
s. EQUIPMENT-5-DOWN,
t. EQUIPMENT-5-UP.

The critical event manager was able to identify and extract the important event of the "noise" and alert that the event occurred and the actions to be taken to solve the problem.

2. **Pattern detection among many unimportant events:** They consist of sending one hundred events (noise) and a small number of important events (LINEPROTO-5-UPDOWN and OSPF-5-ADJCHG) inserted in a non-consecutive manner for detection of "flapping" in the network, so that four consecutive intermittencies in one minute determine the "flapping." The critical event manager was able to identify and extract the temporal sequence of "noise" events, determine the "flapping" in the network, alert the event that occurred, and turn off the interface in "flapping" to mitigate the problem.

6 Conclusions

This paper has covered the issue of fault management in IP communication networks from the perspective of software agents. From this work, it has been proposed that the development of network management systems using the advantages provided by software agent engineering is feasible. In addition, in the academic field, there are several methodologies for the design of systems using agents of software; in this particular case, it has been verified that it is valid to develop this type of application using the "*Ingenias*" methodology. The critical event manager detected, alerted, and took action almost in real time. The length of this process was shorter than the work done by a human being, which has a positive impact on reducing the time to detect and correct failures in IP networks.

References

1. Barba, A.: Gestión de Red. Alfaomega (2001)
2. Millán, R: Communications World Nº 189, IDG Communications S.A., pp. 54, (2004)
3. Hodeghatta U.: Challenges of Implementing Network Management Solution. Int. J. Distrib. Parallel Syst. (IJDPS) **2**(5), 67 (2011) (India)
4. Cuesta, A.: Sistema de monitorización y correlación de eventos en gestión de redes. Departamento de Ingeniería Telemática. Escuela Técnica Superior de Ingeniería Universidad de Sevilla, pp. 2 (2016)
5. Cisco Systems.: Homepage https://www.cisco.com/c/en/us/support/docs/availability/high-availability/15114-NMS-bestpractice.html. Last accessed 01 Oct 2019
6. Wooldridge, M., Jennings, N.R.: Intelligent agents: theory and practice. Knowledge Eng. Rev. **10**(2), (1995)
7. Dignum, V., Dignum, F.: Homepage http://www.cs.uu.nl/research/techreps/UU-CS-2005–015.html. Task and Social Coordination in Agent Organization. Last accessed 03 March 2019
8. Pavón, J., Gómez-Sanz, J.: Agent-oriented software engineering with "Ingenias". Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science), pp. 394–403. Czech Republic (2003)
9. Ingenias. Homepage http://ingenias.sourceforge.net/. Last accessed 11 Oct 2019
10. Gómez-Sanz, J., Fuentes-Fernández, R., Pavón, J. García-Magariño, I.: Ingenias development kit: a visual multi-agent system development environment, pp. 1675–1676. Estoril, Portugal, (2008)
11. Telecom Italia Lab. Homepage http://jade.tilab.com/. Last accessed 01 Oct 2019
12. Esper. Homepage http://www.espertech.com/esper/esper-downloads/. Last accessed 11 Oct 2019
13. RFC 3164. Homepage https://tools.ietf.org/html/rfc3164. Last accessed 11 10 2019
14. RFC 5424. Homepage https://tools.ietf.org/html/rfc5424. Last accessed 11 Oct 2019
15. Akinyokun, O.C., Ekuewa, J.B., Arekete, S.A.: Development of an agent-based system for monitoring software resources in a network environment. Artif. Intell. Res. 3(3) (2017) (Sciedu Press)
16. Raya, K., Gaxiola, C., Castañón-Puga, M.: Network management based on multi-agent architecture (2014)
17. Isaza, G., Mejía, M.H., Castillo L., Morales A., Duque N.: Network management using multi-agents system. Distrib. Comput. Artif. Intell. J. (2012)

A Multispectral Image Compression Algorithm for Small Satellites Based on Wavelet Subband Coding

Joel Telles and Guillermo Kemper

Abstract This article proposes a lossy compression algorithm and scalable multispectral image coding—including blue, green, red, and near-infrared wavelengths—aimed at increasing image quality based on the amount of data received. The algorithm is based on wavelet subband coding and quantization, predictive multispectral image coding at different wavelengths, and the Huffman coding. The methodology was selected due to small satellites' low data rate and a brief line of sight to earth stations. The test image database was made from the PeruSat-1 and LANDSAT 8 satellites in order to have different spatial resolutions. The proposed method was compared with the SPIHT, EZW, and STW techniques and subsequently submitted to a peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) evaluation; it showed better efficiency and reached compression ratios of 20, with a PSNR of 30 and an SSIM of approximately 0.8, depending on the multispectral image wavelength.

Keywords Image compression · Wavelet transform · Small satellites · Entropy coding

1 Introduction

Currently, there are various space programs encouraging the creation of small satellites. These programs are aimed at training human resources, conducting experiments and acquiring images of the Earth. Some international programs promoting the development of small satellites are the American Institute of Aeronautics and Astronautics

J. Telles (✉)
Peruvian National Institute of Research and Training in Telecommunications (INICTEL-UNI), Peruvian University of Engineering, Av. San Luis 1771, Lima, Peru
e-mail: jtelles@inictel-uni.edu.pe

G. Kemper
Peruvian University of Applied Sciences, Av. Prolongación Primavera 2390, Santiago de Surco, Lima, Peru
e-mail: guillermo.kemper@gmail.com

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_17

(AIAA) and the American Society of Astronautics (AAS) through the Texas CanSat Competition, which is a collaborative effort between students and faculty at universities and high schools and high power rocketry enthusiasts to build, launch, test, and recover prototypes of miniaturized satellites. In Europe, the Competition is promoted by the European Space Agency (ESA); e.g., in Norway, it is organized by the Norwegian Center for Space-Related Education (NAROM), and, in France, by the National Center for Space Studies (CNES). In Japan, the Competition is organized by the University Space Engineering Consortium (UNISEC) [1].

There are mainly two limitations to the development of the image acquisition module in small satellites: (1) they have a line of sight to earth stations for a short period of time; (2) they have a low data rate.

The previous work has developed image acquisition modules by using JPEG image encoders with data rates of 1200 bits/s [2–4]. One study proposed a lossless thermal image compression method for nanosatellites using predictive techniques, the discrete wavelet transform (DWT), and a fixed-length entropy encoder. However, tests were performed on grayscale non-satellite images with a 512 $\times$ 512 resolution and compression ratio of 1.42–1.57 [5]. In [6], after a comparison between the Huffman, DWT, and fractal compression techniques, it was concluded that the fractal encoder showed better quality according to the peak signal-to-noise ratio (PSNR), followed by the DWT and Huffman encoders. The study was performed on six non-satellite images using PSNR as metrics. The study in [7] used the 9/7 biorthogonal wavelet to compare EZW, JPEG, and JPEG 2000 encoders on Lena and Barbara grayscale test images with a 512 $\times$ 512 resolution (8 bpp). When comparing the EZW encoder and the JPEG encoder, results showed that EZW performed significantly better according to the PSNR evaluation for the same number of bits per pixel (bpp). When compared with the JPEG 2000 encoder, they showed similar performance. In the same work, another comparative study was conducted on Lena, Barbara, and Goldhill grayscale test images with a 512 $\times$ 512 resolution, using a 9.7 biorthogonal wavelet with six levels of decomposition. The results of the comparisons made between the SPIHT, EZW, and JPEG 2000 methods showed that SPIHT had a higher compression and quality rate when compared to EZW and JPEG2000.

This article proposes a lossy compression algorithm and scalable multispectral image coding—including blue, green, red, and near-infrared wavelengths—aimed at increasing image quality based on the amount of data received. The algorithm is based on wavelet subband coding and quantization, predictive coding of multispectral images at different wavelengths, and the Huffman coding. The test image database was made from the PeruSat-1 and LANDSAT 8 satellites in order to have different spatial resolutions. The proposed method was compared to the SPIHT [8], EZW [9], and STW [10] techniques.

The article is organized as follows: Sect. 2 details the procedure of the proposed algorithm; Sect. 3 shows the results obtained, and Sect. 4 contains the discussion and conclusions.

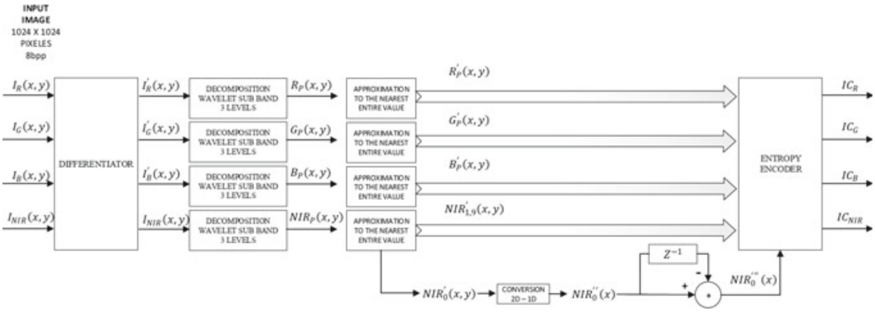


Fig. 1 Block diagram of the proposed encoder

2 Proposed Algorithm

2.1 *Block Diagram*

The block diagram of the proposed algorithm is shown in Fig. 1. A description of the processing steps can be found below.

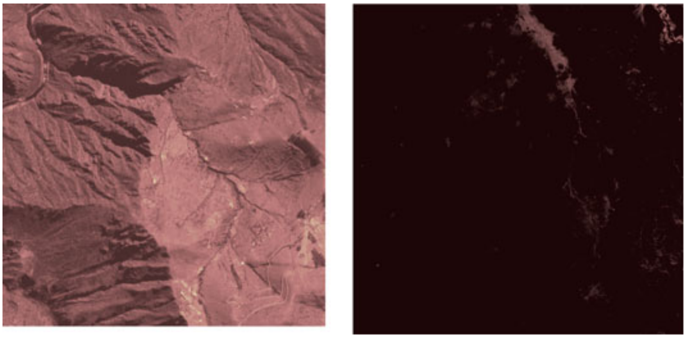
2.2 *Experimental Data—Input Images*

PeruSat-1 satellite images have four bands: red (0.63–0.7 µm), green (0.53–0.59 µm), blue (0.45–0.50 µm), and near-infrared or NIR (0.752–0.885 µm). The spatial resolution of the multispectral bands is 2.8 m per pixel. 50 PeruSat-1 satellite images from different geographical areas (e.g., tropical rainforest, desert, agricultural areas, urban areas, and images with clouds) were used. The resolution of the test images is 1024 × 1024 pixels at 8 bpp [11].

LANDSAT 8 satellite images have four bands: red (0.64–0.67 µm), green (0.53–0.59 µm), blue (0.45–0.51 µm), and near-infrared or NIR (0.85–0.88 µm). The spatial resolution of the multispectral bands is 30 m per pixel. 50 LANDSAT 8 satellite images from different geographical areas were used. The resolution of the test images is 1024 × 1024 pixels at 8 bpp [12].

2.3 *Differentiation*

Image $I_{\mathrm{NIR}}(x, y)$ was taken as a reference because it has a higher pixel value compared to the R, G, and B input images. This block differentiates image $I_{\mathrm{NIR}}(x, y, n)$ from the other images, as shown in (1). This procedure allows for reducing the entropy of the difference images. Figure 2a shows a PeruSat-1 image



(a) PerúSat-1 Image (b) Difference Image

Fig. 2 **a** PerúSat-1 image **b** difference image

in its blue component and respective entropy value of 6.54, and Fig. 2b shows the difference image and its entropy value of 0.61.

$$
\begin{aligned}
I'_R(x, y) &= I_{\mathrm{NIR}}(x, y) - I_R(x, y) \\
I'_G(x, y) &= I_{\mathrm{NIR}}(x, y) - I_G(x, y) \\
I'_B(x, y) &= I_{\mathrm{NIR}}(x, y) - I_B(x, y)
\end{aligned}
\tag{1}
$$

2.4 *Wavelet Subband Decomposition*

The wavelet families that were used in the study are shown in Table 1. To define the wavelet that was to be used, the test satellite images were evaluated and their entropy was measured after each wavelet subband decomposition. The wavelet with the lowest average entropy was the 5.5 biorthogonal. The results are shown in Fig. 3.

After the differentiation stage, images $I'_R(x, y)$, $I'_G(x, y)$, $I'_B(x, y)$, and $I_{\mathrm{NIR}}(x, y)$ were submitted to the wavelet decomposition tree (Fig. 4) for subband generation. The periodic convolution method was used on the input image [13] so that convolution operations did not produce larger sizes than expected; thus, wavelet coefficients did not increase with convolution.

Table 1 Wavelet family

Daubechies	DB1, DB2, DB3, DB10, DB20
Coiflets	Coiflet 1, Coiflet 2, Coiflet 3, Coiflet 4, Coiflet 5
Biorthogonal	Bior1.1, Bior1.3, Bior1.5, Bior2.2, Bior2.4, Bior2.6, Bior2.8, Bior3.1, Bior3.3, Bior3.5, Bior3.7, Bior3.9, Bior4.4, Bior5.5, Bior6.8

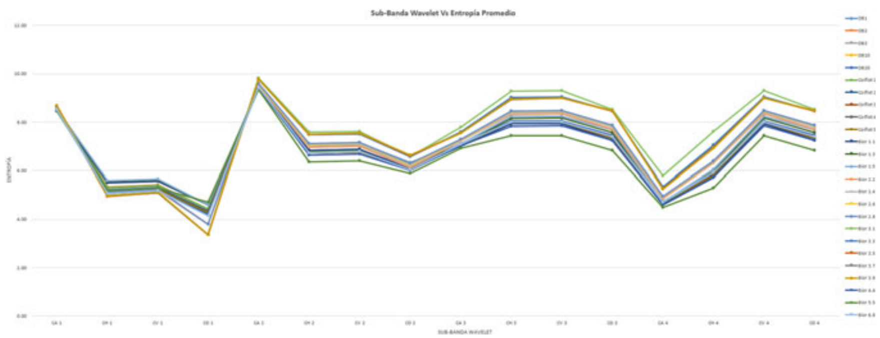


Fig. 3 Entropy evaluation of wavelet families on satellite images

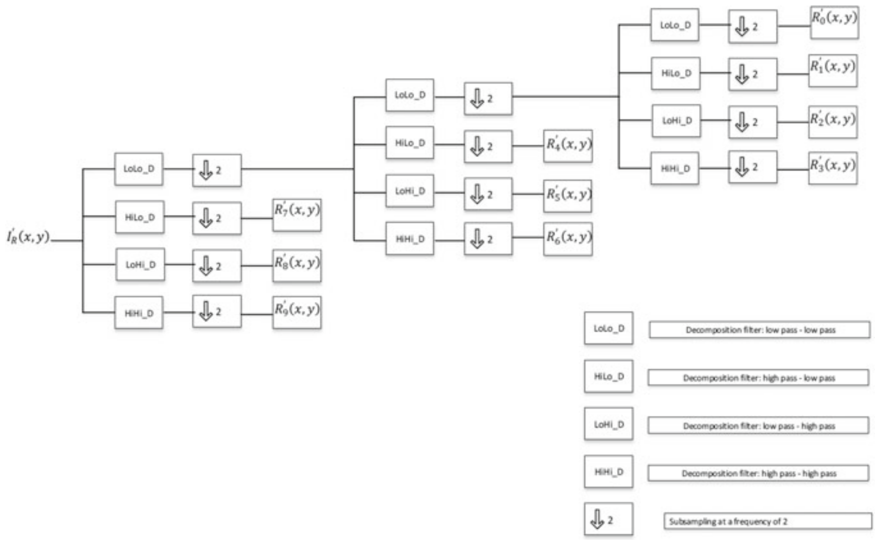


Fig. 4 Level-3 wavelet subband decomposition tree

2.5 Rounding to the Nearest Integer

Wavelet coefficients $R_P(x, y)$, $G_P(x, y)$, $B_P(x, y)$, and $\mathrm{NIR}_P(x, y)$ were divided by ten and rounded to the nearest integer, as shown in (2). It should be noted that the operation is reversed in the reconstruction stage.

$$R'_P(x, y) = \mathrm{Round}\left(\frac{R_P(x, y)}{10}\right)$$

$$G'_P(x, y) = \mathrm{Round}\left(\frac{G_P(x, y)}{10}\right)$$

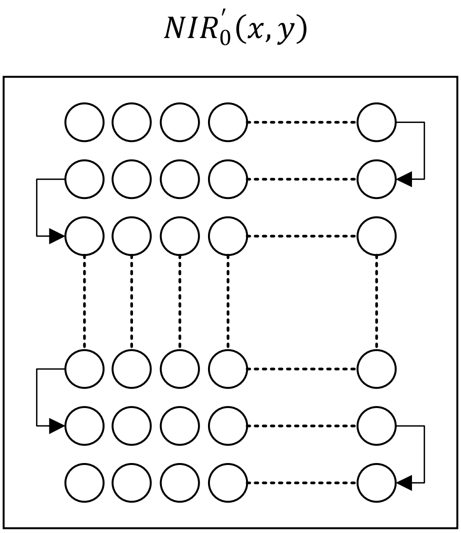


Fig. 5 2D-1D conversion

$$
\begin{aligned}
B'_P(x, y) &= \text{Round}\left(\frac{B_P(x, y)}{10}\right) \\
\text{NIR}'_P(x, y) &= \text{Round}\left(\frac{\text{NIR}_P(x, y)}{10}\right)
\end{aligned} \tag{2}
$$

2.6 2D-1D Conversion

The $\text{NIR}'_0(x, y)$ coefficients, i.e., the rounding coefficients, were converted to vector, as shown in Fig. 5. Subsequently, the signal passed through a retarder and the subtraction was performed, as shown in Fig. 1.

2.7 Entropy Coding

Entropy coding exploits the statistical redundancy of signals $R'_P(x, y)$, $G'_P(x, y)$, $B'_P(x, y)$, and $\text{NIR}'_P(x, y)$; the values of these coefficients are represented by variable size codes generated by the Huffman method. The codes are assigned from a table generated from the input image probability distribution. In this case, the tables were generated from the Laplacian probability distribution, as shown in (3). This distribution is the closest to the subband histograms. A variance of 3.8 was used.

$$L_X(x) = \frac{1}{\sqrt{2\sigma_X}} e^{\left[\frac{-\sqrt{2}|x|}{\sigma_X}\right]} \tag{3}$$

3 Results

3.1 Metrics for Image Quality Measurement

Digital images are prone to distortions during lossy compression. It is, therefore, necessary to use quality assessment metrics. The metrics used in this work were peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

3.1.1 Structural Similarity Index (SSIM)

The structural similarity index (SSIM) is used to compare local patterns of pixel intensities; that is, it quantifies the visibility of errors between a reference image and a distorted image. This index attempts to reproduce the human visual system [14].

3.1.2 Peak Signal-to-Noise Ratio (PSNR)

The peak signal-to-noise ratio (PSNR) is shown in (4), where *max* is the maximum value of a pixel in the image, while M and N are the numbers of rows and columns in the image. A high PSNR value points to small distortions. This value is related to the mean square error (MSE), which is calculated by averaging the squared intensity differences of distorted and reference images. PSNR does not always correlate with the perceived visual quality.

$$\begin{aligned} \text{MSE} &= \frac{1}{M * N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(I_{\text{original}}(x, y) - I_{\text{distorsion}}(x, y)\right)^2 \\ \text{PSNR} &= 10 \log_{10}\left(\frac{\max^2}{\text{MSE}}\right) \end{aligned} \tag{4}$$

3.2 Compression Ratio Measurement

The compression ratio is the division of the size of the original or uncompressed image by the size of the compressed image, as shown in (5).

$$\mathrm{FC} = \frac{\text{Size of original image}}{\text{Size of compressed image}} \tag{5}$$

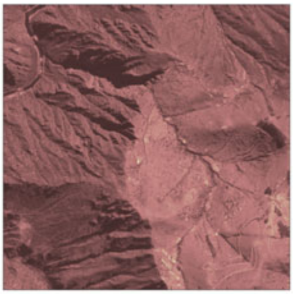
3.3 Comparison of Results

A comparison was performed between the results of the SPIHT [8], EZW [9], and STW [10] techniques and those of the proposed algorithm on the PeruSat-1 images. The proposed algorithm proved to be better than the SPIHT, EZW, and STW techniques for compression ratios of approximately 4–20. A visual comparison is shown in Fig. 6.

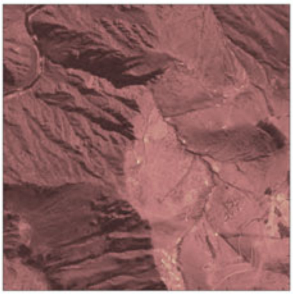
Subsequently, another comparison was made between the results of the SPIHT [8], EZW [9], and STW [10] techniques and those of the proposed algorithm on the LANDSAT 8 images. Once again, the proposed algorithm proved to be better than the SPIHT, EZW, and STW techniques for compression ratios of approximately 2–15, and significantly more efficient when using the SSIM metrics.

4 Discussion

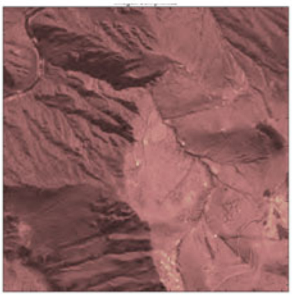
In the proposed algorithm, the NIR image is transmitted with the highest possible quality, which is why its compression ratio is in the order of 2–5. However, when the algorithm is applied to the R, G, and B images, it achieves a high compression ratio and better quality, as evaluated by PSNR and SSIM. It was critical to perform a quality assessment using metrics to evaluate different types of distortions. PSNR is used in the previous works as the only image quality measurement value; however, as shown in Fig. 6, sometimes it does not perfectly correlate with the human perception of images. For this reason, the SSIM metrics proves very useful. On the other hand, biorthogonal wavelet families have a linear phase, which is an advantage. It is an important property in digital image processing, as it allows us to obtain a reconstructed image with less distortion.



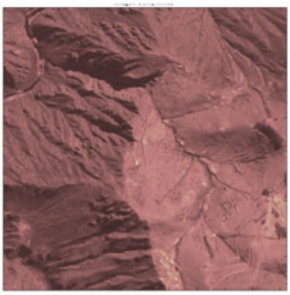
(a) Original Image



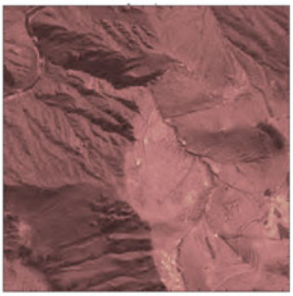
(b) EZW: TC=20, PSNR 21.30, SSIM 0.55



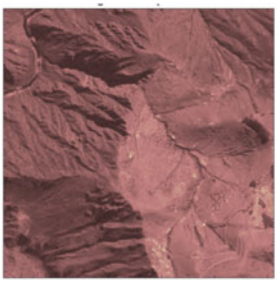
(c) EZW: TC=64, PSNR 18, SSIM 0.32



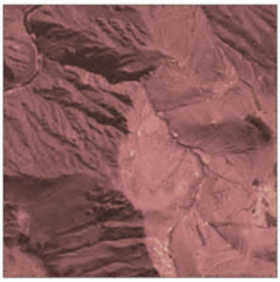
(d) SPIHT: TC=28, PSNR 20, SSIM 0.53



(e) SPIHT: TC=88 PSNR 17 SSIM 0.31

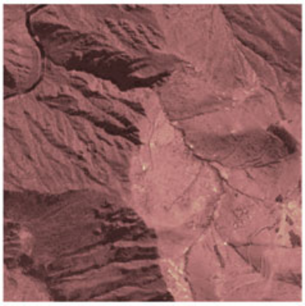


(f) STW:TC=19,PSNR21, SSIM0.57



(g) STW TC=62 PSNR 18 SSIM 0.33

Fig. 6 Image comparison


(h) Proposed Method: TC=21 PSNR 31.6 SSIM 0.81

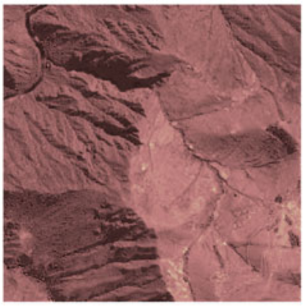

(i) Proposed Method: TC=81 PSNR 11 SSIM 0.71

Fig. 6 (continued)

5 Conclusion

As demonstrated by the PSNR and the SSIM evaluation, the proposed algorithm applied to PeruSat-1 images proves to be more efficient than the SPIHT, EZW and STW methods for compression ratios of approximately 4–20. When only the SSIM measurement metrics are considered, the proposed algorithm shows better performance.

When evaluated by the same quality metrics, the proposed algorithm applied to LANDSAT 8 images proves to be better than the SPIHT, EZW, and STW techniques for compression ratios of approximately 2–15. When only the SSIM metrics are used, it is significantly more efficient.

Wavelet subband coding allows for progressive decoding; that is, it allows us to visualize the image improvement based on the amount of data received, which is important given small satellites' low data rate.

For future work, the implementation of a lossless satellite image compressor using the whole wavelet transform is proposed.

Acknowledgements The authors would like to thank the Peruvian Aerospace Research and Development Commission (CONIDA) for providing the PeruSat-1 images used to conduct testing.

References

1. Aeronautics, N., Administration, S.: Small Spacecraft Technology State of the Art. Ames Research Center, Moffett Field, California (2018)
2. Canales, M., Bedon, H., Estela, J.: Design of a peruvian small satellite network. IEEEAC# 1672 (2009). https://doi.org/10.1109/aero.2010.5447020
3. Miyagusuku, R., Arias, K., Villota, E.: Hybrid magnetic attitude control system under cubesat standards. In: IEEE Aerospace Conference (2012). https://doi.org/10.1109/AERO.2012.6187239

4. Miyagusuku, R., Chicchon, M., Rojas, K., Villota E.: Low cost validation test bed for small satellites attitude determination and control. In: 3rd Nano-Satellite Symposium (2011)
5. Deigant, Y., Akshat, V., Raunak, H., Pranjal, P., Avi, J.: A Proposed method for lossless image compression in nano-satellite systems. In: IEEE Aerospace Conference (2017)
6. Praisline, R., Perumal, B., Pallikonda, M.: Comparison of image compression techniques using huffman coding, DWT and fractal algorithm. In: International Conference on Computer Communication and Informatics (ICCC) (2015)
7. Renau, V.: Image compression and restoration with nonlinear techniques. Ph.D. Thesis, University of Valencia (2011)
8. Said, A., Pearlman, W.: A new fast and efficient image codec based on set partitioning in hierarchical trees. IEEE Trans. Circ. Syst. Video Technol **6** (1996)
9. Shapiro, J.: Embedded image coding using zerotrees of wavelet coefficients. IEEE Trans. Signal Proc. **41**(12) (1993)
10. Kaur, P., Kaur, S.: An approach for combination of technique spatial-orientation trees wavelet (STW) and adaptively scanned wavelet difference reduction (ASWDR). Int. J. Comput. Sci. Mob. Comput. **6**(11), 64–69 (2017)
11. Morales, G., Arteaga, D., Huaman, G., Telles, J., Palomino W.: Shadow detection in high-resolution multispectral satellite imagery using generative adversarial networks. In: IEEE XXV International Conference on Electronics, Electrical Engineering and Computing (INTERCON) (2018)
12. Agustin Codazzi Geographical Institution: Description and Correction of Landsat 8 Products. Colombia (2013)
13. Kemper, G., Iano, Y.: An audio compression method based on wavelets Subband coding, IEEE Latin Am. Trans. **9**(5) (2011)
14. Alwan, N., Hussain, M.: Image quality assessment for different wavelet compression techniques in a visual communication framework. Model. Simul. Eng. article ID 818696 (2013). https://doi.org/10.1155/2013/818696

Patch Antenna Array Operating at 5G New Radio (NR) Frequency Band with the Ability of Beam Steering

Dennis Noel Arizaca-Cusicuna and Jorge Luis Arizaca-Cusicuna

Abstract This article presents the design of an antenna array operating at 3.7 GHz band for applications of the future 5G technology. This arrangement has as its basic element a patch antenna designed in microstrip technology on the FR4 substrate and will allow switching the beam in order to change the radiation direction.

Keywords Microstrip antenna · Antenna array · 5G · Mobile communication

1 Introduction

The International Telecommunication Union (ITU) has defined 5G networks and artificial intelligence as the main fields of innovation necessary for the generation of smart societies [1]. 5G technology brings the promise of improving the end-user experience, offering new applications and services capable of reaching speeds of several gigabits, as well as significantly increasing the quality of operation and reliability [2, 3]. The 5G specifies the spectrum in three fundamental frequency ranges to provide extended coverage and support all forms of use. The three ranges are: below 1 GHz, between 1 and 6 GHz and above 6 GHz. Within these, the band 1–6 GHz offers a good combination between the benefits of coverage and capacity [4], but the Association GSM expects the spectrum of the 3.7 GHz band to form the basis of many of the first 5G services [5], in particular, the provision of improved mobile broadband, and this is because this frequency range is practically agreed on a global scale and, therefore, is capable of promoting the economies of scale necessary for low-cost devices. According to the GSM Association, in the medium term, there will be an allocation of 100 MHz in the 3.7 GHz band.

D. N. Arizaca-Cusicuna (✉) · J. L. Arizaca-Cusicuna
Universidad Nacional de San Antonio Abad del Cusco, Cusco, Peru
e-mail: 111205@unsaac.edu.pe

J. L. Arizaca-Cusicuna
e-mail: jorge.arizaca@unsaac.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_18

To meet the requirements of this technology, it will be necessary to improve the quality of transmission and reception in communication systems. One solution will be to focus the radiation in one direction, in which the interference signals are canceled and the desired signals are improved to increase the carrier/interference ratio and that is why the use of beamforming networks associated with an array antennas will play an important role in the development of wireless communications [6].

2 Patch Antenna Array and Feed Network Design

The proposed antenna array begins with the design of a single radiating element, in this case, a patch antenna, which has shown a balance in technological design issues and commercial criteria: low cost, compact size, radiation efficiency, high gain, good performance in high-frequency bands, among others. Groups of these antennas are joined by a feed network, thus forming an array of 16 elements. Both the patch antenna and the feed network system are evaluated using CST Studio electromagnetic field simulation software [7].

2.1 Patch Antenna Design

The proposed antenna uses the FR4 ($\varepsilon_r = 4.3$) as a substrate, which is available and low-cost material. For the design of the antenna, the following parameters have been considered: the operating frequency ($f = 3.7$ GHz), the appropriate material with its dielectric constant for the substrate (ε_r) and the thickness of the substrate ($H = 1.6$ mm), that will allow us to calculate the dimensions of the patch using the following equations [8].

The width of the antenna can be calculated with:

$$W = \frac{C}{2f}\sqrt{\frac{2}{\varepsilon_r + 1}} = \frac{\lambda}{2}\sqrt{\frac{2}{\varepsilon_r + 1}} \tag{1}$$

And the length with:

$$L = L_{\mathrm{eff}} - 2\Delta L = \frac{C}{2f\sqrt{\varepsilon_{\mathrm{eff}}}} - 2\Delta L \tag{2}$$

Using

$$\Delta L = 0.412\frac{(\varepsilon_{\mathrm{eff}} + 0.3)\left(\frac{W}{H} + 0.264\right)}{(\varepsilon_{\mathrm{eff}} - 0.258)\left(\frac{W}{H} + 0.8\right)} \tag{3}$$

And

$$\varepsilon_{\text{eff}} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + 12\frac{H}{W}\right)^{-\frac{1}{2}} \tag{4}$$

where:

W	width of the patch antenna.
L	length of the patch antenna.
f	resonance frequency.
C	speed of light.
ε_r	dielectric constant of the substrate.
ΔL	length extension.
H	thickness of the substrate.
ε_{eff}	effective dielectric constant of the substrate.

The resulting dimensions were optimized with the software and are shown in detail in Table 1.

Figure 1 shows the antenna that has been designed for the 3.7 GHz frequency and the quarter-wave transformer method that it uses to couple the impedance of the patch with the transmission line.

The simulation results are shown in Fig. 2 where we can see that return loss of the proposed patch antenna at 3.7 GHz is −44.45 dB with a bandwidth of 143 MHz taking as criteria value of −10 dB to define bandwidth.

Also, the 3D and 2D radiation pattern can be obtained with the software simulator, and Fig. 3 shows us that the maximum gain obtained is 3.46 dB in the radiation direction *z*-axis, which is considered a good value in a compact rectangular patch antenna.

Table 1 Patch antenna dimensions

Parameter	Dimension (mm)
W	23.32
L	18.107
Ws	34.504
Ls	28.68
W1	0.58
L1	11.49
W2	2.82
L2	4.48

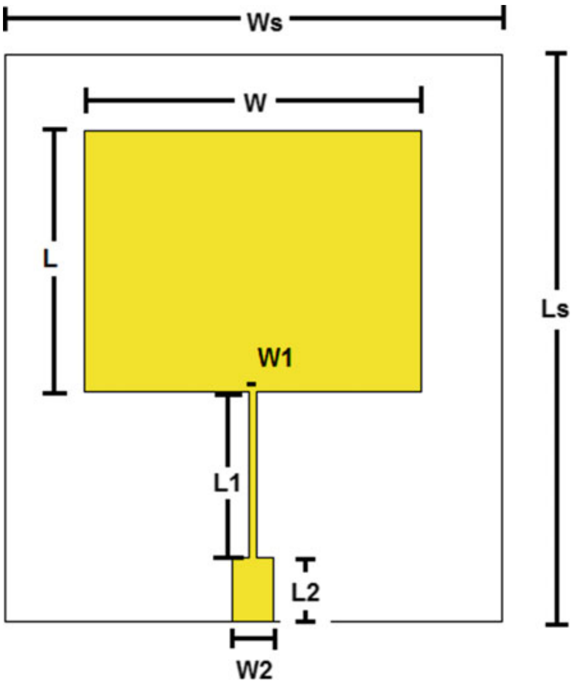


Fig. 1 Patch antenna dimension

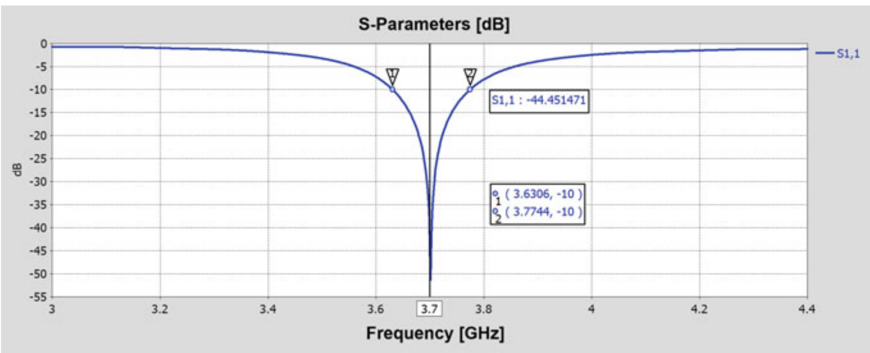


Fig. 2 Return loss of single patch antenna

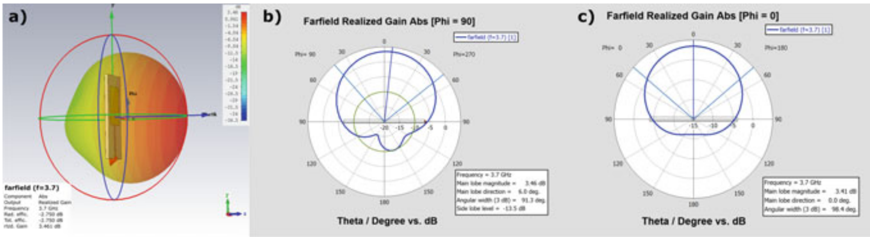


Fig. 3 **a** 3D radiation pattern, **b** radiation pattern in YZ plane, **c** radiation pattern in XZ plane

2.2 Feed Network Design

Patch antennas are commonly used in arrangements to improve their performance, increase their gain, directivity and add functionalities such as electronic beam scanning that would hardly be achieved with a single element. For the feed network system of this arrangement, two types of power dividers were chosen: the T-junction power divider and the Wilkinson power divider. Both were designed on FR4 material because they will share the substrate with the previously analyzed antennas.

T-Junction Power Divider. The T-junction power divider is a three-port network used to feed and split power. These can be modeled as the junction of three transmission lines [9] as shown in Fig. 4 whose impedance characteristics are reduced to:

$$\frac{1}{Z_1} + \frac{1}{Z_2} = \frac{1}{Z_0} \tag{5}$$

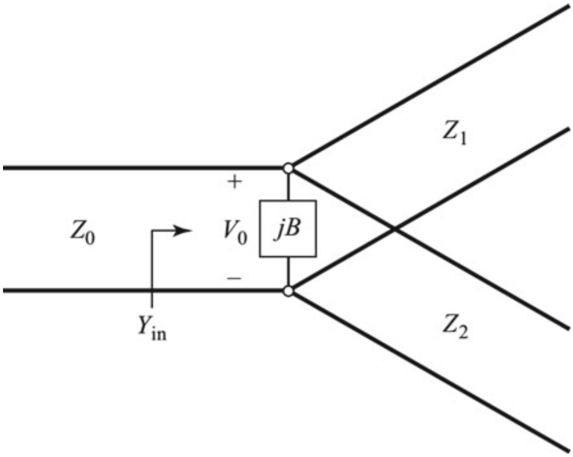


Fig. 4 Transmission line model of Wilkinson divider

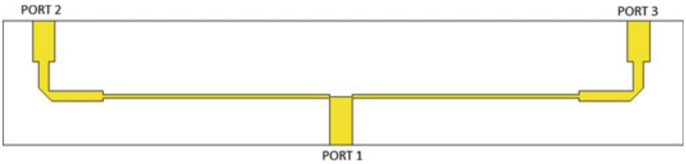


Fig. 5 T-junction divider designed

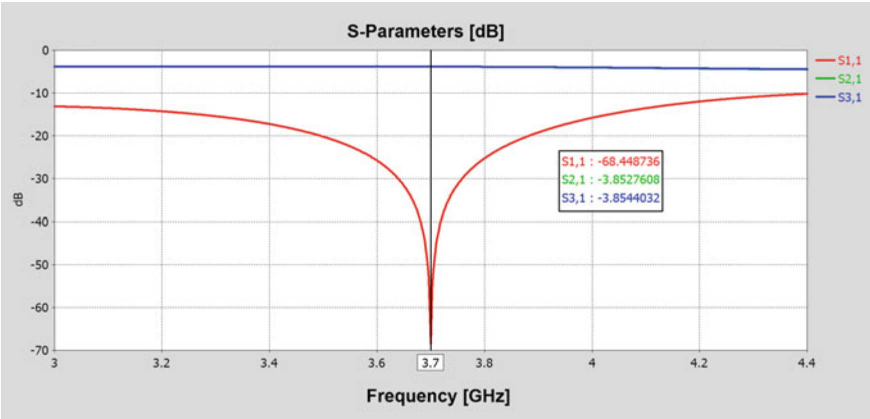


Fig. 6 S-parameters of T-junction divider

Figure 5 shows the model designed in the simulation software. This model is coupled to two quarter-wavelength transformers in order to obtain an impedance of 50 Ω in all ports.

The simulation results are shown in Fig. 6. In this image, we can see that the reflection coefficient S11 is −68.448 dB indicating a good level of coupling at the input port. Similarly, the parameters S21 with a value of -3852 dB and S31 with a value of −3854 are shown. These results suggest that the input power is transferred equally to the output ports 2 and 3.

Wilkinson Power Divider. For the union of radiating elements, a Wilkinson power divider of uniform distribution (3 dB) was used, whose equivalent transmission circuit is shown in Fig. 7. And also, in Fig. 8, we can observe its design in the electromagnetic simulation software.

Figure 9 shows the results obtained in the simulation software of Wilkinson power divider, where we can see that the reflection coefficient S11 at 3.7 GHz is −62.35 dB which represents an excellent level of coupling at port 1. It is also shown that the transmission coefficients S21 and S31 are −3.58 dB and −3.64, respectively, which indicates that the input power is equally distributed between port 2 and port 3.

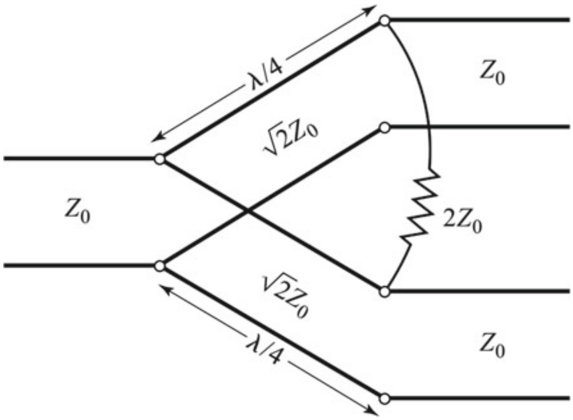


Fig. 7 Transmission line model of Wilkinson divider

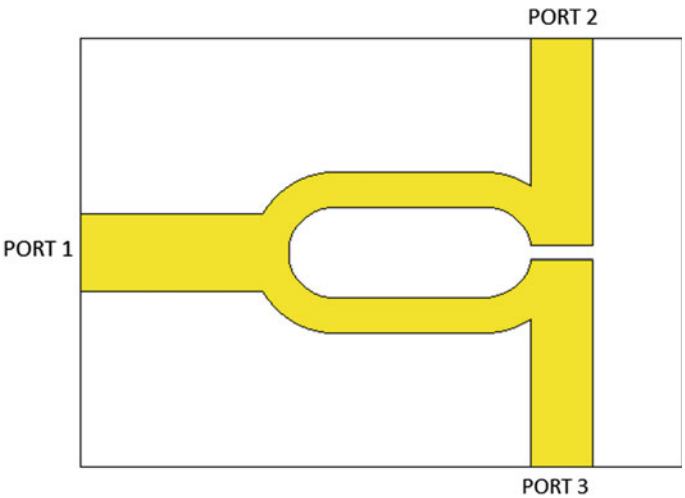


Fig. 8 Wilkinson divider designed

2.3 Antenna Array Design

The previously designed elements are part of the proposed antenna array. This arrangement consists of 16 radiating elements with approximately a half wavelength spacing in order to achieve a switched beam system. Figure 10 shows the design in the electromagnetic simulation software. As can be seen, a feed network is used to form four columns of four antennas each one, resulting in an arrangement of 16 elements with four inputs, which will serve to introduce signals with different phase, so that the main lobe will be oriented to different addresses according to the input ports.

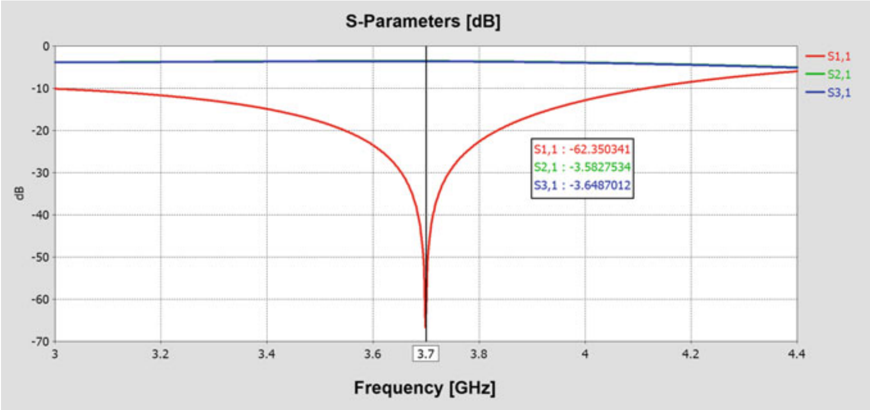


Fig. 9 S-parameters of Wilkinson divider

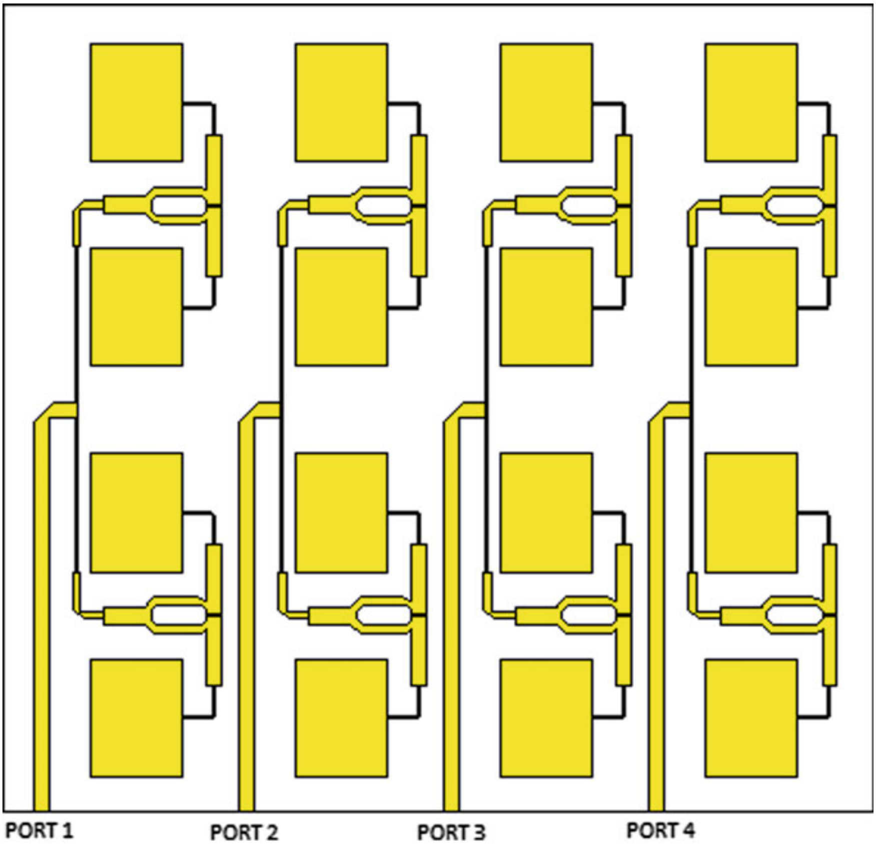


Fig. 10 Antenna array designed

3 Results

The proposed antenna array was simulated in the CST Studio electromagnetic simulation software. The results are shown below.

As it is an antenna array formed by four identical groups with independent input power, the return losses of one of the columns are analyzed obtaining a reflection coefficient S11 under 34 dB Fig. 11. The array operates in a potential band proposed for the fifth mobile generation (5G) and shows a bandwidth of 249 MHz taking −10 dB as a reference value.

Figure 12 presents the 3D and 2D radiation pattern of the proposed 4 × 4 array, obtaining a maximum gain of 11.3 dB for the operating frequency. These results satisfy our requirement of an antenna for 5G applications since a high gain value will compensate for the attenuation suffered by wave propagation.

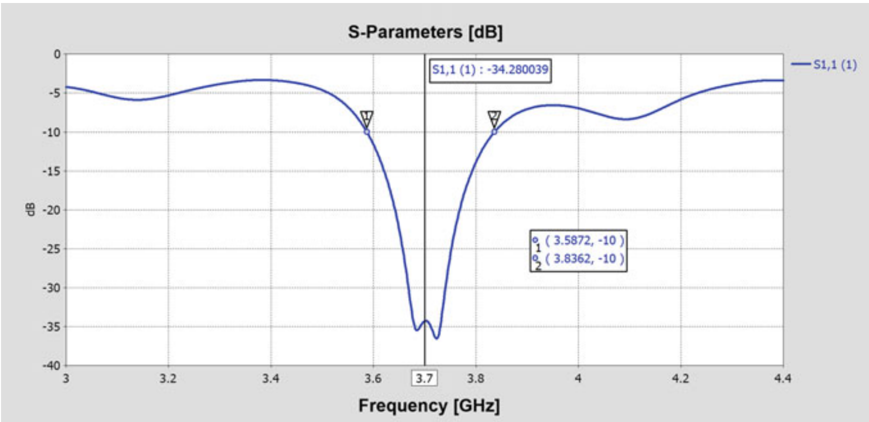


Fig. 11 S-parameters of one column of the array antenna

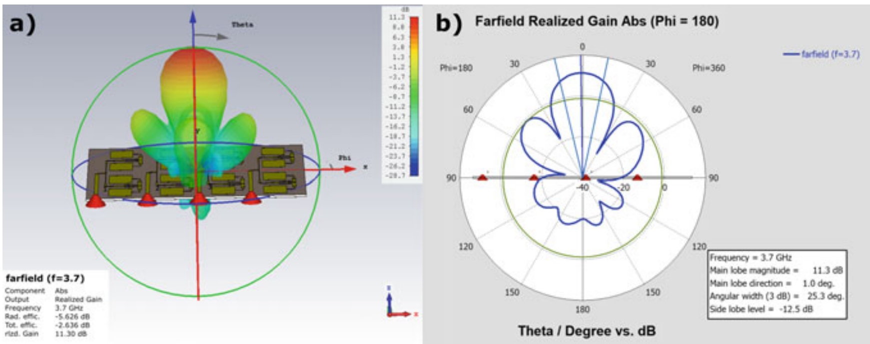


Fig. 12 **a** 3D radiation pattern of antenna array, **b** YZ plane radiation pattern of antenna array

In order to test the beam address, signals with different phases were introduced into the array inputs using the electromagnetic simulation software. The offsets used and the direction of the theoretical main lobe are detailed in Table 2 [10], and the results are shown in Fig. 13.

The directions of the main lobe correspond to the expected result according to Table 2 [5]. However, in future work, it is necessary to improve the accuracy and reduce the presence of side lobes.

Table 2 Main lobe direction according to input ports

Input port	Direction 1 (deg.)	Direction 2 (deg.)	Direction 3 (deg.)	Direction 4 (deg.)
1	135	45	90	0
2	90	180	−45	45
3	45	−45	180	90
4	0	90	45	135
Phase difference	−45	135	−135	45
Main lobe direction	15	−45	45	−15

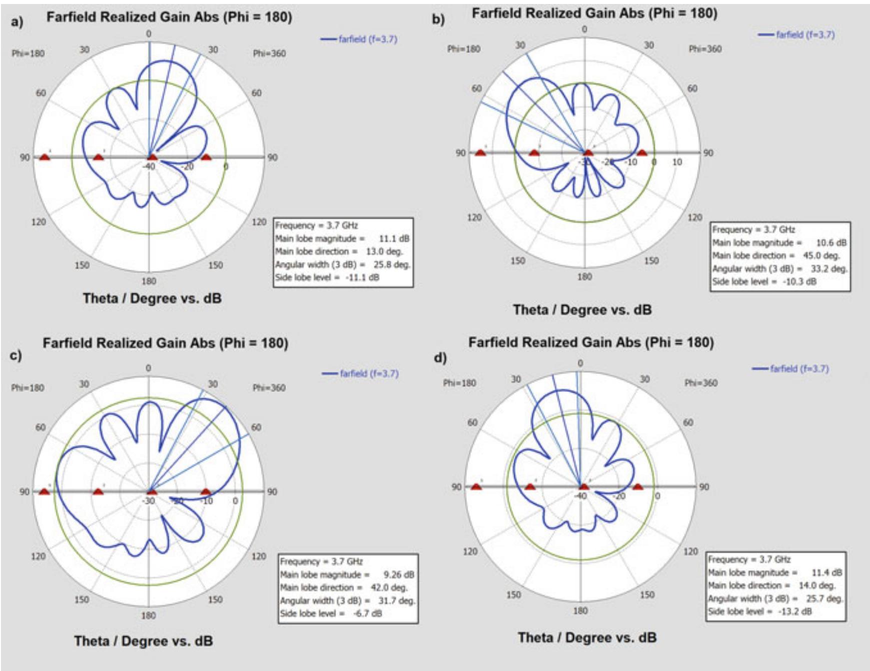


Fig. 13 Radiation patterns in XZ plane with the main lobe in different directions according to Table 2

4 Conclusions

This work designs and simulates an array of 4 × 4 elements antennas that operate in a 5G band and allows the beam to be oriented in a specific direction depending on the phase shifts introduced in each antenna array input ports. The results show a bandwidth greater than 200 MHz, a maximum gain of 11.3 dB and different beam directions were obtained. These suggest the good performance of our arrangement for 5G applications in the 3.7 GHz band. In future works, it is intended to improve the accuracy of the results and achieve a scalable system with the addition of a phase shifter circuit.

References

1. International Telecommunication Union: Setting the Scene for 5G: Opportunities & Challenges. ITU. Geneva, Switzerland (2018)
2. Andrews, J.G., Buzzi, S., Choi, W., Hanly, S.V., Lozano, A., Soong, A.C., Zhang, J.C.: What will 5 g be? IEEE J. Sel. Areas Commun. **32**(6), 1065–1082 (2014)
3. Agiwal, M., Roy, A., Saxena, N.: Next generation 5 g wireless networks: a comprehensive survey. IEEE Commun. Surv. Tutorials **18**(3), 1617–1655 (2016)
4. HUAWEI Technologies Co., L.: 5 g spectrum-public policy position. HUAWEI Technologies., Shenzhen, P.R. China (2018)
5. Association, G., et al.: 5 g spectrum gsma public policy position. GSM Assoc., London, UK, Tech. Rep (2018)
6. Haroun, I., Lin, T.Y., Chang, D.C., Plett, C.: A compact 24–26 GHz ipd-based 4 × 4 butler matrix for beam forming antenna systems. In: 2012 Asia Pacific Microwave Conference Proceedings, pp. 965–967. IEEE (2012)
7. CST Studio Suite (2018.00) [Software]. (2018) Retrieved from https://www.3ds.com/products-services/simulia/products/cst-studio-suite/
8. Balanis, C.A.: Antenna theory: analysis and design. Wiley (2016)
9. Pozar, D.M.: Microwave engineering. Wiley (2012)
10. Cabrera-Botero, J.A., Páez-Rueda, C.I.: Switched antenna pattern with reduced area for the industrial, scientific and medical 2.45 GHz band. Ingeniería y Universidad **18**(1), 121–140 (2014)

A Development of a Mobile Application for Monitoring Siemens S7-1200 Controller Variables Through Firebase

Alexander Diaz, Steven Rivera, Leonardo Vinces, and Christian Del Carpio

Abstract Production systems need to be monitored because failure to do so can cause a delay in production due to a lack of remote monitoring. There are solutions that provide this information, but they are limited, because they are owned and very expensive solutions. This paper proposes the development of a mobile application for the monitoring of S7-1200 programmable controller variables through Firebase. The goal is to use open-source-type tools to develop a real-time variable monitoring application of a programmable controller capable of hosting a Web server, and the sent data to the Firebase cloud, in order to avoid modifying the driver software and the data stored in memory. The implemented system of sending, handling and presenting variables is developed with tools of open source type, since they have as an advantage to have greater freedom in programming, unlike proprietary solutions or services. The tests were performed on the Siemens S7-1200 controller with a plastic injection molding machine.

Keywords Raspberry Pi · Programmable controller · Web server · Windows 10 IoT

A. Diaz · S. Rivera · L. Vinces (✉) · C. Del Carpio
Peruvian University of Applied Sciences, Av. Prolongación Primavera 2390, Santiago de Surco, Lima, Peru
e-mail: nikolai.vinces@gmail.com

A. Diaz
e-mail: u201415388@upc.edu.pe

S. Rivera
e-mail: u201311863@upc.edu.pe

C. Del Carpio
e-mail: pcelcdel@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_19

1 Introduction

Manufacturing in Peru has a high share of the country's economy, making up 75% of the industry according to the National Society of Industries (SNI), along with an increase of 3.7% in March 2019 according to the Ministry of Production of Peru. As a result, the main Peruvian industrial branches are currently increasing production [1].

In this way, the increase in product orders generates companies engaged in these items to increase their production through more specialized machines. Indeed, most commercial machines have an additional infrastructure that facilitates the production of their parts. For example, commercial machines have a programmable automaton, but not all have the same characteristics, such as having the ability to monitor information online.

The importance of having real-time programmable controller internal information helps in better decision-making; preventive and predictive maintenance; better traceability and quality in the product, this is always because we have data, allows us the best decision to continue with production. For example, to make plastic parts in a partially automated extruder machine, it is necessary to be aware of the amount of plastic needed and the number of parts manufactured by the machine, causing a lack of attention in certain parts of the such as part quality or wear of some parts of the machine. However, this can be avoided in the following ways, first by manufacturing the product slower and the second having more staff in the process, for verification of these incidents, but in both cases leads to delays and high costs in production. In addition, it is worth mentioning that this type of company dedicated to the field of injection has 4, 5 or more machines working in parallel. So, one of the main problems is not having the data in real time, and the lack of this monitoring forces to perform this task manually, resulting in a delay in production.

The companies dedicated to the injection industry in Peru have different types of electrical, hydraulic, pneumatic and mechanical machines, which do not provide the variables of their respective processes, causing delay and production disorder. However, there are companies that offer monitoring assistance, in real time through Web pages or mobile applications [2, 3]; but these are limited to delivering expensive and complicated tools that only allow a limited number of variables and do not fit the needs of the user. In view of this need, a method of real-time monitoring of the variables of a controller is proposed.

In similar work, it is proposed to use data acquisition modules separate from the programmable controller, which are connected by means of a switch module for communication with the embedded system, this is quite practical for communication with different controllers at the same time with the embedded system, in addition to using the conventional Ethernet connection that allows easy communication [4]. However, this application requires a great cost in extra hardware and software, and data acquisition devices are not robust enough for industrial applications.

Debbag Y. and Yilmaz E. N. on the offering to use the CP 1242-7 GPRS module for long distances and with an Internet connection, which is one of the most robust and used implementations in the industry [5], but is excessively expensive if you want to use it in applications of little scope and budget.

Britto F., Foncesa I. and Torres Farinha J., in sending the variable information of an OMRON Programmable Logic Controller using Industrial Ethernet protocols, for the purpose of coupling low-cost electronic platforms such as Arduino or Raspberry Pi. However, the acquired information is only maintained locally, limiting its display only for devices that are on the same network as the programmable controller [6].

Another document introduces the development of Android-based monitoring of the digital and/or analog inputs and outputs of a programmable controller with the use of the RS-485 industrial communication port and an Arduino with the Bluetooth module connected, with the purpose of visualizing logical controller data in a low-cost system designed in real time and without cable [7]. However, the Bluetooth module is limited for office, home, and short-range use, because of the number of noise sources in the industry is not applicable to such environments.

In this way, the proposed system is to acquire the variables of a programmable controller, with the ability to create and host a Web server in JavaScript Object Notation (JSON) format, connected by Ethernet to an embedded system with internet availability. The Siemens S7-1200 controller is used for the increased demand which has on the local market and an embedded Raspberry Pi 3B system for versatility in programming custom applications. Therefore, the embedded system reads the Web server from the controller and sends the data to the Firebase real-time database. In fact, Firebase is a node service provided by Google in the cloud, used for storing and hosting a real-time database; achieving effective monitoring, management, visualization of historical data and remote control [8]. Then, a mobile app made in Xamarin connected to the Internet is developed that reads the database generated in Firebase database. This will allow you to visualize at any time the acquired variables of the process. For the demonstration of the developed system, the variables that were monitored were from a plastic injection molding machine, these variables were, temperature, the opening of the machine and the number of parts manufactured. In this way, it is proposed to maintain order and a real-time knowledge of the process variables of a machine that has a built-in programmable controller.

2 Description of the Proposed Algorithm

Figure 1 describes the block diagram that allows the sending and monitoring of programmable controller variables to Firebase through an embedded system.

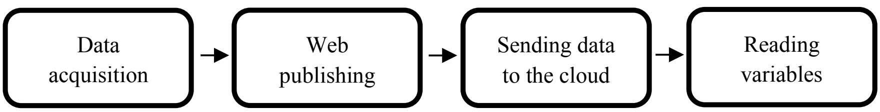


Fig. 1 Description of the proposed method

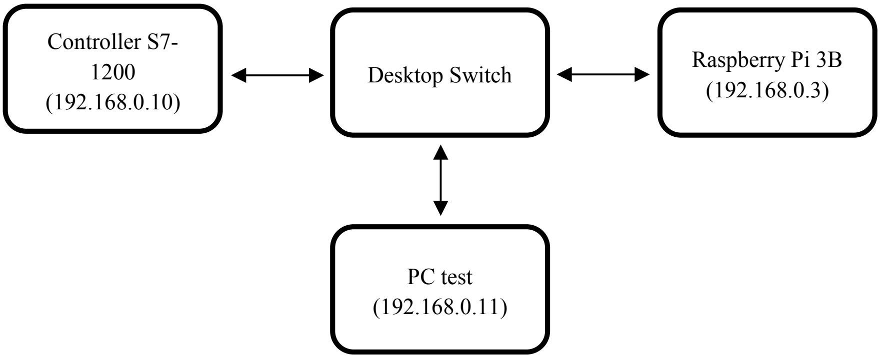


Fig. 2 Networking of the S7-1200 controller, CPU and Raspberry

2.1 Data Acquisition

For the acquisition of data from the Siemens S7-1200 controller, it was networked via a desktop switch with a Raspberry Pi 3B and a personal computer (PC), which features Windows 10 IoT [9] and TIA Portal 13 [10] software, respectively. In Fig. 2, the connection and static IP address of each networked component are observed.

The PC is incorporated into the network momentarily for the purpose of displaying the status of the programmable controller variables, in this case, are variables of type Double Word (Dword), which represent the temperature, and type Dint, which represent the number of parts manufactured by the machine.

2.2 Web Publishing

With the information acquired in *Dword* variables within the Siemens controller, the programmable controller Web server is designed in JSON format. Indeed, the controller's Web server works with Automation Web Programming (AWP), which are commands developed by Siemens that allow you to access the driver application in the daily job without having to modify the Driver program or affect process data [11]. In addition, the JavaScript Object Notation (JSON) language is a lightweight data exchange format primarily used to sort an unordered set of variable pairs [12]. The string designed on the Web server consists of the following variables (Table 1):

Table 1 Programmable controller variables

No. variables	Sensor	Type
4	PT100	Real
1	Photosensor	Dint

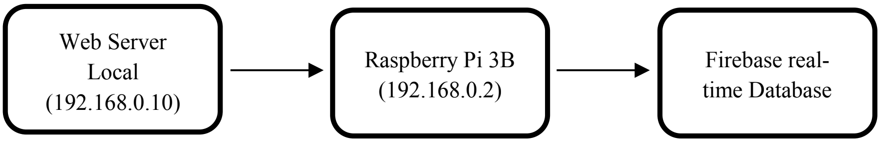


Fig. 3 Data transmission to Firebase database

Subsequently, to generate the Web server package is activated and loaded into a DB of the programmable controller, resulting in the local network extension that is then acquired by the embedded system. In this way, you have the following Web address: http://192.168.0.1/awp/Control/index.htm.

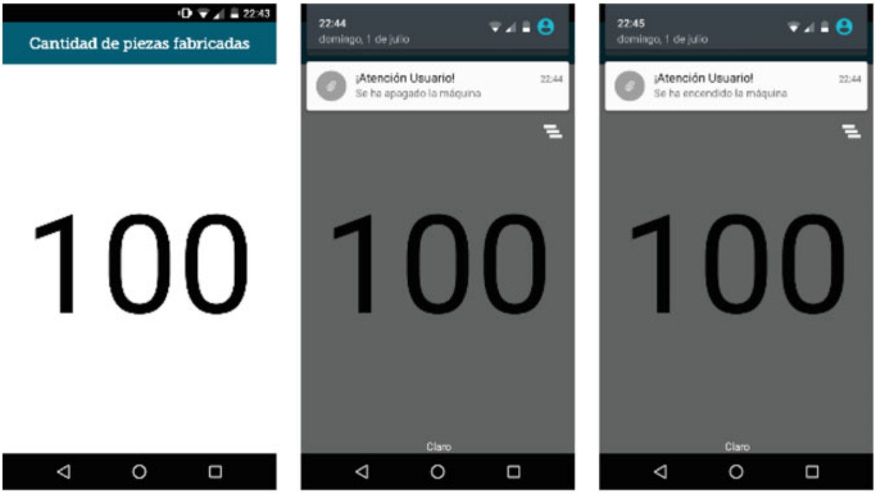
2.3 Sending Data to the Cloud

With the Web server on a local network, the Raspberry Pi 3B is programmed in the Universal Windows Platform (UWP) environment of Visual Studio 2017, with library support, Windows 10 IoT for embedded system detection, NewtonSoft for the deserialization of the JSON format, and finally the FireSharp library for reading and writing variables to the Firebase database.

The operation is as follows: The Raspberry Pi 3B reads the source code of the Web server every second and is then deserialized into a list consisting of five variables of the float type. Therefore, the IFirebaseClient function on the embedded system is used to upload the variables to the cloud in node format, where each node represents a variable. This updates each node in the cloud every second. In Fig. 3 The static IP address and the sending of data developed by the Raspberry from the programmable controller to the Firebase database environment are observed.

2.4 Reading Variables

With variable information in the database, a mobile app is designed in the Visual Studio 2017 Xamarin environment, with support from the NewtonSoft and FireSharp library. In this case, maintaining a constant real-time data refresh uses the FireSharp function called EventStreamResponse that allows asynchronous and constant reading of a database node for each time it is updated. This way, in Fig. 4 the application developed is observed, where it indicates the number of parts produced by the plastic injecting machine, in addition to visualizing that the machine has been turned on and off.



(a) (b) (c)

Fig. 4 Application mobile cloud information receiver Firebase. In **a** the number of parts manufactured is observed, in (**b**) and **c** the notification representing the start and off the machine, respectively

3 Results

The two tests that will be seen on the Siemens S7-1200 1214C AC/DC/RLY [13] controller of a plastic injection molding machine are plotted in Figs. 5 and 6, using as variables the 4 PT100 sensors attached to the thermal wells of each belt resistor that the machine has, and the photosensor of the cooling line of this same machine for the final counting of manufactured parts. For the first test, polyethylene production was manufactured for 2 h and 30 min and monitored with the implemented system.

In Fig. 5a, the variable behavior of the number of parts manufactured in real time is observed with respect to the time it was sampled and uploaded to the cloud by the Raspberry Pi 3B, where for each piece of polyethylene manufactured, an average load time of 3.76 is 64 s to get into the cloud.

For the second test, the temperature of each belt resistance of the injector machine was stabilized. In Fig. 5b, you see the behavior of the temperature variations of the resistors in real time with respect to the time it was sampled and uploaded to the cloud by the Raspberry Pi 3B, where you have an average load time of 3.4117 s to get into the cloud.

Likewise, there is a data load time to the cloud, approximately 3 s, which is acceptable for monitoring variables in the mobile application. In addition, the amount of time it takes for the system to upload information to the cloud is a variation due to the available internet connection speed.

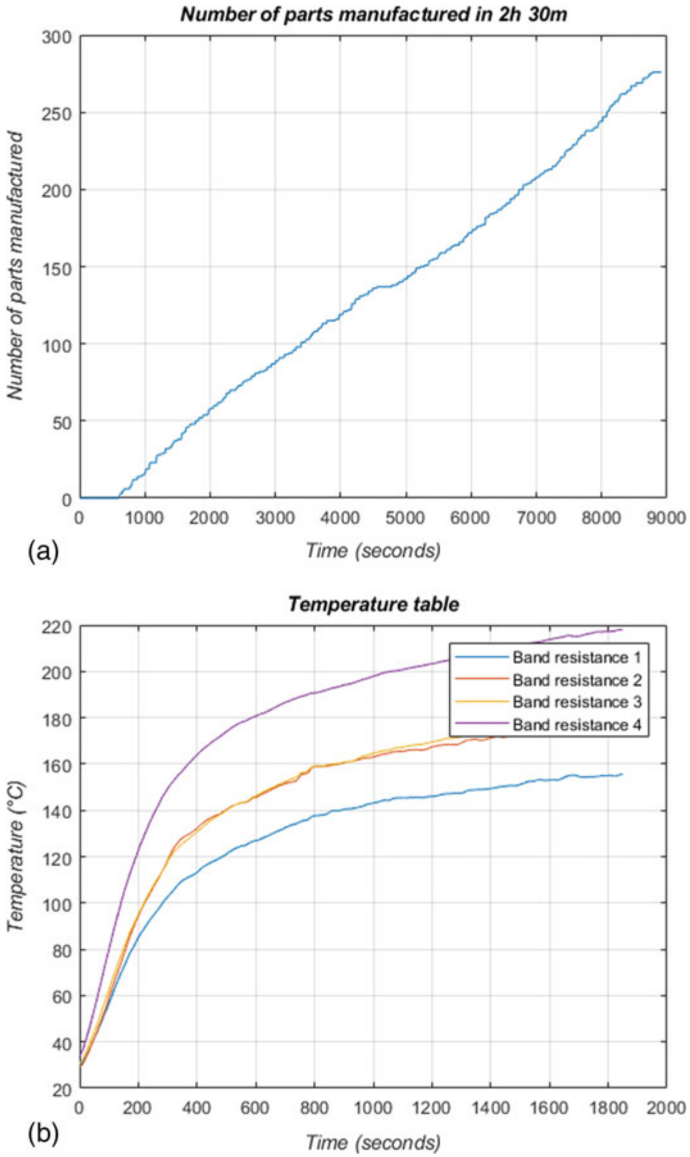


Fig. 5 Real-time PLC variable sampling result. In **a** the factory parts in 2 h 30 min sampled by Raspberry Pi 3B and **b** temperatures of the four ceramic band resistances sampled by the Raspberry Pi

In the same way as proprietary monitoring modules, the proposed method responds appropriately in the same data transfer protocol, but with a slight delay, this being acceptable for the cost difference.

4 Conclusion

In conclusion, it was possible to visualize in a mobile application the variables of a Siemens S7-1200 PLC stored in Firebase by a Raspberry Pi 3B. In addition, among the achievements, adequate time was obtained for the acquisition and writing of data in the cloud through the Raspberry, although it is observed that the amount of time depends on the proximity of the system embedded to the Wi-Fi router.

Also, it is expected that in future works, the Raspberry or other derivatives of the same brand can be changed but at a lower price, along with a better direct reading of data, in order to avoid delay time.

References

1. Ministry of Production of Peru: Produce: Manufacturing Sector Grew 3.7% in March 2019. http://ogeiee.produce.gob.pe/index.php/shortcode/oee-documentos-publicaciones/boletines-industria-manufacturera/item/781-2018-enero-reporte-de-produccion-manufacturera. Last accessed 2019/09/22
2. Industry 4.0/IoT—Products and Solutions. https://new.siemens.com/global/en/company/topic-areas/future-of-manufacturing/digital-enterprise.html. Last accessed 2019/05/13
3. Siemens Special Solutions with the Digital Enterprise Suite. https://new.siemens.com/global/en/company/topic-areas/future-of-manufacturing/digital-enterprise.html. Last accessed 2019/05/13
4. Antony, J., Sharma, S., Mabato, B., Chitranshi, G.: Distributed data acquisition and control system based on low cost Embedded Web servers. In: Blue Eyes Intelligence Engineering & Sciences Publication Pvt. Ltd, International Journal of Instrumentation, Control & Automation (IJICA), vol. 1, pp. 1–4, (2011)
5. Debbag, Y., Yilmaz E.N.: Internet based monitoring and control of a wind turbine via PLC. In: 3rd International Istanbul Smart Grid Congress and Fair (ICSG), pp. 1–5 (2015)
6. Brito, F., Fonseca, I., Torres, J., Maciel, F.: Integrating low cost platforms on electrical power. In: International Conference on Nextgen Electronic Technologies, pp. 260–262 (2017)
7. Kanmaz, A., Örenbaş, H., Şahin, S.: Android based PLC data monitoring interface with industrial data communication port. In: 26th Signal Processing and Communications Applications Conference (SIU), pp. 1–4 (2018)
8. Wu-Jeng, L., Chiaming, Y., You-Sheng, L., Shu-Chu, T., ShihMiao, H.: JustIoT Internet of Things based on the Firebase real-time database. In: IEEE International Conference on Smart Manufacturing, Industrial & Logistics Engineering (SMILE), pp. 43–47 (2018)
9. Windows for Internet of Things. https://developer.microsoft.com/en-us/windows/iot. Last accessed 2019/05/13
10. TIA Portal—An Overview of the Most Important Documents and Links. https://support.industry.siemens.com/cs/document/65601780/tia-portal-an-overview-of-the-most-important-documents-and-links-controller?dti=0&lc=en-WW. Last accessed 2019/05/13

11. Siemens Recipes with SIMATIC. https://cache.industry.siemens.com/dl/files/612/94681612/att_852408/v3/94681612_S7-1200_Recipe_DOKU_v11_en.pdf. Last accessed 2019/05/13
12. An Introduction to JavaScript Object Notation (JSON) in JavaScript and .NET. https://msdn.microsoft.com/en-us/library/bb299886.aspx. Last accessed 2018/11/30
13. Siemens SIPLUS S7-1200 CPU 1214C AC/DC/RLY. https://support.industry.siemens.com/cs/pd/567214?pdti=pi&dl=en&lc=en-WW. Last accessed 2019/11/30

Implementation of a Real-Time Communication Library Between Smart TV Devices and Android Devices Based on WebSocket for the Development of Applications

Check for updates

Gian Gabriel-Caycho, Miguel Blas-Perez, and Alfredo Barrientos

Abstract The smart device industry has grown in recent years, so the vast majority of people have a smart TV and at least one smartphone in their homes. The transformation of the digital era seeks to improve the user experience and facilitate the use of different smart devices. Due to this, it is proposed to develop a communication library that allows multiple Android devices to be connected to an LG WebOS Smart TV and thus promote the development of applications for Smart TVs, avoiding implementation from scratch and speeding up the deployment of new applications that generate great value for users. In this article, a comparison was made of the main communication protocols existing and compatible with the devices involved. During the development stage, the communication library based on the WebSocket protocol was implemented. Also, a retro videogame was implemented that used the communication library to connect the LG WebOS smart TV with the Android Smartphone. Finally, the results obtained and the opportunities for improvements found for the development of future applications will be shown.

Keywords WebOS · WebSocket · Android · Real-time communication · Smart TV

1 Introduction

At present, the vast majority of people have at least one smart device for daily use, which allows them to interact with other devices on their network or through the Internet, so the world of IoT has been growing with the passage of years. The use of Smart TVs has become very common in homes around the world because it has

G. Gabriel-Caycho (✉) · M. Blas-Perez · A. Barrientos
Universidad Peruana de Ciencias Aplicadas (UPC), Lima 15023, Peru
e-mail: u201512251@upc.edu.pe

M. Blas-Perez
e-mail: u201317356@upc.edu.pe

A. Barrientos
e-mail: pcsiabar@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_20

become a fundamental device. The emerging Smart TV model has marked a new generation of screens due to its connection with other devices such as smartphones or tablets [1]. In addition, the growth of Smart TVs has led companies to look for new proposals that replace conventional remote control and allow these devices to be handled through smartphones [2]. In this way, we propose the development of a real-time communication library capable of connecting multiple Android devices with a Smart TV and that they can exchange information quickly. Likewise, it is proposed to use the communication library developed as a basis for future development projects and in this way developers can implement and deploy their applications in a more agile way, avoiding implementation from scratch.

2 State of the Art

2.1 Communication Protocols

There are different investigations where the different communication protocols that are used in IoT are compared as in [3] that compare the Web performance of messaging protocols such as Message Queue Telemetry Transport (MQTT), Advanced Message Queuing Protocol (AMQP), Extensible Messaging and Presence Protocol (XMPP), and Data Distribution Service (DDS) by measuring the latency of the arrival of messages to a sensor. Also, protocols are used within an HTML5 (HyperText Markup Language) Web site to measure latency, performance, and coding. Finally, it is mentioned that the use of the MQTT protocol and the WebSocket protocol is recommended for the implementation of Web applications that use JavaScript and HTML5. In the investigation of [4], they focused on the transmission of data through the different protocols used in the world of IoT where MQTT, CoAP, AMQP, and WebSocket were considered. A comparison of the main characteristics between the protocols was made considering security, architecture, and communication protocol as the most important points.

The authors in [5] analyze different communication protocols using different scenarios, different applications, and different devices because each protocol can be used in the best way in a specific scenario. Therefore, choosing a protocol depends on the scenario and the environment in which it will be used. When applications need to update data frequently it is recommended to use the publication/subscriber protocols. The WebSocket protocol is ideal for use in applications that use Web browsers because it provides full-duplex communication asynchronously between the server and the client allowing for a continuous session until those involved finish the need for communication.

2.2 Applications with Bidirectional Communication

Teamwork or collaborative work requires platforms capable of supporting multiple requests from different users in real time. In [6], an algorithm for collaborative graphics editing is developed using HTML5 Canvas, WebSocket, JQuery, and Node.js. To display a graphical user interface and perform the tests of the proposed model, a Web page is developed that allows collaborative editing of multiple users in real time because each modification is reflected for all users. Collaborative work tools allow optimizing the analysis time to improve the quality of the results. In [7], a Web site called OnGraX is developed that allows changes to be made to a graphic, which can be modified by different people in real time. In addition, a chat is included for conversations between users during the session. HTML5, JavaScript, WebSocket, and Java were used to implement the prototype. For the development of the world of IoT, it is important to establish a correct architecture that meets the requirements of the environment in which it will be used.

The work done by the authors of [8] shows a Web solution, which consists of the monitoring of actions in real time. During the tests, examples are shown that allow us to see the importance of having a flexible system for changes. In addition, it is mentioned that the proposed system was developed using the MQTT protocol and is prepared to be able to relate to other systems with which it can communicate in real time for the exchange of information.

3 Implementation of the Solution

According to the investigation of the different communication technologies we have selected the WebSocket protocol. Our project seeks to improve the user experience by combining the use of smartphones with Smart TVs, which is why this section has been divided into two stages: implementation of the communication library and development of a retro videogame.

3.1 Implementation of the Communication Library

For this stage, the work environment where the development and proof of concept will be carried out was defined. The main tools that were used were the following: Visual Studio Code, WebOS TV IDE, WebOS TV CLI, Google Chrome, Chromium, and Android Studio. To perform all the necessary tests of the development, LG WebOS Smart TV, and an Android Smartphone were used.

Technology and construction. For the development of the library, JavaScript was used for the WebOS part. To obtain the private IP of the Smart TV, an interrupted version of the WebRTC protocol was implemented, which attempts to communicate with the network devices. The IP obtained must be cut, since we only care about the first 3 octets of the IP, which serves to identify the network in which the device is located and then travel through the 255 IP's candidates to be a WebSocket server. During the IP route, a list of servers will be filled, which will be stored temporarily and will be displayed where necessary for users. In the tests that will be carried out in the next part, the list will be shown on the main screen of the applications so that users can then choose which server they want to connect to. On the other hand, for the part of the library that will be implemented in the Android device, the JAVA programming language will be used, in which the Java_WebSocket and GSON library will be used. The library in Android can be used as a client and server so the library was divided into two classes which contain all the necessary functions for each scenario in which it is required to create a server or a WebSocket client. Next, the proofs of concept of the communication library are presented, which are divided into two parts because it must be implemented in an Android application and an LG WebOS Smart TV. So, an APK application was developed that will implement the communication library and act as the WebSocket server while the IPK application will act as the client and be placed on the LG WebOS Smart TV.

Proof of concept—Android. The Android application will have two main options: Create a server and create a client that will connect to another server. The main control has five buttons to send messages to the application on the LG WebOS Smart TV ("up," "down," "left," "right," "Send a message to TV"). The "Send a message to TV" option allows you to write a text that can be displayed on the Smart TV. The buttons will act as commands to move the object to be displayed in the IPK application.

Proof of concept—WebOS. In this section, three applications developed with JavaScript, HTML, and the communication library were developed. The first application receives the commands of the mobile application to be able to move a box in a section of the screen, and the movement is slowly done by frames. The second application shows the continuous movement of a box after receiving a command from the APK application. The third application was developed to validate the behavior of Smart TV by simulating a small game. This application shows a box that must move through obstacles.

3.2 Development of a Retro Video Game

This section will show the defined architecture and the two parts of the development where a mobile application was developed for the smartphone, and the WebOS application that will be used on the Smart TV. The video game is called "Aliens On Earth" and the theme used is that of the "Prince of Persia" of Nintendo. The graphics will be of a retro style and the control of the application will be like that of a Nintendo Nes30.

3.2.1 Architecture

Logical architecture. The logical architecture diagram is shown in Fig. 1. For the part of the LG Smart TV, the WebOS operating system is shown as the main layer, then the communication library, the Phaser Framework, the JavaScript language, the HTML language, the CSS for styles are shown and finally the video game "Aliens On Earth." On the other hand, we have the Android Smartphone part. The main layer is the Android operating system, then the communication library, the GSON library for the JSON files, the Java WebSocket library are shown and finally the video game "Aliens On Earth" is shown.

Physical architecture. The logical architecture diagram is shown in Fig. 2, which shows the three main parts that allow interaction between devices. The diagram represents the communication of devices in a local network. Our central point is the Router, which is responsible for creating the local network. Also, LG WebOS Smart TV device and Android device are shown, which will be communicated through the WebSocket protocol.

Android application. The Android application works as the server, so the WebOS application must be connected to start the video game. The main interface of the application is shown in Fig. 3, where it can be seen that a retro style is used. This logo shows the logo, the name of the videogame and the "Iniciar sala y control" button.

After pressing the "Start room and control" button, the command will be displayed to control the character that will be shown in the video game.

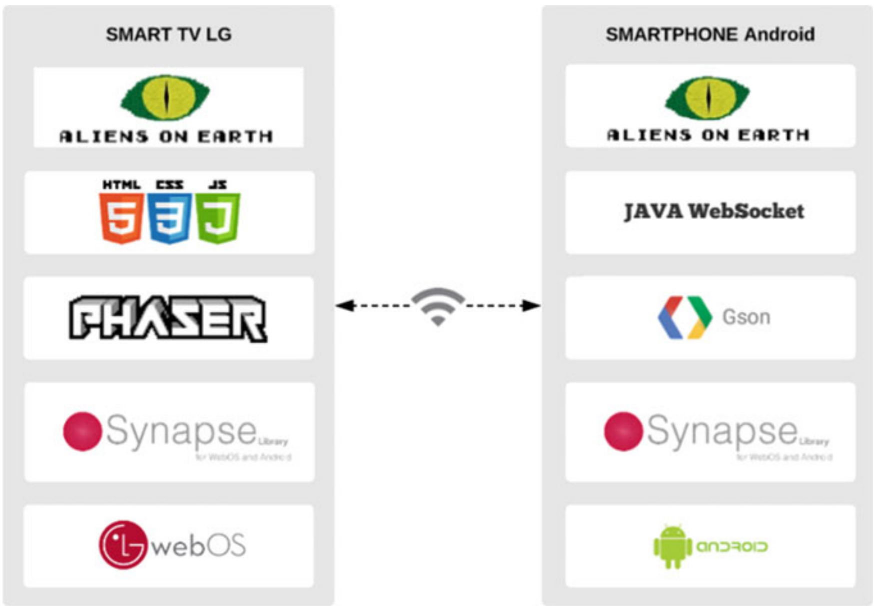


Fig. 1 Logic architecture diagram

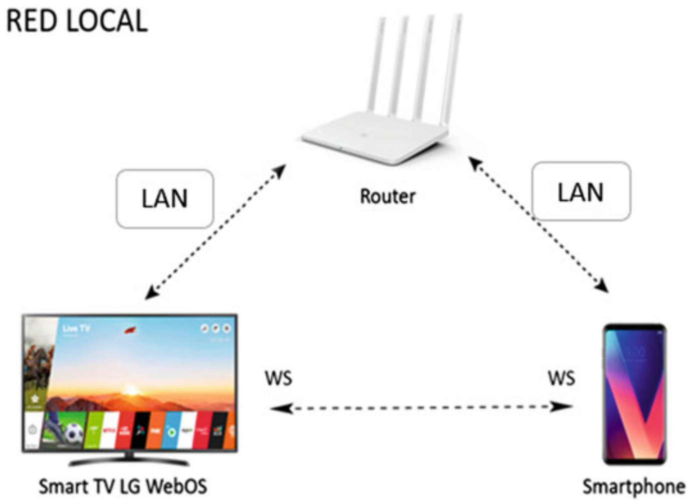


Fig. 2 Physical architecture diagram

Fig. 3 Main interface of the mobile application

WebOS application. The WebOS application works like the WebSocket client and to start the video game it is necessary to follow the steps shown in the main interface, as shown in Fig. 4. In step 1, the user must scan the QR code to be able to download the Android application from Google Play. After downloading and installing the application, the application interface will appear as shown in Fig. 3. Finally, using the Magic Remote of the LG WebOS Smart TV, you must press the "Refresh" button and the WebSocket servers available for that the user can connect. When the Smart TV manages to connect to the WebSocket server, the main videogame interface will be shown as shown in Fig. 5. In which three options will be shown. The "Play" button is responsible for starting the game and displaying Level 1. The "Instructions" button will display a menu containing the description of each button that must be used to control the character and move between levels. Finally, we have the "Exit" button, which is responsible for closing the connection to the WebSocket server and return to the main menu as shown in Fig. 5.

The video game consists of five levels as shown in Fig. 6 and the difficulty increases as it passes between each of them. In each level, there are different obstacles, enemies, potions, control points, levers, and a portal to move to the next level. The character of the video game must be controlled through the retro video game controller shown in the mobile application. To improve the game experience, a background sound and sounds that play when the character acts were added. Also, the video game can be paused when the user needs it without losing the connection to the Android device.

Fig. 4 Main interface of the WebOS application

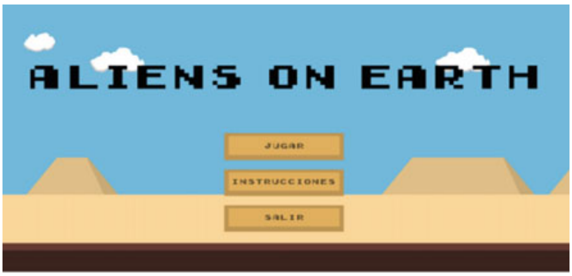


Fig. 5 Main videogame interface

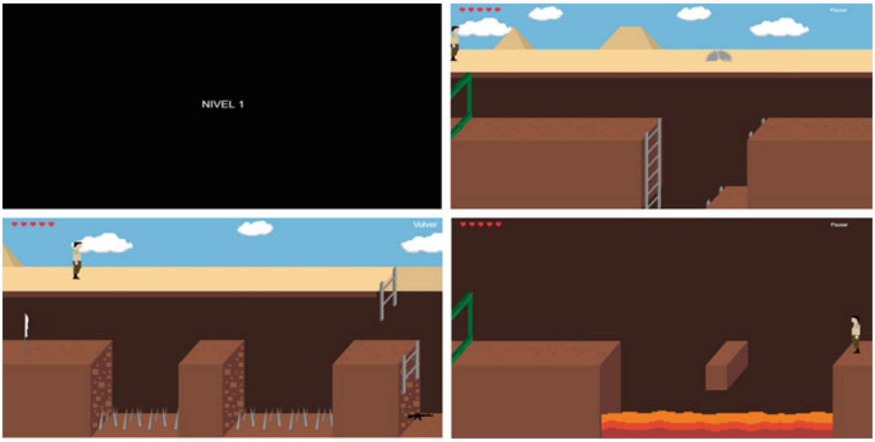


Fig. 6 Video game levels

4 Conclusions

The analysis of the different communication protocols helped us choose the WebSocket protocol, which acts correctly in the environment in which we have used it.

After carrying out the different proofs of concept we have seen that the communication library fulfilled the objective established at the beginning of the project. The communication between the devices involved worked correctly and met each of the requirements to be used in the video game developed.

5 Discussion and Future Work

Finally, we have been able to identify some improvement points that can be developed in the future projects. In the following applications that are developed with the communication library for IOS devices could be used because our research only covered Android devices. In addition, as we have seen that multiple devices can be used, applications that involve the interaction of several smartphones with a Smart TV can be developed allowing more users to interact with an application or a multiplayer video game.

References

1. Francisco, M., Jaime, C., Jhon, I., Gustavo, R., Davinia, H.: Smart TV-smartphone multiscreen interactive middleware for public displays. Sci. World J. **2015**, 534949 (2015)
2. Hae-Jong, J., Hwa-Young, J.: A study on the control of smart device set-top box using image based reverse mirroring technology. Multimedia Tools Appl. **78**(5), 5523–5534
3. Zoran, B., Jelica, P., Veljko, M.: Web performance evaluation for Internet of Things applications. IEEE Access **4**, 6974–6992 (2016)
4. Abdullah, O., Selvakumar M.: A brief review of messaging protocol standards for Internet of Things (IoT). J. Cyber Secur. Mobility **8**(1), 1–14 (2018)
5. Muneer, B., Mohammed, S., Wail, M., Dua', A.: Internet of things' business impact and its application layer protocol in embedded systems. Int. J. Intell. Enterprise **4**(1–2), 143–154 (2017)
6. Chunxue, W., Langfeng, L., Changwei, P., Yan, W., Naixue, X., Changhoon, L.: Design and analysis of an effective graphics collaborative editing system. Eurasip J. Image Video Process. (1), 50 (2019)
7. Björn, Z., Andreas, K.: OnGraX: a web-based system for the collaborative visual analysis of graphs. J. Graph Algorithms Appl. **21**(1), 5–27 (2017)
8. Pedro, D., Nuno, V., Luis, P.: An ideal IoT solution for real-time web monitoring. Cluster Comput. **20**(3), 2193–2209 (2017)

Synchronization Reduction of the Conjugate Gradient Detector Used in Massive MIMO Uplink

Juan Minango, Yuzo Iano, Pablo Minango, David Allan Ibarra, Andrea Flores, and Gabriel Gomes de Oliveira

Abstract In massive multiple-input multiple-output (M-MIMO) systems, with the increase of the number of receiving antennas at the base station (BS), linear detectors are able to achieve near-optimal performance due to M-MIMO channel property of asymptotic orthogonality. However, these detectors involve exact matrix inversion with high complexity. To avoid the matrix inversion, the iterative conjugate gradient (CG) algorithm for signal detection has been recently proposed. However, this algorithm is highly synchronous. In this paper, we propose a modified CG (M-CG) algorithm to reduce the synchronization steps in order to shorten the computational time. Simulation results show the efficiency of the proposed algorithm.

Keywords Massive MIMO · Conjugate gradient · Matrix inversion · MMSE · BER

J. Minango (✉) · D. A. Ibarra
IFTH-SENESCYT, Quito, Ecuador
e-mail: jcarlosminango@gmail.com

D. A. Ibarra
e-mail: david.allan1986@gmail.com

Y. Iano · P. Minango · A. Flores · G. G. de Oliveira
University of Campinas, Campinas, SP, Brazil
e-mail: yuzo@decom.fee.unicamp.br

P. Minango
e-mail: pablodavid218@gmail.com

A. Flores
e-mail: carolina@decom.fee.unicamp.br

G. G. de Oliveira
e-mail: oliveiragomesgabriel@ieee.org

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_21

1 Introduction

Massive multiple-input multiple-output (M-MIMO) systems are considered as an emerging technology for wireless communications due to its high spectral efficiency [1–4]. Unlike the traditional small-scale MIMO (e.g., at most eight antennas in LTE-A [5]), M-MIMO systems with a very large number of antennas at base station (BS) to simultaneously serve a relatively small number of users (in the orders of tens) in the same frequency band have been proposed [6]. Theoretical results have proved that M-MIMO systems provide higher peak data rates and energy efficiency by orders of magnitude with respect to conventional MIMO [7]. However, realizing the attractive merits of M-MIMO systems in practice faces several challenging problems, one of which is the signal detection at the uplink.

Linear detectors, as zero-forcing (ZF) and minimum mean square error (MMSE), achieve near-optimal performance in M-MIMO systems due to the channel property of asymptotical orthogonality [6], which makes them more attractive. Unfortunately, these detectors involve large matrix inversion, whose complexity is cubic with respect to the number of users. To avoid matrix inversion, the iterative conjugate gradient (CG) algorithm for signal detection has proposed in [8], where it has been demonstrated that the CG algorithm requires low computational complexity while achieving near-optimal performance in M-MIMO systems. However, the CG algorithm is highly synchronous, that is, the CG iterative procedure must be stopped until all the data, which are necessary to move to the next step of the procedure, have been computed and collected.

The performance of an algorithm can be measured in two main ways: (i) the amount of time that the algorithm takes to produce the final solution and (ii) the number of floating-point operations (*flops*) performed to arrive at the final solution. The first one is fairly self-explanatory, and the second one represents the total number of all the elementary mathematical operations performed. Remark that the relation between the computational time and the number of flops performed is not always direct [9].

In this paper, we propose a modified CG (M-CG) algorithm for M-MIMO detection, which reduces the synchronization steps in order to shorten the computational time in comparison with CG. The remainder of this paper is organized as follows. Section 2 briefly introduces the uplink of an M-MIMO system. Section 3 introduces the CG algorithm used for signal detection and presents the proposed M-CG algorithm. Simulation results are presented in Sect. 4. Finally, conclusions are drawn in Sect. 5.

Notation: Vectors and matrices are represented by the lower and uppercase boldface, respectively, $\mathbf{x}$ and $\mathbf{X}$. $(\cdot)^T$, $(\cdot)^{-1}$, $|\cdot|$, $\|\cdot\|$, and $\langle\cdot\rangle$ denote the transpose, matrix inversion, absolute value, matrix norm, and inner product, respectively. $\mathbf{I}_N$ denotes the $N \times N$ identity matrix, and $\mathbb{E}\{\cdot\}$ represents expectation. $\Re e\{\cdot\}$ and $\Im m\{\cdot\}$ are the real and imaginary parts of a complex number, respectively.

2 System Model

In this paper, we consider the uplink of an M-MIMO system equipped with M which receive antennas at BS and K single-antenna users, where $K \ll M$, e.g., $M = 128$, $K = 16$. The information bits from all users K are modulated by taking symbols from a set of a constellation alphabet Λ. Let $\mathbf{x}_c \in \Lambda$ denote the $K \times 1$ transmitted signal vector of complex values and $\mathbf{H}_c \in \mathbb{C}^{M\times K}$ denotes the flat Rayleigh fading M-MIMO channel matrix, whose entries are assumed to be independently and identically distributed (i.i.d.) with zero mean and unit variance.

Then, the $M \times 1$ received signal vector $\mathbf{y}_c$ at the BS is given by $\mathbf{y}_c = \mathbf{H}_c\mathbf{x}_c + \mathbf{n}_c$, where $\mathbf{n}_c$ is a $M \times 1$ additive complex white Gaussian noise (AWGN) vector with each element assumed to be i.i.d. complex normal random variables with mean zero and variance N_0. The per-user uplink transmits power for user i which is defined as $\mathbb{E}\left\{|x_i|^2\right\} = E_s$, the average uplink SNR is given by KE_s/N_0 [8], and the complex-valued system model can be also directly converted to a corresponding real-valued system model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{y} = \left[\Re e\{\mathbf{y}_c\}, \Im m\{\mathbf{y}_c\}\right]^T$ is the $2M \times 1$ real equivalent received signal vector $\mathbf{x} = [\Re e\{\mathbf{x}_c\}, \Im m\{\mathbf{x}_c\}]^T$ represents the $2K \times 1$ real equivalent transmitted signal vector and $\mathbf{H}$ denotes the $2M \times 2K$ equivalent channel matrix given by $\mathbf{H} = \begin{bmatrix} \Re e\{\mathbf{H}_c\} & -\Im m\{\mathbf{H}_c\} \\ \Im m\{\mathbf{H}_c\} & \Re e\{\mathbf{H}_c\} \end{bmatrix}$, and $\mathbf{n} = [\Re e\{\mathbf{n}_c\}, \Im m\{\mathbf{n}_c\}]^T$ is the $2M \times 1$ real equivalent noise vector. In this paper, we use the real-valued system model given by (1).

2.1 MMSE Detection

It has been proved that the MMSE linear detector achieves near-optimal performance for the uplink of M-MIMO systems [6], and the estimate of the transmitted signal vector $\mathbf{x}$, denoted as $\hat{\mathbf{x}}$, coming from K different users is given by

$$\hat{\mathbf{x}} = \left(\mathbf{H}^T\mathbf{H} + \sigma^2\mathbf{I}_{2K}\right)^{-1}\mathbf{H}^T\mathbf{y} = \mathbf{W}^{-1}\tilde{\mathbf{y}}, \tag{2}$$

where $\tilde{\mathbf{y}} = \mathbf{H}^T\mathbf{y}$ is the matched-filter output of $\mathbf{y}$, and the MMSE weighting matrix $\mathbf{W}$ is denoted by

$$\mathbf{W} = \left(\mathbf{G} + \sigma^2\mathbf{I}_{2K}\right), \tag{3}$$

where $\mathbf{G} = \mathbf{H}^T\mathbf{H}$ represents the Gram matrix, and $\sigma^2 = N_0/E_s$ is the transmit SNR. Note that the direct computation of $\mathbf{W}^{-1}$ requires a cubic complexity with respect to the number of users, that is $O\left(K^3\right)$.

3 Conjugate Gradient for M-MIMO Signal Detection

In this section, we first introduce the conventional conjugate gradient (CG) algorithm used in [8] for low-complexity M-MIMO signal detection, which is highly synchronous. Then, we propose a modified CG (M-CG) algorithm to reduce the need for synchronization steps.

3.1 Conventional Conjugate Gradient Algorithm

In order to avoid the computation of the matrix inversion of $\mathbf{W}$ given by (3), the MMSE detector can be interpreted as solving the following linear equation [3]:

$$\mathbf{W}\mathbf{x} = \tilde{\mathbf{y}} \tag{4}$$

The authors of [8] have proposed a low-complexity M-MIMO signal detection based on the CG algorithm. Basically, CG is an efficient iterative algorithm that tries to minimize the quadratic function:

$$\phi(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{W}\mathbf{x} - \mathbf{x}^T\tilde{\mathbf{y}} \tag{5}$$

where $\mathbf{W}$ is symmetric $\left(\mathbf{W}^T = \mathbf{W}\right)$ and positive definite $\left(\mathbf{x}^T\mathbf{W}\mathbf{x} > 0\right)$ matrix [3]. The minimum value of $\phi(\mathbf{x})$ given by (5) is $\tilde{\mathbf{y}}^T\mathbf{W}^{-1}\tilde{\mathbf{y}}/2$, achieved by setting $\hat{\mathbf{x}} = \mathbf{W}^{-1}\tilde{\mathbf{y}}$. Thus, minimizing (5) and solving $\mathbf{W}\mathbf{x} = \tilde{\mathbf{y}}$ given by (4) are equivalent problems. CG algorithm iteratively computes the solution $\hat{\mathbf{x}}$, where each iteration requires a low computational complexity of $O\left(K^2\right)$. Algorithm 1 summarizes CG for M-MIMO signal detection.

Algorithm 1 The CG for M-MIMO signal detection
1: input:
2: $\mathbf{H}$ and $\mathbf{y}$
3: **initialization**:
4: $\tilde{\mathbf{y}} = \mathbf{H}^T\mathbf{y}$ and $\mathbf{W} = \mathbf{H}^T\mathbf{H} + \sigma^2\mathbf{I}_{2K}$
5: $\hat{\mathbf{x}}_0 = 0$, $\mathbf{r}_0 = \tilde{\mathbf{y}}$, and $\mathbf{p}_0 = \mathbf{r}_0$
6: **for** $k = 0, \ldots, 2K - 1$ **do**

(continued)

(continued)

Algorithm 1 The CG for M-MIMO signal detection
7: $\alpha_k = \|\mathbf{r}_k\|^2 / \left(\mathbf{p}_k^T \mathbf{W} \mathbf{p}_k\right)$
8: $\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \alpha_k \mathbf{p}_k$
9: $\mathbf{r}_{k+1} = \mathbf{r}_k - \alpha_k \mathbf{W} \mathbf{p}_k$
10: $\beta_{k+1} = \|\mathbf{r}_{k+1}\|^2 / \|\mathbf{r}_k\|^2$
11: $\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \beta_{k+1} \mathbf{p}_k$
12: **end for**
13: **output**:
14: $\hat{\mathbf{x}}_{2K}$

From Algorithm 1, we note that a search direction vector $\mathbf{p}_i$ is chosen at each iteration, which is $\mathbf{W}$-orthogonal to and linearly independent from all the other previous search direction vectors. $\mathbf{W}$-orthogonality means that $\mathbf{p}_i^T \mathbf{W} \mathbf{p}_j = 0, \forall j < i$. Once a suitable direction vector $\mathbf{p}_k$ was found, a step in that direction is taken $\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \alpha_k \mathbf{p}_k$ (see line 8) using an α_k that minimizes the quadratic function $\phi\left(\hat{\mathbf{x}}_{k+1}\right)$. Therefore, a sequence $\left\{\hat{\mathbf{x}}_0, \hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_{2K}\right\}$ is generated and $\hat{\mathbf{x}}_k \rightarrow \mathbf{x}$ as $k \rightarrow 2K$, where $\mathbf{x}$ is the exact solution to (4). Theoretically, the CG iterative procedure (lines 6–12) can reach the exact solution after $2K$ iterations, because of the search direction vectors $\{\mathbf{p}_0, \mathbf{p}_1, \ldots, \mathbf{p}_{2K}\}$ covers the whole search space [9]. However, usually, the CG iterative procedure can be terminated early while still obtaining a solution close to the exact one, which leads to (often significantly) lower computational complexity.

3.2 *Modified Conjugate Gradient Algorithm*

Based on the algorithm 1, we note that CG is highly synchronous. Thus, the computation of $\mathbf{p}_k$ at the k-th iteration depends on having already computed β_k and $\mathbf{r}_k$, d β_k depends on $\mathbf{r}_k$, $\mathbf{r}_k$ depends on α_{k-1}, and α_{k-1} depends on $\mathbf{p}_{k-1}$. Accordingly, we propose a way to compute the search direction vector $\mathbf{p}_k$ without directly relying on already having $\mathbf{r}_k$.

By assuming that we are on the k-th iteration, so we already have k $\mathbf{W}$-orthogonal search direction vectors $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k\}$. Then, the next search direction vector can be computed as follows:

1. Compute: $\mathbf{e} = \mathbf{W}\mathbf{p}_k$.
2. $\mathbf{W}$-orthogonalize $\mathbf{e}$ and $\mathbf{p}_k$ to get $\mathbf{e}'$ as follows: $\mathbf{e}' = \mathbf{e} - \frac{\mathbf{p}_k^T \mathbf{W} \mathbf{e}}{\mathbf{p}_k^T \mathbf{W} \mathbf{p}_k} \mathbf{p}_k$.
3. $\mathbf{W}$-orthogonalize $\mathbf{e}'$ and $\mathbf{p}_{k-1}$ to get $\mathbf{p}_{k+1}$ as follows: $\mathbf{p}_{k+1} = \mathbf{e}' - \frac{\mathbf{p}_{k-1}^T \mathbf{W} \mathbf{e}}{\mathbf{p}_{k-1}^T \mathbf{W} \mathbf{p}_{k-1}} \mathbf{p}_{k-1}$,

where $\mathbf{p}_{k+1}$ is automatically $\mathbf{W}$-orthogonal to $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{k-2}\}$ because $\mathbf{W}\mathbf{p}_k$ is $\mathbf{W}$-orthogonal to $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{k-2}\}$. We can prove the $\mathbf{W}$-orthogonality of $\mathbf{W}\mathbf{p}_k$ with $\mathbf{p}_j$, $j < k - 1$ as follows:

Lemma 1 *The set of vectors* $\mathcal{K}_k = \left\{\tilde{\mathbf{y}}, \mathbf{W}\tilde{\mathbf{y}}, \mathbf{W}^2\tilde{\mathbf{y}}, \cdots, \mathbf{W}^{k-1}\tilde{\mathbf{y}}\right\}$ *that form the k-th Krylov subspace are linearly independent.*

Proof. By supposing that the set of vectors $\mathcal{K}_k$ are not linearly independent for some k and that some vector $\mathbf{W}^i\tilde{\mathbf{y}}$ for $i \in \{0, 1, 2, \cdots, k-1\}$ can be expressed as a linear combination of the vectors in $\left\{\mathbf{W}^i\tilde{\mathbf{y}}\right\}_{j=0,2,\cdots,i-1} = \mathcal{K}_i$. This implies that $\mathbf{W}^{i+1}\tilde{\mathbf{y}} = \mathbf{W}\mathbf{W}^i\tilde{\mathbf{y}}$ can also be expressed in terms of vectors in $\mathcal{K}_i$ and that $span(\mathcal{K}_{i+1}) = span(\mathcal{K}_i)$. On the other hand, the CG algorithm chooses $\hat{\mathbf{x}}_i$ as a linear combination of vectors in $\mathcal{K}_k$ such that the residual vector $\mathbf{r}_k$ is minimized. As the CG iterative procedure solves the problem in a finite number of iterations $2K$, then, $\mathbf{r}_{2K} = 0$. Furthermore, as the span of $\mathcal{K}_k$ is the same as of $\mathcal{K}_i$, thus, $\mathbf{r}_i$ must be a zero vector and the solution has already been found at the i-th iteration, consequently, the vectors in the k-th Krylov subspace $\mathcal{K}_k$ are linearly independent.

Theorem 1 *By assuming that the final solution has not yet been found at the k-th iteration of the CG iterative procedure. Then* $\mathbf{W}\mathbf{p}_k$ *is* $\mathbf{W}$*-orthogonal to* $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{k-2}\}$*, where* $\mathbf{p}_i$ *is the search direction at the i-th iteration.*

Proof By taking any $\mathbf{p}_i$, with $i \in \{1, 2, \ldots, k-2\}$. We can express $\mathbf{W}\mathbf{p}_i$ in terms of $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{k-1}\}$ as $\mathbf{W}\mathbf{p}_i = w_1\mathbf{p}_1 + w_2\mathbf{p}_2 + \cdots + w_{k-1}\mathbf{p}_{k-1}$, where $w_i \in \mathbb{R}$. This can be done because $\mathbf{p}_i$ span the Krylov subspace $\mathcal{K}_i = \left\{\mathbf{p}_1, \mathbf{W}\mathbf{p}_1, \ldots, \mathbf{W}^{i-1}\mathbf{p}_1\right\}$. Since the set of direction vectors $\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{k-1}\}$ are linearly independent based on Lemma 1, the expansion is unique and therefore $w_i = 0$ for $j > i + 1$. Straightaway $\mathbf{p}_i, \mathbf{W}\mathbf{p}_{k\mathbf{W}} = \mathbf{p}_i^T\mathbf{W}\mathbf{W}\mathbf{p}_k = \left(\mathbf{W}\mathbf{p}_i\right)^T\mathbf{W}\mathbf{p}_k$ due to the matrix $\mathbf{W}$ is symmetric, thus, we have that

$$\begin{aligned}\left(\mathbf{W}\mathbf{p}_i\right)^T\mathbf{W}\mathbf{p}_k &= (w_1\mathbf{p}_1 + w_2\mathbf{p}_2 + \cdots + w_{k-2}\mathbf{p}_{k-2})\mathbf{W}\mathbf{p}_k \\ &= w_1\mathbf{p}_1\mathbf{W}\mathbf{p}_k + w_2\mathbf{p}_2\mathbf{W}\mathbf{p}_k + \cdots + w_{k-1}\mathbf{p}_{k-1}\mathbf{W}\mathbf{p}_k = 0 \end{aligned} \tag{6}$$

due to $\mathbf{p}_i$ which is $\mathbf{W}$-orthogonal to $\mathbf{p}_j$ when $i \neq j$. Thus, $\mathbf{W}\mathbf{p}_k$ is $\mathbf{W}$-orthogonal to $\mathbf{p}_k$.

Finally, we use this new search direction vector $\mathbf{p}_{k+1}$ to update $\hat{\mathbf{x}}_{k+1}$ as follows:

$$\alpha_k = \frac{\|\mathbf{p}_{k+1}\|^2}{\mathbf{p}_{k+1}^T\mathbf{W}\mathbf{p}_{k+1}}, \tag{7}$$

$$\hat{\mathbf{x}}_{k+1} = \hat{\mathbf{x}}_k + \alpha_{k+1}\mathbf{p}_{k+1}. \tag{8}$$

One of the key advantages of the proposed M-CG algorithm is the fact that the search direction vectors can be computed independently of any $\hat{\mathbf{x}}_i$ and $\mathbf{r}_i$. Thus, the computations can be done in parallel, and the search direction vectors are used as soon as they are available to minimize the residual vector. In this way, the M-CG algorithm runs in a slightly more parallel way and can shorten the computational time in comparison with the CG algorithm.

4 Simulation Results

Figure 1 compares the bit error rate (BER) performance against the average received SNR (in dB) between the proposed M-CG and CG algorithm reported in [7] for the $M \times K = 128 \times 16$ M-MIMO system, where k denotes the number of iterations. During each channel use, all the K users transmit their uncoded information 64-QAM modulation symbols simultaneously. The BER performance of the MMSE with exact matrix inversion is included as the benchmark for comparison. It is clear that the BER performance of both algorithms improves with the number of iterations, but the proposed M-CG outperforms the CG algorithm when the same number of iteration is used. For three iterations, we note that the M-CG algorithm is able to achieve the near-optimal BER performance of the MMSE. Note that M-CG performance is superior to CG performance because the computation of the search direction vector $\mathbf{p}_k$ implies a better orthogonalization between search direction vectors, and consequently, a faster convergence rate is achieved.

In Fig. 2, we show the computational time comparison between the M-CG and CG algorithm for the $M \times K = 128 \times 16$ M-MIMO system with 64-QAM modulation, where three iterations were considered (see Fig. 1). The computational time of the MMSE detectors with matrix inversion is included as the benchmark. From this figure, we can clearly notice that the M-CG algorithm obtains a computational time reduction of about 66.42% in comparison with CG, which shows the shortened computational time of our proposed M-CG algorithm for M-MIMO signal detection.

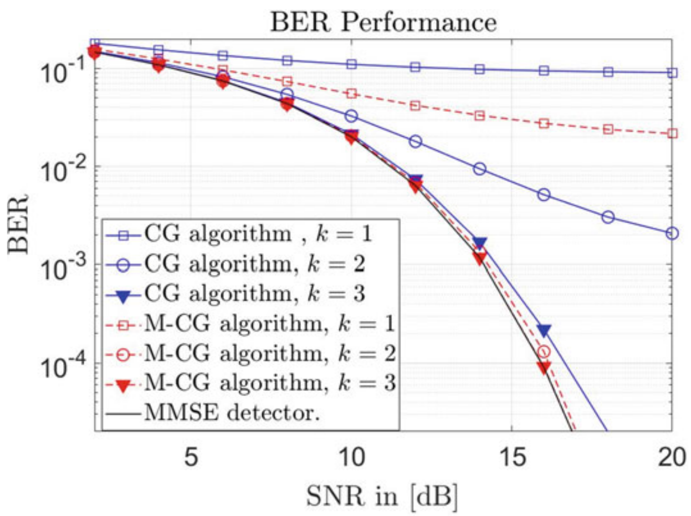


Fig. 1 BER performance comparison in the $M \times K = 128 \times 16$ M-MIMO system

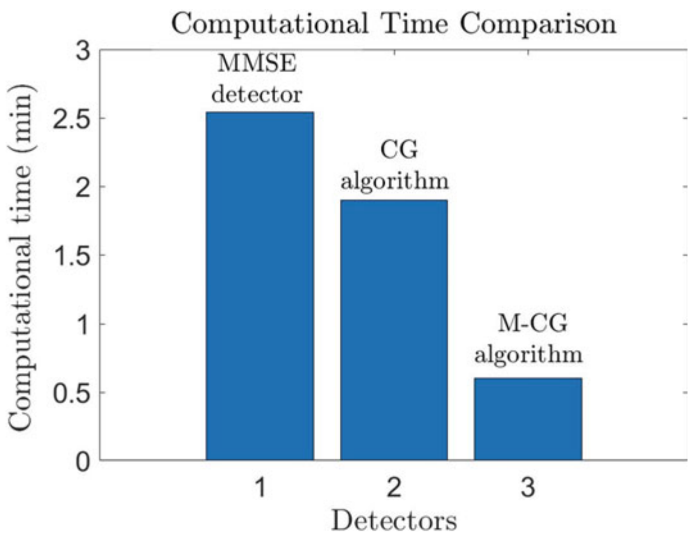


Fig. 2 Computational time comparison in the $M \times K = 128 \times 16$ M-MIMO system

Table 1 Computational time reduction for different M-MIMO relations M/K

M/K	Computational time for CG (min)	Computational time for M-CG (min)	Computational time reduction (%)
2	54.404	23.647	56.53%
4	42.314	16.653	60.64%
8	21.018	7.379	64.89%
16	19.973	6.783	64.03%
32	18.042	5.497	69.53%

Table 1 shows the comparison in terms of the computational time reduction between the proposed M-CG and CG algorithm for different M-MIMO system relations M/K, where three iterations were considered for the convergence of both algorithms. The computational time of both algorithms is given in minutes. We note that the M-CG algorithm achieves a reduction of more than 50% in comparison with the CG algorithm in all cases. Furthermore, as the M-MIMO relation M/K increases, the percentage of computational time reduction increases too, thus, for a $M/K = 32$, the M-CG algorithm achieves a computational time reduction of around 69.53% in comparison with CG algorithm.

5 Conclusion

In this paper, a modification to the conjugate gradient algorithm for M-MIMO signal detection has been proposed, which reduces some synchronization steps in order to shorten the computational time and allow parallel implementation arriving at the final solution quickly. Simulation results verify the efficiency of the synchronization reduction of the proposed M-CG algorithm for M-MIMO signal detection.

References

1. Minango, J., de Almeida, C.: Low complexity zero forcing detector based on Newton Schultz iterative algorithm for massive MIMO systems. IEEE Trans. Veh. Technol. **67**(12), 11 759–711 766, Dec 2018
2. Thompson, J., Ge, X., Wu, H.C., Irmer, R., Jiang, H., Fettweis, G., Alamouti, S.: 5G wireless communication systems: prospects and challenges (Guest Editorial). IEEE Commun. Mag. **52**(2), 62–64 (2014)
3. Minango, J., de Almeida, C.: Optimum and quasi-optimum relaxation parameters for low-complexity massive MIMO detector based on Richardson method. Electron. Lett. **53**(16), 1114–1115 (2017)
4. Minango, J., de Almeida, C.: A low-complexity linear precoding algorithm based on Jacobi method for massive MIMO systems. In: 2018 IEEE 87th VTC Spring, June 2018, pp. 1–5
5. Zhang, S., Kong, D., Mellios, E., Hilton, G., Nix, A., Thomas, T.A., Ghosh, A.: Impact of BS antenna number and array geometry on single-user LTE-A data throughputs in realistic macro and pico cellular environments. IEEE Wireless Communication and Newton Conference (WCNC), pp. 1464–1469, Mar 2015
6. Rusek, F., Persson, D., Lau, B.K., Larsson, E.G., Marzetta, T.L., Edfors, O., Tufvesson, F.: Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Signal Process. Mag. **30**(1), 40–60 (2013)
7. Minango, J., Altamirano, C.D., de Almeida, C.: Performance difference between zero-forcing and maximum likelihood detectors in massive MIMO systems. Electron. Lett. **54**(25), 1464–1466 (2018)
8. Yin, B., Wu, M., Cavallaro, J.R., Studer, C.: Conjugate gradient-based soft-output detection and precoding in massive MIMO systems. In: IEEE GLOBECOM. IEEE, pp. 3696–3701 (2014)
9. Minango, J., de Almeida, C., Daniel Altamirano, C.: Low-complexity MMSE detector for massive MIMO systems based on damped Jacobi method. In: 2017 IEEE PIMRC, pp. 1–5, Oct 2017

Analysis of AMQP for Industrial Internet of Things Based on Low-Cost Automation

Carlos A. Garcia, Jose E. Naranjo, and Marcelo V. Garcia

Abstract Internet of things (IoT) is a new trend increasingly used in industrial applications, and one of the key technologies is the low-cost automation used as the controller during the manufacturing process. This approach gives new insights that enhance decision-making processes and allow a business advantage. This research work uses a low-cost board like Raspberry Pi to control a FESTO industrial process integrating shop-floor communications using the AMQP (Advanced Message Queuing Protocol) protocol. Most of the recent researches are conducted in the use of low-cost automation on high-level abstract models, and this paper shows specific architectural frameworks in a simulated process.

Keywords Shop-floor communications · Low-cost automation · AMQP protocol · Internet of things (IoT)

1 Introduction

Today, the application of the IoT becomes a competitive advantage, as it describes a scenario in which objects are identified and connected to the Internet, enabling remote control of critical or relevant situations for a domain, through sensors and strategically distributed actuators [1]. However, in order to detect such situations, it is necessary to communicate, store, analyses, and process efficiently a large amount of information generated each day by these intelligent devices [2].

Several communication models are compatible with messaging technologies systems that define how information is exchanged between producer and consumer.

C. A. Garcia · M. V. Garcia (✉)
Universidad Tecnica de Ambato, UTA, Ambato 180103, Ecuador
e-mail: mv.garcia@uta.edu.ec; mgarcia294@ehu.eus

C. A. Garcia
e-mail: ca.garcia@uta.edu.ec

J. E. Naranjo · M. V. Garcia
University of Basque Country, UPV/EHU, Bilbao 48013, Spain
e-mail: jnaranjo0463@uta.edu.ec

© Springer Nature Switzerland AG 2021
Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_22

The producer/consumer method has different problems, for example, when there is no consumer of the message, the method cannot manage the semantics for the queues and distribution of the information among several consumers. A protocol that manages the complexity of semantics is required in industrial systems; for this reason, the scope of standard protocols for messaging systems becomes the key to the efficiency of cyber-physical production systems (CPPS) developed in low-cost devices [3].

The protocol AMQP (Advanced Message Queuing Protocol) has a set of communication standards that control the whole messaging process in brokers. This is an open protocol that allows messaging interoperability between industrial systems, ease the information exchange implementing the machine-to-machine (M2M) architecture. This protocol is defined both in the network layer protocol and a high-level architecture for the message within brokers [4].

AMQP is an oriented to messages protocol and defines a group of messages capabilities which should be available by implementation of AMQP compliant server. AMQP focuses on process-to-process communication over IP/TCP networks. AMQP is into the compact protocol schema because AMQP is a binary protocol; all that is sent over AMQP is a binary data. A binary protocol avoids sending useless data across the communication channel. This protocol must be tested in industrial environments to calculate the efficiency of the message size and communication rates. This protocol must be used in low-cost CPPs systems [5].

The paper shows a method for shop-floor integration using low-cost automation for CPPS architecture using the AMQP protocol. This paper uses like low-cost device the Raspberry Pi (RPI) board. The goal of this paper is to identify the characteristics and limits in industrial simulated environments across IoT architectures. The results show that this protocol has low latency and fluctuations, low message losses during a period of time where massive AMQP messages from the controller are sending over the architecture [6].

The paper outline is structured as follows: Sect. 2 shows the related literature; Sect. 3 presents the case study used for the development of the CPPS architecture and AMPQ protocol transmission. In Sect. 4, the architecture and methodology of the communication system are shown. Later in Sect. 5, the implementation of the shop-floor integration in the Festo™ Testing station is presented; the analysis of results obtained in the communication architecture is addressed in Sect. 6; finally, the conclusions and ongoing work are stated in Sect. 7.

2 Background Literature

This section states the related literature about the use of protocols based on messaging protocols. Furthermore, the use of AMQP is analyzed in industrial environments where is used for solving issues in communications networks.

The research work addressed by Subramoni et al. [4] shows the use of AMQP designed a suite of benchmarks for AMQPpsilas Direct, Fanout, and Topic Exchange types. The authors of this research teste the performance of this protocol with a simulated stock exchange application. The results presented show an appreciable bottleneck as far as scalability is concerned. For this reason, the authors use other protocols such as remote direct memory access but to integrate information in industrial systems is considered AMQP protocol for future works.

Luzuriaga et al. [7] highlight that message-oriented middleware (MOM) is a software infrastructure that supports sending and receiving messages between distributed systems. The authors test two types of MOM protocols, the AMQP and MQTT (Message Queuing Telemetry Transport) protocol. This research shows an evaluation in scenarios characterizing their behavior of latency, message loss, and jitter values. The conclusions are that AMQP protocol is useful to build scalable, reliable, and advanced clustering messaging infrastructures over a communication channel.

Nowadays, the researchers compare in an IoT scenario the AMQP and MQTT protocols. This kind of protocol has a routing feature, and this is the process by which an exchange decides which queues to place the message on. Unlike MQTT, the AMQP has reliability and security features. Reliability offers service quality levels (QoS) for message delivery: format destabilization (unreliable) and settlement format (reliable). This protocol can be configured to ensure the messages are delivered to the subscriber. In conclusion, AMQP is used because of supports far better use of security resources than MQTT [8, 9].

The paper addressed by Wang et al. [10] shows the use of Wireshark as a network protocol analyzer. This software presents the structure of different TCP/IP protocols in a comprehensive way. Specifically, when this software is used to analyze AMQP protocol, all the fields of the packet header and layer could be monitored. This is a tool that networks administrators use to implement traffic analysis tasks.

In the same focus, which previously we reported the integration based on AMQP protocol with distinct technologies within the industrial environment [11, 12].

All of the researches mentioned above show the methodologies to compare the MOM protocols but there are not papers focusing on the transmission of data in industrial systems, where the speed and loss of data transmission are key technologies in the use of this type of protocol at industrial level. For this reason, this paper presents the development of AMQP in low-cost boards integrating the data from a programmable logic controller (PLC) and transmitting it to a Web interface.

3 Case Study

Due to industry quality control methods, which are often inappropriate, and products are exposed to errors as a result of malfunction and improper use of materials used by workers for such work, therefore, it becomes the main constraint to improving and finding solutions. The goal of the case study integrates a shop-floor communication using a communication stack of AMQP protocol for low-cost CPPS. This case study classifies the quality of the pieces in the Festo™ testing workstation using an inspection analysis model called "pass–fail report." This report is sending to an AMQP broker that monitors the process in real time.

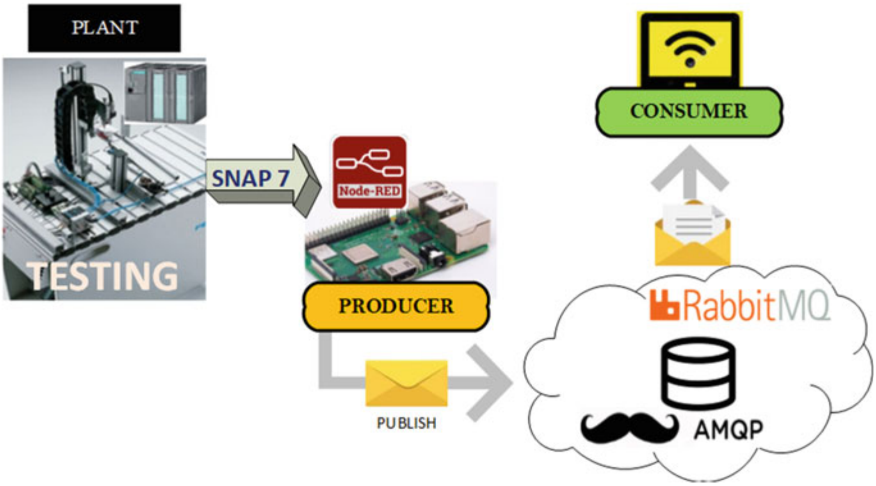


Fig. 1 Study case communication system architecture

The testing station of FESTO is in charge of detecting the material color and the height of workpieces and classifies them. The classification criteria used is if the height of the workpieces is 22.5 mm (black workpieces), these are rejected. If the height of the workpieces is 25 mm (red and silver workpieces), these are accepted. This information is sending to a Web interface using an AMQP broker. In the Web interface, the user sees in real time how many rejected workpieces are in the process. The interface will report these issues and send alert messages to the operator of this process.

This interface permits to the user take control of the testing process using concepts of digitalization and lean workflow. This open-source tool gives a manufacturing dashboard to increase the quality and profitability of the process. The communication to the process using an AMQP protocol permits to develop an industrial IoT architecture. This type of system is aimed to show detailed production and maintenance information that is vital in forecasting and capacity planning. Figure 1 shows the proposed study case.

4 Communication Architecture Proposed

Industrial CPPSs are devices with high computing capacity. The new trend is used CPPSs to monitor and control the industrial process in real time. This system uses the producer/consumer architecture using the AMQP protocol. The CPPS is designed using a low-cost board like Raspberry Pi (RPI). The RPI reads the memory data from a programmable logic controller (PLC). This PLC controls all the testing process. The RPI is considered as the hardware of the CPPSs system.

To read the PLC memory where all tags of the process are, the RPI uses Snap 7 which is an open-source stack for Ethernet communication suite for interfacing natively with Siemens S7 PLCs. Once the tags are read, these values are sending to an AMQP protocol using the Node-RED software.

Node-RED uses the node called node-red-contrib-amqp to send the values of the tags to the AMQP broker. The AMQP consumer is into the Web interface which works as a dashboard to show the data.

To implement AMQP, the authors use the software RabbitMQ™, this an open-source tool to create message brokers. This tool allows creating a lot of a different kind of brokers, since brokers in embedded devices until large-scale brokers. RabbitMQ™ permits to create instances at the cloud level to develop an IoT architecture. With these open-source tools, the author of the papers uses AMQP communication protocol at the industrial level. All of this aimed to analyze the performance and latency of the protocol, and if the result is good for the industrial plant.

5 Implementation Proposed

5.1 Shop-Floor Data Integration on Festo™ MPS Testing

The simulated factory process using the Festo™ MPS Testing platform permits integration data from the field level using the AMQP protocol. This is possible due to the use of low-cost devices as RPI. The RPI reads the PLC memory using the Snap 7 open-source stacks. Snap 7 uses Ethernet communication to read and write data and specifically uses the Profinet protocol. Once all the values of inputs/outputs are in the RPI memory, it uses the Node-RED software to create AMQP nodes which are used to send the data to the broker. The nodes are developed in JavaScript Object Notation (JSON) format to ease the integration of the data.

The Node-RED node verifies some characteristics like IP address, port, rack, and slot configured in the CPU. The Node-RED software flow works by passing messages between nodes. The flow has two properties, msg-topic, and msg-payload, in the transmission of cyclically loaded values every time a variable changes its value.

5.2 Communication Flow in Node-RED

These messages are passed in as an object called msg. The body of the message is into a property called msg.payload. If the client identifies the IP network and port, it can access through a web browser.

Snap7 server and AMQP broker are considered by Node-RED as components, and this software communicates them using an abstract connection. As explained before, the body of the message is in the payload property. This is important for the message topic transfer component. The property msg.topic is a string that contains the routing-key of the AMQP protocol message. The topic property takes the value of the user final destination of the message. The communication message flow continues across the AMQP broker if the component that accepts the message load has configured the property msg.amqpMessage.

The instance of the AMQP output data verifies for the first time the user ID, JSON scheme, and password of the message. Once the information of the output node is verified, all the consumer that wants to connect to the broker identifies data in the payload property.

Figures 2 and 3 present the design of the communication architecture. In the Node-RED dashboard, the process starts with the nodes to communicate with the Snap 7 server. In this figure, there are two configured inputs whose aim is to read the memory from PLC where the sensor's values are stored. The aim is to measure the height and detect the presence in the input of the machine the workpieces. These signals are sending to the AMQP broker implemented using RabbitMQ.

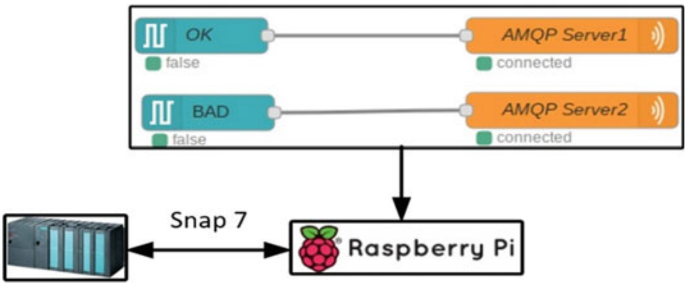


Fig. 2 Data integration on CPPS using Snap7 and AMQP server

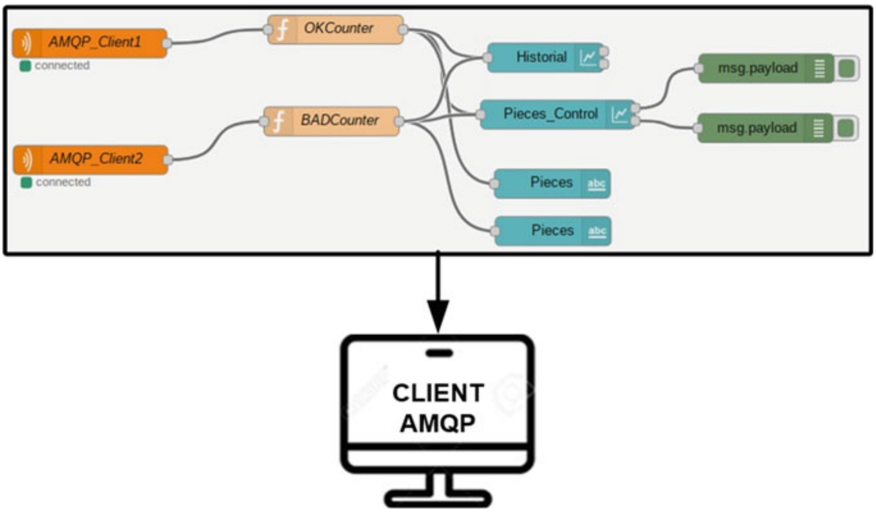


Fig. 3 Data visualization on client side

The nodes which are created with Node-RED send messages to the broker. The broker has inside an exchange, and this part of the broker manages the message to the queues. The functions are: (i) append messages to many queues. (ii) get discarded messages. (iii) The rules for the exchange type. Next, the data from the queues are sending to the consumers. In this paper, the Web interface shows the workpieces accepted and rejected in a statistical chart. A historical chart shows the elements that pass through the process are stored in a timeline.

6 Discussion of Research Outcomes

The measuring module is made by an analog height measuring sensor. This value is reading by the RPI using the Snap 7 library. Later, this value is produced by the Node-RED using the AMPQ protocol. An IoT dashboard is presented in the research. This dashboard shows 100 workpieces measured. The result of the measure is presented as a rejected or accepted product. On the left side, there is a pie chart where the blue part states the defective products, while the light blue part shows the non-defective products (see Fig. 4). On the right side of Fig. 4, there is a historical time-series graph displays data during one day of production. This type of graph measures trends over time. The x-axis shows the specific time where the workpieces are measured. This historical chart shows the products rejected and accepted over the day. This helps the industrial operator to decision making in the plant.

Figure 4 shows additionally the IP address where is the RPI acquiring the data. These two graphical charts permit that the user sees all the possible difficulties in the process. With this information, the operator can take corrective or preventive maintenance in the manufacturing processes.

The AMQP packet frame is analyzed using the software Wireshark (See Fig. 5). Wireshark is a network tool that displays in a real-time chart the amount of bandwidth used by the AMQP producer/consumer activity. The default AMQP port is 5672. The chart shows that the AMQP protocol is efficient for transporting the messages over the TCP/IP network. There is not a high bottleneck in the transmission. The load for RPI is low for this reason, and this protocol is suitable to be installed in the low-cost boards.

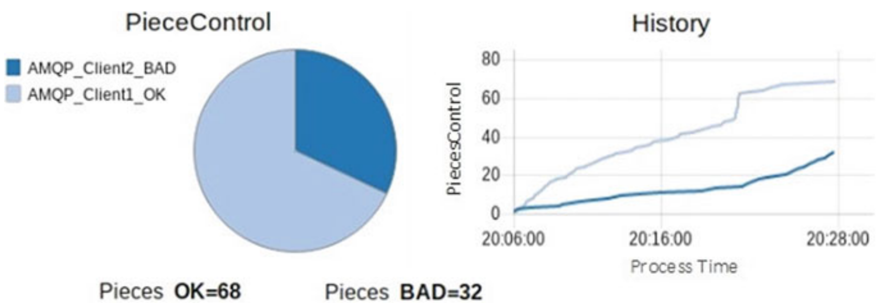


Fig. 4 Screen with statistical data

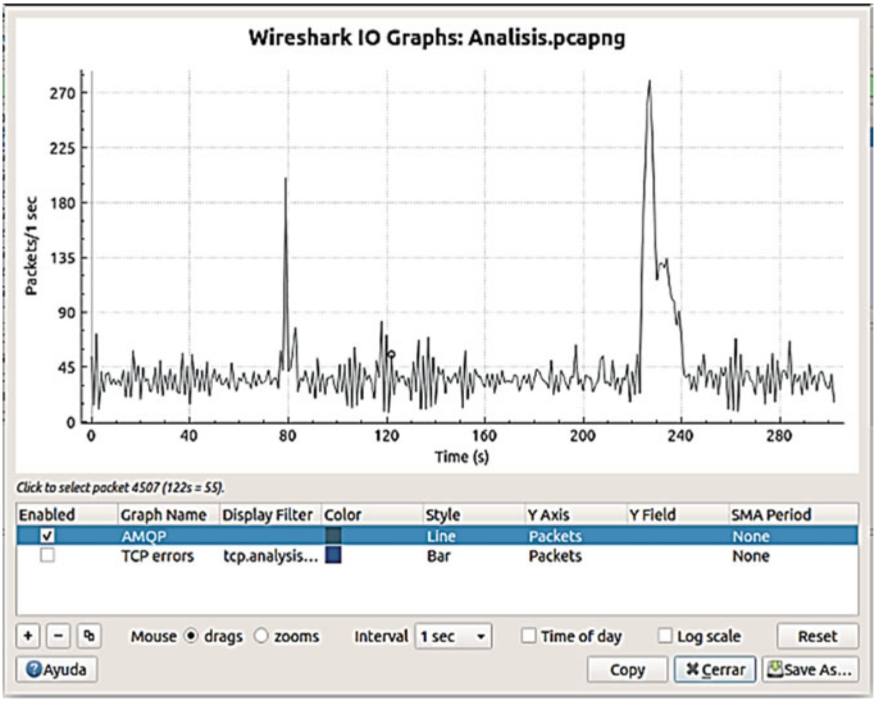


Fig. 5 AMQP analyzer over architecture proposed

7 Conclusions and Ongoing Work

The low-cost shop-floor integration via the AMQP protocol is presented in this research. The proposed architecture has a low bandwidth consumption during the test using the Wireshark software. The Web interface shows in real time the accepted and rejected workpieces. This interface is an IIoT dashboard that permits the operator of the simulated factory process to take a good decision when evaluating the production process.

The AMQP protocol is a light MOM middleware that permits the communication of the data with high-quality control using a secure message routing, good management of queues and reliable delivery of the message to the consumer. This protocol was successfully used with a board with low capabilities as RPI without performed an overheating of its.

The traffic measured across the network has fewer errors because the messages forwarded to the broker or producer objects are not present in the Wireshark chart. The AMQP model used in this paper has message acknowledgments, and this type of message is delivered to the broker notifying the reception of the message produced.

As future ongoing works will be the evaluation of these results with other application contexts and next determine according to characteristics in the communication, their ability is to adapt to other industrial applications. This dashboard used for Industrial IoT applications could be implemented using other industrial protocols like OPC UA or XMPP.

Acknowledgements This work was financed by Universidad Tecnica de Ambato (UTA) and their Research and Development Department (DIDE) under project CONIN-P-256-2019.

References

1. Garcia, M.V., Irisarri, E., Pérez, F., Estévez, E., Marcos, M.: An open CPPS automation architecture based on IEC-61499 over OPC-UA for flexible manufacturing in Oil&Gas Industry. IFAC-PapersOnLine **50**, 1231–1238 (2017)
2. Mizouni, R., Serhani, M.A., Dssouli, R., Benharref, A., Taleb, I.: Performance evaluation of mobile web services. In: 2011 IEEE Ninth European Conference on Web Services [Internet]. Lugano, Switzerland: IEEE; 2011 [cited 2019 Aug 14], pp. 184–91. Available from: http://ieeexplore.ieee.org/document/6061092/
3. Garcia, M.V., Pérez, F., Calvo, I., Morán, G.: Building industrial CPS with the IEC 61499 standard on low-cost hardware platforms. In: 19th IEEE International Conference on Emerging Technologies and Factory Automation, ETFA 2014 (2015)
4. Subramoni, H., Marsh, G., Narravula, S., Lai, P., Panda, D.K.: Design and evaluation of benchmarks for financial applications using Advanced Message Queuing Protocol (AMQP) over InfiniBand. In: 2008 Workshop on High Performance Computational Finance [Internet]. Austin, TX, USA: IEEE; 2008 [cited 2019 Aug 14], pp. 1–8. Available from: http://ieeexplore.ieee.org/document/4745404/
5. Vinoski, S.: Advanced message queuing protocol. IEEE Internet Comput. **10**, 87–89 (2006)
6. FESTO Systems: FESTO MPS 200 Systems [Internet], [cited 2020 Jan 15]. Available from: https://www.festo-didactic.com/int-es/learning-systems/mps-sistema-de-produccion-modular/mps-200/?fbid=aW50LmVzLjU1Ny4xNC4yMC44NTY
7. Luzuriaga, J.E., Perez, M., Boronat, P., Cano, J.C., Calafate, C., Manzoni, P.: A comparative evaluation of AMQP and MQTT protocols over unstable and mobile networks. In: 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC) [Internet]. Las Vegas, NV, USA: IEEE; 2015 [cited 2019 Aug 14], pp. 931–6. Available from: http://ieeexplore.ieee.org/document/7158101/
8. Bezerra, D., Roque Aschoff, R., Szabo, G., Fawzi Hadj Sadok, D.: An IoT protocol evaluation in a smart factory environment. In: 2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE) [Internet]. Joao Pessoa: IEEE; 2018 [cited 2019 May 16], pp. 118–123. Available from: https://ieeexplore.ieee.org/document/8588537/
9. Cai, Y., Qi, D.: Control protocols design for cyber-physical systems. In: 2015 IEEE Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) [Internet]. Chongqing, China: IEEE; 2015 [cited 2019 Aug 14], pp. 668–671. Available from: http://ieeexplore.ieee.org/document/7428638/
10. Wang, S., Xu, D.S., Yan, S.L.: Analysis and application of Wireshark in TCP/IP protocol teaching. In: 2010 International Conference on E-Health Networking Digital Ecosystems and Technologies (EDT) [Internet]. Shenzhen, China: IEEE; 2010 [cited 2019 Aug 14], pp. 269–72. Available from: http://ieeexplore.ieee.org/document/5496372/

11. Caiza, G., Llamuca, E.S., Garcia, C.A., Gallardo-Cardenas, F., Lanas, D., Garcia, M.V.: Industrial shop-floor integration based on AMQP protocol in an IoT environment. In: 2019 IEEE Fourth Ecuador Technical Chapters Meeting (ETCM), Guayaquil, Ecuador, pp. 1–6 (2019)
12. Llamuca, E.S., Garcia, C.A., Naranjo, J.E., Garcia, M.V.: Cyber-physical production systems for industrial shop-floor integration based on AMQP. In: 2019 International Conference on Information Systems and Software Technologies (ICI2ST), Quito, Ecuador, pp. 48–54 (2019)

An Interval Insight to Adomian Decomposition Method for Ordinary Differential Systems by Considering Uncertain Coefficients with Chebyshev Polynomials

Navid Razmjooy, Mehdi Ramezani, and Vania V. Estrela

Abstract Generally, parameters in the mathematical models of engineering problems are considered deterministic, although, in practice, there are always some uncertainties in the model parameters. Uncertainty can make an accurate or even wrong representation for the analyzed model. There is a wide reason which causes the uncertainties, like measurement error, inhomogeneity of the process, etc. This problem leads researches to analyze the problem from a different point of view. When the uncertainty is present in the process, traditional methods of exact values cannot solve the problem with no inaccuracies and mistakes. Interval analysis is a method that can be utilized to solve these kinds of problems. In this paper, an interval Adomian decomposition method combined with Chebyshev polynomial is introduced. The proposed interval Adomian method is then validated through ODE systems. The simulation results are applied on four practical case studies, and the results are compared with interval Euler and Taylor methods. The final results show that the proposed methodology has good accuracy to find the proper interval and to effectively handle the wrapping effect to sharpen the range of non-monotonic intervals.

Keywords Interval analysis · Adomian decomposition method · Chebyshev polynomials · Ordinary differential equations · Uncertainty

---

N. Razmjooy (✉) · M. Ramezani
Department of Electrical and Control Engineering, Tafresh University, Tafresh 39518 79611, Iran
e-mail: navid.razmjooy@hotmail.com

M. Ramezani
e-mail: ramezani@aut.ac.ir

V. V. Estrela
Department of Telecommunications, Fluminense Federal University (UFF), Niteroi, RJ, Brazil
e-mail: vania.estrela.phd@ieee.org

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_23

1 Introduction

Generally, during the mathematical modeling of a practical phenomenon, the corresponding parameters have considered as exact values. However, the parameters of these phenomena have some uncertainties. These uncertainties can be generated from different reasons like neglecting some nonlinear terms on the model, simplifications, etc. These uncertainties lead the researcher to solve problems in the wrong way, and consequently, the final result will be wrong. There are different ways to illustrate these uncertainties [1]. Uncertainties can be modeled by probabilistic variables, fuzzy variables, interval variables, etc. But the most proper method is to use the interval arithmetic. In the interval arithmetic, uncertainties stand throughout a definite lower and upper bounds. In other words, although the uncertainty quantity is unknown, an interval can be defined to them. Ordinary differential equations (ODEs) include a wide range of applications like system modeling, optimal control, etc. There are different techniques that are introduced to solve these types of systems.

In the recent decades, decomposition methods have been shown as an effective, easy, and accurate method to solve a great deal of linear and nonlinear, ordinary, partial, deterministic, or stochastic differential equations by approximation. They have also rapid convergence to achieve accurate solutions [2–5].

Among these methods, the Adomian decomposition method has been transformed into a popular technique for solving functional differential equations like ordinary differential equations [6], differential-algebraic equations [7], nonlinear fractional differential equations [8], delay differential equations [9], etc.

From the above, it is clear that the Adomian decomposition method is a proper method for solving the ODE systems. Now, what happened if these systems have some uncertainties. In this study, an improved domain decomposition method is introduced to achieve a proper and robust solution. The main idea is to find a proper interval bound which keeps the system stable even if the parameters are changed in the considered interval uncertainty.

We also benefit from the Chebyshev orthogonal polynomial for simplifying the complicated source terms to achieve a more compressed solution rather than the Taylor series.

2 Interval Arithmetic

When a mathematical model of an engineering system is built, there are always some simplifications; although simplification reduces the system complication, it makes some natural uncertainties on the model. In other words, some uncertain coefficients are appeared in the model [10]. Hence, utilizing normal methods for modeling or solving these types of systems cause some problems. However, an uncertainty coefficient has an unknown quantity, but it is bounded and can be considered in an interval. Interval arithmetic provides a set of methods to keep track of these uncertainties during the computations [11]. The interval set for an interval number can be described as:

$$\begin{gathered} X = \left[\underline{\boldsymbol{x}}, \overline{\boldsymbol{x}}\right] = \{x | x \in R \cup \{-\infty, \infty\},\ \underline{\boldsymbol{x}} \le x \le \overline{\boldsymbol{x}}\}, \\ I(R) = \{X | X = \left[\underline{\boldsymbol{x}}, \overline{\boldsymbol{x}}\right],\ \underline{\boldsymbol{x}}, \overline{\boldsymbol{x}} \in R,\ \underline{\boldsymbol{x}} \le \overline{\boldsymbol{x}}\}, \end{gathered} \tag{1}$$

where X defines an interval integer over $I(\mathbb{R})$ and $\underline{\boldsymbol{x}}$ and $\overline{\boldsymbol{x}}$ are the lower and upper bounds, respectively. The midpoint value, the width of the interval number, and the radius of an interval can be defined as:

$$\boldsymbol{x_c} = \left(\overline{\boldsymbol{x}} + \underline{\boldsymbol{x}}\right)/2,\ \ \boldsymbol{x_r} = \boldsymbol{x_w}/2\,, \tag{2}$$

$$\boldsymbol{x_w} = \overline{\boldsymbol{x}} - \underline{\boldsymbol{x}} \tag{3}$$

The basic interval arithmetic operations are described so that the interval guarantees the reliability of interval results. The main interval arithmetic operations between two interval numbers are:

$$\boldsymbol{X} + \boldsymbol{Y} = \left[\underline{\boldsymbol{x}} + \underline{\boldsymbol{y}},\ \bar{\boldsymbol{x}} + \bar{\boldsymbol{y}}\right] \tag{4}$$

$$\boldsymbol{X} - \boldsymbol{Y} = \left[\underline{\boldsymbol{x}} - \bar{\boldsymbol{y}},\ \bar{\boldsymbol{x}} - \underline{\boldsymbol{y}}\right] \tag{5}$$

$$\boldsymbol{X} \times \boldsymbol{Y} = \left[\boldsymbol{min}\left\{\underline{\boldsymbol{x}}\underline{\boldsymbol{y}},\ \bar{\boldsymbol{x}}\underline{\boldsymbol{y}},\ \underline{\boldsymbol{x}}\bar{\boldsymbol{y}},\ \bar{\boldsymbol{x}}\bar{\boldsymbol{y}}\right\},\ \ \mathbf{max}\left\{\underline{\boldsymbol{x}}\underline{\boldsymbol{y}},\ \bar{\boldsymbol{x}}\underline{\boldsymbol{y}}, \underline{\boldsymbol{x}}\bar{\boldsymbol{y}},\ \bar{\boldsymbol{x}}\bar{\boldsymbol{y}}\right\}\right] \tag{6}$$

$$\boldsymbol{X}/\boldsymbol{Y} = \boldsymbol{X} \times 1/\boldsymbol{Y}, \tag{7}$$

$$1/\boldsymbol{Y} = \left[1/\bar{\boldsymbol{y}}, 1/\underline{\boldsymbol{y}}],\ \ 0 \notin [\underline{\boldsymbol{y}}, \bar{\boldsymbol{y}}\right]$$

$$\boldsymbol{X^n} = \left\{ \begin{array}{c} \left[0, \mathbf{max}\left(\underline{\boldsymbol{x}}^{\boldsymbol{n}}, \bar{\boldsymbol{x}}^{\boldsymbol{n}}\right)\right],\ \boldsymbol{n} = 2\boldsymbol{k}\ \ \ , 0 \in [\boldsymbol{x}] \\ \left[\mathbf{min}\left(\underline{\boldsymbol{x}}^{\boldsymbol{n}}, \bar{\boldsymbol{x}}^{\boldsymbol{n}}\right), \mathbf{max}\left(\underline{\boldsymbol{x}}^{\boldsymbol{n}}, \bar{\boldsymbol{x}}^{\boldsymbol{n}}\right)\right],\ \boldsymbol{n} = 2\boldsymbol{k}, 0 \notin [\boldsymbol{x}] \\ \left[\underline{\boldsymbol{x}}^{\boldsymbol{n}}, \bar{\boldsymbol{x}}^{\boldsymbol{n}}\right],\ \boldsymbol{n} = 2\boldsymbol{k} + 1 \end{array} \right\} \tag{8}$$

The interval function F is an inclusion function of $f(F) \subset F(X)$ if $\forall X \in I(\mathbb{R})$. The main objective of this study is to find an interval function F from f to achieve an interval form of our method.

3 Chebyshev-Based Adomian Decomposition Method

In the Adomian decomposition method, the unknown function (i.e., $y(x)$) is decomposed into an infinite series $y(x) = \sum_{i=0}^{\infty} y_i(x)$, where $y_0, y_1, \ldots$ is evaluated recursively. It is important to know that if the function has nonlinearity $(N(y(x)))$, it should be obtained by the following equation:

$$N(y(x)) = \sum_{n=0}^{\infty} A_n \tag{9}$$

where $A_n = A_n(y_0(x), y_1(x), \ldots, y_n(x))$ are the Adomian polynomials:

$$A_n = \frac{1}{n!}\frac{d^n}{d\lambda^n} N\left(\sum_{i=0}^{\infty} \lambda^i y_i(x)\right)|_{\lambda=0} \qquad n = 0, 1, 2, \ldots \tag{10}$$

Consider an ordinary differential equation as follows:

$$Ly + Ry + Ny = g(x), \tag{11}$$

where N describes the nonlinear operator, L defines the highest invertible derivative, R is the linear differential operator less order than L and g represent the source term. By applying the inverse term "L^{-1}" into the expression $Ly = g - Ry - Ny$, we have:

$$y = \gamma + f - L^{-1}(Ry) - L^{-1}(Ny), \tag{12}$$

where the function f describes the integration of the source term and γ is the given conditions. By considering the last equation, the recurrence relation of y can be simplified as follows:

$$\begin{cases} y_0 = \gamma + f, \\ y_1 = -L^{-1}\big(Ry_0\big) - L^{-1}\big(Ny_0\big) \\ \vdots \\ y_{k+1} = -L^{-1}\big(Ry_k\big) - L^{-1}\big(Ny_k\big), \quad k \geq 0 \end{cases} \tag{13}$$

Adomian decomposition theoretical convergence can be found in [12]. If the series converges to the considered solution, then

$$y = \lim_{M\to\infty} \tilde{y}_M(x), \tag{14}$$

where $\tilde{y}_M(x) = \sum_{i=0}^{M} y_i(x)$ [2].

In [13], a new improved version of the decomposition method is introduced using a Chebyshev approximation method. The illustrated method has overcome to the Taylor series inaccuracy to expand the source term function. The advantage of the modified approach is verified through several illustrative examples. Since, in this paper, we expand the source term in the Chebyshev series:

$$\boldsymbol{g}(\boldsymbol{x}) \approx \sum_{i=0}^{M} \boldsymbol{a}_i\, \boldsymbol{T}_i(\boldsymbol{x}), \tag{15}$$

where $T_i(x)$ represents the first kind Chebyshev polynomial and can be evaluated as follows:

$$\begin{aligned} \boldsymbol{T}_0(x) &= 1, \\ \boldsymbol{T}_1(x) &= x, \\ \boldsymbol{T}_{k+1}(x) &= 2\,x\,\boldsymbol{T}_k(x) - \boldsymbol{T}_{k-1}(x), \quad k \geq 1 \end{aligned} \tag{16}$$

Since we have:

$$\begin{cases} \boldsymbol{y}_0 = \gamma + \boldsymbol{L}^{-1}\left(\sum\limits_{i=0}^{n} \boldsymbol{a}_i\, \boldsymbol{T}_i(\boldsymbol{x})\right), \\ \boldsymbol{y}_{k+1} = -\boldsymbol{L}^{-1}\big(R\,\boldsymbol{y}_k\big) - \boldsymbol{L}^{-1}\big(N\,\boldsymbol{y}_k\big), \quad k \geq 0 \end{cases} \tag{17}$$

4 Interval Adomian Decomposition with Uncertainty

Consider the following ordinary differential equation (ODE) with considered interval initial conditions:

$$\begin{aligned} F\big(X, \boldsymbol{Y}', \ldots, \boldsymbol{Y}^{(n)}, \Delta\big) = 0, \\ Y(\boldsymbol{x}_0) = \boldsymbol{Y}_0, \ldots, \boldsymbol{Y}^{(n)}(\boldsymbol{x}_0) = \boldsymbol{Y}_n, \end{aligned} \tag{18}$$

where $\Delta = (\delta_1, \delta_1, \delta_2, \delta_3, \ldots, \delta_m)^T$ are uncertain parameters and $\delta_k \in \left[\underline{\delta}_k, \overline{\delta}_k\right],\ k = 0, 1, \ldots, (n+1)$.

The main purpose of this paper is to introduce an interval version of the Adomian decomposition method for solving the ODE problems with uncertainties. To do this, let us consider a standard form as below:

$$\begin{gathered} \underbrace{\boldsymbol{\delta}_n \boldsymbol{Y}^{(n)} + \boldsymbol{\delta}_3 \boldsymbol{Y}'' + \ldots + \boldsymbol{\delta}_2 \boldsymbol{Y}' + \boldsymbol{\delta}_1 \boldsymbol{Y}}_{\boldsymbol{R}\{\boldsymbol{y}(\boldsymbol{x})\}} + \alpha N\{Y(X)\} = \beta\, G\,(X) \\ \alpha \in \left[\underline{\boldsymbol{\alpha}}, \overline{\boldsymbol{\alpha}}\right], \quad \beta \in \left[\underline{\boldsymbol{\beta}}, \overline{\boldsymbol{\beta}}\right], \end{gathered} \tag{19}$$

The first step is to convert the system into the uniform mode,

$$\underbrace{\boldsymbol{\delta_n Y^{(n)}} + \boldsymbol{\delta_3 Y''} + \ldots + \boldsymbol{\delta_2 Y'} + \boldsymbol{\delta_1 Y}}_{\boldsymbol{R\{y(x)\}}} + \boldsymbol{\tilde{\alpha} N\{Y(X)\}} = \boldsymbol{\tilde{\beta}\, G\,(X)} \tag{20}$$

where the term (~) shows the division of the coefficients with δ_n. For instance, $\tilde{\beta}$ is $[\beta]/[\delta_n] = [\beta] \times \left[\frac{1}{\overline{\delta}_n}, \frac{1}{\underline{\delta}_n}\right]$.

δ_n is assumed non-singular, i.e., $0 \notin \delta_n$. From the previous section, by using the given conditions we obtain:

$$\boldsymbol{Y} = \boldsymbol{P_n} + \boldsymbol{F} - \boldsymbol{L}^{-1}(\boldsymbol{RY}) - \boldsymbol{L}^{-1}(\boldsymbol{Ny}), \tag{21}$$

where $P_n = \left[\underline{p}_n, \overline{p}_n\right]$ is an interval polynomial which has been achieved from the initial conditions $P_n = Y_0 + Y_1 x + \ldots + Y_n x^n$. If the given conditions have the exact value, then $P_n = [p_n, \; p_n]$; these kinds of exact values in the interval analysis are the so-called degenerative intervals. The function F describes the interval integration of the source term. For computing the inverse operator of the source term, we have:

$$\boldsymbol{F} = \left[\boldsymbol{L}^{-1}(\boldsymbol{\beta g(x)})\right] = [\underline{\boldsymbol{\beta}}, \overline{\boldsymbol{\beta}}] \times [\boldsymbol{L}^{-1}\left(\underline{\boldsymbol{g}}(\boldsymbol{x})\right), \boldsymbol{L}^{-1}(\overline{\boldsymbol{g}}(\boldsymbol{x}))\Big], \tag{22}$$

By assuming $g = g_c + \delta_g$ where g_c and δ_g comprise the certain and uncertain terms of g,

$$\left[\boldsymbol{L}^{-1}\left(\underline{\boldsymbol{g}}(\boldsymbol{x})\right), \boldsymbol{L}^{-1}(\overline{\boldsymbol{g}}(\boldsymbol{x}))\right] = [\boldsymbol{L}^{-1}\left(\boldsymbol{g_c}(\boldsymbol{x}) + [\boldsymbol{\delta_g}]\right)\Big] \tag{23}$$

A similar operation can be also utilized for achieving the linear and nonlinear differential operators. Since the total proposed interval Adomian method can be formulated as follows:

$$\begin{cases} \boldsymbol{Y}_0 = [\gamma] + \left[\underline{\boldsymbol{\beta}}, \overline{\boldsymbol{\beta}}\right] \times [\boldsymbol{L}^{-1}\left(\underline{\boldsymbol{g}}(x)\right), \boldsymbol{L}^{-1}(\overline{\boldsymbol{g}}(x))\Big], \\ \boldsymbol{Y}_{k+1} = \left[-\boldsymbol{L}^{-1}\left(R\underline{\boldsymbol{y}}_{\boldsymbol{k}}\right) - \boldsymbol{L}^{-1}\left(N\underline{\boldsymbol{y}}_{\boldsymbol{k}}\right), -\boldsymbol{L}^{-1}\left(R\overline{\boldsymbol{y}}_{\boldsymbol{k}}\right) - \boldsymbol{L}^{-1}\left(N\overline{\boldsymbol{y}}_{\boldsymbol{k}}\right)\right], \quad k \geq 0 \end{cases} \tag{24}$$

And the source term can be achieved as:

$$G\,(\boldsymbol{x}) = \sum_{i=0}^{n} [\boldsymbol{a}_i]\,[\boldsymbol{T}_i(\boldsymbol{x})] = \left[\underline{\boldsymbol{\beta}}, \overline{\boldsymbol{\beta}}\right] \times \boldsymbol{L}^{-1}\left(\sum_{i=0}^{n}[\underline{\boldsymbol{a}}_i]\,[\underline{\mathbf{T}}_i(\boldsymbol{x})\right], \sum_{i=0}^{n}[\overline{\boldsymbol{a}}_i]\,[\overline{\boldsymbol{T}}_i(\boldsymbol{x})]), \tag{25}$$

5 Illustrative Examples

To demonstrate the effectiveness of the proposed method, we give four different examples of linear and nonlinear ordinary differential equations. The algorithms are performed by MATLAB R2017b.

Case study 1

Consider for $-1 \le x \le 1$

$$\boldsymbol{Y}''(x) - [\delta]\,Y(x) = 0, \quad [\delta] \in [1, 2]$$
$$Y([0]) \in [1], \quad \boldsymbol{Y}'([0]) \in [\beta] = [-0.1, 0.1],$$

The above equation shows a differential equation where uncertainty in the coefficient term ($[\delta]$), the initial condition $[\beta]$ is also uncertain and appeared the only thing we know is that it stands in an interval.

The purpose of the solution is to find a region that includes all different values within the represented interval. According to the formula,

$$\boldsymbol{LY} + [\boldsymbol{\delta}]\boldsymbol{RY} = 0$$

We can define that $L = \frac{d^2}{dx^2}$, $\breve{R} = [\delta]\,Y(x)$, $NY = G(x) = 0$. By applying the inverse operator $L^{-1} = \int\limits_0^x \int\limits_0^x \{.\}\,\mathrm{d}x\mathrm{d}x$ into the main equation,

$$\boldsymbol{Y}\,(\boldsymbol{x}) = [1, 1] + [-0.1, 0.1]\,\boldsymbol{x} + [\boldsymbol{\delta}] \int\limits_0^{\boldsymbol{x}} \int\limits_0^{\boldsymbol{x}} \sum_{\boldsymbol{n}=0}^{\infty} \boldsymbol{Y}_{\boldsymbol{n}}(\boldsymbol{x})\,\mathbf{d}\boldsymbol{x}\,\mathbf{d}\boldsymbol{x},$$

and finally, the recurrence relation below can be utilized to achieve the $Y(x)$:

$$\boldsymbol{Y}_0(\boldsymbol{x}) = [1] + [\boldsymbol{\beta}]\,\boldsymbol{x},$$

$$\boldsymbol{Y}_1(\boldsymbol{x}) = [\boldsymbol{\delta}] \int_0^{\boldsymbol{x}}\int_0^{\boldsymbol{x}} \boldsymbol{Y}_0(\boldsymbol{x})\,\mathbf{dx}\,\mathbf{dx} = [1,2]\,\tfrac{\boldsymbol{x}^2}{2!} + [-0.2, 0.2]\tfrac{\boldsymbol{x}^3}{3!},$$

$$\boldsymbol{Y}_2(x) = [^{\delta} \int_0^{\boldsymbol{x}}\int_0^{\boldsymbol{x}} \boldsymbol{Y}_1(\boldsymbol{x})\,\mathbf{dx}\,\mathbf{dx} = [1,4]\,\tfrac{\boldsymbol{x}^4}{4!} + [-0.4, 0.4]\,\tfrac{\boldsymbol{x}^5}{5!},$$

$$\vdots$$

$$\boldsymbol{Y}_{\boldsymbol{n}}(x) = [^{\delta} \int_0^{\boldsymbol{x}}\int_0^{\boldsymbol{x}} \boldsymbol{Y}_{\boldsymbol{n}-1}(\boldsymbol{x})\,\mathbf{dx}\,\mathbf{dx} = [1, 2^{\boldsymbol{n}}]\,\tfrac{\boldsymbol{x}^{2n}}{(2\boldsymbol{n})!} + 2^{\boldsymbol{n}}[-0.1, 0.1]\,\tfrac{\boldsymbol{x}^{(2n+1)}}{(2\boldsymbol{n}+1)!}$$

Since the final solution can be achieved by:

$$\boldsymbol{Y}(\boldsymbol{x}) = [\mathbf{cosh}(\boldsymbol{x})] + 0.1\left(\boldsymbol{x} + 2\frac{\boldsymbol{x}^3}{3!} + \cdots + 2^{\boldsymbol{n}}\frac{\boldsymbol{x}^{(2\boldsymbol{n}+1)}}{(2\boldsymbol{n}+1)!}\right)[-1, 1].$$

By calculating the problem in the time interval between 0 and 2 and the same step size $h = 0.4$, the minimum and maximum values of y at each step are obtained and given in Table 1. In this case, we also applied a random value in the considered interval and the results showed that the random solution is placed in the interval solution. We also compared the proposed method by the interval Euler [14] and Taylor methods [15]. Table 1 shows more details of this comparison. As can be seen, the interval space in the proposed interval Adomian method achieves generally narrower intervals than the others.

Case study 2

In this example, we consider a problem with more uncertainty both in the linear differential operator and in the source term. Consider for $-1 \le x \le 1$

$$\boldsymbol{Y}'(x) - [\boldsymbol{\delta}_1]\,xY(x) = [\boldsymbol{\delta}_2]\left(6 - 3\,\boldsymbol{x}^2\right),$$
$$Y(0) = 0,\quad 0.5 \le \boldsymbol{\delta}_1 \le 1\,,\quad 0.5 \le \boldsymbol{\delta}_2 \le 1\,.$$

Table 1 Region bound achieved by the interval Adomian method and a limited random input for case study 1 which stands in the considered region

Time	Interval Adomian method	Interval Euler's method [13]	Interval Taylor method [14]	Random value
0	[1, 1]	[1, 1]	[1, 1]	1.00
0.4	[1.081, 1.164]	[1.02, 1.16]	[1.04, 1.21]	1.081
0.8	[1.337, 1.711]	[1.20, 1.68]	[1.25, 1.81]	1.337
1.2	[1.810, 2.819]	[1.57, 2.69]	[1.66, 3.01]	1.810
1.6	[2.576, 4.838]	[2.17, 4.48]	[2.34, 5.19]	2.577
2	[3.755, 8.378]	[3.09, 7.56]	[3.4, 9.08]	3.762

where $L = \frac{d}{dx}$, $\breve{R} = [\delta_1]\, xY(x)$,$NY = 0$, and $\breve{g}(x) = [\delta_2]\left(6 - 3\,x^2\right)$. In this problem, we have:

$$\boldsymbol{Y}(\boldsymbol{x}) = 0 + [\boldsymbol{\delta}_1] \int\limits_0^{x} \sum_{n=0}^{\infty} \boldsymbol{x}\boldsymbol{Y}_n(\boldsymbol{x})\, \mathbf{d}\boldsymbol{x} + [\boldsymbol{\delta}_2] \int\limits_0^{x} \left(6 - 3\boldsymbol{x}^2\right) \mathbf{d}\boldsymbol{x}.$$

Now using the Adomian decomposition method, we get

$$\begin{aligned}
&\boldsymbol{Y}_0(\boldsymbol{x}) = [\boldsymbol{\delta}_2]\left(-\boldsymbol{x}^3 + 6\,\boldsymbol{x}\right) = \left[-\boldsymbol{x}^3/2 + 3\,\boldsymbol{x},\ -\boldsymbol{x}^3 + 6\,\boldsymbol{x}\right],\\
&\boldsymbol{Y}_1(\boldsymbol{x}) = [\boldsymbol{\delta}_1] \int_0^{x} \boldsymbol{x}\boldsymbol{Y}_0(\boldsymbol{x})\, \mathbf{d}\boldsymbol{x} = ([\boldsymbol{\delta}_1]\times[\boldsymbol{\delta}_2]) \int_0^{x} \boldsymbol{x}\left(-\boldsymbol{x}^3 + 6\,\boldsymbol{x}\right) \mathbf{d}\boldsymbol{x}\\
&= [0.25,\, 1]\left(-\boldsymbol{x}^5/5 + 2\,\boldsymbol{x}^3\right) = \left[-0.05\,\boldsymbol{x}^5 + 0.5\,\boldsymbol{x}^3,\ -0.2\,\boldsymbol{x}^5 + 2\,\boldsymbol{x}^3\right]\\
&\boldsymbol{Y}_2(\boldsymbol{x}) = [\boldsymbol{\delta}_1] \int_0^{x} \boldsymbol{x}\boldsymbol{Y}_1(\boldsymbol{x})\, \mathbf{d}\boldsymbol{x} = \left([^{\delta_1}\times[\boldsymbol{\delta}_2]\right) \int_0^{x} \boldsymbol{x}\left(-\boldsymbol{x}^5/5 + 2\,\boldsymbol{x}^3\right) \mathbf{d}\boldsymbol{x}\\
&= [0.125,\, 1]\left(-0.03\boldsymbol{x}^7 + 0.4\,\boldsymbol{x}^5\right) = \left[-0.004\,\boldsymbol{x}^7 + 0.05\,\boldsymbol{x}^5,\ -0.03\boldsymbol{x}^7 + 0.4\,\boldsymbol{x}^5\right]\\
&\vdots\\
&\boldsymbol{Y}_n(x) = [^{\delta_1} \int_0^{x} \boldsymbol{x}\boldsymbol{Y}_{n-1}(\boldsymbol{x})\, \mathbf{d}\boldsymbol{x}
\end{aligned}$$

So, we have:

$$Y(x) = \left[-0.03x^7 + 0.2\,x^5 + x^3 + 6\,x - + \cdots,\ -0.004\,x^7 + 3\,x - + \cdots\right].$$

From Table 2, we can say that the interval Euler method fails the interval in the "time = 0.4". It is also obvious that the proposed interval Adomian method has a narrower interval rather than the interval Taylor method.

Case study 3

Now, we consider a problem with complicated source terms. Consider for $-1 \le x \le 1$,

Table 2 Region bound achieved by the interval Adomian method and a limited random input for case study 2 standing in the considered region

Time	Interval Adomian method	Interval Euler's method [13]	Interval Taylor method [14]	Random value
0	[0, 0]	[0,1]	[0, 0]	0
0.4	[1.826, 2.466]	[2.23, 2.44]	[1.20, 2.46]	1.826
0.8	[3.850, 5.371]	[3.55, 5.26]	[2.40, 5.38]	3.850
1.2	[6.457, 9.323]	[4.98, 9.11]	[3.60, 9.54]	6.457
1.6	[10.271, 15.026]	[6.59, 15.42]	[4.80, 16.84]	10.271
2	[15.943, 22.743]	[8.51, 27.81]	[6.00, 32.52]	15.943

$$\boldsymbol{Y''}(x) - [\delta]\, Y(x) = \boldsymbol{e^x}\,,$$
$$Y(0) = 1,\ \boldsymbol{Y'}(0) = 1,\ \ 0.5 \le \delta \le 1\,,$$

where $L = \frac{d^2}{dx^2}$, $\breve{R} = [\delta]\, Y(x)$, $NY = 0$, $g(x) = e^x$. By letting $M = 4$ and using the Chebyshev approximation, we have:

$$\boldsymbol{g}(\boldsymbol{x}) = 0.0432\, \boldsymbol{x}^4 + 0.1772\, \boldsymbol{x}^3 + 0.4998\, \boldsymbol{x}^2 + 0.99\,74\, \boldsymbol{x} + 0.9892$$

Error ($|g(x) - e^x|$) for this approximation is 0.0006397, while using the Taylor series gives us an error of about 0.0099. From the main equation, we have:

$$\boldsymbol{Y}(\boldsymbol{x}) = 1 + \boldsymbol{x} - [\boldsymbol{\delta}] \int\limits_0^{\boldsymbol{x}} \int\limits_0^{\boldsymbol{x}} \sum_{\boldsymbol{n}=0}^{\infty} \boldsymbol{Y_n}(\boldsymbol{x})\, \mathbf{d}\boldsymbol{x}\, \mathbf{d}\boldsymbol{x} + \int\limits_0^{\boldsymbol{x}} \int\limits_0^{\boldsymbol{x}} \left(\boldsymbol{e^x}\right) \mathbf{d}\boldsymbol{x}\, \mathbf{d}\boldsymbol{x}$$

Now using the Adomian decomposition method, we get

$$\begin{aligned}
Y_0(x) &= 1 + x + \int\limits_0^x \int\limits_0^x g(x)\mathrm{d}x\,\mathrm{d}x = 1 + x + 0.4946x^2 + 0.1662x^3 \\
&\quad + 0.0417x^4 + 0.0089x^5 + 0.0014x^6 \\
Y_1(x) &= -[\delta] \int\limits_0^x \int\limits_0^x Y_0(x)\mathrm{d}x\,\mathrm{d}x = [-2.5e^{-5}x^8 - 2.119e^{-4}x^7 - 0.0014x^6 \\
&\quad - 0.0083x^5 - 0.0412x^4 - 0.1667x^3 - 0.5x^2, \\
&\quad - 1.25e^{-5}x^8 - 1.059e^{-4}x^7 - 0.0007x^6 - 0.00415x^5 \\
&\quad - 0.0206x^4 - 0.0833x^3 - 0.25x^2], \\
&\ \ \vdots \\
Y_n(x) &= [^{\delta_1} \int\limits_0^x x Y_{n-1}(x)\mathrm{d}x,
\end{aligned}$$

And finally using $\tilde{y}_M(x) = \sum\limits_{i=0}^{M} y_i(x)$, the solution has been achieved.

From Table 3, it is obvious that the interval Euler and Taylor methods fail their interval and do not include the random value, but the interval Adomian method includes the value.

We also apply Chebyshev approximation to the source term of the interval Adomian method. As can be seen from Fig. 1b, Chebyshev polynomial gives us a tighter interval and from the interval arithmetic, it has better performance rather than the Taylor series. Furthermore, it is important to know that sometimes lower

Table 3 Region bound achieved by the interval Adomian method and a limited random input for case study 3 standing in the considered region

Time	Interval Adomian method	Interval Euler's method [13]	Interval Taylor method [14]	Random value
0	[1, 1]	[1, 1]	[1, 1]	1
0.4	[1.112, 1.126]	[1.16, 1.50]	[1.21, 1.54]	1.112
0.8	[1.408, 1.439]	[1.73, 2.32]	[1.87, 2.44]	1.409
1.2	[1.934, 1.986]	[2.87, 3.64]	[3.20, 3.91]	1.934
1.6	[2.773, 2.858]	[4.91, 5.71]	[5.59, 6.23]	2.774
2	[4.070, 4.202]	[8.38, 8.92]	[9.70, 9.86]	4.071

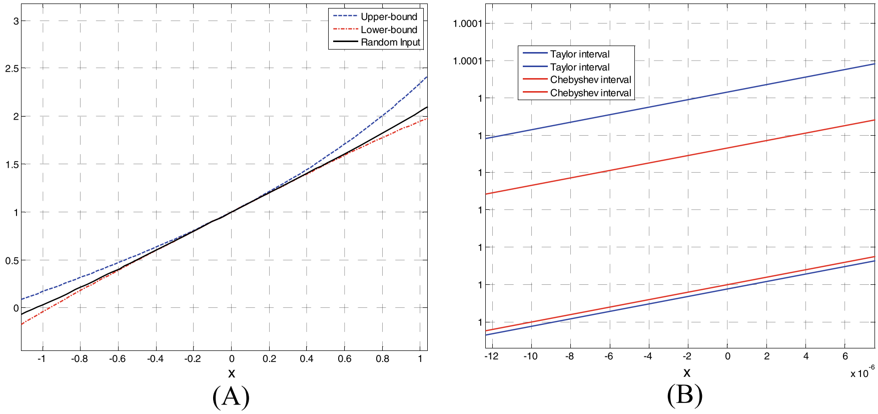


Fig. 1 **a** The region bound achieved by the interval Adomian method and a limited random input which stands in the considered region and **b** comparison of the solution using Taylor and Chebyshev approximation for case study 3

and upper bounds have crossover with each other in some ODEs. In this situation, we should consider the general bound in between them as the reliability region.

Case study 4

Consider the nonlinear ordinary differential equation for $0 \le x \le 1$

$$\boldsymbol{Y}''(x) - [\boldsymbol{\delta}_1]\, x\boldsymbol{Y}^2(x)\boldsymbol{Y}'(x) - [\boldsymbol{\delta}_2]\, \boldsymbol{Y}^3(x) = \boldsymbol{e}^{\boldsymbol{x}} - 2\boldsymbol{e}^{3x} - 3x\boldsymbol{e}^{3x}\,,$$
$$Y(0) = 1, \quad \boldsymbol{Y}'(0) = 1, \quad 1 \le \boldsymbol{\delta}_1, \boldsymbol{\delta}_2 \le 2\,,$$

where $L = \frac{d^2}{dx^2}, \breve{R} = 0,\ N_1 Y = Y^2(x)Y'(x),\ \ N_2 Y = Y^3(x),\ \ g(x) = e^x - 2e^{3x} - 3xe^{3x}$

By considering fourth-order Chebyshev approximation to the source term,

$$\boldsymbol{g}(\boldsymbol{x}) \approx -32.612\,\boldsymbol{x}^4 - 45.0884\,\boldsymbol{x}^3 - 14.2512\,\boldsymbol{x}^2 - 1.7918\,\boldsymbol{x} + 7.1530$$

Since

$$Y(x) = 1 + x + [\delta_1] \int_0^x \int_0^x x\, N_1(x)\, dx\, dx + [\delta_2] \int_0^x \int_0^x N_2(x)\, dx\, dx + \int_0^x \int_0^x g(x)\, dx\, dx$$

The Adomian polynomials for $N_1(x)$ and $N_2(x)$ are (Table 4):
So, following the same illustrated approach,

$$Y_0(x) = 1 + x + 3.5765\, x^2 - 0.2986\, x^3 - 1.1876\, x^4 - 2.2844\, x^5 - 1.0871\, x^6$$
$$Y_1(x) = [\delta_1] \int_0^x \int_0^x Y_0(x)^2 Y'_0(x)\, dx dx + [\delta_2] \int_0^x \int_0^x Y_0^3(x)\, dx\, dx = [x^2 + 2\, x^3 + 2.8\, x^4 + \cdots, 2x^2 + 4x^3 + \ldots]$$
$$\vdots$$
$$y_n(x) = [\delta_1] \int_0^x \int_0^x Y_{n-1}^2(x) Y'_{n-1}(x)\, dx\, dx + [\delta_2] \int_0^x \int_0^x Y_{n-1}^3(x)\, dx\, dx$$

Table 5 shows that interval Euler and Taylor methods fail the interval from "time = 0.8", but the proposed method is total including the random value.

Table 4 Adomian polynomials for nonlinear terms

$N_1(x)$	$N_2(x)$
$A_0(x) = Y_0^2 Y_0'$	$A_0(x) = Y_0^3$
$A_1(x) = Y_0^2 Y_1' + 2Y_0 Y_1 Y_0'$	$A_1(x) = 3Y_0^2 Y_1$
$A_2(x) = Y_0^2 Y_2' + 2Y_0 Y_1 Y_1' + 2Y_0 Y_2 Y_0'$	$A_2(x) = 3Y_0^2 Y_2 + 3Y_0 Y_1^2$
$\vdots$	$\vdots$

Table 5 Region bound achieved by the interval Adomian method and a limited random input for case study 4 standing in the considered region

Time	Interval Adomian method	Interval Euler's method [13]	Interval Taylor method [14]	Random value
0	[1, 1]	[1, 1]	[1, 1]	1
0.4	[1.398, 1.401]	[1.06, 1.37]	[0.99, 1.33]	1.399
0.8	[1.780, 1.819]	[0.27, 0.95]	[-0.67, 0.04]	1.781
1.2	[2.092, 2.307]	Divergent	Divergent	2.092
1.6	[2.234, 2.961]	Divergent	Divergent	2.235
2	[2.042, 3.941]	Divergent	Divergent	2.043

6 Conclusions

The interval Adomian decomposition is introduced for solving differential equations with uncertainties. This approach provides a robust approximation of the solution. The main advantage of this approach over traditional numerical methods is that the proposed method is the first time which used the interval arithmetic to provide a robust result for ODEs with uncertain coefficients. In addition, for increasing the system accuracy, Chebyshev polynomial is utilized for the expansion of the source term.

References

1. Wu, J., Gao, J., Luo, Z., Brown, T.: Robust topology optimization for structures under interval uncertainty. Adv. Eng. Softw. **30**(99), 36–48 (2016)
2. Adomian, G.: Solving Frontier Problems of Physics: The Decomposition Method. Kluwer Academic Publishers, Boston, MA (1994)
3. Razmjooy, N., Ramezani, M.: Solution of the Hamilton jacobi bellman uncertainties by the interval version of adomian decomposition method. Int. Rob. Auto J. **4**(2), 113–117 (2018)
4. Razmjooy, N., Mehdi, R.: Interval structure of Runge-Kutta methods for solving optimal control problems with uncertainties. Comput. Methods Differential Equations **7**(2), 235–251 (2019)
5. Razmjooy, N., Roujin, M., Mohammad, A., Youness, A.: Optimal control of self-balancing robot in the presence of uncertainties based on interval analysis. In: The 27th Annual International Conference of Iranian Society of Mechanical Engineers, vol. 27, issue 1, pp. 15–20 (2019)
6. Biazar, J.: Solution of systems of integral–differential equations by Adomian decomposition method. Appl. Math. Comput. **168**, 1232–1238 (2005)
7. Çelik, E., Bayram, M., Yelolu. T.: Solution of Differential-Algebraic Equations (DAEs) by Adomian decomposition method. Int. J. Pure Appl. Math. Sci. **3**(1), 93–100 (2006)
8. Jafari, H., Daftardar-Gejji, V.: Solving a system of nonlinear fractional differential equations using Adomian decomposition. J. Comput. Appl. Math. **196**, 644–651 (2006)
9. Evans, D., Raslan, K.: The Adomian decomposition method for solving delay differential equation. Int. J. Comput. Math. **82**(1), 914–923 (2004)
10. Guo, G., Razmjooy, N.: A new interval differential equation for edge detection and determining breast cancer regions in mammography images. Syst. Sci. Control Eng. **7**(1), 346–356 (2019)
11. Lesnic, D.: Convergence of Adomian decomposition method: periodic temperatures. Comput. Math Appl. **44**, 13–24 (2002)
12. Hosseini, M.M.: Adomian decomposition method with Chebyshev polynomials. Appl. Math. Comput. **175**, 1685–1693 (2006)
13. Razmjooy, N., Ramezani., M.: Analytical Solution for Optimal Control by the Second kind Chebyshev Polynomials Expansion. Iranian Journal of Science and Technology, 2017, Article in press
14. Su Huan, C., Wu, J.: Interval optimization of dynamic response for structures with interval parameters. Comput. Struct. **82**(1), 1–11 (2004)
15. Araya, I., Gilles T., Bertrand, N.: A contractor based on convex interval taylor. In: International Conference on Integration of Artificial Intelligence (AI) and Operations Research (OR) Techniques in Constraint Programming, pp. 1–16. Springer, Heidelberg (2012)

Estimation of Spatial-Temporal Gait Parameters Using a Momentary-Contact-Sensors-Based System: A Preliminary Evaluation

Wellingtânia Domingos Dias, Meinhard Sesselmann, Iury Cardoso Brito, Ivo Oliveira Capanema, Maria Emilia de Abreu Chaves, Renata Kirkwood, and Rudolf Huebner

Abstract Nowadays, there are many technologies commercially available to perform gait analysis. However, most of those technologies present limitations and, therefore, their applicability in clinics is reduced. The measuring system proposed in this study consists of hardware, an algorithm recorded in microcontrollers and a human–machine interface software installed on a computer. This device was validated through tests, in which an individual walked barefoot and in footwear on a certain surface while her steps were mapped and the spatial–temporal gait parameters were quantified. The measured values were compared with other gait analysis methods such as podogram and filming. The results demonstrated that the measured spatial–temporal parameters are similar to others mentioned in the literature, and

W. D. Dias (✉) · M. Sesselmann · I. C. Brito · I. O. Capanema · M. E. de Abreu Chaves · R. Huebner
Bioengineering Laboratory, Post Graduate, Department of Mechanical Engineering, Federal University of Minas Gerais, 6627 Presidente Antônio Carlos Ave, Belo Horizonte, MG 31270-901, Brazil
e-mail: wellingtaniad@gmail.com

M. Sesselmann
e-mail: meinhard@ufmg.br

I. C. Brito
e-mail: iurycb6@gmail.com

I. O. Capanema
e-mail: ivoliveirabh@hotmail.com

M. E. de Abreu Chaves
e-mail: mariaemilia.ufmg@gmail.com

R. Huebner
e-mail: rudolf@ufmg.br

R. Kirkwood
Programm in Rehabilitation, Federal University of Minas Gerais, 6627 Presidente Antônio Carlos Ave, Belo Horizonte, MG 31270-901, Brazil
e-mail: renata.kirkwood@gmail.com

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_24

they are correlated with the values obtained in the podogram and filming methods. It is concluded that the measuring system presents adequate functionality and connectivity, capable of mapping the gait cycle, and quantifying the spatial–temporal parameters.

Keywords Spatial–temporal parameters · Gait analysis · Measuring system · Gait

1 Introduction

Gait can be described as a sequence of complex and coordinated movements between various body joints, especially the lower limbs [1]. The gait analysis provides essential information about the functional capabilities of individuals, through kinematic and kinetic measurement of the body segments. Among the relevant measures, spatial–temporal parameters are widely used in the clinical context. These include speed, step and stride length, step and stride time, cadence, and support base width. These parameters quantitatively describe the main gait events, and they are used to identify deviations from normal gait, define the therapeutic intervention, and evaluate the results [2].

There are several methods to perform gait analysis, being the most common clinical observation. The observational analysis allows us to know the speed (by timing the time spent to walk a known distance) and the cadence (by counting how many steps are performed in a given unit of time). On the other hand, it is possible to identify the length of the step and stride using podogram methodology, which is defined as an imprint of the foot sole on a paper mat. The foot support pattern, when associated with a stopwatch, determines the speed, cadence, and the total walking cycle time of an individual.

Methods of gait analysis have evolved with the influence of technology. Advances in the field of engineering have increased the reliability of these methods, improved the acquisition speed, and the storage of the obtained data. Thus, this has allowed the information to be accessible for clinical interpretation. The technological development makes possible the creation of special surfaces with pressure sensors. They allow the evaluation of spatial and temporal gait parameters, without the need to adhere to or couple sensors in the patient's body [3, 4]. Methods that include optical systems of motion capture and force platforms have become gold standards [5].

Nevertheless, there are some limitations in most technologies that perform gait analysis. These are related to portability, acquisition cost, physical space and time, difficult handling and maintenance, and reduced applicability in health centers [6]. It is difficult (rare) to see, especially in clinics and doctor's offices, the use of such (excessively costly) equipment. This is because clinical problems can be adequately managed using simpler and cheaper resources that have comparable reliability in results [7].

In this context, the development of new gait analysis devices aims to reduce costs and, hence, facilitate access to a larger number of health professionals as well as increase the accuracy and efficiency of the analysis. Therefore, the objective of this study is to develop and validate a device capable of mapping the footsteps of an individual as well as measuring the spatial–temporal gait parameters automatically, obtained by momentary contact sensors.

2 Methods

The measuring system for mapping the footsteps and the quantification of spatial–temporal gait parameters was developed and validated at the Bioengineering Laboratory of the Federal University of Minas Gerais.

2.1 Development of the Measuring System

The measuring system of gait consists of hardware, an algorithm programmed in microcontrollers and a human–machine interface (HMI) software installed on a computer. The hardware consists of one or more modules representing the electromechanical part of the entire measuring system, which operates individually or together. Each module (see Fig. 1) consists of five overlapped layers that form a sandwich structure, which allows preserving the lifespan of the electronic circuit, as

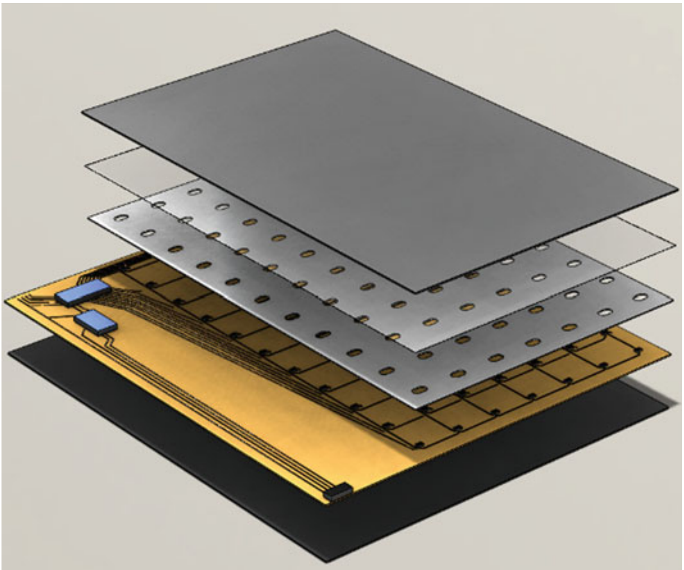


Fig. 1 Representative scheme of the applied sandwich structure

well as to protect it (as a whole) to guarantee adequate robustness. The main layer consists of an electronic circuit that includes microcontrollers, momentary contact sensors and diodes arranged in a matrix, fixed in areas and distances pre-established empirically. The layout of the electronic circuit was designed from software for creating and generating the layout of printed circuit boards.

The connection between the modules occurs at the upper and lower ends through USB male–female connectors. Two columns of the modules make up a satisfactory width of the measuring system. They are connected to each other at the upper or lower end by a male–female USB extender cable. These columns form a “U” shaped bus for control, power, and data transfer. The spatial resolution of the contact area of the plantar region is defined by the constant spacing between the momentary contact sensors that make up the matrix. An external source is used as a power supply.

The algorithm was developed using the Arduino IDE platform and programmed in the microcontroller of each module. This algorithm performs a sweep of all sensors and, consequently, reads their logic level continuously.

The human–machine interface software was developed using MATLAB. During the measurement of an individual’s gait on the system, this software provides the functionality to request the scanning of the sensors for the acquisition and transmission of data to the main microcontroller. Then, the information is transmitted to the computer through serial communication using the USB port. In addition, the HMI software performs the processing of information and displays the results of the mapping of the gait cycle and the quantification of spatial–temporal parameters. Figure 2 shows this operation cycle of the measuring system in a flowchart.

The sweep routine of the momentary-contact sensors is essential to perform data acquisition for further processing. The rows and columns of the sensor matrix are each connected to a pin of the secondary microcontroller. Initially, one column is enabled in the sensor matrix. Then, each matrix row is verified looking for a state change. If this change is identified, it is stored a list in the memory of the secondary microcontroller. This process is continually repeated on each column while the sweep is active. At the end of each sweep, a list informing any state changes is generated and sent to the main microcontroller, which collects all the lists and transfers them to the computer for further processing and generation of results via HMI.

The momentary-contact-sensors matrix that integrates the electronic circuit of the measuring system is one of the most important parts because it provides essential information to data analysis. In this way, the implementation of a matrix sweep routine to identify sensor failures was fundamental. If any sensor fails or presents problems, the HMI software will print a warning containing the coordinates (row, column) in which the failed sensor is located. Therefore, preventive maintenance may avoid information loss during data acquisition. This matrix sweep for failure identification occurs at the start (at the first reading), when the measuring system is powered on and connected to the computer. Then, if the HMI indicates that it is working properly, the system is ready to start the activities.

Whenever a state change of any of the momentary contact sensors occurs, it is associated with a proper timestamp. In order to calculate the time of the gait cycle,

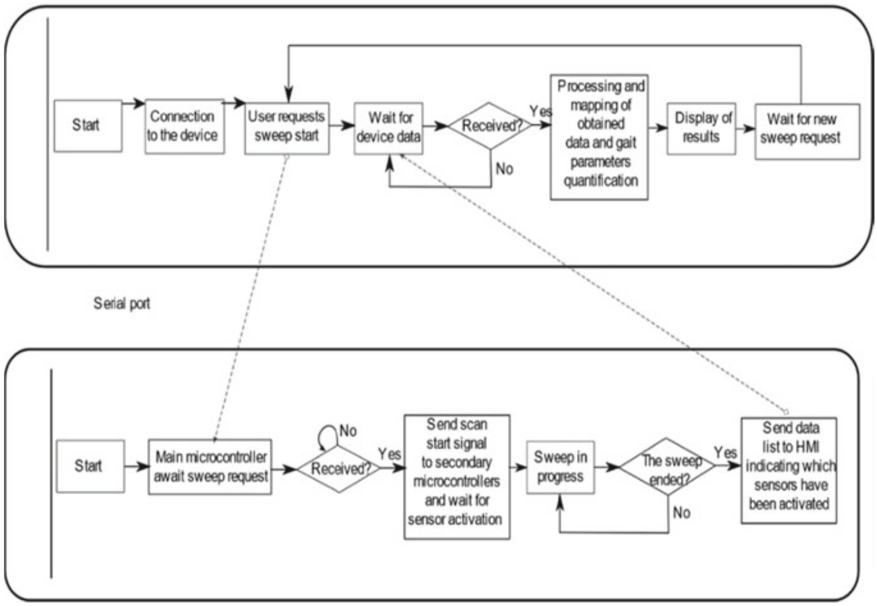


Fig. 2 Flowchart of the operating cycle of the measuring system

only the first activation is considered (when the first heel touches the system surface) and the last one (when the last heel touches the system surface).

By analyzing the time stamp of those events, it is possible to obtain the time interval (Δt):

$$\Delta t = \left(t_{\text{final}(Bx,y)} - t_{\text{initial}(A(x,y))}\right) \tag{1}$$

The distances for step, stride, and base width are calculated by the HMI through the Euclidean distance metric defined by:

$$D(x, y) = \sqrt{\left((x1-x2)^2 + (y1-y2)^2 + (\cdots) + (yn-yn)^2\right)} \tag{2}$$

where $x = (x1, x2, \ldots, xn)$ and $y = (y1, y2, \ldots, yn)$ are the coordinates (x, y) of activated sensors within the sensor matrix.

The speed (v) will be determined by the quotient of the variation of the distance (see Eq. 2) with respect to time (see Eq. 1), according to the equation:

$$v = D(x, y)/\Delta t \tag{3}$$

The identification of the first step (distance between each foot's initial contact) is performed from the first detected activation. Then, the algorithm checks the distance between the current activation and the previous one, defining the step distance.

For a new step detection since the first activation, the algorithm checks the distance between the current activation and the previous one considering a corresponding distance of three sensors. The same applies to the analysis of the accommodation period or intermediate states in which the individual is balanced on the supporting leg. For gait analysis is examined the distance between the heels and the mean support (the moment when the foot is totally in contact with the surface) is considered in the image generation of the mapping the footsteps.

Now, for stride detection, that is the distance between the initial contact of the same foot (right to right, left to left), the same procedure is adopted as in the identification of step distance (right to left, left to right). In addition, the distances between the current activation and the first one of the last and penultimate identified steps are also checked.

Base width is the horizontal distance between the initial contact of the feet in one step (right to left, left to right). The measurement is performed by calculating the mean of distances obtained during the gait cycle on the measuring system.

2.2 Measuring System Validation

The functionality of the measuring system was evaluated by the measurement of the spatial–temporal gait parameters, which are step length, stride length, cadence, and speed. The parameters were measured using three methods: application of the proposed measuring system, podogram, and filming.

The measuring system consists of an area of approximately 3.08 m (length) by 0.74 m (width). The separating distance between sensors is 4 cm. The algorithm programmed in the microcontrollers of the electronic circuit of the measuring system performs the sweep in the sensor matrix after being requested by the HMI. As the individual walks over the measurement system, the pressed sensors are activated, causing their logic state to change from high to low in the circuit. Then, the information of the logical state of the sensors is saved as arrays of data elements “n”, allowing the representation of maps of the individual’s gait cycle and the quantification of spatial–temporal parameters.

Now, podogram is a method composed of a demarcated paper track for a plantar print of feet. In addition to the track, there is a stopwatch to take the start and end time of the individual on the demarcated surface.

In turn, filming consists of positioning a camera at a predetermined distance from the area to be analyzed. Then, a video is recorded of the individual’s gait on the demarcated and predetermined surface. Subsequently, the analysis is performed through proper software.

The instruments of the three methods used in the measurement of the spatial–temporal parameters were positioned on a leveled and flat surface, without disturbing factors. Firstly, the proposed measurement system was positioned. Then, the demarcated paper track of 7.08 m was placed on it. For the respective filming of the tests, a camera was mounted perpendicularly to them and at a distance of 2.0 m from them. Two types of tests were performed: 1—barefoot individual and 2—footwear wielding individual (with walking shoes). All tests were performed by the same healthy individual, being female and 44 years of age.

Data acquisition for the three methods occurred simultaneously, triggered by the user, over a length of 7.08 m, the central region of 3.08 m corresponding to the length of the proposed measuring system, and the region of interest used for means of comparison. Within this region, the proposed measuring system generated results automatically. However, from the podogram and filming methods, the spatial–temporal parameters were calculated manually.

The individual walked at her preferred speed, starting at a distance from the measuring system with the intention to cause a natural tread in the region of interest, which foot initiates on the surface of the proposed measuring system is arbitrary. To eliminate the acceleration and deceleration component, the individual started the walk two meters in front of and finished two meters behind the region of interest at a natural speed. The time was obtained by equation "1" implemented as a software routine in order to sweep through the sensors matrix. For the purpose of comparison of the methods described above, the gait was considered only obtained within an area sensor matrix of 3.08 m (length) and 56 cm (width), coinciding with the active surface of the proposed measuring system. Thus, when requested by the user, the HMI displays as a visual result the image of the mapping the footsteps referring to the gait cycle in gif form.

The impressions of footsteps from the podogram method were analyzed to determine the spatial variables, such as step length, stride length, and base width, according to the protocol described in the literature [8]. The distance between the two heels impressions of different feet (from right to left) was linearly measured by a measuring tape in order to determine the step length. The distance between the impressions of two heels of the same foot (from the right to the right) was also measured in a linear way, by tracing at the base of the heels, in order to determine the stride length. Finally, cadence and speed were determined from the timed clocked.

For the analysis of video recording, the marking frame by frame of points (those associated with a position as well as time) was done manually using the motion analysis software. As a result, data sets (distance vs. time) are generated that can be analyzed. In order to provide reliable results, a metric reference was placed within the region of interest and all spatial measures obtained from the images were according to such standard. The timestamp stored naturally within each frame was used as a time standard. All measures containing the relevant information for the gait analysis, such as demarcated track length, initial and final time in seconds, right and left step length, and right and left stride length in meters were determined by using these standards. It is worth mentioning that the tests carried out did not have the objective of analyzing the individual gait for clinical purposes, but rather, validate the functionality of the developed measuring system as proposed.

3 Experimental Results and Analysis

Figure 3a, b shows the images obtained from test 1 (barefoot individual) and test 2 (individual with footwear), respectively, after the gait on the surface of the measuring system. These images represent the mapping of the gait cycle of the individual. These results are consistent with the study of Mohammed et al. [9], in which images of the plantar mapping of the phases of the gait cycle were generated from pressure sensors placed inside the shoes of two individuals.

When comparing the same individual, barefoot and using footwear, it is possible to affirm that there are differences between the spatial–temporal parameters of the gait. These differences are justified because footwear interferes with the functional capacity of the human foot, which corresponds to changes in gait [10, 11].

Besides the images, the values of spatial–temporal parameters of the gait obtained in tests 1 (barefoot individual) and 2 (individual footwear) were calculated from the measuring system (Table 1).

The values of the spatial–temporal parameters of the gait obtained by the measuring system in one 44-year-old female using footwear, correspond to the stride length of 124 cm, step length of 62 cm, and cadence of 123 steps/min (Table 1). Novaes et al.; de Almeida Busch [12] analyzed the gait of 11 females, in the age group of 40–49 years and using footwear, through the 10 m walk test. In this study, the authors find the following spatial–temporal gait values: stride length of 127 cm, step length of 63 cm, and cadence of 120 steps/min. For the same gender and age, the values of our study resemble those found by Novaes et al. [12]. This fact proves the applicability of the developed measuring system in this paper.

Fig. 3 **a** Test 1 (barefoot individual) and **b** test 2 (individual footwear)

Table 1 Results of test 1 (barefoot) and test 2 (footwear) by the measuring system

Spatial–temporal gait parameters	Test 1	Test 2
Mean stride length (cm)	123 ± 0.18	124 ± 0.18
Mean step length (cm)	62 ± 0.18	62 ± 0.18
Cadence (steps/min)	124 ± 0.5E-3	123 ± 0.5E-3
Speed (cm/s)	0.613 ± 0.5E-3	0.608 ± 0.5E-3

Table 2 Results of test 1 (barefoot) and test 2 (footwear) by podogram

Spatial–temporal gait parameters	Test 1	Test 2
Mean stride length (cm)	122.3 ± 0.18	126.0 ± 0.18
Mean step length (cm)	61.9 ± 0.18	63.5 ± 0.18
Cadence (steps/min)	124 ± 0.22E-3	122 ± 0.15E-3
Speed (cm/s)	0.762 ± 0.037E-3	0.749 ± 0.037E-3

Table 3 Results of test 1 (barefoot) and test 2 (footwear) by filming

Spatial–temporal gait parameters	Test 1	Test 2
Mean stride length (cm)	141.5 ± 0.03	141.4 ± 0.03
Mean step length (cm)	70.7 ± 0.03	70.0 ± 0.18
Cadence (steps/min)	120 ± 0.0011	115 ± 0.002
Speed (cm/s)	0.738 ± 0.5E-3	0.710 ± 0.037E-2

In another study, Pau et al. [13] analyzed the gait of 40 healthy individuals, using footwear, being 11 males and 29 females, in the age range of 26–64 years, with a mean of 41.3 years. The spatial–temporal parameters of gait were evaluated using a wearable accelerometer located at the bottom of the spine. The authors obtained the following values from their study: step length of 131 cm, the cadence of 113 steps/min, and speed of 1.24 m/s. According to Table 1, it is possible to evidence that these values differ from the ones found in our study. This is because the variety of gender and age may interfere with gait parameters [14], as well as the characteristics of the studied population and even the conditions during data collection.

The values of spatial–temporal parameters of tests 1 (barefoot individual) and 2 individuals with footwear) extracted from the podogram method are shown in Table 2. The same comparison of results obtained from the filming method is demonstrated in Table 3.

One can infer from Tables 1, 2, and 3 that the values of the spatial–temporal parameters measured by the proposed measuring system are consistent with the values obtained by the podogram and filming methods. This fact very well proved its functionality.

Regarding the performance of the proposed measuring system, the results obtained from the developed software using the Arduino and MATLAB platforms were satisfactory. It allowed for user-friendly and robust handling of the system in order to automatically acquire and calculate the necessary parameters that characterize the gait, besides its subsequent analysis. The developed algorithm was able to estimate the spatial–temporal parameters relevant to this study, namely the length and the time of the step and the stride, the cadence and the speed, showing uncertainties equal or inferior to classic non-automated techniques of measurement.

The developed device combines versatility with practicality, easy handling, modularity, and mobility. The tests demonstrated that it has adequate connectivity, allowing easy adaptation and extension of the system, when adding a number of modules according to the needs of the user.

4 Conclusion

Based on the results found, it can be stated that the measuring system is reliable in performing all the functions necessary in order to automatically map the gait cycle and quantifying its spatial–temporal parameters. The modularity and, thus, high mobility of this prototype make it a perfect tool for human gait analysis in adult and elderly individuals.

However, the measuring system needs to be validated for clinical purposes. Preliminary results obtained for the proposed spatial–temporal parameters demonstrate the importance of this study. Therefore, it is proposed that software refinements be made to identify the variability of spatial–temporal parameters of human gait, as well as the development of modules with more robust software algorithms, inserting new interfaces. After implementation, a new evaluation of the robustness of hardware will be required.

It is suggested to continue with experiments containing larger sample sizes of individuals in order to adequately describe healthy gait parameters and their possible variability, especially gait speed in individuals between 60 and 65 years of life. Further, the development of this measuring system may include clinical application in reliably quantifying habitual gait variations as a resource for rehabilitation. Some refinements in the project will make it possible to provide an on-board measuring system in homes and nursing homes to monitor the health of the individual from the analysis of daily gait in the home environment. This can be considered as a promising tool for its large-scale application in research laboratories as well as in doctor's offices and clinics.

Acknowledgements The authors thank the Federal University of Minas Gerais (UFMG) and the Brazilian National Council for Scientific and Technological Development (CNPQ) for all financial support.

References

1. Norkin, C.C., Levangie, P.K.: Joint Structure and Function: A Comprehensive Analysis. F.A. Davis Company, Philadelphia, USA (2011)
2. Bugané, F., Benedetti, M.G., Casadio, G., Attala, S., Biagi, F., Manca, M., Leardini, A.: Estimation of spatial-temporal gait parameters in level walking based on a single accelerometer: validation on normal subjects by standard gait analysis. Comput. Methods Programs Biomed. **108**(1), 129–137 (2012). https://doi.org/10.1016/j.cmpb.2012.02.003
3. Muro-de-la-Herran, A., Garcia-Zapirain, B., Mendez-Zorrilla, A.: Gait analysis methods: an overview of wearable and non-wearable systems, highlighting clinical applications. Sensors **14**(2), 3362–3394 (2014). https://doi.org/10.3390/s140203362
4. Taborri, J., Palermo, E., Rossi, S., Cappa, P.: Gait partitioning methods: a systematic review. Sensors **16**(1), 66 (2016)
5. Najafi, B., Khan, T., Wrobel, J.: Laboratory in a box: wearable sensors and its advantages for gait analysis. In: IEEE EMBS, pp. 6507–6510 (2011). https://doi.org/10.1109/iembs.2011.6091605

6. McDonough, A.L., Batavia, M., Chen, F.C., Kwon, S., Ziai, J.: The validity and reliability of the GAITRite system's measurements: a preliminary evaluation. Arch. Phys. Med. Rehabil. **82**(3), 419–425 (2001)
7. Whittle, M.W.: Gait Analysis—An Introduction. Butterworth Heinemann Ltd, Oxford, Great Britain (1991)
8. Rose, J., Gamble, J.G.: Human Walking. Williams & Wilkins, Baltimore, USA (2006)
9. Mohammed, S., Samé, A., Oukhellou, L., Kong, K., Huo, W., Amirat, Y.: Recognition of gait cycle phases using wearable sensors. Rob. Auton. Syst. **75**, 50–59 (2016). https://doi.org/10.1016/j.robot.2014.10.012
10. Franklin, S., Grey, M.J., Heneghan, N., Bowen, L., Li, F.X.: Barefoot vs common footwear: a systematic review of the kinematic, kinetic and muscle activity differences during walking. Gait Posture **42**(3), 230–239 (2015). https://doi.org/10.1016/j.gaitpost.2015.05.019
11. Hollander, K., Heidt, C., Van der Zwaard, B.C., Braumann, K.M., Zech, A.: Long-term effects of habitual barefoot running and walking: a systematic review. Med. Sci. Sports Exercise **49**(4), 752–762 (2017)
12. Novaes, R.D., Miranda, A.S., Dourado, V.Z.: Usual gait speed assessment in middle-aged and elderly Brazilian subjects. Braz. J. Phys. Ther. **15**(2), 117–122 (2011)
13. Pau, M., Corona, F., Pilloni, G., Porta, M., Coghe, G., Cocco, E.: Texting while walking differently alters gait patterns in people with multiple sclerosis and healthy individuals. Mult. Scler. Relat. Disord. **19**, 129–133 (2018). https://doi.org/10.1016/j.msard.2017.11.021
14. De Almeida Busch, T., Duarte, Y.A., Nunes, D.P., Lebrão, M.L., Naslavsky, M.S., dos Santos Rodrigues, A., Amaro, E.: Factors associated with lower gait speed among the elderly living in a developing country: a cross-sectional population-based study. BMC Geriatrics **15**(1), 35 (2015)

Energy Quality Monitoring, Testing, and Measuring System Through Mobile Networks

Felipe Henrique Ribeiro Dos Santos, Vicente Idalberto Becerra Sablón, and Annete Silva Faesarella

Abstract This article describes the development of a methodology for the communication of Weg's MMW02-M multimeter with a computer using the serial drive. Its communication happens through Modbus RTU protocol; thus, it is possible to extract and store the necessary information to make the analysis of the quality of the electric power based on Module 8—PRODIST dictated by the Brazilian National Electricity Agency (ANEEL). This data will be stored in an SQL database and processed by two softwares, Elipse E3 and Visual Studio, which have the function of processing and make available for consultation. After that, Visual Studio processes the data and sends it to a Web site, where it is available for anyone to view at any time. With real-time information, it is possible to analyze the current electrical demand, view the behavior of the power factor, active power, reactive power, three-phase voltage, phase current balancing, sags, and elevations.

Keywords Elipse E3 · Energy · Laboratory · Multimeter · Visual studio

1 Introduction

Power quality has become a major topic that is increasingly being addressed by companies. In recent years, many standards have grown around this subject, such as the International Organization for Standardization (ISO) 50001 [1].

The Brazilian National Electric Energy Agency (ANEEL) is also increasingly strict regarding the quality of electricity, in which modules were created to delimit certain parameters in some variables. One of them is Module 8—PRODIST, which reports mainly the harmonic requirements and their relationship and interaction with other variables present in the power grid.

When it comes to electricity, there is a very complex issue to look at, as it is not just the energy that is available at the power outlets of homes, businesses, or industries. It has some elements that, if they are wrong or in the wrong proportions, can shake

F. H. R. Dos Santos (✉) · V. I. B. Sablón · A. S. Faesarella
Electrical Engineering Course, University of São Francisco, Itatiba Campus, San Francisco, USA
e-mail: felipe@cseeletrica.com.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_25

and damage the sinusoid. As Sadiku [2] explains, the perfect sinusoid is one free of imperfections or interferences, presenting the 60 Hz in all its extension [2].

The main sinusoid in the system is the voltage wave amplitude. They are expressed in Hertz, and the consequences of amplitude deviations are large, including the risk of damaging some sensitive electronic equipment, even the general shutdown of transformer power. For an electrical system to be in good condition, it must have no disturbances or anomalies, and be based on a pure sine waveform, i.e., without changes in amplitude and frequency. Therefore, when it comes to low-quality energy, it is considered that there are deviations or disturbances in the sinusoid.

As already mentioned, the low quality of electricity causes many problems, and another disadvantage of this phenomenon is the increase of electricity consumption, making a substandard installation to demand more, when it comes to the consumption aspect, than an installation is in good condition. The Joule effect is also present when conditions are not good [3].

In order to verify the quality of energy, it is necessary to analyze several parameters, among them:

(a) Three-phase voltage: The presence of voltage difference between phases is analyzed;
(b) Three-phase current: The presence of phase unbalances is analyzed;
(c) Three-phase power: The power values in the phases are analyzed;
(d) Power factor: is one of the main aspects, for which the ratio of active power to apparent power is analyzed. Such a parameter must always be above 0.92.
(e) THD I: The current harmonic index is analyzed;
(f) THD U: The harmonic index existing in the voltage is analyzed;

In this context, it is necessary to seek solutions to solve the problems related to power quality and measurement of such parameters in real time. For this, multimeters are used in communication with computers, in which they play all the role of bringing information to the software, where it is processed and exported to an online graphic.

This study aims to develop a test platform to analyze and evaluate the performance and quality of electricity through software or some online hosting platform, which will allow the visualization of some parameters contained in electricity.

2 Theoretical Framework

In order to better understand the operation of the project and the whole set, it is important to have some knowledge and notions beforehand. It is necessary to understand basically the composition of the electric power, the communication methods adopted, the programming languages, and the other components.

2.1 Electric Energy and Its Variables

In electricity, many factors are present that can make major changes to the power grid and load, depending on their characteristics. Certain elements may even be grounds for fines by power utilities as they dramatically increase electrical consumption and "pollute" the original waveform, also known as the sine wave [4].

2.2 Voltage, Current, and Three-Phase Power

Sadiku [2] explains that voltage and current fluctuation is a random, repetitive or sporadic variation of its effective value. Thus, the determination of the quality of the potential difference of a distribution system bus with regard to voltage fluctuation aims to evaluate the discomfort caused by the effect of light flicker on the end consumer who has his lighting points fed at low voltage [2].

For variations to be within the maximum limits established by the regulatory body, the variation from phase to phase may not exceed 7% between them.

2.3 Power Factor

Most loads from consumer units consume inductive reactive energy, such as motors, transformers, discharge lamp reactors, induction furnaces, among others. Inductive loads require electromagnetic field for their operation, so their operation requires two types of power:

(a) Active power: the one that definitely performs work generating heat, light, movement, etc., being measured in kW.
(b) Reactive power: used only to create and maintain the electromagnetic fields of inductive loads. It is measured in kVAr.

Cogo [5] explains that when using alternating current, there are two types of power to consider: active or peal power and reactive power. The first is the one that does the expected work, for example, the production of light in a lamp or the generation of heat in an electric shower or the movement of the axis of an engine. Reactive power is used for the production and maintenance of electromagnetic fields in inductive loads and electric fields in capacitive loads [5].

The Brazilian Legislation, ANEEL, establishes that the minimum admitted PF is 0.92. This means that the consumer units must turn active power at least 92% of apparent power.

If the consumer's electrical installations are less than 92% efficient, the concessionaire will be charged a tax under FP. This tax is charged directly to the bill delivered to the consumer. If fines are being levied by the power utility, it is the customer's responsibility to resolve the issue. There are several ways to correct a low PF and therefore fail to pay a tax.

2.4 THD

The THD, or total harmonic distortion, indicates the disturbances of voltage and current waves, the level of distortion generated by harmonics. Such a phenomenon can have harmful effects on electrical installations and on quality, causing problems or instabilities in the systems, and therefore, ANEEL establishes some maximum acceptable values for the parameters:

Tolerable THD levels for voltage:

- Up to 5%: Normal levels may be considerable;
- 5–8%: Indicates significant harmonic distortion. Anomalous equipment malfunctions may occur;
- Above 8%: Reveals an important harmonic distortion. Anomalous equipment malfunctions are likely. An in-depth analysis and an attenuation system are required.

THD tolerable levels for current:

- Up to 10%: Normal levels can be considerable;
- 10–50%: Indicates significant harmonic distortion. Anomalous equipment malfunctions may occur;
- Above 50%: Reveals an important harmonic distortion. Anomalous equipment malfunctions are likely. An in-depth analysis and an attenuation system are required.

THD can be generated by many components, but the biggest villains are the equipment that implements electronic power systems. To power these devices, the equipment has a power supply switching system with an input rectifier that absorbs harmonic currents. Computers, variable speed controllers (frequency inverters), and fluorescent lamps are the main generating agents of the phenomenon [3].

2.5 The Software

To develop this project, we use three software, which is: Elipse E3, SQL Server, and Visual Studio 2017. All of them are free and can be downloaded without difficulty.

Elipse E3 is a software developed by a Brazilian company: Elipse. It has an HMI/SCADA base, i.e., platforms for supervisory monitoring and control. It has simple interfaces, but requires knowledge to operate and configure the communication protocols existing in it. It works with a block architecture [6].

Microsoft SQL Server was developed by Microsoft, a US company focused on creating operating systems and software. Such a program has the function of managing databases, creating and editing tables and their data. This is a security feature where only authenticated people or machines can make changes to the data, thus gaining more reliability and security for the project. In this work, he will do all the management of the data read by the electric meter.

Already Visual Studio 2017, also developed by Microsoft, is a software aimed at compilation and development of programs. It can work with several programming languages, including C, C++, C#, F#, ASP.NET, and others. In this project, VStudio has the function of collecting the last read data from the SQL database and exporting it to the online server.

2.6 PhpMyAdmin

PhpMyAdmin is known to be a free and open-source Web application. It allows database management, table creation and editing, and table alteration. Created in 1995, this language is characterized as versatile and easy to program [7].

For this project, the information is hosted at https://br.000webhost.com/, but there are cases where the server is directly linked to the phpMyAdmin directory.

One of the reasons for using it in the project is that it does not depend on operating licenses, is completely free, and is available in any browser. Through it, it is also possible to graph the parameters, thus facilitating the visualization of the analyzed variables.

2.7 RS485 Communication Protocol

Recommended Standard-485 (RS-485) is a type of communication protocol in which machines and meters exchange data. It is one of the most used media and can reach speeds of approximately 50 Mbps. From this type of communication, an open protocol, Modbus remote terminal unit (Modbus RTU), is used to take readings from the multimeters connected to the computer [8].

The Modbus protocol works with a master and a slave, in which the master sends a command to the slave and he responds with the read values. Such communication is via the Recommended Standard network-485 (RS-485). From the output of the equipment, a cable is connected to the RS-485/USB converter, where it is connected to the computer, where the information is read, removed, and processed by the programs.

3 Methodology

For this project, an integrated system for reading and storing data extracted from a multimeter of electrical quantities was developed. For this, the Weg MMW02-M multimeter connected to an RS-485 to USB converter was used, which in turn Elipse E3 software communicates and extracts data, saving it into an offline SQL database. The Visual Studio program captures them and generates a link to write them to the phpMyAdmin server, which is later read by a Web site and generates the complete graph [9].

In Step 1, the multimeter data acquisition is done through its connection to the power grid and the CTs, as Fig. 1 illustrates in more detail.

In the test of this project, there were no CTs, so only the voltage and frequency variables were analyzed, as shown in Fig. 2.

Because readings are performed at a 127 V regime, only phase A to neutral voltage is available.

It is also important to make the correct configuration of the device, because if it is parameterized incorrectly, it becomes impossible to connect to Step 2. For this purpose, a 9600 bps 8N1 Baud was adopted, with the multimeter addressed in 1.

In Step 2, the Elipse program is charged with making the informed readings. In this example, three-phase voltage, the three-phase voltage to neutral, THD reading of phase A voltage to neutral, three-phase current, and three-phase power are scanned. As can be seen in Fig. 3, the values of V2, V3, U1, U2, U3, and P are null because the CTs were not connected to the meter.

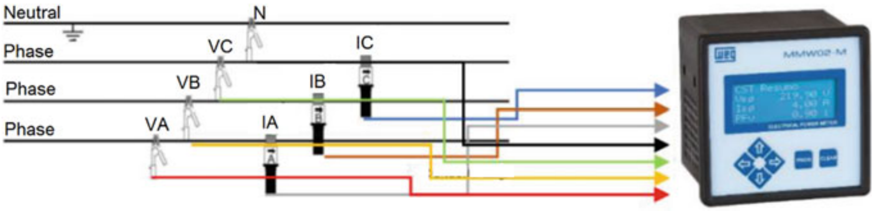


Fig. 1 MMW02-M electrical connection

Fig. 2 Electrical meter display

Nome	Dispo...	Item	P1/N1...	P2/N2...	P3/N3...	P4/N4...	Ta...	Var...	Valor
Driver1			0	0	0	0			
VT			1	10	0	3	1	1000	
Element							0		9 38,61518
V1			1	10	0	5	1	1000	
Element							0		9 115,8455
V2			1	10	0	7	1	1000	
Element							0		Null
V3			1	10	0	9	1	1000	
Element							0		Null
Hz			1	10	0	67	1	1000	
Element							0		9 60,0124
THD-AN			1	10	0	601	1	1000	
Element 1							0		9 4,488369

Fig. 3 Ellipse E3 taking measurements

E3TimeStamp	VT	V1	V2	V3	Hz	U1	U2	U3	P	TH
07/05/2019 00:02:24	0	0	0	0	0	0	0	0	0	0
07/05/2019 00:03:24	38,8910865...	116,673263...	0	0	59,9542274...	0	0	0	0	4,63006591...
07/05/2019 00:04:24	38,8205451...	116,461631...	0	0	59,9785461...	0	0	0	0	4,64217853...
07/05/2019 00:05:24	38,8657341...	116,597198...	0	0	59,9959373...	0	0	0	0	4,65409564...
07/05/2019 00:06:24	38,8376083...	116,512825...	0	0	60,0066375...	0	0	0	0	4,67165803...
07/05/2019 00:07:24	38,8042068...	116,412620...	0	0	60,0263977...	0	0	0	0	4,53208827...
07/05/2019 00:08:25	38,7041816...	116,112541...	0	0	60,0140342...	0	0	0	0	4,48943233...
07/05/2019 00:09:25	38,7853736...	116,356124...	0	0	60,0111846...	0	0	0	0	4,44716358...
07/05/2019 00:10:25	38,7722320...	116,316696...	0	0	59,9905815...	0	0	0	0	4,53967905...

Fig. 4 SQL database

In Step 3, after acquiring the values, the program saves them to the SQL database. The same is fed every minute with new information, and the Elipse triggers the information in the database to save in this period. This time is fully adjustable and can be ten minutes or one hour, depending on customer preference. Figure 4 illustrates the saved information.

In the fourth step, the Microsoft Visual Studio software is responsible for designing a URL that will fire the latest data to the PHP online server. The basis of this generated link is the latest database information, and the last readings are taken every minute. Inside Visual Studio, a programming code, written in C# language, is generated, which one will do all this process of collecting information and generating the URL automatically, without requiring manual authentication. It is happening through this URL that the values will go to phpMyAdmin. Inside it, and on the hosting site, there are some functions that are the key to the proper functioning of the entire system.

In the fifth and last Step, the index.php file collects data from the online server and generates a graph. In the graph, the power factor, active power, reactive power, and apparent power values are not explicit because these variables are too low or too high, thus changing the accuracy of the readings. You can access this data through the main database dashboard, either online or offline, and download the complete table with all measurements.

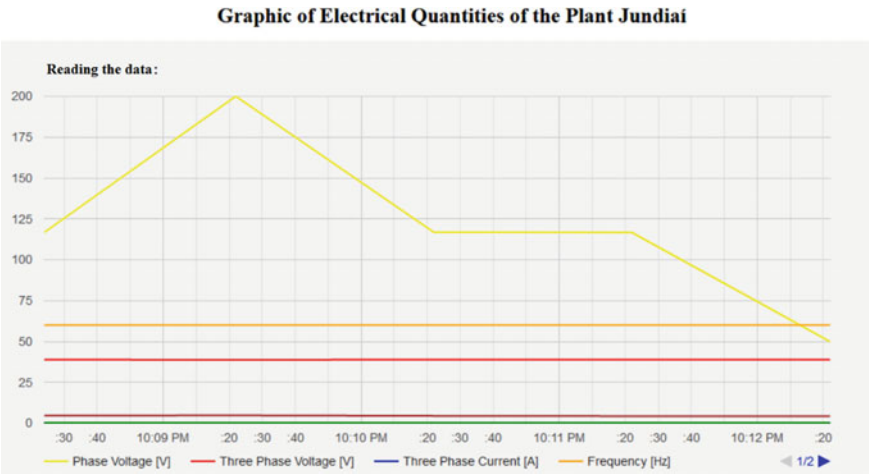


Fig. 5 Generated graph

To have access to the database and its values, you must enter the page domain, access the desired table, and select the export option. Here, you can select several different file extensions to save or generate a spreadsheet through online banking.

In Fig. 5, you can see the graphic generation by the site.

4 Results and Discussion

It was noted that for the operation of the project, no equipment is required or require a lot of investment, just need to purchase the Weg MMW02-M multimeter and three TC's, since all software was downloaded for free from the suppliers' Web site. It was also possible to observe a good data transfer speed in Elipse E3, without data loss or operating instability. This situation was also present in the SQL Database, where at no time were there any records of crash or loss of information. In Visual Studio, however, there was a slight problem with the programming code. The same was solved by replacing some codes. And for the phpMyAdmin server and the site, during the entire testing period, no jitter points were generated either, as it depends on the third part server. It was possible to validate the application and found that it obtained great performance.

With a project in operation, it is possible to get more easily measure the electrical quantities in plants or substations that are difficult to access. The most important results obtained from this research are:

(a) Easily evaluate the quality of electricity;
(b) Easy reading electrical data of long-distance installations;
(c) Availability of data via online platforms;

(d) Critical point alarms when the electrical installations are presenting poor quality.
(e) Easy visualization of the curves of the configured electrical quantities.

5 Conclusion

All project steps were successfully completed. Multimeter measurements were easily performed. Data acquisition by E3 took some work until the settings were adjusted, and the data writing to the SQL database was completed without much work, but the Visual Studio programming code had a certain difficulty, but it was fixed, and the program runs correctly, and finally, the graph generation on the online platform was completed with difficulty the medium degree as there were some script changes due to the installed version of PHP on the local server.

It was also concluded that it was possible to perform all work requests, first reading the data at the installation site, then communicating with the Elipse E3 software, then storing the data in the SQL Bank and generating the programming scripts in Visual Studio, and finally, the generation of graphics on the online platform, thus making it possible to better evaluate the quality of electricity, facilitate the reading of these parameters in long-distance installations, still making this data available on online platforms and making it available by the alarms and curves of the electrical variables selected in the program.

References

1. Agência Nacional De Energia Elétrica: Prodist Módulo 8: Procedimento De Distribuição De Energia Elétrica No Sistema Elétrico Nacional—Qualidade Da Energia Elétrica. Available: https://www.aneel.gov.br/documents/656827/14866914/m%c3%b3dulo8_revis%c3%a3o_9_limpo.pdf/d689590a-231d-b8ad-b354-bd6e0a4cd401. Accessed 14 May 2019
2. Sadiku, M., Alexander, C.K.: Fundamentos De Circuitos Elétricos. AMGH, São Paulo (2013)
3. Rocha, J.E.: Qualidade Da Energia Elétrica. Curitiba (2016)
4. Riedel, S.A., Nilsson, J.W.: Circuitos Elétricos. Pearson Universidades, São Paulo (2015)
5. Cogo, J.R., Siqueira, J.B.: Capacitores De Potência E Filtros De Harmônicos. Ciência Moderna, São Paulo (2018)
6. Fernandes, T.: Implantação De Um Weblab Para Ensino De Controle Em Curso De Engenharia (2011)
7. Welling, L., Thomson, L.: Php e Mysql – Desenvolvimento Web. Campus, São Paulo (2005)
8. Embarcados. Available: https://www.embarcados.com.br/protocolo-modbus. Accessed 14 May 2019
9. Weg. Available: https://www.weg.net/institutional/br/pt/. Accessed 14 May 2019

Analysis of Electrical Adequacy in Riverside and Road Schools in the Amazon

Thainá Reis Brasil Ribeiro and Rubens de Andrade Fernandes

Abstract The Amazon region is composed of riverside and indigenous communities, which have a way of life integrated with the dynamics of nature. Through governmental social projects, these communities now have access to technological and social development. Electricity is one of the factors driving development in these communities, as well as education. However, due to the lack of inspection and maintenance processes, local facilities remain in a precarious condition and constant physical wear, especially in local schools. This paper aims to present an analysis of the current adequacy of the electrical installations in some schools to contribute to the improvement of the teaching quality and the supply of electric energy in these localities.

Keywords Riverside communities · Indigenous communities · Energy efficiency · Electrical installations · Sociotechnological development

1 Introduction

Brazil, one of the world's references in natural resources, has much of its energy use from hydroelectric, thermoelectric, nuclear, wind and solar plants, reaching an installed capacity of 150,338 MW of electricity in 2016 [1]. However, much of Brazil's rural environment does not have the benefit of electricity. This not only affects the technological development of the interior of the region, but also the social and economic aspects of the population. This includes the riverside and indigenous communities of the Amazon region.

The "Luz para Todos" program aims to reduce social and technological inequalities through universal access to electricity in rural areas [2]. The goal is to promote electricity distribution services for the social inclusion of low-income families. This has also made it possible to implement other rural development programs, including the installation of photovoltaic solar systems in these communities [3].

T. R. B. Ribeiro (✉) · R. de Andrade Fernandes
State University of Amazonas, Manaus, Amazonas 1200, Brazil
e-mail: trbr.eng@uea.edu.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_26

However, indigenous and riverine populations are in hard-to-reach regions, making it even more difficult to supply electricity and implement the program with quality. With the poor economic and technological performance of rural Amazonia, these communities lose visibility and often lose social relevance for the urban community.

In seeking alternative sources of income or livelihoods, many rural households live off agriculture. As this livelihood is explored in the Amazon, fauna, and flora are affected by burning and land demarcation, which imposes an unsustainable livelihood and compromises one of the most important ecosystems on the planet [4].

One of the social sectors most affected by the quality of program implementation is education. Many of them are located near state highways to facilitate access to electricity. In addition to the low indicators of electrical performance quality, rural school facilities have malfunctions in their structure, compromising the health of the students and teachers of the institutions.

The purpose of this paper is to present a case study regarding the evaluations of the electrical structural part of four municipal roads, riverside or indigenous schools near the municipality of Manaus, in the state of Amazonas. The intention is to propose, along with the responsible bodies, notes of improvements regarding the input of energy, lighting, and adequate cooling for the comfort of students, teachers, and staff. Besides, it is also expected to increase students' school performance and also add visibility to the riverside community, often overlooked by the urban community, seeking to highlight its social relevance through the real changes in schools visited socially, economically, and environmentally.

2 Materials and Methodology

To carry out the research regarding the analysis of electrical structure in rural schools, it was granted from the appropriate permissions of the responsible authorities of the Municipal Department of Education, SEMED, of the city of Manaus.

The study conducted in this work was based on the Brazilian standard ABNT NBR 5410, which establishes the requirements to be met in low voltage electrical installations, to ensure the safety of people and animals, the proper functioning of the installation and the conservation of aggregate goods [5]. From it, it is possible to obtain the minimum specifications of materials, instruments, and equipment of the installation, as well as the definition of the electrical structure of the installation and its respective dimensioning.

2.1 Recommendations for Elaboration of Electrical Projects

Following NBR 5410 and the guidelines of the Municipal Department of Education, SEMED, of the city of Manaus, all school units must have a low voltage electrical project made with a minimum power forecast for the lighting circuits, general-purpose and specific use outlets to be installed.

Table 1 Provisions on electricity distribution boards of school units

Item	Provisions
1	All school units that have more than one distribution board must have a three-phase busbar box, with neutral and ground, or general distribution board
2	For schools with more than one floor, a minimum of one three-phase distribution board per floor should be installed, and this distribution board should only supply circuits contained on the floor where it is installed
3	Each classroom shall independently contain its lighting circuits, sockets, and air conditioners, with minimum cross-sectional cables of 2.5 mm^2, 2.5 mm^2, and 6 mm^2, respectively
4	Install a junction box near the switchboards and the switchboard. Its dimensions should be at least $80 \times 80 \times 80$ cm

The types of electricity supply to schools in rural areas, whether riverside, indigenous or road, are defined according to the installed load, the demand, the type of network and shall have a substation or generator compatible with the installed load [6, 7].

2.2 Electricity Distribution Boards

The distribution board is one of the main elements in the electrical installations of education centers, as they concentrate the protection devices and the terminal circuits that feed the loads. There are indispensable criteria for configuring the distribution boards of SEMED-linked schools, which are listed in Table 1 [6].

2.3 Power Quality

Regarding the quality of the performance of electrical parameters, the procedures for distribution of electricity in the national electrical system, known as PRODIST, define in Module 8 the terminology that characterizes the phenomena and establishes the indicators and limits related to the conformity of electric voltage.

When it comes to the voltage levels in the power supply to the interior communities, the points of connection to the distribution network between distributors and with the consuming units should be evaluated.

Voltage readings can be classified as adequate, precarious or critical, based on the deviation of the reading voltage value from the reference voltage, as defined in Table 2 [8].

Table 2 Connection points in nominal voltage or under 1 kV (220/127 V)

Service voltage (SV)	Rated voltage (V)	Reading voltage range (V)
Adequate	220	$202 \leq SV \leq 231$
	127	$117 \leq SV \leq 133$
Precarious	220	$191 \leq SV \leq 202$ or $231 \leq SV \leq 233$
	127	$110 \leq SV \leq 117$ or $133 \leq SV \leq 135$
Critical	220	$SV < 331$ or $SV > 403$
	127	$SV < 191$ or $SV > 233$

2.4 Photovoltaic Solar Power Generation

The direct conversion of solar energy into electrical energy occurs by the effects of radiation on semiconductor materials, which we call the photovoltaic effect. Photons contained in sunlight are converted into electrical energy through the use of solar cells. These systems can be configured in connection with the on-grid distribution system or installed in isolation from the grid, known as off-grid systems.

Off-grid systems are an excellent solution to problems related to the isolation of rural and riverside communities in the Amazon region and can be deployed as a primary or secondary generator.

There are also hybrid systems integrating photovoltaic panels and diesel generator sets, such as the system implemented in 2001 in the municipality of Nova Mamoré in the state of Rondônia, Brazil [9].

3 Case of Study

For the implementation of this work, the electrical structure of four rural schools was analyzed, with support from SEMED's engineering department, among them, two road schools and two riverside schools.

These schools work in a single shift (morning or afternoon) and serve from two to seven classes of children attending grades 2–9. The verified data were cataloged through reports and summarized in the following subsections.

3.1 Abílio Alencar Municipal School

Following is the information obtained from the Abílio Alencar municipal school, whose facade is shown in Fig. 1:

- Address: Torquato Tapajós Highway, Km. 35—AM 010 (Manaus—Itacoatiara);
- Classification: road school;

Fig. 1 Facade of Abílio Alencar municipal school

- Supply type: four-wire (three phases and one neutral), medium voltage, 75 kVA substation;
- Service status: the phase-to-phase and phase-to-neutral voltage measurements were in an adequate state;
- Installed power: 40.6 kW;
- Average power consumption: 5.323 kWh;
- Electrical distribution board: two electrical distribution boards without the presence of a general distribution board or busbar;
- Power generator: The school has no additional power generation system;
- Protection and cabling system presented in Table 3.

Table 3 Cabling and protection data of Abílio Alencar municipal school

Item	Protection element	Cabling
Power supply input	Fuses, 300 A	Features 50 mm^2 cross-sectional cabling from the input branch to the power meter input and 70 mm^2 at the meter output
Distribution Board 01	Breaker, 125 A	Powered through 16 mm^2 rigid cables without grounding
Distribution Board 02	Breaker, 150 A	Powered through 50 mm^2 rigid cables without grounding

3.2 Prof. Paulo César da Silva Municipal School

In this subsection information about the Prof. Paulo César da Silva municipal school is presented, whose facade is shown in Fig. 2:

- Address: Nova Esperança Community, Igarapé do Tiú - Rio Negro;
- Classification: riverside school;
- Supply type: four-wire (three phases and one neutral), low voltage, 25 kVA substation;
- Service Status: It was observed that the service tension presented a critical level, as shown in Fig. 3;
- Installed power: 21.6 kW;
- Average power consumption: 1500 kWh;

Fig. 2 Facade of Prof. Paulo César da Silva municipal school

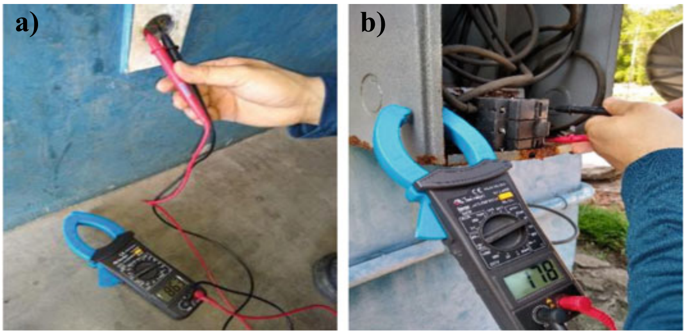


Fig. 3 Measurement of phase-neutral (**a**) and phase-to-phase (**b**) electrical voltage

Table 4 Cabling and protection data of Prof. Paulo César da Silva municipal school

Item	Protection element	Cabling
Power supply input	Breaker, 100 A	Powered through 35 mm^2 rigid cables, without grounding
Distribution Board 01	Breaker, 50 A	Powered through 10 mm^2 rigid cables, without grounding
Distribution Board 02	Breaker, 100 A	Powered through 35 mm^2 rigid cables, without grounding
Distribution Board 03	Breaker, 63 A	Powered through 16 mm^2 rigid cables, without grounding

- Electrical distribution board: three electrical distribution boards without the presence of a general distribution board or busbar;
- Power generator: The school has a disused 24 kVA generator set due to lack of maintenance;
- Protection and cabling system presented in Table 4.

3.3 Nossa Senhora do Carmo Municipal School

Here will be presented the information of the municipal school Nossa Senhora do Carmo, where the facade is shown in Fig. 4:

- Address: Paraná da Eva—Amazon River;
- Classification: riverside school;

Fig. 4 Facade of Nossa Senhora of the Carmo municipal school

Table 5 Cabling and protection data of Nossa Senhora do Carmo municipal school

Item	Protection element	Cabling
Power supply input	Breaker, 63 A	Powered through 16 mm^2 rigid cables, without grounding
Distribution Board 01	Breaker, 50 A	Three-phase, 100 A, busbar frame, powered by 10 mm^2 rigid cables without grounding

- Supply type: three-wire (two phases and one neutral), low voltage, 25 kVA transformer;
- Service status: At the time of the visit, it was contacted that only one phase was energized in the distribution board; as a result of this, several types of equipment were not working;
- Installed power: 8.3 kW;
- Average power consumption: 342 kWh;
- Electrical distribution board: only one electrical distribution board in the school;
- Power generator: The school has a disused 24 kVA generator set due to lack of maintenance;
- Protection and cabling system presented in Table 5.

3.4 Epitácio Nunes da Fonseca Municipal School

Here will be presented the information of the municipal school Eptácio Nunes da Fonseca, where the facade is shown in Fig. 5:

Fig. 5 Facade of Eptácio Nunes da Fonseca municipal school

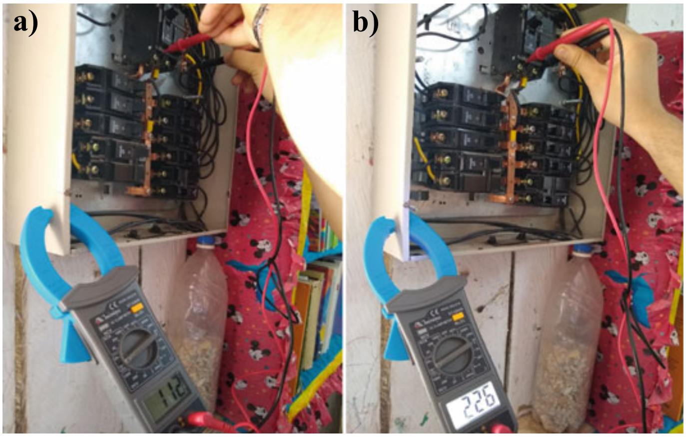


Fig. 6 Measurement of phase-neutral (**a**) and phase-to-phase (**b**) electrical voltage

Table 6 Cabling and protection data of Eptácio Nunes da Fonseca municipal school

Item	Protection element	Cabling
Power supply input	Breaker, 50 A	Powered through 10 mm^2 rigid cables, without grounding
Distribution Board 01	Breaker, 50 A	Powered through 10 mm^2 rigid cables, without grounding

- Address: Nova Luz Community—BR 174—Km. 21, branch Bom Destino;
- Classification: road school;
- Supply type: three-wire (two phases and one neutral), low voltage, 25 kVA transformer;
- Service Status: the phase-neutral voltage measurements were in a precarious state, as shown in Fig. 6;
- Installed power: 7.5 kW;
- Average power consumption: 1682 kWh;
- Electrical distribution board: only one electrical distribution board in the school;
- Power generator: The school has no additional power generation system;
- Protection and cabling system presented in Table 6.

4 Results Obtained

With the data obtained and the applicable technical standards, NBR-5410 and the Construction and Rental Guide, already mentioned, it is possible to determine the precariousness of the electrical installations described herein.

Due to the difficulty of access and communication between schools and the support provided by the capital, many changes in the initial electrical design of these schools were made by the community itself in an attempt to meet emerging demands, losing the standard technical identity. This compromises both the safety of people entering the environment and the preservation of locally available material goods. For the community, for the most part, it is not endowed with people qualified and qualified to perform electrical maintenance.

Turning to the field of education, there are constantly burned out lamps, badly sized air conditioners, faulty or nonexistent, exposed cabling, unprotected sockets, among others.

It is noted that most of these schools do not have grounding systems and the switchboards are connected without the presence of a busbar or a general distribution board. Also, protective device sizing is often oversized and cabling is constantly undersized.

Another worrying factor is the level of electrical voltage that reaches the entrance branch of these schools since they participate in the "Luz para Todos" program, are supplied by very long transmission lines and poor and critical voltage levels are a worrying reality, but common. It was unanimous, among the four schools, the reports about constant falls, power surge and equipment burn, which indicate bad parameters of the quality of electrical performance.

To solve problems with power outages, rural schools must have an independent power generation system. In the case of schools supported by SEMED, diesel generator sets are used, but the vast majority of these types of equipment are not regularly serviced as there is no local labor specialized in this service and end up being rendered unusable or poorly functioning.

5 Solution Proposal

Based on the previous topic, it is possible to cite some possible solutions according to the current scenario of these rural schools.

5.1 Solutions for Problems Related to the Electrical Structure

It is plausible to design a new electrical project taking into account the particularities of each school and to implement training programs to enable the local workforce to effectively solve future electrical problems and to carry out minor electrical reforms when necessary to improve the quality of service to students.

Replace damaged equipment and maintain a stock of equipment that is more sensitive to electrical surges, such as light bulbs.

5.2 *Solutions for Power Outages and Power Surges*

Knowing that rural schools operate in the mornings or afternoons, that is, they operate in the period of abundant sunshine, a solar generator can be used to maintain the loads in the event of a power outage or surge.

The off-grid photovoltaic generator is ideal for powering loads in remote locations where there is no utility grid or when power surge is too frequent. In this way, this type of system could replace diesel generators in both rural schools served by "Luz Para Todos" and unattended schools. This generator type has a modular architecture and can be extended to countless situations allowing flexibility and wide coverage, adapting to the needs of the school.

6 Conclusion

The four schools analyzed in this project are used as a reference to understand the reality of the 51 riverside schools and 36 road schools present in the interior of the Amazon, where we observed problems in the supply of energy, under dimensioning of the lighting and cooling of the school environment, oversizing of protective devices, lack of grounding system, among others.

It is necessary to restructure the electrical installations and basic equipment used within the building environment, so that students have a comfortable and safe environment to study. Besides, training plans should be implemented for the local population, generating jobs, and facilitating the maintenance of these systems. Investment in alternative energy sources can enable sustainable independent generation in indigenous and riverside environments, as well as improve the power quality parameters.

It is hoped to raise awareness among the scientific community and the authorities of the current scenario of electric power distribution in this environment to improve social, economic, and technological development in isolated communities in the Amazon.

References

1. Assad, V.C.D., Batista, O.E.: Scenario and perspective of distributed generation in Brazil. In: 2018 Simposio Brasileiro de Sistemas Eletricos (SBSE), pp. 1–5, Niteroi (2018)
2. Freitas, G., Silveira, S.F.R.: Programa Luz Para Todos: uma representação da teoria do programa por meio do modelo lógico. Planejamento e Políticas Públicas, n 45 (2015)
3. Diniz, A.S.A.C., Franca, E.D., Camara, C.F., Morais, P.M.R., Vilhena, L.: The important contribution of photovoltaics in a rural school electrification program. In: 2006 IEEE 4th World Conference on Photovoltaic Energy Conference, pp. 2528–2531, Waikoloa, HI (2006)
4. Andrade, D.F.C., Gama, J.R.V., Ruschel, Melo, A.R., L.O., Avila, A.L., Carvalho, J.O.P.: Post-fire recovery of a dense ombrophylous forest in Amazon. Anais da Academia Brasileira de Ciências, vol. 91, n. 2, Rio de Janeiro (2019)

5. Associação Brasileira De Normas Técnicas. NBR 5410: Electrical installations of buildings – Low voltage. 2 ed., n 209, Rio de Janeiro (2004)
6. Secretaria Municipal De Educação. Guia de Contrução e Locação 2018. 2 ed., n44, Manaus (2018)
7. Eletrobrás. NDEE-02: Fornecimento De Energia Elétrica Em Baixa Tensão (Edificações Individuais). 1 ed., n 157 (2014)
8. Agência Nacional De Energia Elétrica. Procedimentos De Distribuição De Energia Elétrica No Sistema Elétrico Nacional - PRODIST: Modulo 8 - Qualidade Da Energia Elétrica. 10 ed., n 88, Brasil (2018)
9. ANEEL. Atlas de Energia Elétrica. 2 ed., n 246, Brasil (2005)

An Overview of Magnetic Field Visualization Methods and Applications for Macroscale Objects

Leandro Bachega Trujilho and Osamu Saotome

Abstract There are different techniques and devices for visualizing magnetic fields in macro-, micro-, and nanoscale. A non-exhaustive literary overview of macroscale magnetic field visualization methods is presented. Theoretical and experimental techniques are briefly described and contrasted over computational cost, implementation, and usability criterion. Finite element methods (FEM) have a higher computational cost that finite volume methods (FVM) and that meshfree approximation method. Scanning systems have a low computational cost as the input dataset size is small, and there is no need to solve differential equations to reconstruct the fields. Augmented reality (AR) systems have the largest computational cost of all methods. FEM and FVM simulations are simple to implement in 2D and can become quite complex in 3D. All experimental setups are not trivial to build. Scanning systems requires the construction of a linear X,Y mechanical table and a controller design to synchronize sensor readings with probe movement. They are slow, mapping a single surface can take 1 h (Tumanski in Handbook of magnetic measurements (Series inSensors), 1st edn. CRC Press, 2016). Magnetic cameras need hardware design, embedded and application software development. They are simple to use and provide real-time imagery of the field. Finally, AR systems depend on head-mounted devices, 3D FEM models, image processing algorithms to identify fiducials and synchronism of real and virtual objects (Matsutomo et al. in IEEE Trans Magn 53(6):1–4, Article No 8102004, 2017). Despite the high computational cost and the difficulty to build, it provides an immersive interactive experience for the user.

Keywords Magnetic field visualization · Magnetic field intensity maps · Magnetic cameras

L. B. Trujilho (✉) · O. Saotome
Department of Electronics and Computer Engineering, Aeronautics Institute of Technology, São José dos Campos, SP 12228-900, Brazil
e-mail: leandro.trujilho@yahoo.com

O. Saotome
e-mail: osaotome@ita.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_27

1 Introduction

Magnetism is a well-researched and extensive subject. The present overview compiles theoretical and experimental magnetic field imagery techniques with a focus on methods that can be applied to visualize the field from macroscopic scale magnetized objects. The goal is to summarize, centralize, and list relevant developments updating the reader with the latest field advancements and highlighting the main characteristics and drawbacks of each method. Micro- and nanoscale magnetic fields are out of the scope of this work and are covered in detail by Hopster and Oepen [1]. Field visualization has applications in nondestructive testing (NDT), localization and tracking of metallic objects, human–machine interface and can be used for educational purposes.

2 Simulation and Theoretical Models

2.1 Methods with Mesh

Finite Element Method. The classical way of simulating a static magnetic field emanating from a permanent magnet is via the finite element method (FEM). This method requires the definition of the magnet area and the boundaries surrounding it so that a mesh can be generated within the limits. The mesh is formed by nodes and regions called elements where magnetostatics partial differential equations (PDEs) are numerically solved. Some advantages include simulation software availability such as finite element method magnetics (FEMM) [2] which is easy to use for simple geometries, the capability to handle complex geometries and dynamic fields, it is a popular method that can accommodate material properties and can implement higher-order elements. However, it is computationally expensive due to the necessity of mesh creation and solving the equations in each node and element. For instance, a simple simulation of a bar magnet in FEMM in 2D within a radius of 1 cm as shown in Fig. 1 generates 4359 nodes and 8356 elements. The size of the mesh grows fast in 3D.

Finite Volume Method. The finite volume method converts the divergence term of the PDEs into surface integrals using the divergence theorem and evaluates the fluxes at the surfaces of the magnetized object. This eliminates the need for free space mesh and boundary conditions, limiting the mesh only to the active surface region. Therefore, it is less computationally intensive than FEM, but the functions that approximate the solution cannot be easily made into higher orders which are a disadvantage. There are several integral formulations for solving magnetostatics problems available such as the magnetic moment method that is based on uniform magnetization on each element, B-Facet which is based on interpolation of magnetic flux, and the volume integration method, based on the interpolation of the magnetic vector potential on edge elements. They are listed and contrasted by Le-Van et al. [3].

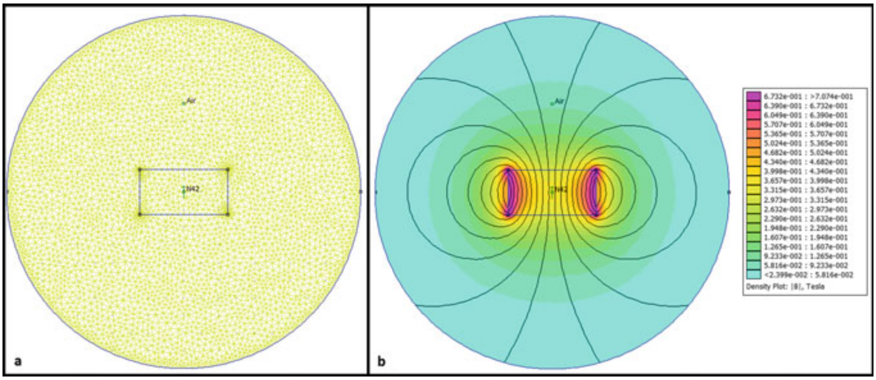


Fig. 1 **a** Mesh on a 1 cm radius surrounding a bar magnet. **b** Field solved in all elements. Magnetic intensity B is 0.08 T at 0.5 cm distance northbound

2.2 Meshfree Methods

Algebraic approach. The dipole equations can be used directly to display the distribution of all the three components of the magnetic field. Equations (1) and Fig. 2 were extracted from Lima et al. [4]:

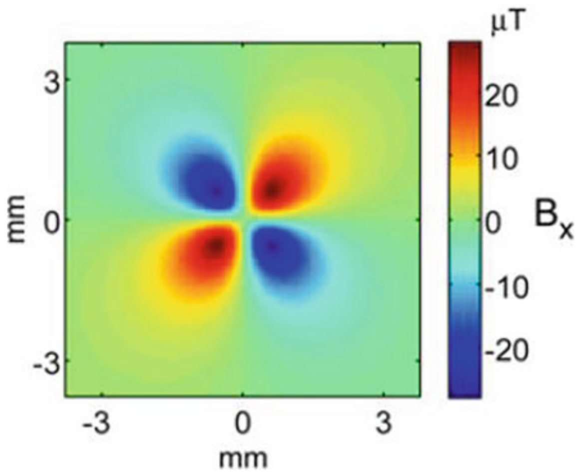


Fig. 2 Simulated *x* component field map calculated on a 128 × 128 horizontal square grid at 1 mm above the sample target dipole aligned in the *y*-direction [4]

$$
\begin{aligned}
B_x(x, y, z) &= \frac{\mu_0 m}{4\pi} \frac{3xz}{\left(x^2 + y^2 + z^2\right)^{\frac{5}{2}}} \\
B_y(x, y, z) &= \frac{\mu_0 m}{4\pi} \frac{3yz}{\left(x^2 + y^2 + z^2\right)^{\frac{5}{2}}} \\
B_z(x, y, z) &= \frac{\mu_0 m}{4\pi} \left[\frac{3z^2}{\left(x^2 + y^2 + z^2\right)^{\frac{5}{2}}} - \frac{1}{\left(x^2 + y^2 + z^2\right)^{\frac{3}{2}}} \right]
\end{aligned}
\tag{1}
$$

Ren et al. [5] wrote that the dipole model is not suitable for the near field and enhanced it by adding and tuning coefficients to fit the field distribution against measured data via minimization of a sum of the squared error function, the least squares method. There are several statistical learning algorithms used in regression, such as the Gauss–Newton, gradient descent, and the Levenberg–Marquardt.

Machine Learning. Another approach is to learn and reconstruct the field from sensor data. Magnetic fields do not diverge and follow a Gaussian distribution. Therefore, kernelized learning algorithms that use the positive-definite radial basis kernel are well-suited to reproduce magnetic fields.

This kernel returns a Gaussian distribution between two points and has the capability of keeping symmetry when transformed spatially. That allows orthogonal projection into higher dimensions via inner product and since the components of the magnetic field are dependent on each other [4], it is also appropriate for component estimation. Macedo and Castro [6] showed that a matrix-valued RBF kernel (2) can be used to learn divergence-free vector fields, and Wahlström et al. [7] used this kernel to estimate the field of a metallic table in a room using Gaussian process. Being able to choose the kernel is an advantage of these methods as it considers expert knowledge. Gaussian process is a nonparametric technique and uses the entire training dataset for estimation, and thus, the computational costs go up with the number of samples. Many machine learning libraries are freely available online.

For further details on meshfree approximation techniques, the reader can refer to Chen et al. work [8] which is comprised of a 20 years survey on meshfree methods covering a broad range of applications.

$$
K(P, Q) = \sigma^2 \mathrm{e}^{-\frac{\|P-Q\|^2}{2l^2}} + \sigma \tag{2}
$$

3 Experimental Systems with Visual Effects

In 1820, Hans Christian Oersted demonstrated that a current-carrying wire could deflect a compass needle, linking electricity and magnetism. Shortly after, in 1845 Michael Faraday observed that a beam of polarized light rotated proportionally to the magnetic field, connecting magnetism with optics. In 1851, he also showed that iron filings formed lines when placed in the vicinity of a permanent magnet. The next advancement took place in 1963 when Stephen Papell invented the ferrofluid [9] which allowed 3D visualization of magnetization.

The advent of sensors for measuring magnetic intensities gave rise to other approaches to display the field. Henrichsen [10] provided an introductory overview of measurement and visualization methods, calibration concerns, and instrumentation accuracy per field strength as summarized by Fig. 3.

Nuclear magnetic resonance (NMR) is the standard for calibrating magnetometers. The most important issue for Hall effect and magneto-resistive calibration is temperature effects on the measurements.

Tumanski [11] wrote a detailed review of modern magnetic field sensors. Figure 4 illustrates the range of field measurements per sensor type.

Hall effect and magneto-resistive sensors—AMR and GMR are available off the shelf—are cheap and sensitive enough to for macroscale measurements involved in metallic target localization and tracking [12] and non-destructive testing applications [13, 14]. A detailed work of magnetic field measurements can be found in [15].

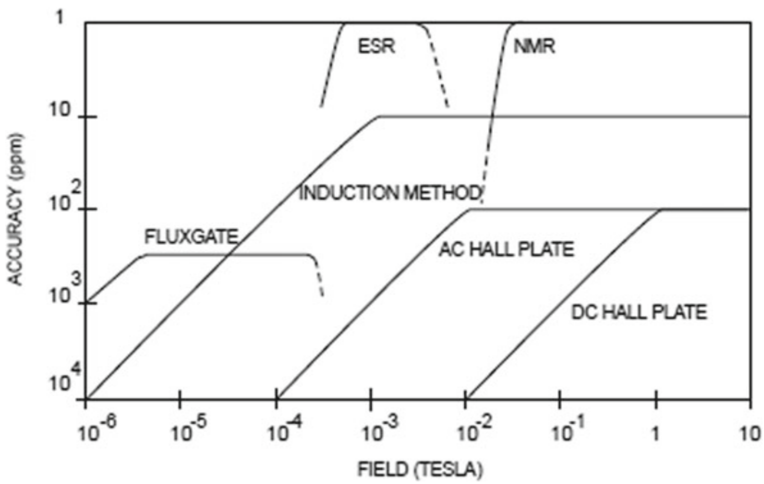


Fig. 3 Measurement methods: accuracies and ranges [10]

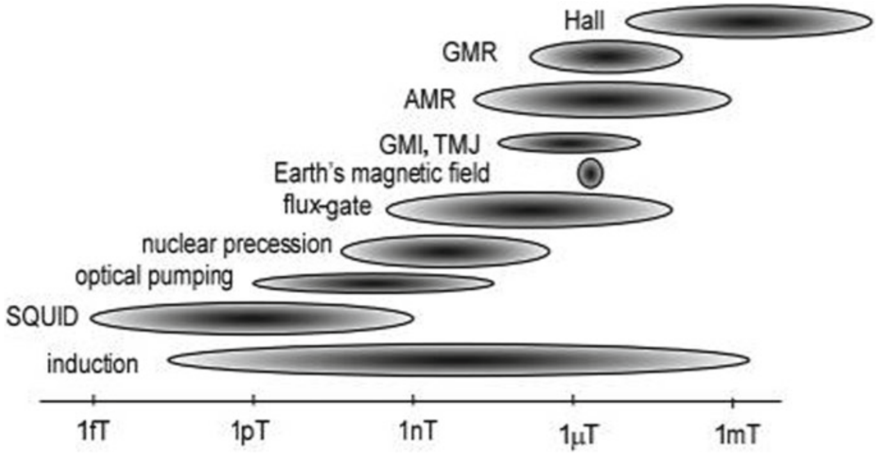


Fig. 4 Magnetic sensors field intensity range per sensor type [11]

3.1 *Scanning Devices*

Scanning systems as illustrated by Fig. 5 consist of a magnetic sensor that is mounted on the extremity of support and moved linearly following a raster pattern by step motors on the top of a target magnetized object. Sensor data is captured in every point that the probe moves, and intensity height maps are produced by a personal computer when the scan is complete. The advantage of such a system is the map resolution, but a single map takes a long time to be produced. Tumanski [15] reports that to scan an area of 10 by 10 cm with steps of 0.5 mm takes around one hour to complete. Building scanning devices is a complex task as it involves mechanical movements and synchronization with sensor reading. If not done properly vibration can affect the plot result.

Such systems can detect cracks on aluminum plates [14] and imperfections on steel plates [15] as in Fig. 6.

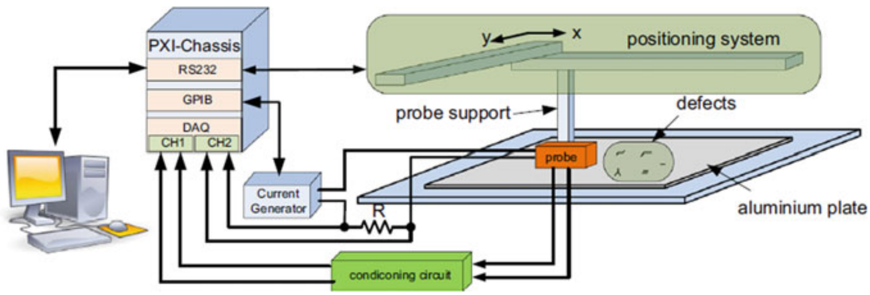


Fig. 5 NDT scanning setup by Pasadas et al. [14]

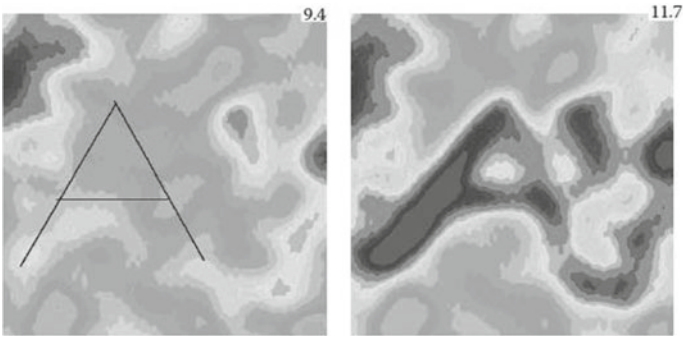


Fig. 6 Magnetic field height map of a steel sheet before and after writing the letter "A" on the surface [15]

3.1.1 Magnetic Cameras

The concept of a magnetic camera is the inverse of the scanning system. An array of sensors is held still, and the target magnetized object can freely move on the vicinity of the camera. The data from the sensors are serialized and sent into a computer that plots the field. The advantage of cameras is real-time imagery. However, the image resolution depends on the number of sensors, and due to physical space constraints, the system suffers from a lack of data. Nonlinear regression techniques such as the least square method or standard curve smoothing algorithms as B-Splines can be used to address it. The cost of the system also depends on the amount and quality of the sensors. Commercially available MniCube 3D from MagCam Fig. 7 has 16.384 on-chip microscopic Hall sensors with a spatial resolution of 0.1 mm $\times$ 0.1 mm costs around USD 58.00000 and it is described in Koen Vervaeke work [16].

Fig. 7 MiniCube 3D magnetic camera [16]

Fig. 8 Metrolab MFC3045 NMR magnetic camera [17]

Metrolab MFC3045 camera [17] has 32 NMR probes with a resolution of 0.01 ppm and costs around USD 31.50000 Fig. 8. Both the MiniCube and the MFC3045 capture all the 3 components of the field, *Bx*, *By*, and *Bz*.

Tuan et al. [18] created a system comprised of 25 analog A1302 Hall effect sensors arranged in a 5 × 5 grid with the Hall element aligned in parallel with the table, a camera Fig. 9, and a personal computer with custom software that draws the surface plot of the field by connecting discrete measurement data points from the sensors via triangular mesh, as shown in Fig. 10.

This system only captures the vertical component of the field and leaves the horizontal components unexplored.

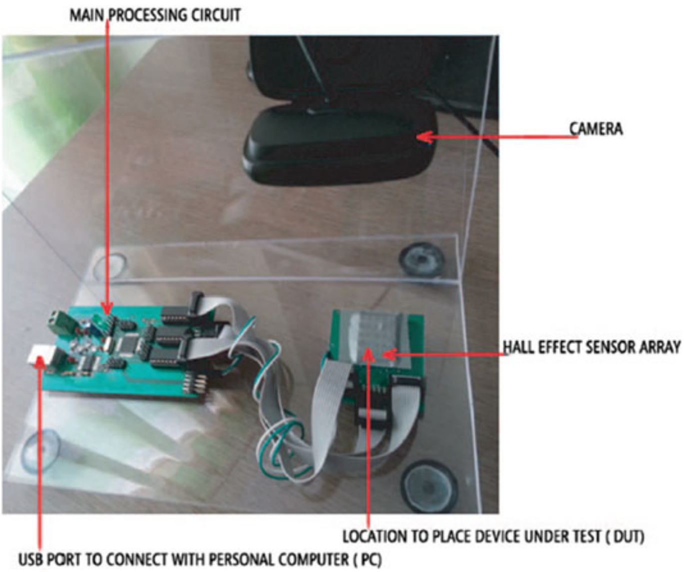


Fig. 9 Complete magnetic camera system [18]

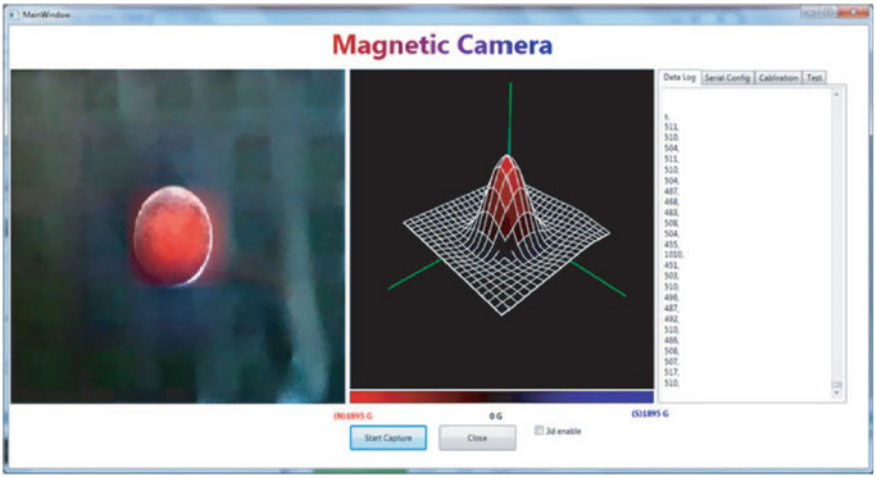


Fig. 10 Circular magnet north pole and its real-time height map [18]

We created a similar system without using a camera and displaying only the field intensity height map in real time. Two boards were devised, one with 16 A1302 surfaces mounted (SMT) sensors that capture the vertical component of the field, Fig. 11, and another with 25 trough hole (TH) mounted sensors which instruments the horizontal component Fig. 12. Sensors from both boards were arranged with a 1 cm distance from each other. Figures 13 and 14 illustrate the field distribution from the same south pole surface of a 1 cm cube magnet from vertical and horizontal perspectives.

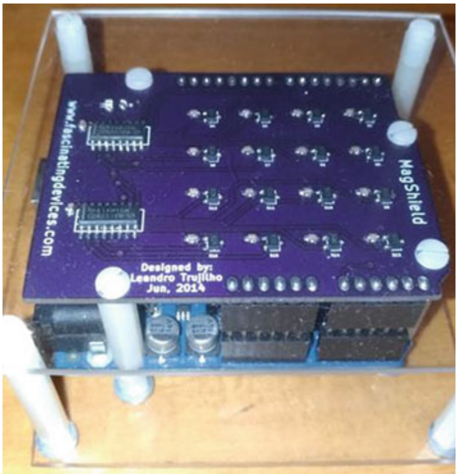


Fig. 11 SMT Board 1—Hall elements are aligned in parallel with the table. Captures the vertical component of the field

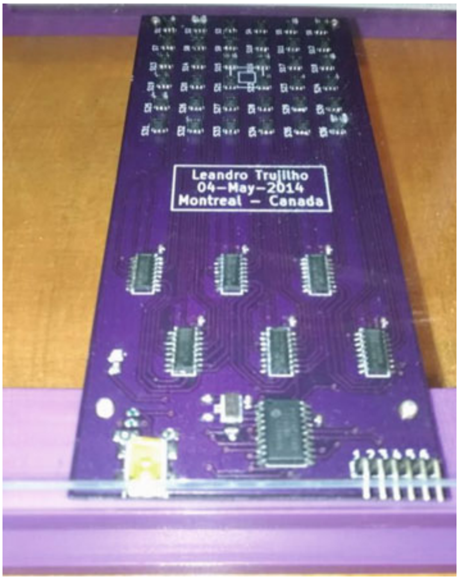


Fig. 12 TH Board 2—Hall elements are aligned orthogonally with the table. Captures the horizontal component of the field

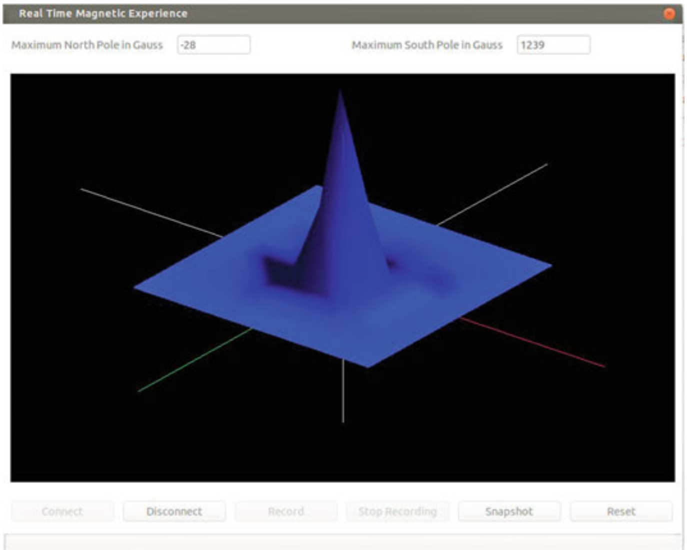


Fig. 13 South pole of a cube magnet on top of SMT board 1, showing the B field vertical component with a peak of 123.9 mT at approximate 4 mm height

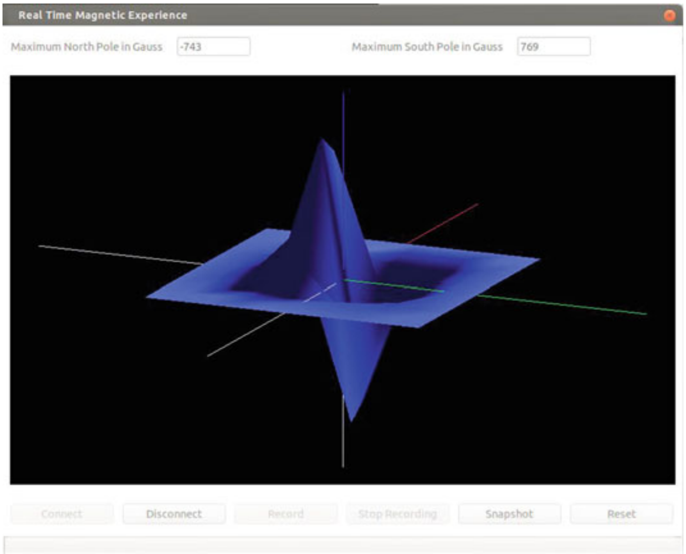


Fig. 14 South pole of a cube magnet on top a TH board 2, showing the B field horizontal component with peaks of −74.3 and 76.9 mT at approximate 4 mm height

Magnetic cameras are complex to build as it requires custom hardware and software design; however, the user experience simply consists of moving the magnet on top of the boards.

3.1.2 Augmented Reality

Augmented reality (AR) environment blends virtual objects with real-world objects. Buchau et al. [19] developed an AR system using 3D FEM to display magnetic field lines for educational purposes. As stated previously, FEM requires a mesh to compute field values and is computationally intensive. In 3D, the number of nodes grows fast. Figure 15 shows 3D the field lines of a bar magnet and Fig. 16 shows the field of a Helmholtz coil.

Matsutomo et al. presented in [20] an immersive real-time system capable of displaying the 3D field interference originated from multiple magnetic sources. They also described six steps for achieving it using a head-mounted display (HMD) and fiducial markers as in Fig. 17a, d. The first step was to register the markers into the visualization system, the second consisted of capturing the markers with a web camera mounted on the HMD, the third identified the markers via image processing techniques, the fourth computed the magnetic field distribution Fig. 17c, the fifth was the generation of the parallax images for the stereoscopic vision on the HMD, and finally the sixth combined the synthetic magnetic field distribution with the magnetized objects images for simultaneous visualization on the HMD as shown in Fig. 17b.

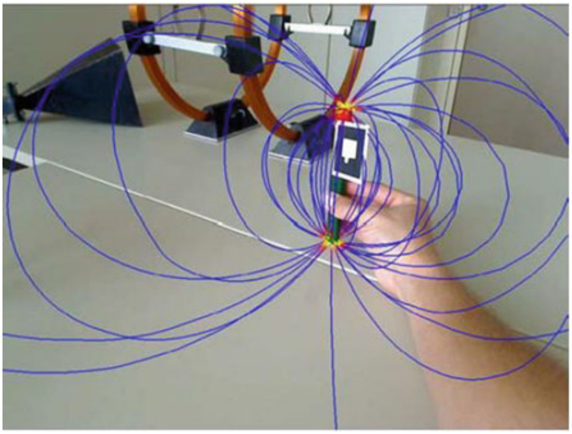


Fig. 15 Visualization of the magnetic flux density of a permanent magnet in an AR environment [19]

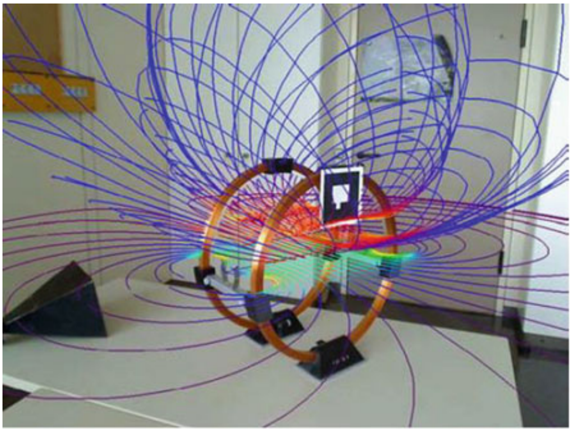


Fig. 16 Visualization of the magnetic field strength of a Helmholtz coil in an AR environment [19]

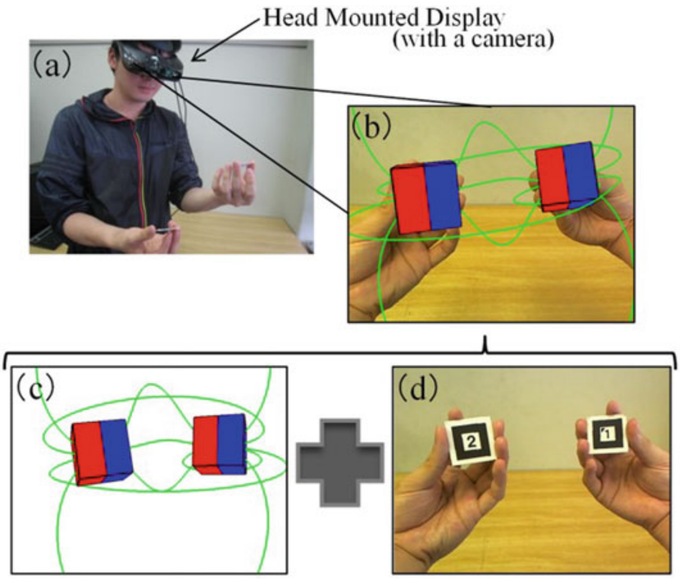


Fig. 17 An immersive real-time visualization system for 3D magnetic field distribution. **a** Overview of the system and the HMD with a camera. **b** The composite image displayed on the user's HMD. **c** The simulated magnetic flux lines in 3D space. **d** The markers which indicate the positions and the type of the objects (e.g., magnet, coil, iron) [20]

4 Summary and Conclusion

A set of simulation and experimental techniques to visualize the magnetic fields of macroscale objects were presented. In a simulation, finite element methods have the benefit of handling complex geometries, but it is computationally expensive because of the free space mesh; finite integral methods have better performance than FEM because the mesh area is restricted only to the active surface, but it is not easy to scale solution functions to higher orders. Algebraic meshfree methods are fast but require equation parameterization and learning to increase precision and accuracy. Nonlinear regression and statistical methods can estimate the data points in between the measured data from scanning systems smoothing the field distribution, but the can suffer from performance issues when large training sets are used. Magnetic cameras display the field in real time with the drawback of low resolution, requiring smoothing algorithms as well. Augmented reality systems provide immersion experience and allow interaction among multiple magnetic sources, but it is the method that has the largest computational cost of all the previous methods.

Acknowledgements Thanks to Ricardo Wandré Dias Pedro for reviewing and providing valuable inputs to this paper on short notice.

References

1. Hopster, H., Oepen H. P.: Magnetic Microscopy of Nanostructures. Springer, Berlin (2004)
2. Meeker, D., Bianchi, N., Gyselinck J., Sabariego R. V., Alberti, L., Pellegrino, G., Cupertino, F.: Electrical machine analysis using free software. IEEE Energy Conversion Congress and Exposition, pp. 1–289 (2017)
3. Le-Van, V., Meunier, G., Chadebec, O., Guichon, J.: A magnetic vector potential volume integral formulation for nonlinear magnetostatic problems. IEEE Trans. Magn. **52**(3), 1–4 (2016)
4. Lima, E.A., Weiss, B.P.: Obtaining vector magnetic field maps from single-component measurements of geological samples. J. Geophys. Res. **114**, B6 (2009)
5. Ren, Y., Hu, C., Xiang, S., Feng, Z.: Magnetic dipole model in the near-field. In: IEEE International Conference on Information and Automation, Lijiang, pp. 1085–1089 (2015)
6. Macedo, I., Castro R.: Learning div-free and curl-free vector fields by matrix-valued kernels. Instituto Nacional de Matemática Pura e Aplicada, Preprint serie A679 (2010)
7. Wahlström, N., Kok, M., Schön, T.B., Gustafsson, F.: Modeling magnetic fields using Gaussian processes. In: IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, pp. 3522–3526 (2013)
8. Jiun-Shyan, C., Michael, H., Sheng-Wei, C.: Meshfree methods: progress made after 20 years. J. Eng. Mech. **143**, 04017001 (2017). https://doi.org/10.1061/(ASCE)EM.1943-7889.0001176
9. Papell, S.S.: Low viscosity magnetic fluid obtained by the colloidal suspension of magnetic particles. US Patent US3215572A (1963)
10. Henrichsen, K.N.: Overview of magnet measurement methods. CERN, Geneva, Switzerland (1998). https://doi.org/10.5170/CERN-1998-005.127
11. Tumanski, S.: Modern magnetic field sensors—a review. Przeglad Elektrotechniczny **89**, 1–12 (2013)
12. Wahlström, N., Gustafsson, F.: Magnetometer modeling and validation for tracking metallic targets. IEEE Trans. Signal Process. **62**(3), 545–556 (2014)
13. Nowicki, M., Szewczyk, R.: Ferromagnetic objects magnetovision detection system. Materials (Basel) (2013)
14. Pasadas, D.J., Ribeiro, A.L., Rocha, T., Ramos, H.G.: 2D surface defect images applying Tikhonov regularized inversion and ECT. NDT & E Int. **80** (2016)
15. Tumanski, S.: Handbook of Magnetic Measurements (Series inSensors), 1st edn. CRC Press (2016)
16. Vervaeke, K.: 6D magnetic field distribution measurements of permanent magnets with magnetic field camera scanner. In: 5th International Electric Drives Production Conference (EDPC), Nuremberg, pp. 1–4 (2015)
17. Metrolab Home page: https://www.metrolab.com/wpcontent/uploads/2015/07/MFC3045_manual_v30_rev10.pdf
18. Tuan, B.A., de Souza-Daw, A., Phuong, H.M., Hoang, T.M., Dzung, N.T.: Magnetic camera for visualizing magnetic fields of home appliances. In: IEEE Fifth International Conference on Communications and Electronics (ICCE), Danang, pp. 370–375 (2014)
19. Buchau, A., Rucker, W., Wössner, U., Becker, M.: Augmented reality in teaching of electrodynamics. COMPEL-Int. J. Comput. Math. Electr. Electron. Eng. **28**(4), 948–963 (2009)
20. Matsutomo, S., Manabe, T., Cingoski, V., Noguchi, S.: A computer aided education system based on augmented reality by immersion to 3-D magnetic field. IEEE Trans. Magn. **53**(6), 1–4, Article No. 8102004 (2017)

Assessment of Engineering Techniques for Failures Simulation in Induction Motors Using Numerical Tool

Gustavo Caiza, Carlos A. Garcia, Jose E. Naranjo, and Marcelo V. Garcia

Abstract Most industrial processes require reliability, safe and continuous operation of electric motors. However, unexpected failures result in production losses, high emergency maintenance costs, damage to other related machines and prolonged shutdowns. Therefore, the development of failure analysis, detection and prevention systems is of high importance for the industrial sector. The following paper proposes the generation of a model of a squirrel cage induction motor in the software tool Simulink capable of simulating: (i) broken bars and broken rotor rings and (ii) failures in the bearings of the motor, based on the Motor Current Signature Analysis (MCSA) technique which is a non-invasive method, and the fast Fourier transform (FFT) algorithm to perform spectral analysis on the current to detect specific components that characterize faults in these conditions.

Keywords Induction motor · Squirrel cage · Damaged bearings · Simulation failures · Broken bars

1 Introduction

The induction motor is the most widely used electric motor in the industry. The squirrel cage variant has the great advantage of being a simple, robust, low-cost

G. Caiza
Universidad Politecnica Salesiana, UPS, 170146 Quito, Ecuador
e-mail: gcaiza@ups.edu.ec

C. A. Garcia · J. E. Naranjo · M. V. Garcia (✉)
Universidad Tecnica de Ambato, UTA, 180103 Ambato, Ecuador
e-mail: mgarcia294@ehu.eus; mv.garcia@uta.edu.ec

C. A. Garcia
e-mail: ca.garcia@uta.edu.ec

J. E. Naranjo
e-mail: jnaranjo0463@uta.edu.ec

M. V. Garcia
University of Basque Country, UPV/EHU, 48013 Bilbao, Spain

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_28

construction engine that requires much less maintenance than any other rotary machine with no brushes, ring collector or delta collector. Due to the induction method of energy production, they are easy to operate. As a result, these engines are quite suitable for large applications as in the oil extraction industry [1–16]. Although these motors are robust and have good reliability, they are not exempt from suffering some faults during their use and because of the application of strong forces in the magnetic cores, bearings, shaft, stator and cage windings, or used in excessive times.

Component failure is usually defined as a condition of capacity reduction related to minimum specification requirements. It is the result of normal wear, poor design or poor specification, incorrect assembly, misuse or a combination of the aforementioned. If a failure is not detected in time or if it is allowed to develop further, it can lead to the collapse of the machine [6]. Nowadays, it is important to consider the implementation of online fault diagnosis strategies to increase the useful life of machine components, thus increasing their availability and the productivity of the plant.

The detection and diagnosis of electrical or mechanical faults in an induction motor involve, in many cases, the interpretation of the frequency spectrum of the motor current, the power or the Park vector, among others [3]. This requires an expert to perform the task, based on the information obtained from the processed signals. In the current days, the techniques of machine learning-based artificial intelligences have taken great importance. Some of the reasons are that they require a minimal interpretation of the studied system and greatly simplifying the diagnostic task [7, 14]. Machine learning-based fault diagnosis techniques have been widely studied and have succeeded in many applications of electrical machines and drives. However, these types of solutions require a great amount of data to generate models capable of identifying specific features.

The research area related to the following work aims to generate an online data analysis system based on machine learning for failure detection in the inductive motors of an oil extraction station. One of the main limitations while developing this artificial intelligence is the amount of data available to train the model. However, with the help of the software tool Simulink, a reconfigurable model of an inductive motor can be generated. With the right understanding of the effects that some of the most common failures have in the operation of this type of motors, the classic model of the induction motor can be modified to also simulate: broken bars, broken rotor rings, short circuits between copper winds and bearing failures. The following work describes some of the advances done in the generation of the proposed simulation model.

This paper is structured as follows: Sect. 2 shows some related works that have been used as the starting point for this research. Section 3 describes the modeling of a squirrel cage induction motor using the finite element method (FEM). Section 4 summarizes some of the most common failures in induction motors and their effects. Section 5 shows the analysis of results, and finally, in Sect. 6, some conclusions and ongoing work are given.

2 Literature Review

The aim of this section is to show how other researches deal with the predict electric motor damage using fault system detection based on the simulation or implementation of Motor Current Signature Analysis (MCSA) technique. In this regard, some related works are shown, paradigms and development technologies, that present the use of the MCSA technique at the electrical industrial level.

The Motor Current Signature Analysis (MCSA) is a classical technique for the valuation of the rotor healthy condition in electric induction motors [1, 2, 5, 10] that depends on acquiring the waveform of the current consumed by the motor during steady-state operation and analyzing this waveform using the fast Fourier transform (FFT). Calculating the amplitude of components and harmonics is amplified by the rotor fault. This novel technique has given promises results for various kinds of induction motors used in a diverse range of industrial applications, where the operation regime is rather stable.

The use of math techniques like Motor Current Signature Analysis (MCSA) and Motor Current Square Signature Analysis (MSCSA) to diagnose and detect faults patterns, described by the stator current spectrum, in induction motors, is presented in [8]. The authors of this research compared different faults in real motors the waveform of the current, the types of signals compared are eccentricity in the air gap and inter-turn short circuit in the stator winding. The results presented in this article demonstrated that the MCSA and MSCSA techniques were suitable for the detect electric motor faults, whose comparison between the techniques showed the suitability of each one.

The main goal of the paper [11] is to present an evaluation of the MCSA method to be applied to detect inductor faults in the motor-driven load. This evaluation is based on transfer functions which are derived from a small-signal model of the induction motor calculating the space phasor per unit representation when determine the shaft speed to load torque transfer function. This technique needs to connect more sensors for voltage and current. This is cumbersome and impractical for many industrial applications.

A research work presented in [4] develop a model capable of providing supplementary information into the transient operation of damaged rotor cages. This is made by computing the rotor's thermal, electrical and mechanical magnitudes under which that damage occurs. This model is upgraded and re-adjusted based on the measurements performed on the stator during the fatigue test. In line with the complexity of the task, some techniques are used according to the degree of detail needed. This paper shows preliminary results that provide a qualitative knowledge of the progress of a bar breakage during a fatigue test.

In this paper, the integration of simulating techniques of the electric motor model and different failures are addressed. Furthermore, this paper develops a new software tool based on Simulink, enabling an easy and fast building of this kind of electric systems.

3 The finite element method (FEM) for Modeling Electric Machines Faults

The finite element method (FEM) is a numerical method to deal with issues of engineering. Once the FEM is used, partial differential equations that shape physical systems with an infinite number of degrees of freedom are simplified to systems of algebraic equations. The FEM approaches the unidentified function over the domain. The basic idea of this method is divided the motor structure into several parts called finite elements (FE), in these small parts can be applied calculation theories corresponding to the scheme adopted. The character of the generality of the method gives the advantage to adapt, with simple modifications, the most complex and varied problems.

Dividing the structure into smaller parts called meshing operation will result in obtaining simple forms for finite elements composing the electric motor. The computational model used in the finite element analysis is approximately got by finite element assembly components, considering the electric faults of the motor. Connecting FE is achieved only at specific points called nodal points or nodes. The simulation made in a computer system can resolve this large set of simultaneous equations. The computer gets from the solution, the behavior of the individual FE.

The FEM gives specific information on the machine's nonlinear effects. This modeling approach is capable of obtaining an accurate and complete description of an electrical machine [9]. The magnetic circuit is defined by a mesh of small elements. The field values are then supposed to be a simple function of position within these finite elements, enabling interpolation of results.

A compromise must be reached between using finer meshes to achieve higher accuracy and the processing resources needed to achieve reasonable simulation times. The FEM is very flexible, especially for new designs incorporating new shapes. However, long-time simulation requirements reduce its attractiveness for a case when a control algorithm needs to be incorporated.

3.1 Simulation of the Induction Motor Using FEM Approach

The first problem found while trying to simulate the induction machine is that the induction motor is defined by nonlinear differential equations. Therefore, it is necessary to perform a transformation to a system of linear differential equations. These transformations are performed to facilitate the analysis of the engine.

The first transformation that is performed during this analysis is the change from a three-phase voltage reference (a, b, c) into a bi-phase voltage reference (α, β). Being the three-phase system represented by the vector in Eq. 1.

$$\overrightarrow{V_S} = \frac{2}{3}\left(V_A + a * V_B + a^2 * V_C\right) \tag{1}$$

it can be represented in the $\alpha\beta$ reference with Eqs. 2 and 3.

$$V_{S\alpha} = \frac{2}{3}\left(V_A - \frac{1}{2}V_B - \frac{1}{2}V_C\right) \tag{2}$$

$$V_{S\beta} = \frac{2}{3}\left(\frac{\sqrt{3}}{2}V_B - \frac{\sqrt{3}}{2}V_C\right) \tag{3}$$

Normally, the motor parameters are measured from the stator winding, as shown in Eqs. 4 and 5.

$$V_{S\alpha} = \frac{\mathrm{d}(\psi_{S\alpha})}{\mathrm{d}t} + \psi_{S\alpha} * w + R_S * i_{S\alpha} \tag{4}$$

$$V_{S\beta} = \frac{\mathrm{d}(\psi_{S\beta})}{\mathrm{d}t} + \psi_{S\beta} * w + R_S * i_{S\beta} \tag{5}$$

Therefore, it is convenient to refer all rotor parameters to the stator winding. As shown in Eqs. 6 and 7.

$$V_{R\alpha} = \frac{\mathrm{d}(\psi_{R\alpha})}{\mathrm{d}t} + \psi_{R\alpha} * (w - w_R) + R_R * i_{R\alpha} \tag{6}$$

$$V_{R\beta} = \frac{\mathrm{d}(\psi_{R\beta})}{\mathrm{d}t} + \psi_{R\beta} * (w - w_R) + R_R * i_{R\beta} \tag{7}$$

where w is the angular velocity of the random reference system and w_R is the angular velocity of the rotor.

From Eqs. 4–7, it can determine the equations of the stator and rotor fluxes with the reference system fixed to the stator ($w = 0$). It also has to be taken into account that the motor to be simulated in the present work is a squirrel cage model; therefore, the rotor voltages $V_{R\alpha}$ and $V_{R\beta}$ are zero. After introducing these variants and solving the equations system, the final fluxes equations are presented in Eqs. 8–11.

$$\psi_{S\alpha} = \int V_{S\alpha} - R_S * i_{S\alpha} \tag{8}$$

$$\psi_{S\beta} = \int V_{S\beta} - R_S * i_{S\beta} \tag{9}$$

$$\psi_{R\alpha} = \int -\psi_{R\beta} * w_R - R_R * i_{R\alpha} \tag{10}$$

$$\psi_{R\beta} = \int \psi_{R\alpha} * w_R - R_R * i_{R\beta} \tag{11}$$

To obtain the equations from which calculate $i_{S\alpha}, i_{S\beta}, i_{R\alpha}$ and $i_{R\beta}$, the analysis process must begin with the Eqs. 12 and 13.

$$\psi_R = L'_R i_R + L_m i_S \tag{12}$$

$$\psi_S = L'_S i_S + L_m i_R \tag{13}$$

where $\left\{L'\right\}\_R = L\_R + L\_m\$$ and $\$\left\{L'\right\}\_S = L\_S + L\_m.$

After decomposing Eqs. 12 and 13 into their $\alpha\beta$ components, solving the new equations for the values of $i_{S\alpha}, i_{S\beta}, i_{R\alpha}$ and $i_{R\beta}$, and solving the new system, it is obtained that (see Eqs. 14–17).

$$i_{S\alpha} = \frac{L_m \psi_{R\alpha}}{L_m^2 - L'_R L'_S} - \frac{L_R \psi_{S\alpha}}{L_m^2 - L'_R L'_S} \tag{14}$$

$$i_{S\beta} = \frac{L_m \psi_{R\beta}}{L_m^2 - L'_R L'_S} - \frac{L_R \psi_{S\beta}}{L_m^2 - L'_R L'_S} \tag{15}$$

$$i_{R\alpha} = \frac{L_m \psi_{S\alpha}}{L_m^2 - L'_R L'_S} - \frac{L_S \psi_{R\alpha}}{L_m^2 - L'_R L'_S} \tag{16}$$

$$i_{R\beta} = \frac{L_m \psi_{S\beta}}{L_m^2 - L'_R L'_S} - \frac{L_S \psi_{R\beta}}{L_m^2 - L'_R L'_S} \tag{17}$$

The generated torque of the motor can be obtained from Eq. 18.

$$M_{mi} = \frac{3p}{2} * L_m \left(i_{S\alpha} i_{R\beta} - i_{S\beta} i_{R\alpha}\right) \tag{18}$$

And the speed can be obtained from Eq. 19.

$$w_R = \int \frac{p}{J}\left(M_{mi} - M_C - \frac{f}{p} w_R\right) \tag{19}$$

Finally, the three-phase currents must be recovered from the bi-phase reference with the Eqs. 20–22.

$$i_a = i_a \tag{20}$$

$$i_b = \frac{1}{2}\left(\sqrt{3} i_\beta - i_\alpha\right) \tag{21}$$

$$i_c = -\frac{1}{2}\left(\sqrt{3} i_\beta - i_\alpha\right) \tag{22}$$

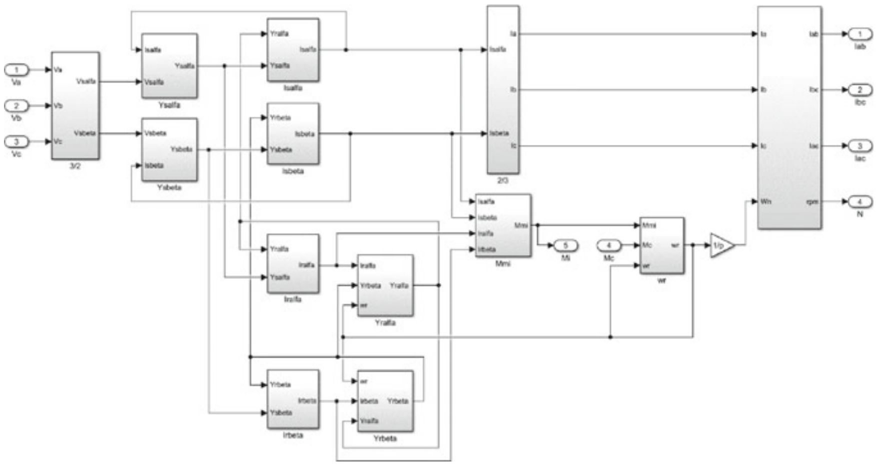


Fig. 1 Simulation of an induction squirrel cage motor in Simulink

Using the expression depicted in Eqs. 2, 3, 8–11, 14–22, a simulation model can be implemented in Simulink as shown in Fig. 1.

4 Simulation Methodology of Failures in Induction Motor

Since induction motors are rotatory machines with simple construction, their most common causes of motor failure are bearing failures, insulation failures and rotor failures. However, as depicted in [12], almost 40–50% of all motor failures are related to their bearings. Table 1 describes the most common failures in inductive motors and their respective effects on the normal operation of the motor.

Table 1 Most common motor failures and their respective effects

Failure	Effects
Breakage of bars and rotor rings	The magnetomotive force of the rotor rotating in the opposite direction due to asymmetric currents
Static eccentricity	Magnetic pulse unbalanced due to the minimum space in the air gap
Dynamic eccentricity	Magnetic pulse unbalanced and periodic due to the minimum variable distance of the air gap
Mixed eccentricity	An unbalanced magnetic pulse influenced by static and dynamic eccentricity
Short circuit between copper turns	Asymmetric magnetomotive force in the stator windings
Damaged bearings	Increase in vibrations

4.1 Broken Bars

A problem that occurs in induction motors is that one or more rotor bars fracture. Despite this damage, the motor remains capable of operating in an apparently normal way. However, one of these little problems can end up with serious consequences since a fractured bar will progressively cause the fracture of more bars. Consecutively, so many bars can break up causing that the motor stops operating normally in a sudden way [15]. The physical effect that one or more broken bars will have in the induction motor is the variation of the rotor resistance. Therefore, this failure can be replicated in the generated model by partially or totally decreasing the nominal value of the rotor resistance while running the simulation.

4.2 Short Circuit Between Copper Turns

A short circuit between turns reduces the ability to produce a balanced magnetic field, which has consequences such as an increase in the vibration of the machine. This effect accelerates the degradation of the insulation and damage to the motor bearings. In most cases, the short circuit between turns includes the following possibilities: “turn-by-turn,” “phase-to-phase” or “phase-to-ground,” causing the motor to collapse.

However, there are cases in which the collapse of the motor occurs immediately after the fault is detected [13]. Like the previous case, this failure can be replicated in the induction motor model by partially or totally decreasing the nominal value of the stator inductance while running the simulation. However, some limitations and execution measurements must be taken into account, due to small values of stator inductance can end up in mathematical indetermination while calculating the simulation.

4.3 Damaged Bearings

Early detection of bearing failures allows the replacement of the bearings, instead of the replacement of the motor. Although the replacement of defective bearings is the cheapest solution among the three causes of failure, it is the most difficult to detect. Bearings are formed by: interior track, balls or rollers, cage and an outer track. The deterioration of each of these elements will generate one or several characteristic frequencies in the frequency spectra that will allow a quick and easy identification. The four possible frequencies of deterioration of a bearing are:

- Ball pass frequency of the outer race (BPFO) is the frequency of step of rolling elements due to a defect in the external track. Physically, it is the number of balls or rollers that pass through a point of the outer track each time the axis makes a complete turn.

- Ball pass frequency of the inner race (BPFI) is the frequency of step of the rolling elements due to a defect in the internal track. Physically, it is the number of balls or rollers that pass through a point of the inner race each time the axle makes a complete turn.
- Ball spin frequency (BSF) is the frequency of deterioration of the rolling elements. Physically, it is the number of turns made by a bearing ball every time the shaft makes a complete turn.
- Fundamental train frequency (FTF) is the rotation frequency of the cage that contains the rolling elements. Physically, it is the number of turns made by the bearing cage each time the shaft makes a complete turn.

These different bearing failures can be simulated by adding some specific torques to the final rotor speed equation.

5 Analysis of Results

This section presents signals of the current simulated in each fault to later be analyzed and interpret their spectral characteristics of incipient faults as in the case of bearing wear, broken bars and short circuits between turns of the same phase.

5.1 Broken Bar Fault Simulated Results

The methodology to simulate a broken bar fault is to produce three-phase current with distortion in a nominal load operation. As can see in Fig. 3, the amplitude in the frequency range from 75 to 110 (Hz) decreases below −40 dBv if this signal is compared with the simulated signal without fault in normal operation. An increase of the spectral amplitude in the fault simulated signal is appreciated at frequencies from 100 (Hz) to 110 (Hz). This characteristic is recursive in the three phases of the electric current simulated. In this kind of electric motor fault, its speed increase by about 5% regarding the nominal speed and the other effect is the decrease of the slippage. Figure 2 shown is cut at 100 (Hz) to get a better analysis of broken bar faults. The broken fault of electric motors produces side peaks around 70 (Hz), 40 (Hz) and 50 (Hz) with an amplitude around −10 dBv, and these peaks appear in the three phases simulated. This is another effect of a broken bar in the squirrel cage motors.

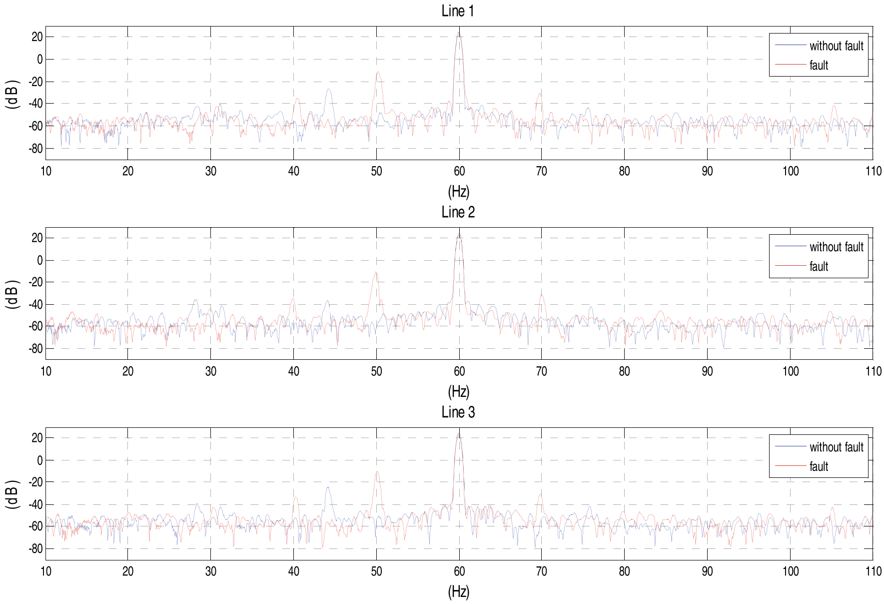


Fig. 2 Presence of broken bars, diagnosis by upper and lower lateral peaks

5.2 *Short-Circuit Fault Simulated Results*

The model developed in MATLAB software is possible to simulate the effects of a short circuit in different turns of the stator in an electric motor. In this case, a short circuit in 28 turns of phase three is simulated. Figure 3 shows that an increase of spectral amplitude over the noise level around the frequency of 180 (Hz) occurs in phase three. A small increase of about 0.53 dBv in phase three of the current simulated over the noise level in the frequency of 300 (Hz) shows that there is a short circuit in this phase. The next step of this simulation methodology is to add a short circuit in 28 turns on phase one, the spectral signal moves about 11 dBv over the fundamental signal without fault, showing an increase of distortion of the three-phase currents. Finally, the spectrum of the fault signal around 180 (Hz) increases around 23.44 dBv over the noise level in all the phases, with which the simulated signals show the existence of a stator fault.

5.3 *Bearing Fault Simulated Results*

Figure 4 shows the spectral amplitude change in the three phases of the system around a range of frequencies from 68 to 110 (Hz); in these frequencies, a decrease in the spectral amplitude is less than −40 dBv. Besides, there is a slight increase in amplitude in the frequency of 300 Hz in regard to the signals of a healthy motor.

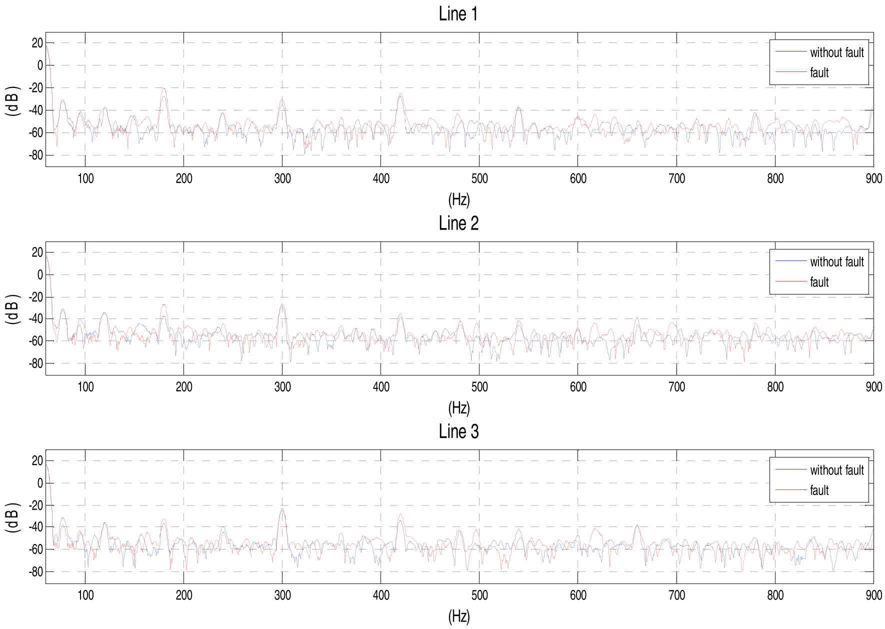


Fig. 3 Distortion of stator currents due to a short circuit in phase 3

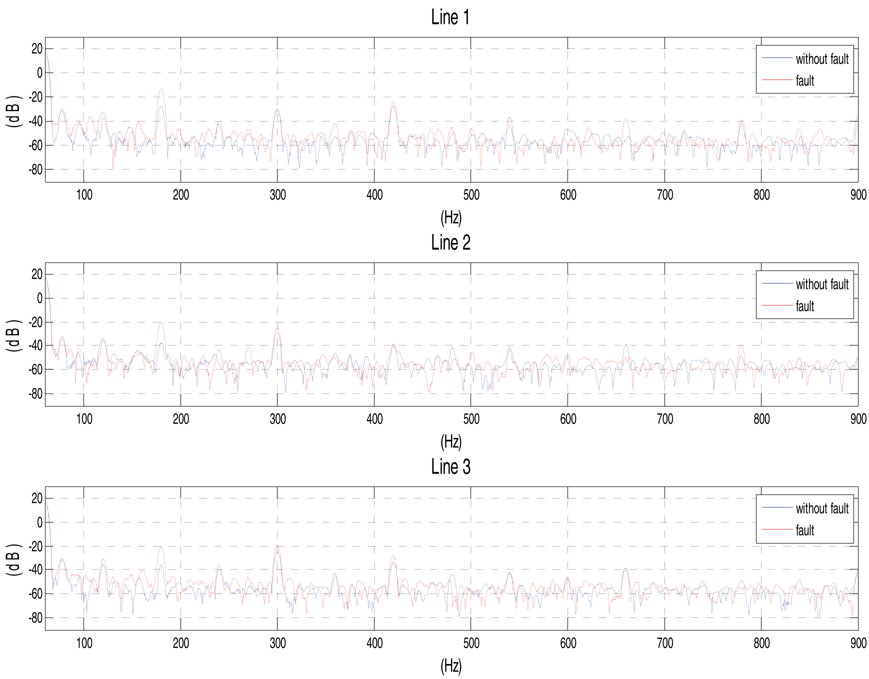


Fig. 4 Presence of bearing fault in all three phases

6 Conclusions and Future Work

With the use of the software tool Simulink, a generic model of an induction motor is generated. This model is designed, so it can also be capable of simulating failures in the motor such as broke bars, short circuit between copper wind and bearing failures. These failures can be replicated in the model by adding portions of code that modify some specific parameters such as rotor resistance, stator inductance and rotor speed.

This model is developed with the aim of that in future works, we can generate the data needed to create a machine learning-based system capable of detecting failures in induction motor, while the whole industrial system is running.

Acknowledgements This work was financed by Universidad Tecnica de Ambato (UTA) and their Research and Development Department (DIDE) under project CONIN-P-256-2019.

References

1. Abhinandan, A.C., Sidram, M.H.: Fault diagnosis of an induction motor through motor current signature analysis, FFT & DWT analysis. In: 2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS) [Internet]. Salmabad: IEEE; 2017 [cited 2019 July 30], pp. 1–7. Available from: https://ieeexplore.ieee.org/document/8277869/
2. Antonino-Daviu, J., Lopez, A.Q., Rubbiolo, M., Climente-Alarcon, V.: Diagnosis of the rotor condition in electric motors operating in mining facilities through the analysis of motor currents. In: 2017 IEEE Industry Applications Society Annual Meeting [Internet]. Cincinnati, OH: IEEE; 2017 [cited 2019 July 30], pp. 1–8. Available from: https://ieeexplore.ieee.org/document/8101834/
3. Bellini, A., Filippetti, F., Tassoni, C., Capolino, G.A.: Advances in diagnostic techniques for induction machines. IEEE Trans. Ind. Electron. **55**, 4109–4126 (2008)
4. Climente-Alarcon, V., Nair, D., Sundaria, R., Antonino-Daviu, J.A., Arkkio, A.: Combined model for simulating the effect of transients on a damaged rotor cage. IEEE Trans. Ind. Appl. **53**, 3528–3537 (2017)
5. Cox, R.W., Parikh, P.K., Cox, R., Parikh, P.: Algorithm for tracking the health of multiple induction motors using bus-level current. In: IECON 2018—44th Annual Conference of the IEEE Industrial Electronics Society [Internet]. Washington, DC: IEEE; 2018 [cited 2019 July 30], pp. 373–378. Available from: https://ieeexplore.ieee.org/document/8591655/
6. Díaz, D., Díaz, R., Amaya, M.C., Palacios, J.: Análisis del cortocircuito entre espiras de un motor de inducción tipo jaula de ardilla mediante la aplicación del método de elementos finitos (mef). Revista Facultad de Ingeniería Universidad de Antioquia, 2010, pp. 16–23
7. Filippetti, F., Bellini, A., Capolino, G.: Condition monitoring and diagnosis of rotor faults in induction machines: state of art and future perspectives. In: 2013 IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD) [Internet]. Paris: IEEE; 2013 [cited 2019 July 11], pp. 196–209. Available from: https://ieeexplore.ieee.org/document/6525180/
8. Fontes, A.S., Cardoso, C.A.V., Oliveira, L.P.B.: Comparison of techniques based on current signature analysis to fault detection and diagnosis in induction electrical motors. In: 2016 Electrical Engineering Conference (EECon) [Internet]. Colombo, Sri Lanka: IEEE; 2016 [cited 2019 July 30], pp. 74–79. Available from: https://ieeexplore.ieee.org/document/7830938/
9. Mohammed, O.A., Abed, N.Y., Ganu, S.: Modeling and characterization of induction motor internal faults using finite element and discrete wavelet transforms. In: 2007 IEEE Electric

Ship Technologies Symposium [Internet]. Arlington, VA, USA: IEEE; 2007 [cited 2019 Oct 1], pp. 250–253. Available from: https://ieeexplore.ieee.org/document/4233830/
10. Morales-Perez, C., Grande-Barreto, J., Rangel-Magdaleno, J., Peregrina-Barreto, H.: Bearing fault detection in induction motors using MCSA and statistical analysis. In: 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) [Internet]. Houston, TX, USA: IEEE; 2018 [cited 2019 July 30], pp. 1–5. Available from: https://ieeexplore.ieee.org/document/8409780/
11. Naha, A., Samanta, A.K., Routray, A., Deb, A.K.: A method for detecting half-broken rotor bar in lightly loaded induction motors using current. IEEE Trans. Instrum. Meas. **65**, 1614–1625 (2016)
12. Nandi, S., Toliyat, H.A., Li, X.: Condition monitoring and fault diagnosis of electrical motors—a review. IEEE Trans. Energy Convers. **20**, 719–729 (2005)
13. Rojas, C., Melero, M.G., Cabanas, M.F., Cano, J.M., Orcajo, G.A., Pedrayes, F.: Finite element model for the study of inter-turn short circuits in induction motors. In: 2007 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives [Internet]. Cracow, Poland: IEEE; 2007 [cited 2019 July 30], pp. 415–419. Available from: https://ieeexplore.ieee.org/document/4393130/
14. Toliyat, H.A. (ed.): Electric machines: modeling, condition monitoring, and fault diagnosis. Taylor & Francis, Boca Raton (2013)
15. Villada, F., Cadavid, D.R.: Diagnostico de Fallas en Motores de Inducción Mediante la Aplicación de Redes Neuronales Artificiales. Información tecnológica [Internet]. 2007 [cited 2019 July 11], p. 18. Available from: https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718–0764200700020 0016&lng=en&nrm=iso&tlng=en
16. Wang, X., Fang, F.: Bearing failure diagnosis in three-phase induction motor by chirp-Z transform and zoom-MUSIC. In: 2011 International Conference on Electrical and Control Engineering [Internet]. Yichang, China: IEEE; 2011 [cited 2019 July 11], pp. 1016–1019. Available from: https://www.ieeexplore.ieee.org/document/6057197/

LQR Predictive Control of SISO Time-Delay Processes Having Output Disturbances

Arturo Rojas-Moreno

Abstract This paper proposes an approach to control continuous-time single-input single-output (SISO) time-delay processes having output disturbances employing a linear quadratic regulator (LQR) predictive controller in the discrete-time domain. To demonstrate the validity of such an approach, this work performs the control of SISO stable with large time delay, unstable, non-minimum phase, oscillatory, and integrating benchmark time–delay processes, in other words, processes that have been used by the researches to demonstrate various published control approaches. In all cases, a pulse train disturbance signal on the output is rejected, while the controlled outputs follow pulse-shape reference signals.

Keywords Linear quadratic regulator (LQR) control · Predictive control · Time-delay process · Integrating process · Non-minimum phase process · Unstable process

1 Introduction

The control of the output of a time-delay process is more difficult, precisely, due to the presence of the time delay. Any loop with a time delay can generate oscillations and become unstable because of the unlimited growth of the phase lag. The control difficulty increases with the presence of a disturbance signal acting on the process output. This paper proposes an approach to control continuous-time SISO time-delay processes of the form:

$$y(s) = G(s)e^{-\tau s}u(s) + d(s) \tag{1}$$

In (1), $\tau > 0$ is a time delay, also known as a transport delay or dead time, s is the Laplace operator, $G(s)$ is either a stable, unstable, integrating, oscillating, or non-minimum phase transfer function, and $d(s)$ is the disturbance signal. The output $y(s)$

A. Rojas-Moreno (✉)
Universidad de Ingeniería y Tecnología—UTEC, Lima 15603, Peru
e-mail: arojas@utec.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_29

in (1) is controlled if such output follows a reference signal or setpoint $r(s)$, thanks to the action of the designed control signal $u(s)$, meeting the design specifications previously established, despite the action of the disturbance signal $d(s)$.

This work develops a controller design procedure to control processes of the form given by (1), employing a discrete-time LQR predictive controller. To validate the proposed design procedure, the designed controllers are applied to benchmark time-delay processes, that is, processes used by researches to demonstrate their control approaches. For the sake of information, a continuous-time LQR controller was employed in [1] to control MIMO stable time-delay processes, while discrete-time PI LQR controllers were used in [2] to control SISO disturbed time-delay processes.

This paper is organized as follows. Section 2 deals with the design of the discrete-time SISO LQR predictive controller. Various time-delay processes having a disturbance signal on the output are controlled in Sect. 3, while Sect. 4 presents some concluding remarks.

2 SISO LQR Predictive Controller Design

This section employs the LQR predictive control developed by Ikonen in [3]. Let us consider an arbitrary SISO continuous-time time-delay process represented by (1). For a sampling time T_s, it is possible to find the corresponding discrete-time process $G(z)$ (z is the shift operator) employing the z-transform operation. We have learned that the order of $G(z)$ grows as the sampling time T_s decreases. The discrete-time state-space representation of $G(z)$ can be obtained by using any available method, resulting

$$\mathbf{x}_d(k+1) = \mathbf{A}_d\mathbf{x}_d(k) + \mathbf{B}_d u(k); \quad y(k) = \mathbf{C}_d\mathbf{x}_d \tag{2}$$

In (2), $\mathbf{A}_d$ of order $n \times n$, $\mathbf{B}_d$ of order $n \times 1$, and $\mathbf{C}_d$ of order $1 \times n$ are the state, control, and output matrices, respectively. Also, $\boldsymbol{x}_d$ of order n, $y(k)$ of order 1, $u(k)$ of order 1, and the integer $k \geq 0$ are the state vector, the output signal, the control force, and the discrete-time index, respectively.

For predictive control, we compute the incremental control $\Delta u(k)$, but the actual control $u(k)$ is applied to the process

$$u(k) = u(k-1) + \Delta u(k) \tag{3}$$

Replacing (3) in (2) produces the following augmented predictive state-space representation

$$\mathbf{x}(k+1) = \mathbf{A}x(k) + \mathbf{B}\Delta u(k); \quad y(k) = \mathbf{C}x(k) \tag{4}$$

$$\mathbf{x}(k) = \begin{bmatrix} \mathbf{x}_d(k) \\ u(k-1) \end{bmatrix}; \mathbf{A} = \begin{bmatrix} \mathbf{A}_d & \mathbf{B}_d \\ 0 & 1 \end{bmatrix}; \mathbf{B} = \begin{bmatrix} \mathbf{B}_d \\ 1 \end{bmatrix}; \mathbf{C} = \begin{bmatrix} \mathbf{C}_d & 0 \end{bmatrix}$$

Consider the following finite-time cost function:

$$J = \sum_{i=1}^{p} \left[r(k+i) - \hat{y}(k+i) \right]^2 + \lambda [\Delta u(k+i-1)]^2 \tag{5}$$

In (5), $r(k+i)$ is the reference signal, also known as a set point or desired output signal, $\hat{y}(k+i)$ is the predicted process output, and λ is a balancing factor, actually a tuning parameter. The minimization of (5) with respect to the sequence $\Delta u(k)$, …, $\Delta u(k+p-1)$), where p is the order of the sequence, results in a sequence of future control signals. The first value of such a sequence is employed for the control law. The optimization is repeated to generate a new sequence of future control signals. Again, the first value of the new sequence $\Delta u(k)$ is used for control, and so on. This form of control is known as the receding horizon control. The predictive outputs $\hat{y}(k+1)$ and $\hat{y}(k+i)$ are formulated as:

$$\hat{y}(k+1) = \mathbf{C}\mathbf{x}(k+1) = \mathbf{C}[\mathbf{A}\mathbf{x}(k) + \mathbf{B}\Delta u(k)] \tag{6}$$

$$\hat{y}(k+1) = \mathbf{C}\mathbf{A}^i\mathbf{x}(k) + \sum_{j=1}^{i} \mathbf{C}\mathbf{A}^{i-j}\mathbf{B}\Delta u(k) \tag{7}$$

Let us define:

$$\hat{\mathbf{y}}(k+1) = \left[\hat{y}(k+1), \ldots, \hat{y}(k+p) \right]^T \tag{8}$$

$$\Delta\mathbf{u}(k) = [\Delta u(k), \ldots, \Delta u(k+p-1)]^T \tag{9}$$

Equation (7) can take the following form

$$\hat{\mathbf{y}}(k+1) = \mathbf{K}_{CA}\mathbf{x}(k) + \mathbf{K}_{CAB}\Delta u(k) \tag{10}$$

$$= \boldsymbol{y}^p + \mathbf{G}\Delta u(k) \tag{11}$$

$$\mathbf{K}_{CA} = \begin{bmatrix} \mathbf{CA} \\ \mathbf{CA}^2 \\ \vdots \\ \mathbf{CA}^p \end{bmatrix}; \mathbf{K}_{CAB} = \begin{bmatrix} \mathbf{CB} & 0 & \cdots & 0 \\ \mathbf{CAB} & \mathbf{CB} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \mathbf{CA}^{p-1}\mathbf{B} & \cdots & \mathbf{CAB} & \mathbf{C}B \end{bmatrix}$$

To make simpler the notation, we assume that $y(k) = \hat{y}(k)$, and $\mathbf{y}(k) = \hat{\mathbf{y}}(k)$. The following procedure is to design a LQR predictive controller:

1. Formulate the time-delay transfer function to be controlled: $G(s)e^{-\tau s}$
2. Determine $G(z)$ from $G(s)e^{-\tau s}$.
3. Compute the state-space model (2) of $G(z)$.
4. Obtain the predictive discrete-time state-space representation given by (4).
5. Set the integer p to calculate $\mathbf{K}_{CA}$ and $\mathbf{K}_{CAB}$ (Eqs. (10) and (11)).
6. Set the integer c to compute $\mathbf{G}$ equal to the c first columns of $\mathbf{K}_{CAB}$.
7. Compute

$$\mathbf{H} = \left[\mathbf{G}^T\mathbf{G} + \lambda\mathbf{I}\right]^{-1}\mathbf{G}^T \tag{12}$$

   In (12), $\mathbf{I}$ is the identity matrix and λ is a tuning parameter.
8. Knowing that $d(k)$ is the disturbance signal, compute

$$y^p(k+1) = \mathbf{K}_{CA}\mathbf{x}(k) + d(k) \tag{13}$$

$$e(k+1) = r(k+1) - y^p(k+1) \tag{14}$$

9. Select the first value of the sequence

$$\Delta\mathbf{u}(k) = \mathbf{H}\mathrm{e}(k+1) \tag{15}$$

10. Apply to the process the control force given by (3).

3 LQR Predictive Control of SISO Time-Delay Processes

All examples in this section follows the design procedure of Sect. 2. The tuning of λ given in (12), as well as parameters c and p, were set using the trial and error method. The reference signal $r(t)$ is a pulse train mounted on a constant. The sampling time T_s employed to obtain $G(z)$ from $G(s)$ is selected, so that $G(z)$ can be modeled by a low-order pulse transfer function without degrading the dynamic behavior of the process model. The disturbance signal $d(t)$ is a pulse train.

Example 1 The integrating processes given by (16), (17), and (18) were controlled by Kaya in [4] based on optimum analytical formulas. This work uses sampling times T_s of 1, 0.1, and 2 s in such processes to obtain $G_1(z)$, $G_2(z)$, and $G_3(z)$, respectively. Figure 1 shows the controlled outputs with the presence of disturbances. As an illustrative example, below is the listing of the code to generate $G_3(z)$ from $G_3(s)$.

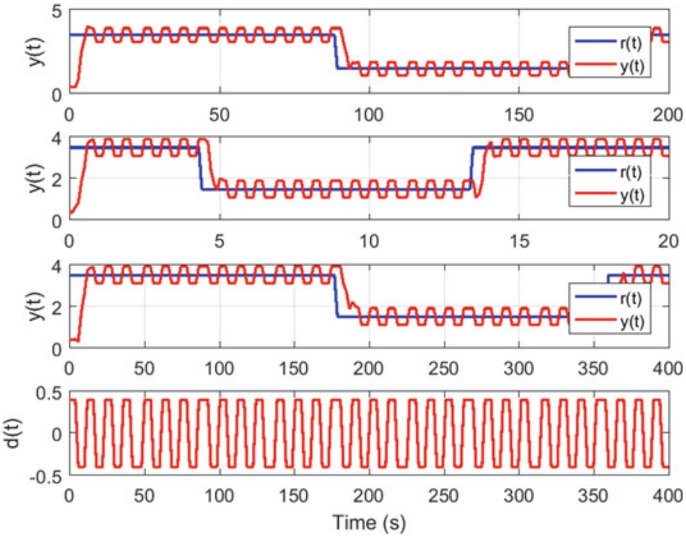


Fig. 1 LQR predictive control of the disturbing processes given by (16) (top graph), (17) (second graph from the top), and (18) (third graph from the top). The bottom graph shows the pulse form disturbed signal $d(t)$

$$G_1(s) = \frac{e^{-3s}}{s(s+1)};\ G_1(z) = \frac{b_4 z + b_5}{z^5 + a_1 z^4 + a_2 z^5} \tag{16}$$

$$G_2(s) = \frac{e^{-0.2s}}{s(0.15s+1)(s+1.2)}$$
$$G_2(z) = \frac{b_3 z^5 + b_4 z + b_5}{z^5 + a_1 z^4 + a_2 z^5} \tag{17}$$

$$G_3(s) = \frac{e^{-5s}}{s(10s+1)(s+1)(0.5s+1)(0.3s+1)}$$
$$G_3(z) = \frac{b_3 z^5 + b_4 z^4 + b_5 z^3 + b_6 z^2 + b_7 z + b_8}{z^8 + a_1 z^7 + a_2 z^6 + a_3 z^5 + a_4 z^4 + a_3 z^3} \tag{18}$$

```
% Generating G3(z) from G3(s)
clear all; close all; clc; s=tf('s');
G3s = 3*exp(-5*s)/s/(10*s+1)/(s+1)/(0.5*s+1)/(0.3*s+1);
% G3s
Ts=2; G3z=c2d(G3,Ts) % G3(z)
%
%                  b3*z^5+b4*z^4+b5*z^3+b6*z^2+b7*z+b8
%  G3z = ----------------------------------------
%           z^8+a1*z^7+a2*z^6+a3*z^5+a4*z^4+a5*z^3
%
     [nu,de] = tfdata(Gz,'v'); n=8; m=1;
 a1 = de(2); a2 = de(3); a3 = de(4); a4 = de(5); a5 = de(6);
 b3 = nu(1); b4 = nu(2); b5 = nu(3); b6 = nu(4); b7= nu(5);
 b8 = nu(6);
 num=[0 0 0 b3 b4 b5 b6 b7 b8]; den=[1 a1 a2 a3 a4 a5 0 0 0];
[Ad, Bd, Cd, Dd]=tf2ss(num, den);
A = [Ad Bd; zeros(1,n) 1]; B=[Bd; 1]; C=[Cd 0];
```

Example 2 The stable and oscillatory processes given by (19), (20), and (21), respectively, were controlled by Mirzal in [5] using a Smith predictor. This work uses a sampling time T_s of 1 s to obtain $G_4(z)$, $G_5(z)$, and $G_6(z)$. Figure 2 depicts the controlled outputs in the presence of the disturbance signal $d(t)$.

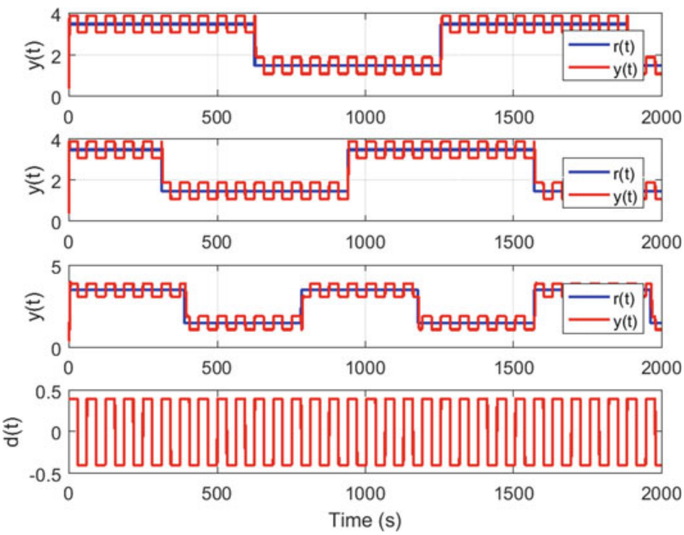


Fig. 2 LQR predictive control of the disturbing processes given by (19) (top graph), (20) (second graph from the top), and (21) (third graph from the top). The bottom graph depicts the pulse form disturbed signal $d(t)$

$$G_4(s) = \frac{e^{-3s}}{s+1}; G_4(z) = \frac{b_4}{z^4 + a_1 z^3} \tag{19}$$

$$\begin{aligned} G_5(s) &= \frac{e^{-3s}}{s^2 + 1.4s + 4} \\ G_5(z) &= \frac{b_3 z + b_4}{z^4 + a_1 z^3 + a_2 z^2} \end{aligned} \tag{20}$$

$$\begin{aligned} G_6(s) &= \frac{e^{-3s}}{s^3 + 3s^2 + 4s + 2} \\ G_6(z) &= \frac{b_4 z^2 + b_5 z + b_6}{z^6 + a_1 z^5 + a_2 z^4 + a_3 z^3} \end{aligned} \tag{21}$$

Example 3 The stable, integrating, and stable with large dead time processes given by (22), (23), and (24), respectively, were controlled by Matusu and Prokov in [6] using a modified Smith predictor. In this work, sampling times T_s of 1.5, 0.5, and 2 s were employed in such processes to obtain $G_7(z)$, $G_8(z)$, and $G_9(z)$, respectively. Figure 3 illustrates the controlled outputs subject to the action of the disturbance $d(t)$.

$$G_7(s) = \frac{3e^{-5s}}{10s + 1}$$

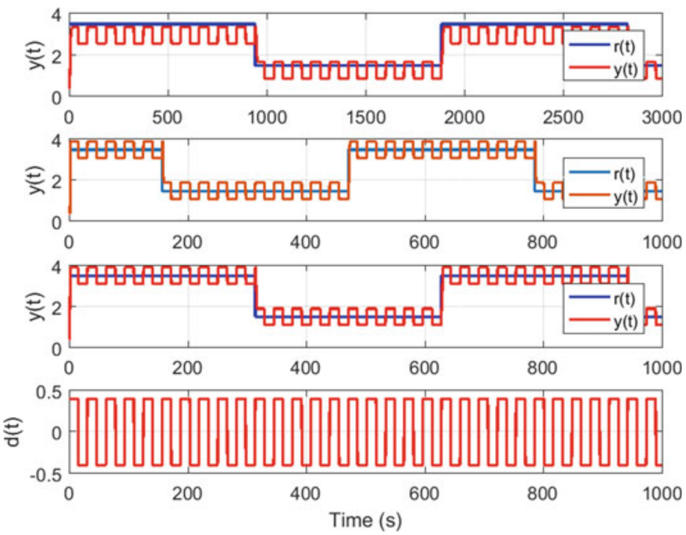


Fig. 3 LQR predictive control of the disturbing processes given by (22) (top graph), (23) (second graph from the top), and (24) (third graph from the top). The bottom graph shows the pulse form disturbed signal $d(t)$

$$G_7(z) = \frac{b_4 z + b_5}{z^5 + a_1 z^4} \tag{22}$$

$$G_8(s) = \frac{e^{-2s}}{s}$$
$$G_8(s) = \frac{b_4 z + b_5}{z^5 + a_1 z^4} \tag{23}$$

$$G_9(s) = \frac{5e^{-10s}}{(10s+1)(2s+1)}$$
$$G_9(z) = \frac{b_6 z + b_7}{z^7 + a_1 z^6 + a_2 z^5} \tag{24}$$

Example 4 The oscillating processes given by (25) and (26), and the non-minimum phase process in (27) were controlled by Tala and Daxini in [7] employing a Smith predictive controller. The unstable process in (28) was controlled by Kumar and Singh in [8] using a PID controller. Sampling times of 1.5, 1.5, 2, and 0.1 s, were used in this paper to obtain $G_{10}(z)$, $G_{11}(z)$, $G_{12}(z)$, and $G_{13}(z)$, respectively. Figure 4 shows the controlled disturbed outputs.

.

$$G_{10}(s) = \frac{e^{-4s}}{s^2 + 6s + 1}$$

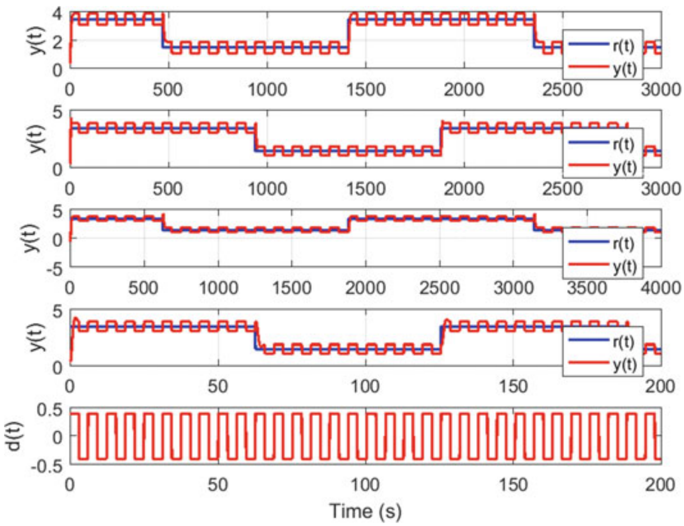


Fig. 4 LQR predictive control of the disturbing processes given by (25) (top graph), (26) (second graph from the top), (27) (third graph from the top), and (28) (fourth graph from the top). The bottom graph shows the pulse form disturbed signal $d(t)$

$$G_{10}(z) = \frac{b_3 z^2 + b_4 z + b_5}{z^5 + a_1 z^4 + a_2 z^3} \tag{25}$$

$$G_{11}(s) = \frac{e^{-4s}}{s^2 + s + 1}$$
$$G_{11}(z) = \frac{b_3 z^2 + b_4 z + b_5}{z^5 + a_1 z^4 + a_2 z^3} \tag{26}$$

$$G_{12}(s) = \frac{(-4s + 1)e^{-4s}}{s^2 + 6s + 1}$$
$$G_{12}(s) = \frac{b_3 z + b_4}{z^4 + a_1 z^3 + a_2 z^2} \tag{27}$$

$$G_{13}(s) = \frac{2e^{-0.3s}}{(3s - 1)(s - 1)}$$
$$G_{13}(z) = \frac{b_4 z + b_5}{z^5 + a_1 z^4 + a_2 z^3} \tag{28}$$

4 Concluding Remarks

This work developed a procedure to control continuous-time SISO time-delay processes having output disturbances using discrete-time LQR predictive controllers. The design of the control algorithm requires the discrete-time state-space representation of the time-delay process. Successful control of continuous-time stable with large dead time, unstable, integrating, oscillatory, and non-minimum phase disturbed processes validates the proposed control approach. The sampling time should be selected to obtain a pulse transfer function of the process model, preferably one with low order, without degrading the dynamic characteristics of such a process. The trial and error method was employed to set the tuning parameters in order to obtain the desired time-response.

For future work, it should be necessary to introduce constraints on inputs and outputs, perturbation signals on outputs and states, and robustness on process parameters.

References

1. Rojas-Moreno, A.: An approach to control MIMO time-delay processes. Glob. Stochastic Anal. (GSA) **4**(2), 463–469 (2017). Available at: https://www.mukpublications.com/gsavol4no2.php
2. Rojas-Moreno A.: An approach to control SISO disturbed time-delay processes. In: IEEE IV Sciences and Humanities International Research Conference (SHIRCON), Lima, Peru, 12–15

Nov 2019, pp. 1–4 (2019)
3. Ikonen, E.: Model Predictive Control and State Estimation, 2017. Available at: https://cc.oulu.fi/iko/SSKM/SSKM2016-MPC-SE.pdf
4. Kaya, I.: I-PD Controller design for integrating time delay processes based on optimum analytical formulas. In: Preprints of the 3rd IFAC Conference on Advances in PID Control, Ghent, Belgium, May 9–11, pp. 575–580 (2018). https://doi.org/10.1016/j.ifacol.2018.06.157
5. Mirzal, A.: Delay compensation using the Smith predictor: a brief review with numerical examples. Int. J. Comput.-aided Mech. Des. Implementation **3**(1), 1–6 (2017). https://doi.org/10.21742/ijcmdi.2017.3.1.01
6. Matusu, R., Prokov, R.: Control of time-delay systems by means of modified Smith predictors: a simple Matlab toolbox. Int. J. Math. Models Methods Appl. Sci. **9**, 668–676 (2015)
7. Tala, A., Daxini, B.: Smith predictive control of time-delay processes (2015). https://doi.org/10.13140/RG.2.1.1166.5765
8. Kumar, S., Singh, V.K.: PID controller design for unstable processes with time delays. Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng. **2**(1), 837–845 (2014)

Information Security Policies Based on the Behavior of IT Users in the Microfinance Sector of Lambayeque-Peru

Ernesto Celi and Regis Díaz

Abstract Financial institutions in Peru are obliged to protect information through security systems that are generally not effective in identifying, evaluating and treating the risks resulting from internal threats, defined as an employee who intentionally or not, accesses the information to negatively affect its confidentiality, integrity or availability. The study developed a conceptual model identifies and evaluates the main factors that influence the behaviors, intentional or not, of IT users in organizations of the microfinance sector of Lambayeque Peru. The evaluated factors are related to the behavior, the influence of the environment and the perception of the control established, to identify in advance the possible risk scenarios that later become an internal threat. This work is based on the theories of planned behavior, reasoned action and deterrence. A survey prepared by the researchers and validated using the Conbach alpha statistic was applied to a sample of 110 IT users, belonging to eight microfinance entities. Through the correlation analysis, it was obtained that the factors related to the intentional behavior obtain a correlation coefficient of 0.695 with respect to the compliance of the information security policies, while the factors related to the non-intentional behavior obtain a coefficient of correlation of 0.564. The model explains 63.9% of the behavior with the identified factors.

Keywords Information security policies · Behavioral theories · IT users

E. Celi (✉) · R. Díaz
Universidad Nacional Pedro Ruiz Gallo, Av. Juan XXIII, Lambayeque, Peru
e-mail: eceli@unprg.edu.pe

R. Díaz
e-mail: rdiazp@unprg.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_30

1 Introduction

To manage security risks, an organization must implement an information security strategy by establishing a global framework that allows the development, institutionalization, evaluation and improvement of an information security program. Information security means protecting information and information systems from access, use, disclosure, alteration, modification, reading, inspection, recording or unauthorized destruction. Therefore, information security has a very important role in supporting the activities of the organization [1].

The procedures, which generally end in a risk management matrix, achieve acceptable levels of effectiveness in mitigating threats specifically from technology (equipment, communication infrastructure, computer applications, database, etc.) and processes (user account management, information backup, application development, change control, etc.). However, they lack techniques and strategies to assess threats from people's behavior, specifically from IT users, being these types of threats, those that have the greatest negative impact and those that have the highest frequency of occurrence, regardless of whether these are intentional or non-intentional.

The effectiveness of the security of organizational information depends on three components: people, processes and technology; then, the "people" component is not properly managed with the procedures and techniques that the financial sector organizations have currently implemented [2]. The regulation of employee behavior is never an easy task because bad actions are difficult to detect through the classic monitoring mechanisms considered in the frames of reference currently used in information security management. The attitude and bad behavior of people in the performance of their work is influenced by a variety of organizational and individual factors [3].

1.1 Theory of Planned Behavior (TPB)

TPB is a theory designed to predict and explain human behavior in specific contexts [4]. The theory proposes a model that can measure how human actions are guided. The occurrence of a particular behavior is predicted, provided that the behavior is intentional [5]. The TPB postulates three intention determinants, conceptually independent. The first is the attitude toward the behavior and refers to the degree to which a person values, favorable or unfavorable, a certain behavior. The second predictor is a social factor, called subjective norm, which refers to the perceived social pressure to perform or not perform the behavior. The third antecedent of intention is the perceived degree of behavior control which, as we have seen previously, refers to the ease or difficulty of performing the perceived behavior, as well as the expected impediments and obstacles. As a general rule, the more favorable the attitude and the subjective norm with respect to a behavior, the greater the control over a perceived behavior; then, the greater the intention of a person to perform that behavior [4].

1.2 Rational Choice Theory (RCT)

Although RCT is generally identified with a paradigm of economic science, it is a general theoretical perspective of the human behavioral sciences, and its scope is that of human interaction, that is, it refers to all kinds of social situations. It assumes that the individual tends to maximize his profit-benefit and reduce costs or risks. Individuals prefer more than good and less than what causes them badly. This rationality has to do with a certain intuition that leads individuals to optimize and improve their conditions [6]. People do not tend to behave in a completely arbitrary way, or in a way simply influenced by feeling; on the contrary, people tend to develop strategies to maximize their interests, to meet their needs as much as possible depending on the resources they have [7].

1.3 Deterrence Theory

Deterrence implies the threat of punishment through some form of sanction. Deterrence is a way of achieving control through fear. Deterrence, in general, is the control of behavior so that a potential offender considers that it is not worth taking a risk with certain behavior, for fear of its consequences [8].

The theory of deterrence postulates that individuals weigh the costs and benefits when deciding whether or not to commit a crime, and choose the crime when it pays [9]. Dissuasion advocates believe that people choose to obey or violate the law after calculating the profits and consequences of their actions.

2 Methodology

The independent variables were defined: intentional behavior of information technology users and non-intentional behavior of information technology users; and the dependent variable, such as information security policies.

The behavior-related factors, which were identified and considered in the model, based on the application of the behavioral theories used, were:

a. Behavioral factors associated with the attitude

   – Integration and commitment: how much it influences, the perception that the individual has of feeling integrated, inserted and adapted with the security system of the organization, in such a way that it implies a personal decision to put into play all the capacities and do more than expected in relation to compliance with information security policies.

- Awareness: assesses the level of awareness that users have to efficiently comply with the safety regulations associated with their functions in their workplace, and that allows mitigating the risks that are present in daily tasks.

b. Behavioral factors associated with influence the environment

- Motivation: evaluates the factors generated by the organization to provoke, maintain and direct the behavior of employees toward compliance with information security policies.
- Training: evaluates the effectiveness of the activities carried out by the organization to achieve specific knowledge in users, in order to increase their expertise in the management of information security controls.

c. Behavioral factors associated with the perception of control

- Technologies for control: evaluates the effectiveness of the technologies implemented by the organization so that processes work based on regulations or parameters established in security policies, specifically related to the control of access to information by users.
- Usability of security tools: evaluates the ability of computer systems to be understood, learned, easily used and attractive to a user, under specific conditions of use.

d. Behavioral factors associated with deterrence

- Deterrence measures: evaluates the action and effect of the strategies implemented by the organization to dissuade its employees from not attempting to violate information security controls.
- Time pressure and work overload: evaluate whether an employee, working in adverse conditions of time pressure or task overload, maintains efficiency in his work and does not make more mistakes than usual, in relation to information security.

2.1 Questionnaire

The research analysis unit was composed of the workers of eight financial entities of the regional microfinance sector. A total of 365 questionnaires were distributed among the workers of the selected entities, of which 133 responded, obtaining a response rate of 36.44%, and 23 cases were eliminated because they were incomplete, resulting in 110 consistent cases. For the collection of data, permission was requested from each of the microfinance entities, applying the questionnaires at times available by the workers. The questionnaire is published in [10].

The questionnaire was applied anonymously since the names or any identity document of the surveyed worker was not requested or registered; the name of the financial institution in which the questionnaire was applied was also not recorded. The questions in the instrument do not allow the worker or the entity to be identified. Still considering that each was informed about the academic objective of the research as well as the guarantee of anonymity giving the choice of participation or not as requested by the Research Ethics Committee of the Graduate School of the Pedro Ruiz Gallo National University.

As the unit of analysis was people, the ethical standards of the institutional research committee and the 1964 Declaration of Helsinki and its subsequent modifications were taken as a reference to carry out the research procedures.

2.2 Reliability and Normality Analysis

Cronbach's alpha test was applied to determine the reliability of the internal consistency of the instrument, resulting in 0.784, considering the instrument as good according to the criteria proposed by George and Mallery [11].

For the test of normality of the data, the Kolmogorov–Smirnov test was used, determining that the data does not follow a normal distribution, so the type of statistic that was applied was a nonparametric type.

3 Results and Discussion

The results of the correlations in the conceptual model are shown in Fig. 1.

3.1 Analysis of the Dimensions of the Variable Intentional Behavior

The hypotheses between intentional behavior and its dimensions were formulated as follows:

H1: High levels of integration and commitment of employees to your organization will reduce internal threats to information security from intentional behavior
H2: Deterrence measures that are reinforced by disciplinary actions will reduce internal threats to information security from intentional behavior
H3: Technology-based control mechanisms will reduce internal threats to information security from intentional behavior.

Considering the application of a nonparametric test, Table 1 presents the Pearson correlation matrix tests performed to determine the relationship between intentional behavior and its dimensions.

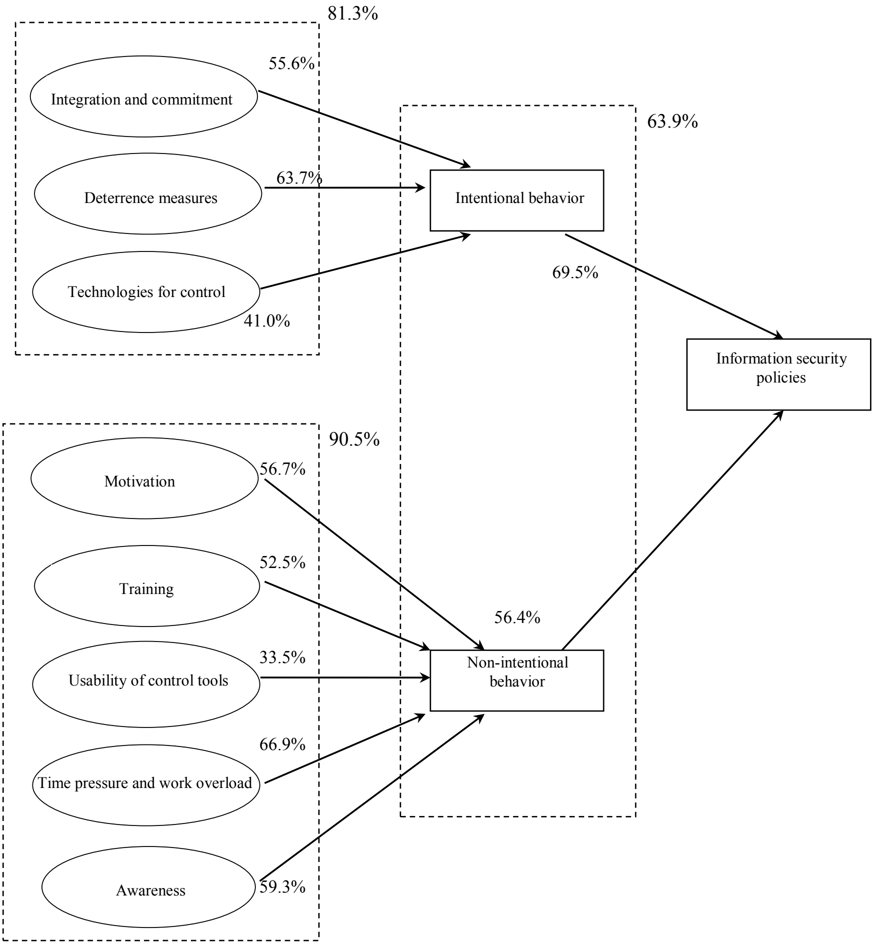


Fig. 1 Conceptual model

Table 1 Correlation of the dimensions of intentional behavior

Variable/dimension	Integration and commitment ($D1$)	Deterrence measures ($D2$)	Technologies for control ($D3$)
Intentional behavior (X_1)	0.556 p-value: <0.0001	0.637 p-value: <0.0001	0.410 p-value: <0.0001

The correlation coefficients indicate that there is a relationship between the dimensions ($D1$) integration and commitment, ($D2$) deterrence measures and ($D3$) technologies for control with the variable intentional behavior (dependent variable X_1).

Therefore, if the levels of integration and commitment of employees in the organization are increased, and deterrence measures are reinforced with disciplinary actions, and control technologies are implemented, then internal threats from intentional behavior will be reduced.

To obtain the coefficient of determination of this relationship, linear regression was applied as follows:

$$X1 = C0 + C1 * D1 + C2 * D2 + C3 * D3 + e \quad (1)$$

The coefficient of determination of the set of dimensions that was obtained was 81.3% with respect to the intentional behavior. This means that intentional behavior can be interpreted through these three dimensions.

3.2 Analysis of the Dimensions of the Variable Non-intentional Behavior

The hypotheses between non-intentional behavior and its dimensions were formulated as follows:

H4: Increasing the intrinsic motivation of the user will reduce internal threats to information security from unintended behavior.
H5: Training users of security tools will reduce internal threats to information security from unintended behavior.
H6: High levels of usability of security tools will reduce internal threats to information security from unintended behavior.
H7: Reducing work-related stress and fatigue levels by adjusting the pressure of time and workload will reduce internal threats to information security from unintended behavior.
H8: Increasing user knowledge will reduce internal threats to information security from unintended behavior.

Considering the application of a nonparametric test, Table 2 presents Pearson's correlation matrix tests performed to determine the relationship between non-intentional behavior and its dimensions.

Table 2 Correlation of the dimensions of non-intentional behavior

Variable/dimension	Motivation (*D*4)	Training (*D*5)	Usability of work tools (*D*6)	Time pressure and work overload (*D*7)	Awareness (*D*8)
Non-intentional behavior (X_2)	0.567 *p*-value: <0.0001	0.525 *p*-value: <0.0001	0.335 *p*-value: <0.0001	0.669 *p*-value: <0.0001	0.593 *p*-value: <0.0001

Correlation coefficients indicate that there is a relationship between the dimensions ($D4$) motivation, ($D5$) training, ($D6$) usability of work tools, ($D7$) time pressure and work overload and ($D8$) awareness with the variable non-intentional behavior (dependent variable X_2). Therefore, if the intrinsic motivation of the user is increased, and user training is conducted on safety tools, and the ability to use safety tools is improved, and the levels of work-related stress and fatigue are reduced by adjusting time pressure and work overload and increasing user awareness and security, then internal threats from non-intentional behavior will be reduced. To obtain the coefficient of determination of this relationship, linear regression was applied as follows:

$$X_2 = C0 + C4{*}D4 + C5{*}D5 + C6{*}D6 + C7{*}D7 + C8{*}D8 + e \qquad (2)$$

The coefficient of determination of the set of dimensions obtained was 90.5% with respect to non-intentional behavior. This means that non-intentional behavior can be interpreted through these five dimensions.

3.3 Analysis of Intentional Behavior, Non-intentional Behavior and Security Policies

Considering the application of a nonparametric test, Table 3 presents Pearson's correlation matrix tests performed to determine the relationship between independent variables, non-intentional behavior and intentional behavior, and the dependent variable security policies.

The correlation coefficients indicate that there is a relationship between the variable's intentional behavior and non-intentional behavior with the variable information security policies. Therefore, if intentional behavior is improved, non-intentional behavior will improve compliance with information security policies.

To obtain the coefficient of determination of this relationship, linear regression was applied as follows:

$$X_f = C0 + C1 * X_1 + C2 * X_2 + e \qquad (3)$$

Table 3 Correlation between the independent variables non-intentional behavior and intentional behavior and the dependent variable Information security policies

Variable/dimension	Intentional behavior (X_1)	Non-intentional behavior (X_2)
Information security policies (X_f)	0.695 p-value: <0.0001	0.564 p-value: <0.0001

The coefficient of determination of the set of variables that was obtained was 63.9% with respect to compliance with information security policies. That is, the variable's intentional behavior and non-intentional behavior allow us to interpret compliance with information security policies.

4 Conclusions

Given the low effectiveness of the tools and methodologies currently used in financial companies to manage the risks that come from the behaviors of IT users are the main source of threats that negatively affect IT assets and businesses, and due to the scarce literature and research on the subject in entities of the Lambayeque-Peru region, this research paper poses the challenge of developing a model that identifies the factors that influence the behavior of internal employees to commit malicious acts and violate security controls.

As a consequence, we propose a set of research directions on the behavior of IT users with other behavioral analysis theories, to find other factors that influence people's behavioral intensities, which have not been identified in this work. Likewise, based on the results obtained, we propose other investigations related to the elaboration of strategies and techniques of behavior management, which allow those responsible for the security of the information of the financial entities, to work this problem more effectively.

References

1. Susanto, H., Almunawar, M.: Information security awareness within business environment: an IT review. In: FBPES AGBEP Ph.D. Colloquium "Bridging the research community towards excellence", pp. 1–24. Universiti Brunei Darussalam (2012)
2. Herath, T., Rao, H.: Encouraging information security behaviors in organizations: Role of penalties, pressures and perceived effectiveness. Decis. Support Syst. **47**(2), 154–165 (2009). https://doi.org/10.1016/j.dss.2009.02.00
3. Dang-Pham, D., Pittayachawan, S., Bruno, V.: Towards a complete understanding of information security misbehaviors: a proposal for future research with social network approach. In: 25th Australasian Conference on Information Systems, pp. 1–10 (2014)
4. Ajzen, I.: The theory of planned behavior. Organ. Behav. Hum. Decis. Process. **50**, 179–211 (1991)
5. Ajzen, I.: From intentions to actions: A theory of planned behavior. J. Kuhl et al. (eds.), Action Control. pp. 11–39. Springer-Verlag Berlin Heidelberg (1985)
6. Vidal de la Rosa, G.: La Teoría de la Elección Racional en las ciencias sociales. Sociología **23**(67), 221–236 (2008)
7. Paramio, L.: Teorías de la decisión racional y de la acción colectiva. Sociológica **20**(57), 13–34 (2005)
8. Elliott, B.: Deterrence theory revisited. In: Road Safety Research, Policing and Education Conference—From Research to Action: Conference Proceedings (2003)

9. Vance, A.: Why Do Employees Violate Is Security policies? Insights from Multiple Theoretical Perspectives. Faculty of Science, Department of Information Processing Science, University of Oulu, Finland (2010)
10. Celi, E., Díaz, R.: Information security policies based on the behavior of information technology users in the microfinance sector of Lambayeque. https://repositorio.unprg.edu.pe/bitstream/handle/UNPRG/1365/BC-TES-TMP-201.pdf?sequence=1&isAllowed=y. Last accessed 20 Oct 2019
11. George, D., Mallery, P.: SPSS for Windows Step by Step: A Simple Guide and Reference, 4th edn. Allyn & Bacon, Boston (2003)

Productive Metrology as a Guide for Dimensional Quality Improvement—A Case Study

Fabio Luis M. Paulon and Crhistian R. Baldo

Abstract Quality is generally associated with adhesion to specifications, absence of defects and minimizing variability. When specifications are unclear or incorrect, manufacturing and verification processes become questionable, and productive metrology can provide the tools and guidance to improve knowledge of manufacturing systems, through SPC and gage R&R studies, and specification endeavors, using GD&T definitions. In a case study, a methodology based on the reduction of uncertainties using the aforementioned methods was proposed, leading to improvements and a better understanding of the measurement systems, as well as suggestions for improving geometrical product specifications.

Keywords Dimensioning and tolerancing · Statistical process control · Measurement system analysis · Measurement uncertainty

1 Introduction

Quality can be defined as a feature of a product or service that enables it to meet needs that were previously presented or implied or a disability-free product or service. Within the context of manufacturing, quality is generally associated with the production of items that physically adhere to specifications [1]. Quality is also associated with costs: reworks, scraps, warranties, claims and damage to the company's image or market position [2]. In this view, a quality product or process not only conforms to specifications, but consistently hits the target with reduced variation. The widespread concept of total quality management gives a strategic dimension linking continuous improvement of management, organizational or production processes to the center of the system with quality. Within the context of dimensional product quality, the responsibilities of associated processes, such as the manufacturing process, specification and metrology process, increase, and metrology plays an essential role, as it can support the other processes, whether in conformity assessment, inspection, process

F. L. M. Paulon (✉) · C. R. Baldo
Federal University of ABC-UFABC, Santo André / SP, Brazil
e-mail: fabio.paulon@ufabc.edu.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_31

control or knowledge generation to support continuous improvement methodologies, such as the PDCA (Plan, Do, Check, Act), DMAIC (Define, Measure, Analyze, Improve, Control) and the NPD (New Product Development) process.

When faced with a quality problem, the big question is: Which processes to work on and how to develop this work? Assuming a specification is accurate, any deviation from it can cause a functional failure, and there is a balance between the cost of the metrology process and the cost of failure [3]. Finding the optimal solution without the overuse or underuse of the metrology infrastructure is a challenge. The correct specification of a product, including all characteristics and tolerances, depends on the knowledge of the product function. Assessing the manufacturing feasibility of a specification requires manufacturing process know-how which can be acquired from sufficiently adequate metrology infrastructure. On the other hand, another problem is: How to select the appropriate metrology package when specifications are unclear? When resources are scarce, there is no room for trial and error methods. The main hypothesis that this work brings is that it is possible to work on dimensional product quality issues with metrology as a guide for decisions. Therefore, the objective of this article is to present and apply the methodology in a case study with dimensional and geometrical product characteristics. In this regard, some important concepts are presented in the remainder of this section.

1.1 Productive Metrology

Productive metrology is "the field of knowledge concerned with measurement to obtain information and, subsequently, knowledge to change something or someone" [3]. This concept challenges the view that metrology is a non-value-added activity, starting from the underlying interaction between better measurement quality, through reduced uncertainties and better information quality, which was shown by Kunzmann et al. [3] using the information theory, to many applications under the manufacturing context, such as conformity assessment, process control, improvement and support for the NPD activities, all achieved through the support of the gained knowledge and know-how. To achieve those results, it is paramount that the following cautions are being taken [3]: (a) The measurand is defined unambiguously and reflects functional requirements; (b) the results are traceable and must have their associated uncertainties; (c) measurement uncertainty is adequate for the specification requirements and follows closely the manufacturing process.

1.2 Measurement Uncertainty

Measurement uncertainty is defined in the ISO-VIM as a nonnegative parameter that characterizes the dispersion of quantity values assigned to a measurand. The relationship between measurements, specifications and functional requirements is thoroughly addressed in terms of uncertainty in the ISO/TS 17450-1: Correlation uncertainty is a measure of how well the specification expresses functional requirements; specification uncertainty is associated with ambiguities in the specification; measurement uncertainty is the sum of the method uncertainty and the implementation uncertainty.

1.3 Geometrical Product Specification

Considering the increasing complexity of manufacturing interactions and supply chains, where there is more distance between designers, productive sectors, tooling workshops, and metrology professionals, communication is key. The well-known mechanical drawing standards are ASME Y14.5 and ISO 1101. Their use permits improved communication of function of the part, by using geometric characteristics. In the ISO-GPS, there is the concept of duality of specification operator and verification operator, stating that both specification and verification must use the same set of operations. These operations are part of the specification itself, and therefore the responsibility of the specification operator, and serve as a guide for implementation in the verification step [4], and the implementation form is the responsibility of the verification operator, who must also manage related uncertainties [5].

2 Proposed Methodology

Following the principles of productive metrology, which is the search for understanding the wholes and, consequently, the reduction of uncertainties, a method was developed approaching from the measurement uncertainty aspects to specification and correlation uncertainties by using basic concepts such as statistical process control (SPC), process capability index (PCI), measurement system analysis (MSA) and geometrical dimensioning and tolerancing (GD&T). The proposed process improvement route follows the opposite direction as product realization. The general idea is depicted in Fig. 1. The main tools are described in the following subsections.

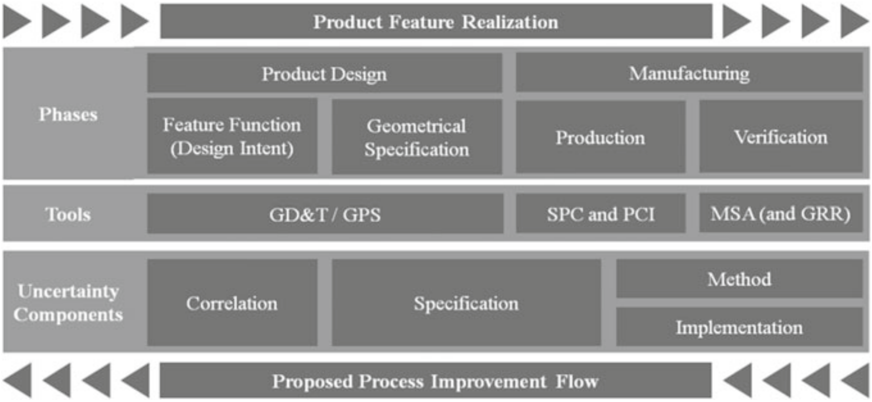


Fig. 1 Improvement analysis route and the main tools considered in the methodology

2.1 SPC and PCI

Statistical control charts are used to view and manage variation within a process characteristic over time. They are useful in identifying changes and discriminating between causes of variation. They are particularly useful as a baseline for process improvements and can also provide information on the measurement system [6]. Control charts are based on the existence of two sources of variation: (a) common causes, which is the random variation present in all processes, and (b) special causes, which is not explained by randomness and always has one or more explanations, such as tool wear, ambient temperature variation, tampering, among others. Control charts can be constructed using individual observations or summary statistics obtained from measurements, and the main purpose is to check if the variable is under statistical control, i.e., the variation comes from common causes only. If the process is out of control, the special cause associated with the out-of-control condition needs to be investigated in order to remove it from the process or incorporate it, if beneficial, into the process. Control charts are useful in two different phases: Phase 1, one has a historical dataset that is evaluated using SPC charts, usually made retroactively, i.e., after data collection to verify whether the process was in a statistical control situation or not; phase 2, the process, which is known to be under statistical control, is monitored to detect changes in this state [7]. When processes are in the predictable state, which could be assessed with SPC, it is possible to compare the variability of the process with the allowance of the engineering tolerance using PCI metrics such as Cp or Cpk, demonstrated in [8].

2.2 MSA and Gage R&R Study

MSA comprises a set of techniques to evaluate the quality of a measurement system. The associated errors can be divided into *accuracy*, difference between the average of a measurement and the true value, which is affected by stability, bias and linearity, and *precision*, capability of a measurement system to reproduce its results, divided into repeatability and reproducibility, usually quantified by gage repeatability and reproducibility (R&R) studies. The R&R study quantifies the variation introduced by the measurement system itself to the measurement results. The R&R index shows the percentage of tolerance range or process variation that is consumed only by the measurement system. The most usual methods are the MSA R&R [9] and the honest R&R, which is explained within the EMP (evaluating the measurement process) approach [10]. Both methods are explored in this work. Particularly, the honest R&R provides insight on process signal attenuation, measurement system attenuation, ability to detect process shifts using SPC and ability to track the process improvements.

2.3 GD&T

This paper will focus on those GD&T definitions related to the function of the product under study. *Datums* are reference features or starting points for spatially locating and orienting part characteristics and are usually defined on features of the type plane and/or axis. Without these reference features, both the fabricator and the inspector must decide how to fix the part according to their own experience and judgment. *Circularity* is a feature that occurs when, on a surface of revolution, all points on this surface, when intersecting a plane perpendicular to the axis, are equidistant from this axis. *Concentricity* relates to the symmetry of an axis relative to the datum axis and does not control either size or shape, but symmetry with respect to the datum. Verification is complex as it requires two indicators at 180° to each other. *Run-out* controls the surrounding surface and is perpendicular to a datum axis. When applied to surfaces built around the datum, it controls circularity and coaxiality.

3 Case Study Description

The engineering drawing of the part considered in this case study is shown in Fig. 2. It is a resin-impregnated carbon–graphite bearing used in submersible water pumps. Those bearings are attached to the pump housing by mechanical interference, indicated as an f7 shaft interference tolerance. The manufacturing process of this part starts from a cylindrical blank, which is obtained after pressing, heat treatment and impregnation processes. The blank is then machined on a CNC lathe. The key characteristics presented in the study are the outside diameter (OD), inside diameter (ID) and concentricity relative to the datum axis A, as presented in Fig. 2.

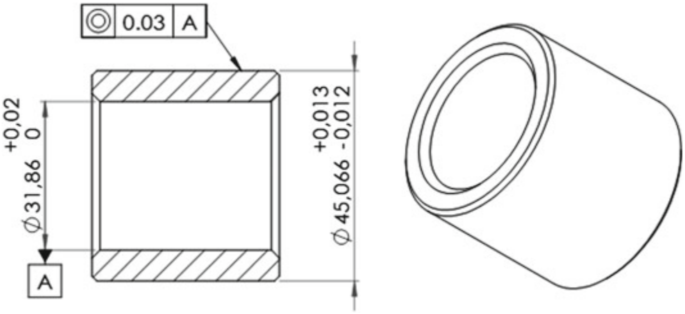


Fig. 2 Simplified geometrical product specification and 3D view of the part

4 Results

4.1 Measurement Improvement

Characteristics such as the OD and ID displayed high variability and therefore insufficient process performance. The first measurements were made using a caliper that has a resolution of 0.01 mm. Initially, control charts for individuals and moving range (X-MR) charts were used over a dataset previously collected. According to the EMP concept [10], there must be at least five categories within the range chart. For ID, there were only four possible (see Fig. 3, left). This means that the measure-

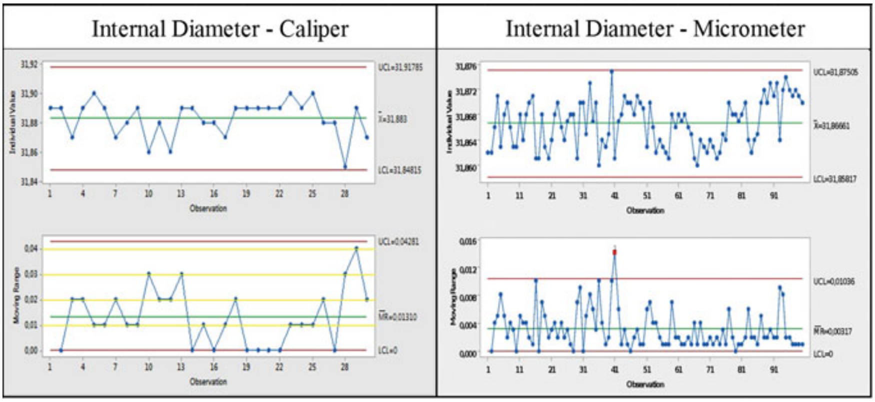


Fig. 3 Control charts for the ID feature using a caliper (left) and a micrometer (right)

ment process is not suitable for SPC. The OD charts showed wider control limits, including six categories in the range chart. Since micrometers with a resolution of 0.001 mm were available, a new dataset was collected using these instruments for OD and ID. Better discrimination ratio of the micrometers has improved drastically the number of possible categories and increased range control limits in the range chart, as shown in Fig. 3 (right). As for the individuals' chart, both OD and ID show correlation because the operator adjusts the process continuously to compensate for the cutting tool wear, which is intense due to the abrasiveness of the carbon–graphite material. In addition, measurement inconsistencies were observed between operators and inspectors, thus requiring the analysis of the measurement process through R&R studies.

4.2 *Preliminary R&R Studies*

The first R&R studies were performed using the ANOVA method. Ten parts were chosen, along with three operators, that measured each part of the study twice. The percentage of variation from R&R was higher than the marginally acceptable range between 10 and 30% in the MSA [9]. The number of distinctive categories (NDC) was the unity for both ID and OD which means this measurement system is unacceptable, as a minimum of five categories is required. To better evaluate the results, average and range charts per operator were plotted (see Fig. 4). Some operators presented higher variation, as observed in the range chart, and significant operator bias could be evidenced in the average chart.

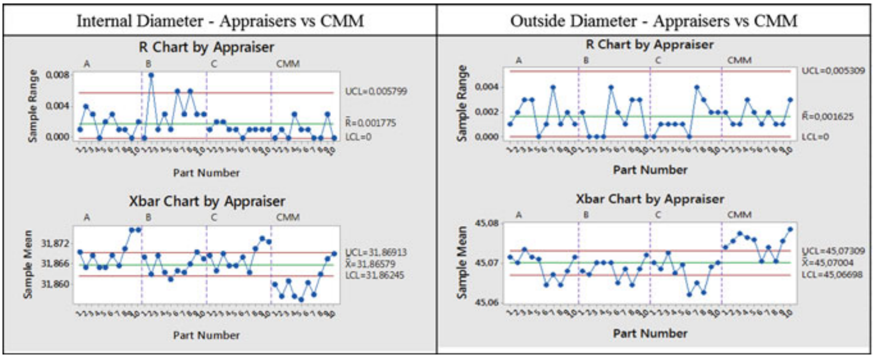


Fig. 4 Average and range charts for OD and ID in the new R&R study

4.3 S.W.I.P.E

To comprehend the measurement variability, a brainstorming based on the S.W.I.P.E (Standard, Workpiece, Instrument, Person and Procedure, and Environment) concept was carried out. Regarding the Standard, no issue was raised because the instruments were regularly calibrated by an accredited metrology laboratory. The Workpiece characteristic that was most concerning is the elasticity of the material, which is greater than that of the measurement instrument, so that the contact force of the measuring equipment could deform the part and provide inaccurate results. The fact that carbon–graphite parts are black does not help to visualize the part during the measurement procedure. As for the Instrument, micrometers, the main claim is that it is unreliable to measure the diameter because the operator is in charge of defining the diameter line, which results in errors of reproducibility and repeatability, and bias is caused by the constant force ratchet. The ergonomics of holding the part in one hand to measure with the micrometer with another hand, as well as the combination of irregular lightning with the black carbo-graphite parts and vibration and distraction from the process are the Environment factor. Finally, for the Person, the main contributors were associated with the operator's familiarity with the instruments, as calipers were the single standard device for such parts, and operational definitions were lacking. Consequently, each operator was taking multiple measurements and trying to guess the correct one.

To reduce the associated measurement uncertainty, a new operational definition of the measurement was made. The diameters would require readings in two different sections, two readings for each section. The highest measurement would be recorded for the OD and the smallest for the ID. The measurements were taken in a place with better lighting and ergonomics. Operators were reminded of parallax errors, and a two-ratchet-click rule was defined, so the measurements were made using a similar contact force between the micrometer spindles and the part.

4.4 New R&R Study and Reference Values from a CMM

A new study was conducted after training and using the new operational definition and knowledge brought by the S.W.I.P.E exercise. The total contribution of variation from R&R reduced considerably: from 39.7 to 32.0% for OD and from 65.0 to 40.2% for ID. The honest R&R approach indicated that the new processes improved the intraclass coefficient (ICC) ratio for both features, which means the measurement systems better transmit the signal from the processes and could be used to track improvements to some extent through SPC. In order to better assess the results of the measurements, the same parts used in the previous R&R studies were measured on a CMM (coordinate measuring machine) housed in a temperature-controlled room. Both OD and ID results exhibited a similar pattern in the average charts, but with a significant offset (see Fig. 4), which is, according to the S.W.I.P.E analysis, caused by the elasticity of the part material. Thus, it is fundamental to take the method bias into account during the specification and evaluation of the conformity of the part.

4.5 Interactions with Geometrical Features and Specifications

The part specification used the ISO 1101 concepts for dimensioning and tolerancing. This means, by default, that the independence principle is assumed for verification, which implies that every measured diameter should meet the specification regardless of form or axial defects. As shown in Fig. 2, the part has a concentricity callout, characteristic which is generally applied to parts that rotate and need some kind of symmetry. In the case of an OD-axis concentric to ID-axis datum, it controls the coaxiality between them. The measurement setup requires two indicators at 180° to each other, from which the median points would be derived. This measurement is affected by probable profile defects from the part, such as circularity, and is more difficult to measure than run-out. Run-out only needs a single feeler gauge in contact with a part rotating 360°, where the highest value is recorded; hence, it is easier to evaluate and controls both circularity and concentricity. Run-out measurements performed on 20 parts showed it was never greater than 0.01 mm. By using the CMM, the average concentricity was 0.004 mm and the maximum value of 0.007 mm; therefore, run-out errors are below the 0.03 mm tolerance. To take advantage of that in the conformance evaluation, run-out could be measured instead of concentricity. By controlling run-out, there is control of the form of the OD, checking the occurrence of local surface points below the lower specification limit that could affect the interference level required in the specification. If an envelope specification could be applied to OD and ID, it would enable the use of go/no-go gauges to evaluate conformance to specification, avoiding possible misclassification of parts due to measurement uncertainty and bias, reducing the compliance uncertainty of the part.

5 Conclusion

The presented results show that the methodology provided insights through all steps, provoking discussions and actions that increased overall knowledge of the product manufacturing process, metrology infrastructure applied, specification and application by a small set of fundamental tools of productive metrology. Control charts were a fundamental data visualization tool that started the analysis process and aided measurement system analysis' R&R by the EMP concept, along with S.W.I.P.E analysis to further improve the measurement system, reducing the overall R&R percentage contribution in the studies. GD&T definitions and concepts improved the knowledge of the part form and function, and permit the suggestion of an improved specification, potentially reducing compliance verification cost and uncertainty.

References

1. Lewis, P., Cooke, G.: Developing a lean measurement system to enhance process improvement. Int. J. Metrol. Qual. Eng. **4**, 145–151 (2013)
2. Taguchi, G., Clausing, D.: Robust Quality. Harvard Bus. Rev. **68**(1), 65–75 (1990)
3. Kunzmann, H., Pfeifer, T., Schmitt, R., Schwenke, H., Weckenmann, A.: Productive metrology—adding value to manufacture. CIRP Ann. **54**(2), 155–168 (2005)
4. Oliveira, A.: Relações entre a especificação de tolerâncias geométricas e a verificação de produto. In: Metrologia (2017)
5. Srinivasa, V.: Standardizing the specification, verification and exchange of product geometry: research, status and trends. Comput. Aided Des. **40**, 738–749 (2008)
6. Diering, M., Hamrol, A., Kujawinska, A.: Measurement system analysis combined with Shewhart's approach. Key Eng. Mater. **637**, 7–11 (2015)
7. Madanhire, I., Mbohwa, C.: Application of statistical process control (SPC) in manufacturing industry in a developing country. Procedia CIRP **40**, 580–583 (2016)
8. Jain, A.: A case study on measurement system analysis (MSA) at a pump company. Int. Res. J. Eng. Technol. **4**(5), 1567–1571 (2017)
9. The Automotive Industries Action Group (AIAG), ASQ.: Measurement Systems Analysis (MSA), Reference Manual, 3rd edn. (2002)
10. Wheeler, D.: EMP (Evaluating the Measurement Process) III: Using Imperfect Data. SPC Press (2006)

Understanding Quantum Computation Concepts and Applications

Reinaldo Padilha França, Ana Carolina Borges Monteiro, Rangel Arthur, and Yuzo Iano

Abstract It is extremely important to point out that through quantum computing, it is possible to obtain results previously unachievable until then by traditional computing. Based on this, this paper aims to give an overview and glimpse of the potential horizon of this area of study that has become increasingly robust in computing over the recent past. It is beyond our scope to address deep concepts of quantum mechanics on which quantum computing is based and reveals. It is rather to form a simple, comprehensive and consistent intellectual scope, not only for specialists but targeted common interest. Thus, this research was built on the latest studies on the subject recognized within the scientific and academic fields.

Keywords Quantum computing · Bit · Qubit · Physical · Overlay

1 Introduction

Quantum computers are hundreds of thousands of times faster than all computers and equivalents that exist nowadays, where such velocity is due to the use of atomic particles in their construction, since they have speed close to that of light, and the processing of information could be done in complete parallelism. To exemplify this quantum parallelism, in theory, one should compare the execution of a quantum

R. Padilha França (✉) · A. C. Borges Monteiro · Y. Iano
School of Electrical and Computer Engineering (FEEC), University of Campinas—UNICAMP, Av. Albert Einstein, 400, Barão Geraldo, Campinas, SP, Brazil
e-mail: padilha@decom.fee.unicamp.br

A. C. Borges Monteiro
e-mail: monteiro@decom.fee.unicamp.br

Y. Iano
e-mail: yuzo@decom.fee.unicamp.br

R. Arthur
School of Technology (FT), University of Campinas—UNICAMP, R. Paschoal Marmo, 1888, Jardim Nova Italia, Limeira, SP, Brazil
e-mail: rangel@ft.unicamp.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_32

computer concomitantly with a “conventional” computer, wherein a chess game scenario, the most powerful conventional computer can analyze 200 million moves per second to overcome an opponent with a high level of expertise. On the other hand, a quantum machine would be able to analyze 1 trillion movements every second, where this processing power will allow problems that could take years to be solved in seconds [1, 2].

In this essence, what establishes the quantum computer faster than a conventional one is, its ability to progress from one probabilistic state to another probabilistic state without having to enumerate all the possibilities involved, doing all of this in one execution. While the conventional would have to perform a high number of processing to achieve the same result with a similar confidence level, remembering that this high processing number serves to reduce the margin of error to an acceptable value (something both traditional and quantum have to do anyway). A system distributed on conventional computers works on a similar principle, but instead of processing on the same processor, as in quantum computing, data is divided into several machines to be processed separately [3].

Thus, a quantum computer is a device that performs calculations establishing direct use of quantum mechanical attributes, generating optimal gain present in the possibility of efficient resolution of some problems that in classical computation would take impractical time, such as factoring in prime cousins, natural numbers (exponential in input size). It is important to note that quantum mechanics limits and rules physical phenomena in the microscopic world, where atoms and molecules exist, and overlap and interference are essential in this process [4].

Such a factor is relative that a computational task to quickly find prime factors for very large integers is beyond the reach of conventional computers, since this is a function that has exponential growth, and one of the most important features of exponential growth is that although it starts slowly can result in huge amounts quickly. Thus, one of the fields of application searched for quantum machines is in cryptography, where they are capable of such a high amount of calculations per second would be unbeatable in the analysis of complex cryptographic keys. That reducing the resolution time of this problem would break vast majority of cryptographic systems today have double bias, constituting authentication systems much safer. Where virtually unbreakable algorithms could be developed with the aid of quantum physics concepts and in the near future have machines with much more processing power, with quantum technology offering a new, more secure channel scheme [5].

To better illustrate the differences between a classical computer and a quantum computer, just think of a car from the 1950s and an automatic car launched in the year 2020, or think of a machine created at the beginning of the first industrial revolution and today’s machines. That work based on deep learning techniques in production line chains. Thus, this paper aims to provide an updated review of the theme showing and approaching its particularities, through a concise bibliographic background, categorizing and synthesizing the potential of technologies.

2 Methodology

This survey carries out a bibliographic review of the main research of scientific articles related to the theme of quantum computing, published in the last five years on renowned bases.

3 Results and Discussion

Throughout the history of mankind, the most spectacular wave of technological innovation has been seen in recent years, and yet there are certain computational problems that the digital revolution cannot yet solve, some of which may be holding back major scientific advances. Today's conventional computers are built on a classic and very limited model of computing, in a way, since their power and speed have been doubling processing year after year, through Moore's Law that signs of exhaustion have been noticed for some time. The great computational advance of recent years has been the miniaturization of electronic components, such as resistors and capacitors, diodes, transistors, present in microprocessors and microcircuits, but this reduction in component dimensions is reaching the limit, although the pace breakneck progress may have started to slow down a bit [6].

What is seen over the last decades, the exponential growth of processors is only accompanied by the reduction of the size of the hardware, which is a practical sense made possible in many ways the technological revolution that is witnessed. Since that silicon, the main component used in the construction of processors is already reaching the physical limits inherent to the material structure itself that, due to the miniaturization of the elements inside the chip, and even with such advancement, they are still not close to solving complex problems [6, 7].

Thus, fruiting the information age, which has been an extremely prosperous time for the world, where the power of computing has led to startling advances in almost every field of human endeavor, greatly contributing to raising the quality of life for most people. The present living age can be called the digital age, with people generating more and more new data and knowledge each year than has been recorded throughout the history of previous humanity, the continual advancement of the power of artificial minds, and the continuing evolution in research of a powerful tool that is quantum computing, which will remove any conceivable boundary that an AI might have between a singularity [8].

To make a simple analogy, today's processors, while powerful and complex, can be understood as extremely small lamps where "off" represents 0 and "on" represents 1; similarly, they are transistors, dealing with a data (bit) stored in one of two possible states: 0 or 1, where their speed and quantity with which they change their state is what allows a computer to perform the calculations. In computing, the "quantum bits," or qubits, this being the quantum version of the classic bits of traditional computing, where they have not exclusively two states, however, an infinity of between 0 and 1. While a bit assumes only 1 or 0, the qubit can assume 1, 0, can be in any overlap of these two states until it is measured, or a combination of both, collapsing to a single value, as long as a small variation is enough to achieve a state change, so more operations can be performed at one time. That is why the machine can do more calculations and solve unsolvable problems compared to today's computers [9].

Due to the laws of quantum mechanics, which includes interlacing, it is described as one of the foundations of quantum computing. Quantum entanglement occurs when two particles are connected (though apart) so that they instantly share their physical states, no matter how large the distance separates them, so what happens in one particle is reflected in the other. A simple example of a clockwise spin on the first particle will be equivalent to a counterclockwise spin on the second particle, with the combined spin of the two is zero, which leads to two quantum systems interacting with each other; they become hopelessly intertwined partners, where the state of one system will give you accurate information about the state of the other system, regardless of whether they are separated by light years and yet provide accurate and instant information about each other, as illustrated in Fig. 1 [10].

In this horizon of knowledge, a quantum bit or qubit is a fundamental unit of information, playing the same role in quantum computing, just as the bit in classical computing, where even though both bits and qubits generate one of the two states (0 or 1) as a result of a calculation, a qubit can be simultaneously in states 0 and 1 before that result, i.e., quantum superposition. Overlapping, unlike conventional transistor bits, qubits are capable of storing three possible information, besides 0 or 1, the qubit can also be 0 and 1 at the same time. The quantum superposition admits that it is possible for an object to exist in several distinct states, where a priori are several of these states simultaneously. In order to illustrate this property, a ball can either be outside or inside a box (classic), or in a quantum superposition of the two states simultaneously, as shown in Fig. 2 [11, 12].

Considering a scenario of a switch (on/off), which is always in a state (on) or another (off), putting in a two-state quantum scenario is something quite different, yet considering that when performing state measurements, it will be indeed on or off, just like a classic system, but between the measurements is the differential, as long as this system can be in an overlap of "on" and "off" states at the same time, no matter how counterintuitive it may seem [13].

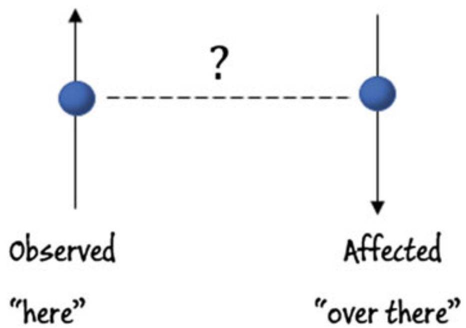


Fig. 1 Quantum entanglement

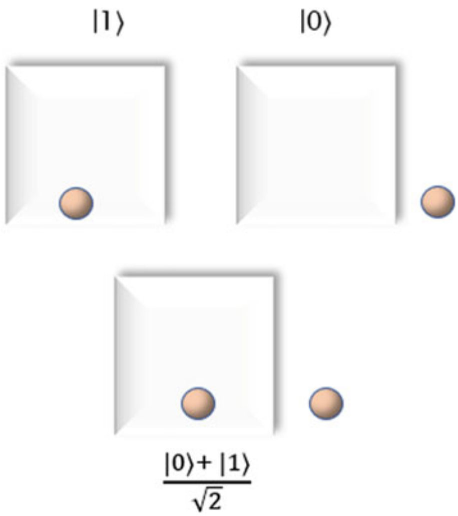


Fig. 2 Quantum superposition

Still taking into account the atomic level, identifying the position of an electron in relation to the nucleus can also be uncertain. Where this third possibility makes storage capacity at the most basic level of a processor dramatically greater, speaking at second scales, quantum computers are capable of performing calculations that conventional ones would take thousands of years [14].

Traditional computers require a lot of resources to calculate properties of more complex molecules, just as modeling the behavior of molecules, where quantum computing can trigger the complexity of molecular and chemical interactions, can even lead to the discovery of new drugs and since it allows many calculations to be performed simultaneously, and also since the molecular dynamics itself is based on quantum physics [15].

Briefly, this is relative to the property of subatomic particles, which can assume different states simultaneously until an observer determines their current state. Since conventional computers are based on classical physics working essentially with two states, represented by 0 and 1, (bits); in the quantum computing, the particle could assume the value 0, 1, or both (qubits), working with an overlap of several different information at the same time, running an algorithm simultaneously on a million data, using an amount in qubits equivalent to your quantity of bits that a conventional computer would use to execute the same algorithm on individual data [16].

This is one of the reasons that allow the quantum computer to perform in fractions of time the work required by conventional computers, greatly increasing the ability to manipulate information due to the quantum state. Thus, in an information context, a bit is a single piece of information, and a qubit is seen as an imaginary sphere. While a bit can only be at one of the poles of this sphere, the qubit, in turn, can occupy any point of the form, so a computational structure with such a system can store an absurd amount of information using less energy, as exemplified in Fig. 3 [16].

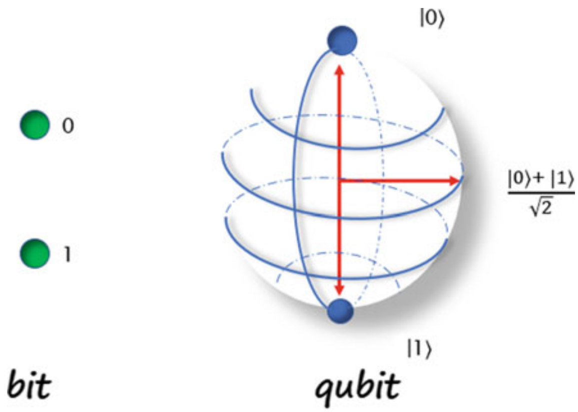


Fig. 3 Bit x qubit

Qubits store information in the form of the orientation, or spin, of an atomic nucleus or electron, where from that atomic spin, taking the photon as an example, the information can be encoded in the polarization, angular momentum or other degrees of freedom of light. The classic bit is through electromagnetic conduction. With the qubit can be represented 25% of one value and 75% of another, imagine a heads or tails toss equivalent to the classical theory, where the bits are absolute, and qubits can represent different values [17].

Correlation is the phenomenon where the quantum states of different objects are interconnected even though the objects are far apart. Putting in a practical setting, a person is in the house and the neighbor outside, although the neighbor does not enter the house, even so, the resident can talk to him (as if by telepathy), wherein a quantum world. This is possible, but in this case, the particles that maintain this "conversation" are via quantum correlation, regardless of the distance they are in, losing their individual characteristics, operating as if they were a single entity, in two different bodies [18].

Still, this quantum experiment, on the other hand, can be interrupted by certain "noises," or any other form of molecular agitation, by increasing temperature, for example, thus weakening the correlation between particles. With the application horizon brought by the quantum correlation, partly because it is possible to break classic encryption used in today's world and partly by a new and more secure encryption method using properties of this technology [19].

There are already several quantum cryptographic standards created, but the physical aspects that these standards adopted to encrypt information are primarily Heisenberg's uncertainty principle and are used to designate the state of an electron, since it is impossible to know the position exactly what an electron occupies in the electrosphere of an atom, and the quantum correlation. Heisenberg's uncertainty principle has practical effects, ensuring that if a qubit communication channel is intercepted, the data will be modified in such a way that the message will reach the other side enough to know that there is a spy on the line. Thus, a quantum computer has the ability to operate on teraflops, where a classic computer can operate on gigaflops [20].

However, there are certain practical disadvantages to the implementation of a quantum computational structure, since they must operate under extreme conditions, since they are very sensitive, and need to be in extremely isolated environments from electrical or magnetic interference, where to ensure very specific operating conditions, such as no noise and no electrical oscillation, plus temperatures so low that they border on zero Kelvin, absolute zero, or -273 °C [21, 22].

There are programming languages with quantum purpose, among the pioneers we can mention Quantum Computing Language (QCL), a derivation of C language, which uses virtual hardware Quantum Random Access Memory (QRAM) to work in a hybrid between quantum and classical, Quantum Flow Chart (QFC) language considered functional, presented in 2004, having a text version called Quantum Programming Language (QPL), and Quantum Modeling Language (QML), a language for both quantum data control, based on linear logic, presenting the if conditional. Still, there is a huge list of simulators and programming languages based on various languages like C/C++, Java, JavaScript, Mathematica, Matlab/Octave, .NET, Python, among others online services providing calculators, compilers, simulators, and toolkits for the development of logic and quantum algorithms [1, 22].

Thus, it is possible to highlight among the many advantages of quantum computing, features such as high processing speed, quantum parallelism, overlap and security. On the other hand, it still has inconvenience as its production difficulty, the issue of information instability, since the complexity of ensuring that the information contained in a physical qubit does not change unexpectedly also represents an unsurpassed obstacle, where mere interaction with the quantum computer produces certain instabilities. Its high sensitivity to the external environment, such as electromagnetic noise, however, small, it can change the qubit state, which makes it very complex to perform calculations and retrieve the results stably, and the room temperature [1, 22].

Thus, quantum computing faces major obstacles to its scalability and consequently to its use as an everyday product. Further research and enhancements should be made on this subject, as quantum computing presents great possibilities for improvements to telecommunications systems [23] and medical diagnostic methods based on digital image processing techniques [24–26]. That is, it has enormous potential to revolutionize many academic and industrial sectors.

4 Conclusions

Every day, there are several major new developments in computing, some endure and are adopted as a standard, others disappear due to various factors, such as not being widely accepted in the market or not being suitable to solve the new challenges that are submitted.

Quantum computing is not such a recent idea in the technology field, where it has been conquering spaces as well as increasingly considered applications, leading to an even more present future in computing.

Having received a lot of attention in recent years, and very clear power of it is its ability to decrypt encryption, but today its potential can be seen in the matter of being able to simulate other quantum systems of interest, such as the dynamics of molecules. Where there are many efforts to produce such an architecture, it still has errors in the logic gates, preventing it from simulating complex systems. Such errors are caused by imperfections in today's architecture; however, as technology develops, these errors will be reduced sufficiently.

References

1. Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., Lloyd, S.: Quantum machine learning. Nature **549**(7671), 195 (2017)
2. Wei, S.J., Long, G.L.: Duality quantum computer and the efficient quantum simulations. Quantum Inf. Process. **15**(3), 1189–1212 (2016)
3. Faist, P., Dupuis, F., Oppenheim, J., Renner, R.: The minimal work cost of information processing. Nat. Commun. **6**, 7669 (2015)
4. Harrow, A.W., Montanaro, A.: Quantum computational supremacy. Nature **549**(7671), 203 (2017)
5. Sergienko, A.V. (ed.): Quantum Communications and Cryptography. CRC press (2018)
6. Radin, M.J.: Reconsidering the rule of law. In: The Rule of Law and the Separation of Powers, pp. 37–76. Routledge (2017)
7. Waldrop, M.M.: The chips are down for Moore's law. Nature News **530**(7589), 144 (2016)
8. Singh, J., Singh, M.: Evolution in quantum computing. In: 2016 International Conference System Modeling & Advancement in Research Trends (SMART), pp. 267–270. IEEE (2016)
9. Shiva, S.G.: Advanced Computer Architectures. CRC Press (2018)
10. d'Espagnat, B.: Conceptual Foundations of Quantum Mechanics. CRC Press (2018)
11. Chiribella, G., Spekkens, R.W. (eds.): Quantum Theory: INFORMATIONAL Foundations and Foils, p. 67. Springer, Dordrecht, The Netherlands (2016)
12. Gomes, L.: Quantum computing: both here and not here. IEEE Spectr. **55**(4), 42–47 (2018)
13. De Wolf, R.: The potential impact of quantum computers on society. Ethics Inf. Technol. **19**(4), 271–276 (2017)
14. Gao, W.B., Imamoglu, A., Bernien, H., Hanson, R.: Coherent manipulation, measurement and entanglement of individual solid-state spins using optical fields. Nat. Photonics **9**(6), 363 (2015)
15. Halpern, N.Y., Crosson, E.: Quantum information in the Posner model of quantum cognition. Ann. Phys. **407**, 92–147 (2019)
16. French, A.P.: An Introduction to Quantum Physics. Routledge (2018)
17. Suter, D., Álvarez, G.A.: Colloquium: protecting quantum information against environmental noise. Rev. Mod. Phys. **88**(4), 041001 (2016)
18. Malek, B., Achour, B., Skander, A.: Quantum cryptography. In: Quantum Communications: New Telecommunications Systems, pp. 121–161. Wiley (2017)
19. Charbon, E., Sebastiano, F., Vladimirescu, A., Homulle, H., Visser, S., Song, L., Incandela, R.M.: Cryo-CMOS for quantum. in: IEEE International Electron Devices Meeting (IEDM), pp. 13–15. IEEE (2016)
20. Gambetta, J.M., Chow, J.M., Steffen, M.: Building logical qubits in a superconducting quantum computing system. NPJ Quantum Inf **3**(1), 2 (2017)

21. Cross, A.W., Bishop, L.S., Smolin, J.A., Gambetta, J.M.: Open quantum assembly language (2017). arXiv preprint arXiv:1707.03429
22. Bravyi, S., Gosset, D., Koenig, R.: Quantum advantage with shallow circuits. Science **362**(6412), 308–311 (2018)
23. Padilha, R., Iano, Y., Monteiro, A.C.B., Arthur, R., Estrela, V.V.: Betterment proposal to multipath fading channels potential to MIMO systems. In: Brazilian Technology Symposium, pp. 115–130. Springer, Cham (2018)
24. Monteiro, A.C.B., Iano, Y., França, R.P., & Arthur, R.: Methodology of high accuracy, sensitivity and specificity in the counts of erythrocytes and leukocytes in blood smear images. In: Brazilian Technology Symposium, pp. 79–90. Springer, Cham (2018)
25. Monteiro, A.C.B., Iano, Y., França, R.P., Arthur, R., Estrela, V.V.: A comparative study between methodologies based on the Hough transform and watershed transform on the blood cell count. In: Brazilian Technology Symposium, pp. 65–78. Springer, Cham (2018)
26. Monteiro, A.C.B., Iano, Y., França, R.P., Arthur, R.: Toxoplasmosis Gondii: from discovery to advances in image processing. In: Brazilian Technology Symposium, pp. 91–101. Springer, Cham (2018)

Emerging Trends in Industry 4.0, Smart and Sustainable Future of Cities

Development of a Robotic System with Movements Based on Computer Vision Detection

Gabriel Gomes de Souza, Bruno Luis Soares de Lima, Rodrigo Vieira dos Santos, and Fábio Jesus Moreira de Almeida

Abstract Nowadays, with the increasing use of robots, and the high level of education required to program and interact with these machines, it becomes necessary to develop platforms that make the human–machine interaction simpler and more precise. In the present work, a control system was developed for robotic manipulators based on computer vision and image processing. An algorithm was developed in Python based on the OpenCV open-source library, which identifies gesture commands captured by a webcam and, from a user interface, transforms them into specific movements to be performed by a prototype manipulator that was designed and built by the authors. The prototype, built-in MDF, has a range of 180 mm and 3 degrees of freedom, and its movements are made from the activation of 3 step motors. The activation of the motors takes place from a firmware that was developed for the Arduino microcontroller, which receives the gestural command that was detected. It was possible to establish critical operating conditions of the developed system in relation to the ambient brightness, distance from the operator to the webcam and the mechanical accuracy of the manipulator's motor response. From the technologies applied in this work, it may be possible to develop control systems by gestures for other robotic mechanisms.

Keywords Robotic mobile · Computer vision · Collaborative robotics

1 Introduction

Over the past few decades, robots have been cooperating with a man in performing numerous tasks, ranging from simple shop floor manipulation tasks to complex tasks such as medical surgeries. In high hazardous applications, for example, remote access through robotic arms or mobile robots is possible, which can mitigate the risks to

G. G. de Souza · B. L. S. de Lima (✉) · R. V. dos Santos · F. J. M. de Almeida
Engineering School, Mackenzie Presbyterian University, São Paulo, Brazil
e-mail: bruno.lima@mackenzie.com.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_33

humans inherent in such applications. In 2016, the Robotics Industries Association (RIA) [1] described four situations in which remote robot access stands out: marine exploration, space exploration, welding and industrial tasks, and disaster response, like landslides and floods. Today's industry is based on large assembly lines for mass production. This leads to the need for large complexes, which tend to grow as society's demand for products increases. Each of the major paradigm shifts that occurred in the history of industry, due to technological changes and innovation, constituted an Industrial Revolution. These revolutions were caused [2] by the mechanization of production (First Revolution), the use of electricity (Second Revolution) and the use of electronics and automation (Third Revolution).

According to the International Federation of Robotics (IFR) statement, 3 million industrial robots will be used in factories around the world by 2020. This means that operational inventory will more than double within seven years [3]. However, one of the negative points of this evolution stems from the need for a high capacity of operators; therefore, only a small portion of professionals are qualified to work. Robot programmers use various programming languages to program robot movements. There are the proprietary languages of robot manufacturers, and basically, every manufacturer has its own [3]. However, once a simple and efficient method of control, combined with gesture detection, is implemented, the programming becomes practicable by any operator. Another promising sphere is the area of medicine. The award-winning "KUKA Medical Robotics" project has demonstrated how this technological field is able to contribute to the mastery of surgeries, imaging and patient rehabilitation due to the development of precise, human-touch manipulators such as the collaborative robot LBR iiwa. Future projection points to the increasingly frequent application of these practices, mainly because they have advantages due to the extreme precision of the movements and the mitigation of human failures, being possible even to perform surgeries remotely by health specialists [3]. Similarly, robotics, within the field of computer vision, has obtained high investments in recent years. The Association for Advancing Automation (A3) found an expense of $709 million in the first quarter of 2018 [4]. This was caused by technological development, which made it possible to integrate industrial machines with cameras and efficient processors, improving their control and interaction with the surrounding environment. Thus, with the current context of increasing automation of industrial processes and digital [5], the development of relatively low-cost remote-controlled robotic platforms with programming interface is become increasingly necessary.

Following this premise, the present work proposes the development of a robotic manipulator and its intelligent control, making it capable of recognizing gesture commands through an algorithm based on the OpenCV computer vision library. The robot will perform movements in response to the specific gestures detected by the developed algorithm, from images captured by a camera. The system to be developed in this work will allow a greater human–robot interaction and the possibility of practical manipulations with the minimum of instructions to the operator.

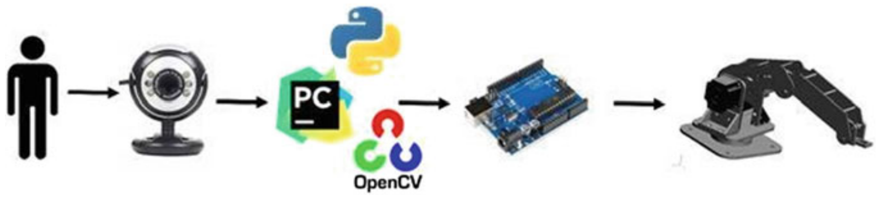


Fig. 1 General design of the proposed system

Figure 1 presents, in general, the system proposed in this work.

The general objective of this work was to develop an intelligent control system of a robotic manipulator controlled by gestures detected by a computer vision algorithm. As shown in Fig. 1, a webcam detects the movements of a standard object manipulated by the operator. From the OpenCV computer vision library, the image detected by the webcam is treated and processed in order to identify the movements of the standard object and to identify the commands to be sent to the Arduino microcontroller board which has a code responsible for receiving the commands and triggering the command, robotic arm.

The following item deals with the computer vision and OpenCV library important for understanding the work developed.

2 Computer Vision and the OpenCV Library

Computer vision is a field dedicated to high-level analysis, modification and recognition of images [3]. Its purpose is to determine what is happening in front of a camera, and to use this understanding to control a computer or robotic system, or to provide people with new images that are more informative or aesthetically useful than the original images.

In computer vision applications, the open-source computer vision library (OpenCV) is widely used because it is an open-source programming library with numerous computer vision algorithms available in the literature and extensive documentation. It was initially developed as a research project by Intel Corporation in 1998 and has been available since 2000 [5].

OpenCV is divided into the following groups: motion analysis and object tracking; image processing; structural analysis; pattern recognition and camera calibration and 3D reconstruction. Computer vision applications need a preprocessing step, encompassing image processing [6, 7]. Most of the time, the images you want to extract some information from need to be converted to some specific format or size and still need to be filtered to eliminate any noise.

In the present work, a computer vision application was implemented, which consists of the development of an algorithm based on the OpenCV library, which allows the recognition of gestures, which made it possible to send specific commands for the movement of the robotic arm.

3 Materials and Methods

3.1 Color Detection and Object Position

The present work deals with the development of a control system for robotic manipulators based on gesture commands, using an object that will be detected by computer vision. This applied technology aims to obtain information from an image using hardware and software, replicating the human vision [8, 9].

The object detection process starts with the use of the OpenCV library, because among its features, this library allows the process of "masking" the image in order to "discretize" it in pixels that will correspond to a certain rule, which will result in splitting the image into regions. The pixels of the pre-established RGB (red, green, blue) color range are scanned, which will produce the binary effect on the image where the rule's pixel will have a value of 1, remaining white, and the pixel non-belonging will have a value of zero, remaining black.

Figure 2 presents a diagram illustrating the process of triggering the robotic manipulator from detecting an image. A green ball was used as a standard object to detect the commands to be sent to the robotic arm.

Initially, the green ball is placed in front of a webcam and is used to determine the command to be executed according to the quadrants shown above. Python's image detection algorithm, based on the OpenCV library, makes it possible to locate the ball's position and its center. Once the chosen quadrant is determined, the algorithm returns the appropriate command to be sent to the microcontroller that will start the motors.

The grouping of these regions will produce image segmentation, a process applied in computer vision to simplify image interpretation and used to locate object contours. This process makes possible the application of the concept of invariant moment of the image, technique present in image processing that allows, through calculations, the definition of the total area of an object and consequently the geometric center of the body referenced in the X- and Y-axes.

Thus, the identification of the position of the object located in the target quadrant will follow a command to be written on the microcontroller serial port, which in turn will drive the motors, leading to specific robot movement.

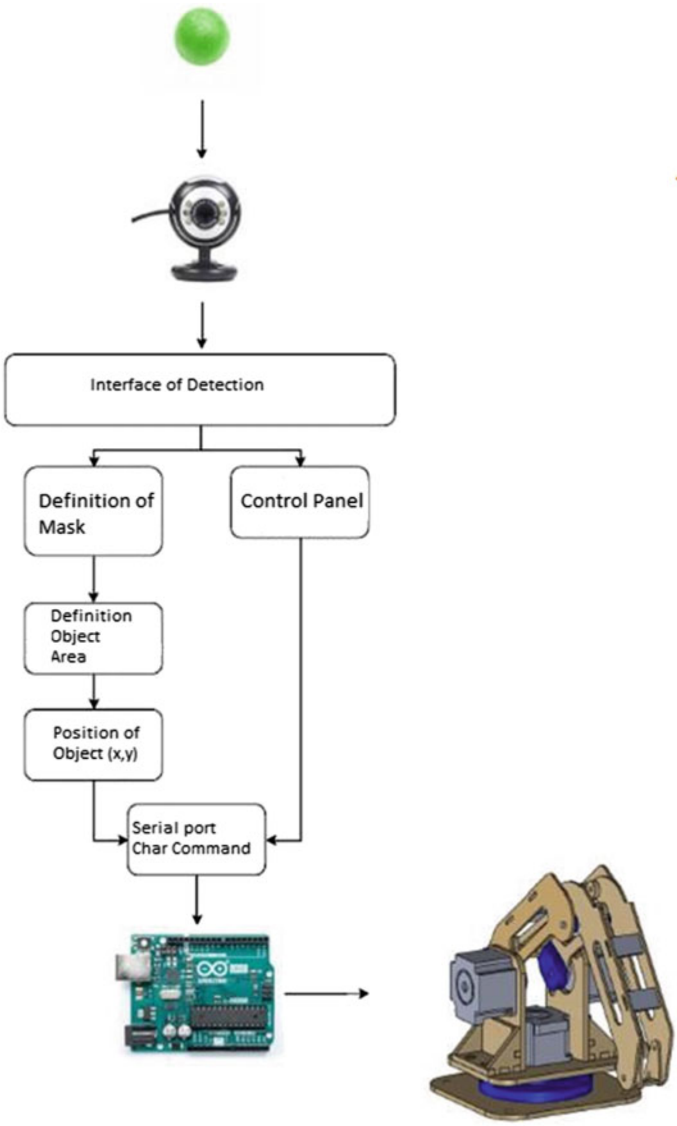


Fig. 2 Overview of image detection process leading to robotic manipulator activation

3.2 Kinematics and Modeling

The main structure of the prototype is formed by the actuator base, arm, forearm, mounting bracket. It also has as its main structure three degrees of freedom: the first degree for manipulator rotation at its base, the second degree for arm movement and the third degree for forearm movement.

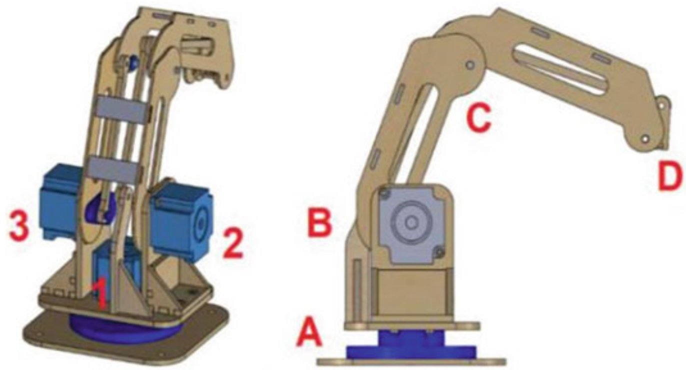


Fig. 3 Modeled robotic manipulator

Figure 3 on the left shows a design of the manipulator, with the location of the motors showing the three degrees of freedom for robot movements.

The three detached motors are for the three degrees of freedom (left) and arm segmentation for torque calculation (right).

4 Algorithm for Image Detection

The developed algorithm collects the commands to move the robotic arm, first identifying the parameterized color that will determine the movement to be detected. After this, the algorithm identifies the image and draws an approximate circle on the board. It then transmits the command to the Arduino microcontroller via the serial port, and then the robotic arm activates. The object to be detected by the algorithm, to determine the movements, is a green ball. The movement of this object relative to the six quadrants displayed on the screen will determine the movements of the robotic arm.

Figure 4 shows the human–computer interface that the operator sends the commands to the system. The interface has quadrants that allow the user to send commands to the arm, and the position of the object within each quadrant defines the command to be sent.

It was defined that the object to be detected has the color green, because this color has an easy perception in most environments. There are numerous combinations of values that are recognized as some kind of green. Therefore, a range of values—a lower bound and an upper bound—is defined within which the color is considered to be green. Then, the radius of the circumference is calculated and its geometric center (centroid) is identified. The program collects numerous radius samples of the identified object, as some of the collected readings may be noisy. Thus, the radius readings are compared to obtain the smallest possible radius for the object.

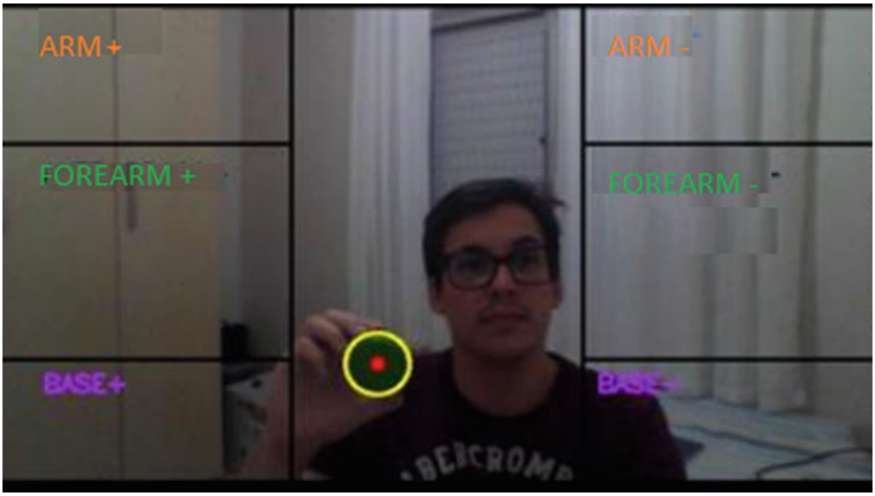


Fig. 4 Scheme of quadrants of motion commands available to the user (author image Gabriel Gomes)

After determining the center of the object, it is decided to inform on the screen the direction of movement made by the user who controls the system through the webcam (only if there is significant displacement made by the ball). The algorithm recognizes commands given by the user from the movement of the green object (ball) in 4 directions, being: north, south, east and west.

A total of four lines are drawn to delimit the quadrants; therefore, there are nine quadrants in all. The quadrants adjacent to the side edges represent the commands to be performed by the robotic arm, totaling seven commands—two for each of the motors present in the arm. Already the three central quadrants emit the variable G to signal the algorithm when the ball is in the three central quadrants.

In the last decision, one of the main aspects of the algorithm development is to identify if the ball is in any of the determined quadrants. If true, a character is transmitted via serial port to the Arduino platform. A communication protocol was established between the image and motion detection algorithm for the microcontroller board responsible for sending the commands to the robotic manipulator motors.

5 Results and Discussion

To perform the concept tests of the robotic arm manipulation system, axis movements were stipulated therein, in order to verify the accuracy of the movements commanded by the computer vision algorithm. Angle deltas have been defined for the movements: 5°, 10°, 20° and 40° on each of the X-, Y- and Z-axes. These angles reflect the need for more precise commanded movements, also with greater pitch and therefore faster.

Figure 5 shows a pair of tests performed, with a 5° angle on the left photographs (85–90° on the Y-axis) and with a 40° angle on the right photographs (130–90° on the X-axis). The starting position of the angle reference is shown to the left in each pair and then to the right in each pair the reference after the movement performed by the robotic manipulator (Fig. 6).

The tests performed to elucidate the positive points, but also the aspects that allow improvements in the project actions, despite the accuracy obtained. Clearances and structural points are critical in successfully executing the desired angle variations. Since the commands were transmitted via the detection algorithm, the instructions were successfully replicated in the robotic manipulator. To verify the functioning of the prototype in relation to ambient light, a factor that affects the command detection algorithm, the tests were performed by checking the ambient light using a P7 photosensitive light sensor from GBK Robotics. This module makes use of the LDR ("light-dependent resistor") photocell sensor, which is a resistor whose resistance varies with the brightness and incident light. Tests were performed by varying the ambient light, measuring it through the described system and verifying the effectiveness in detecting the movements of the object responsible for triggering the robotic arm. The ideal illuminance operating ranges were then obtained for design between 450 and 550 lx, which is observed in conventional industrial installations.

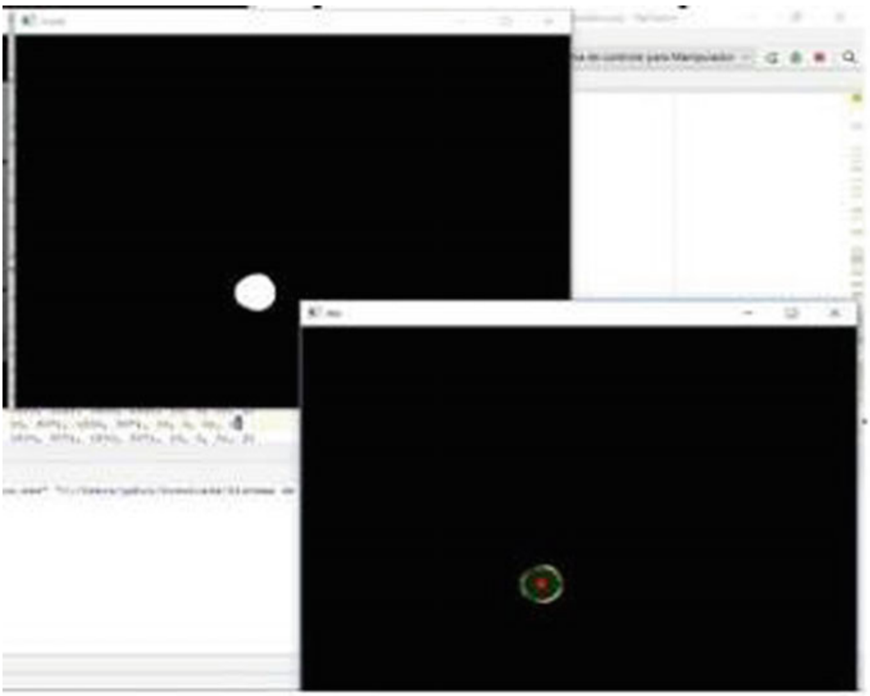


Fig. 5 Color identification mask applied to the detected object

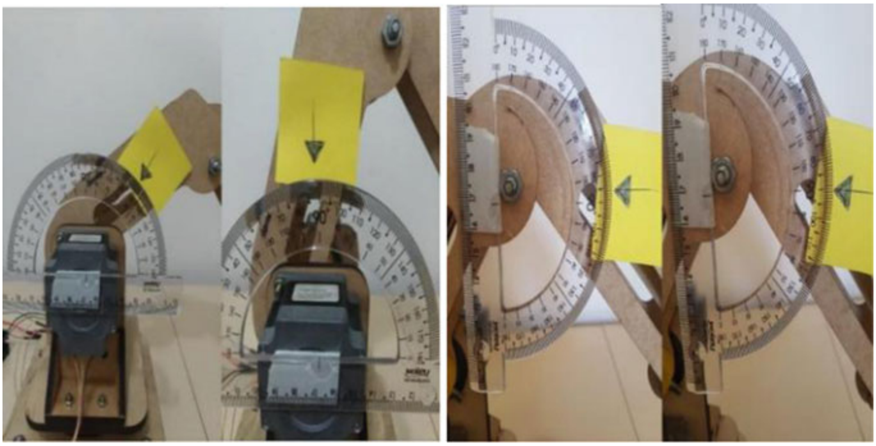


Fig. 6 Demonstration of the movement performed by the stepper motor of the Y-axis, where the movement of the motor is in the position of 85º–90º (left) and movement of 130º–90º in the X-axis (right)

6 Conclusion

Was developed an intelligent control system for a robotic manipulator through the use of computer vision. It was possible to develop the microcontroller firmware to drive the robotic arm and the computer vision algorithm and to build the robotic manipulator itself.

Initially, the proposal was to use Kinect for gesture recognition. Due to difficulties with support due to the discontinuity of Kinect and its libraries, we opted to use a webcam coupled with the OpenCV library. This enabled object detection, contour extraction, and actual tracking and positioning. Thus, it became possible to apply the control of the panel, the user interface of the proposed system, which receives the gesture commands and successfully verifies their replication in the built robotic manipulator.

The work allowed to define a methodology and a set of technologies and development tools that can be applied to the development of other robotic applications with gesture commands.

In the industrial context, where the training of programmers and operators is a challenge, the work proposal may allow the programming or sending commands to a robotic mechanism in a friendly way.

In the future, work may evolve into the context of a collaborative robot where people, operators and other robots can collaborate to perform an industrial task. This will require a system with sensors and actuators to provide safety for the operator and those close to the robots.

References

1. Benesová, A., Tupa, J.: Requirements for education and qualification of people in Industry 4.0. Procedia Manuf. [s.l.] **11**, 2195–2202 (2017). https://doi.org/10.1016/j.promfg.2017.07.366
2. Heer, C.: Robots Double Worldwide by 2020: 3 Million Industrial Robots Use by 2020, IFR.ORG (2018)
3. McMullen, et al.: Demonstration of semi-autonomous hybrid brain-machine interface. IEEE Trans. Neur. Syst. Rehab. Eng. **22**(4) (2014).
4. Arbor, A.: North American Machine Vision Market Grows 19% in First Quarter of 2018, Sets New Records. Association of Advancing Automation (2018)
5. Zezulka, F., et al.: Industry 4.0—An Introduction in the Phenomenon. Ifac-Papersonline [s.l.], **49**(25), 8–12 (2016). https://doi.org/10.1016/j.ifacol.2016.12.002
6. Pulli, K., et al.: Real-time computer vision with OpenCV. Commun. ACM, [s.l.] **55**(6), 61–69 (2012). https://doi.org/10.1145/2184319.2184337
7. Niang, M.Y., et al.: Development of Computer Vision-Based Movement Controlling in Mecanum Wheel Robotic Car, First Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (2019)
8. Hrishikeshan, G.S., Object sorting EYE-BOT based on color sensing using MATLAB for industrial applications. Int. J. Combined Res. Dev. (IJCRD) **4**(6) (2015).
9. Chandramohan, A., et al.: Cost effective object recognition and sorting robot using embedded image processing techniques. Int. J. Innov. Technol. Explor. Eng. (IJITEE) **3**(11) (2014)

Lean Startup in a Commercial Management Model Based on Digital Marketing to Increase Sales in Companies of the Health Food Industry

Rafaela Lazo-Durand, Jose Rojas-Garcia, Carlos Torres-Sifuentes, and Carlos Raymundo

Abstract In 2018, the Ministry of Health of Peru reported that more than 50% of the Peruvian population suffers from obesity. In total, 40% of the people affected by this disease are in the metropolitan area of Lima. Moreover, a report by the newspaper *Gestión* states that, on average, each Peruvian citizen consumes 52 kg of ultra-processed food per year. These alarming figures show that a change in the lifestyle of the citizens of Lima is necessary. It is imperative to increase peoples' awareness of the importance of leading a healthy lifestyle, which in turn means improving eating habits and exercising. Given the increase of people with health problems caused by poor nutrition, the manufacture of health products, such as nut butter, must be natural, without added sugar, hydrogenated oils, added chemicals, or preservatives. Nuts belong to the group of energy foods and eating them provides great health benefits and helps fight obesity. They are a source rich in proteins and healthy fat, provide energy, and contain vitamins and antioxidants, among other benefits. They are also suitable for vegetarians, vegan, celiac, and diabetic people. Therefore, the manufacture and sale of handmade nut butter is a business opportunity in Lima. To prove this, the lean startup methodology was used, and through the development of viable minimum products, the viability of the project was evidenced. The project management bases of the Project Management Institute (PMI) are used as guidelines to develop the project, and digital marketing tools are applied to make the products known to the target market.

R. Lazo-Durand · J. Rojas-Garcia · C. Torres-Sifuentes · C. Raymundo (✉)
Universidad Peruana de Ciencias Aplicadas, Lima 15023, Peru
e-mail: carlos.raymundo@upc.edu.pe

R. Lazo-Durand
e-mail: u201514395@upc.edu.pe

J. Rojas-Garcia
e-mail: jose.rojas@upc.pe

C. Torres-Sifuentes
e-mail: carlos.torres@upc.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_34

Keywords Nuts · Lean startup · Project management · Digital marketing

1 Introduction

This paper aims to develop a minimum viable business plan using the methodology of small business projects. To demonstrate the feasibility of a project for the sales growth of a brand of natural nut butter in Lima, this paper proposes to use the lean startup methodology, which allows companies to implement projects efficiently, minimizing risks and maximizing success opportunities. In addition, the project will follow the tools and techniques established by the Project Management Institute (PMI).

Thus, this paper proposes that natural nut butter, which provides great health benefits and is an important source of vegetable protein and energy, are ideal for diets based on healthy natural foods.

According to a survey conducted by IPSOS, a market research company, among people between the ages of 18 and 70 in urban Peru in 2018, 60% care more about their food, 32% exercise, and 55% care more about their personal [1]. Furthermore, according to the same research company, in a study on the profile of homemakers in Peru, it is concluded that 31% of housewives who follow a diet claim to do so with the aim of eating healthy food, 23% to improve their health, 11% to lower their sugar levels, and 2% to control anxiety and increase energy. It was also noted that 98% of homemakers say it is important to eat food that is good for one's health and 87% agree that it is important to review the nutritional information of the products they buy [1]. In conclusion, Peruvian consumers look for natural, nutritious and healthy food when shopping.

2 State of the Art

In this project, the lean startup methodology will be used for the development of the minimum viable product and its testing. Once approved, the commissioning of the final product will occur and digital marketing tools will be used to reach the target public and begin trading.

Lean Startup. Lean startup is an extremely popular method supported by entrepreneurs, mentors and investors who participate in entrepreneurship courses around the world. Companies using this method enter markets with a minimum of viable products, so failure occurs quickly and at low costs, which allows them to adapt their offers and business models to penetrate the market [2].

This methodology helps companies avoid the large-scale failure that often results from the traditional development based on business plans. Therefore, losses are smaller and, in the event of failure, it is easier to recover. In addition, entrepreneurs can learn from their mistakes and less capital is required [3].

Digital Marketing. According to [4] social media currently plays a significant role in the lives of most people. According to statistics, 93% of Russians using the Internet use social media around three hours a day. For a long-time social media has not only been a platform for communication, but also for users, an additional source of information, a place to exchange opinions, and to seek entertainment, work, and inspiration [5].

Project Management. Projects are essentially risky and project managers (PMs) require various tools to succeed. Some of these tools are tangible, measurable, and certifiable. PMs understand that a project involves the following five process groups [6], which are considered in the latest edition of [7]: initiating, planning, executing, monitoring and control, closing.

Lean startup in a commercial management model in a company from the health food industry. In a sales analysis in a community, healthy food is put for sale at health food stores and its impact on sales is analyzed. Due to the proximity of the health products, people in the community began taking an interest in healthy eating and they preferred healthy food to processed and industrial products [8]. This shows that if people have access to health products, it is easier for them to be interested in a healthy lifestyle.

Digital marketing in a commercial management model in the food industry. Online shopping offers access to items from a global marketplace in e-commerce space, increases customer value and builds sustainable capabilities. Consumers have a tendency to buy items recommended by people they consider trustworthy. Over the past decades, corporations such as Amazon, Alibaba, eBay, and Netflix have become the main drivers of the modern economy. These corporations have stressed the importance of building digital connectivity with their customers [9].

3 Contribution

3.1 Proposed Model

Based on the research and review of scientific articles [2–5], the lean startup methodology will be used together with digital marketing tools to promote advertising. These will follow the project management guidelines described by the PMI for the sales growth project of a company manufacturing and marketing nut butter in Lima. Thus, the following tools will be used in each stage, as described in Fig. 1.

Defining needs. The first step will be the development of market analysis to determine the needs of the target audience. The determined market niche includes people from the A and B socio-economic levels in the city of Lima, aged between 18 and 39, who live in the geographical area known as Modern Lima. According to the National

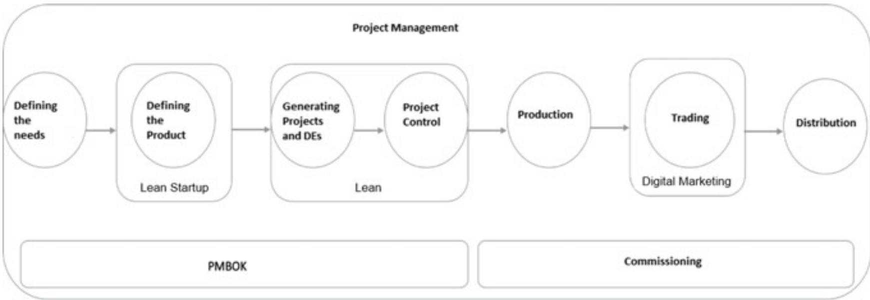


Fig. 1 Project management model

Institute of Statistics and Informatics (INEI), there are 903,000 people from the A and B sectors between the ages of 18 and 39 in the city of Lima.

Defining the products. At this stage, the use of the lean startup methodology is suggested to test the butter currently on the market and receive comments and criticism from current customers to improve and find defects in the products, to correct them, and learn from them. With the use of this methodology, it is possible to know if the people to whom the business is directed to accept the value proposition.

Process Generation. In this stage, lean manufacturing tools will be used to reduce errors in the production process. In doing so, the plan is to use key performance indicators showing the percentage of waste obtained during the process of making the butter and reprocesses. Moreover, the Kanban methodology is planned to eliminate ongoing inventory and overproduction, reduce downtime, and improve delivery time.

Project Control. In this stage, the control and monitoring of the activities designated in the project and their compliance will be carried out.

Commissioning. In this stage, digital marketing tools will be used to reach potential customers through social media so that the brand is positioned in the market. This will be achieved through stories, publications, the sponsorship of influencers related to a healthy and natural lifestyle in social media, and a plan of promotions on significant dates to achieve higher sales.

For this purpose, the use of digital marketing tools such as the following is proposed:

- Advertising on social media.
- Investing in Facebook and Instagram advertisements.
- Using statistical analysis tools.

In Fig. 2, the steps for the implementation of the proposed model are detailed.

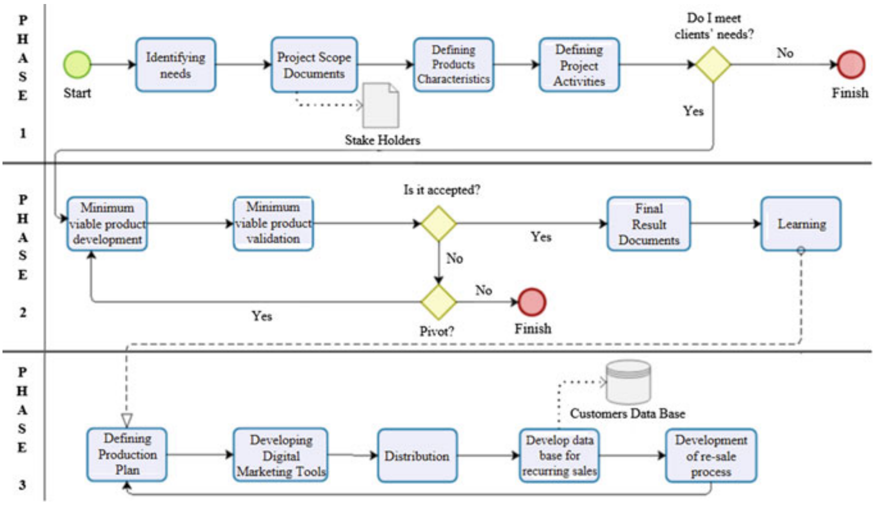


Fig. 2 Implementation guide of the model proposed

Indicators	Calculation	Measurement Unit	Goal	Critical	Tolerable	Optimal
Profitability	$\frac{unit\ sale\ price - u.production\ cost}{unit\ sale\ price} x100$	Percentage	80%	< 70%	70–80%	80–100%
Annual sales growth	$\frac{current\ year\ sales - last\ year\ sales}{last\ year\ sales} x100$	Percentage	15%	< 10%	10–15%	> 15%
Recurring sales	$Average\ ticket\ over\ a\ one - year\ period$	Number of sales per year	12	< 8	9 - 11	> 12

Fig. 3 Indicators

3.2 Indicators

The indicators used before and after the implementation of the methodology are mentioned and explained below (Fig. 3):

4 Validation

4.1 Case Study

The validation was applied to The Butter Factory, a brand that belongs to the Peruvian company, The Superfood Factory S.A.C., located in the city of Lima. It is a company manufacturing, selling, and distributing handmade nut butter. It seeks to manufacture 100% natural products, without preservatives, sugar, or hydrogenated oils, to provide its customers with quality products that contribute to their health.

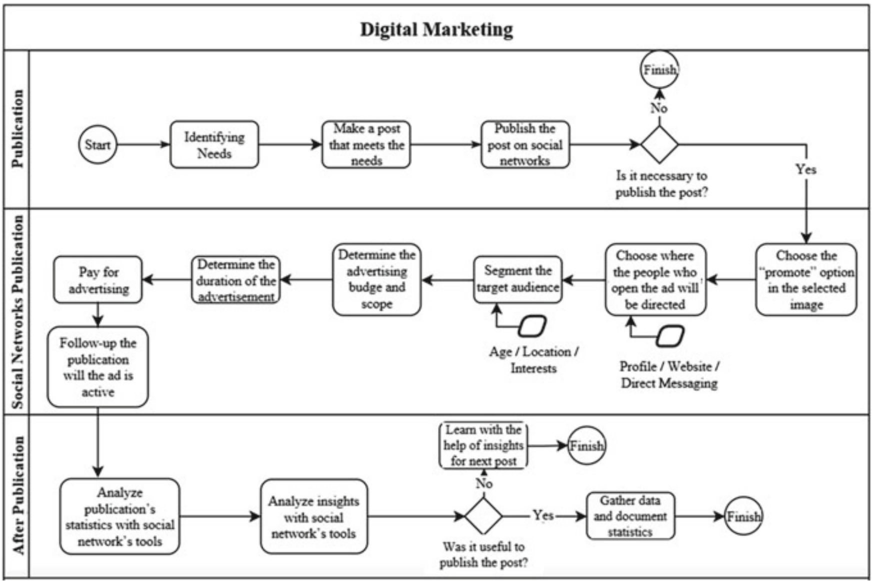


Fig. 4 Digital marketing application (As-Is)

4.2 *Initial Diagnosis*

The Butter Factory is a brand characterized by the various flavors of nut butter it offers (currently, six flavors). Moreover, social media is its only sales channel, and the network management of the company was basic, which did not contribute to the expected sales growth.

Launching new flavors was a simple process but one with a lot of uncertainty on the reactions of the customers. By contrast, the process of applying digital marketing—a vital tool because sales are carried out online—was based on advertising some images on social media and analyzing the results.

The model is shown in Fig. 4 governed the process of applying digital marketing in the company. The results were measured according to the number of new followers on social media owing to the company's investment in advertising and, in turn, the number of purchases.

4.3 *Model Application*

The lean startup methodology helped the company standardize the process of launching new products by using social media surveys to validate proposals; the Kanban methodology helped to sort products as pending, in process, finished, and validated. Once the proposal is validated, small batches called minimum viable products are launched to determine their success or failure based on feedback from a group of strategic clients.

Conversely, a new model was proposed for the application of more efficient digital marketing tools for the success of the sales of the brand, and this model enables the registration and monitoring of returning customers. This model aims to create an e-commerce system to simplify the sales process and make it more efficient (Fig. 5).

As can be seen in Fig. 6, sales of both peanut butter with cookies and cream and sales of pack 4 that includes such butter, were increasing over the months. However, the trend remains variable since there is only one sales channel. In total, the sale of cookie and cream and pack 4 kinds of butter meant a total of S/. 2384.00 for the company in the first 10 months of the year.

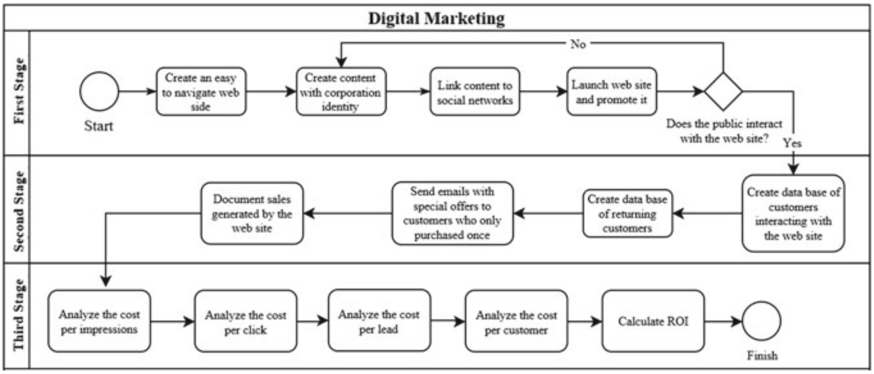


Fig. 5 Digital marketing application proposal

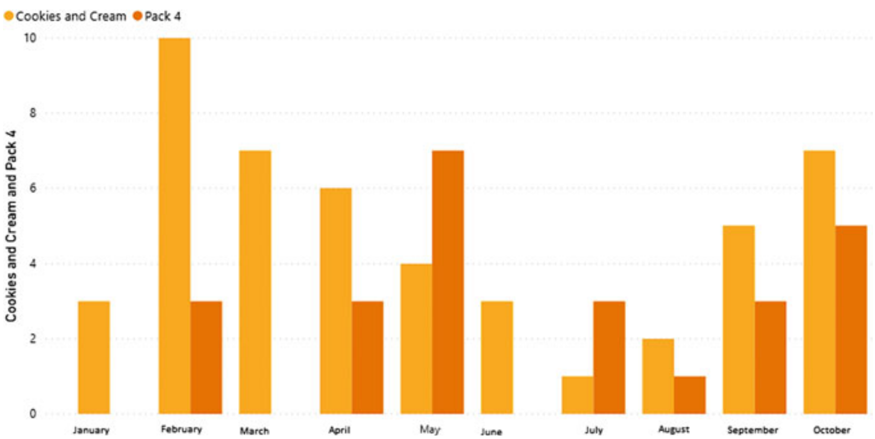


Fig. 6 Sales of cookies and cream flavored peanut butter

5 Conclusions

The uncertainty in the launch of new products decreased considerably when implementing the lean startup methodology. An annual sales growth of 15% is expected with the implementation of e-commerce. The increase in the use of social media entails an opportunity for the company, as social networking services and its delivery service are currently the company's only sales channels. It is recommended that the company continue to innovate in the launch of new flavors to expand its customer base and have a competitive advantage over competitors.

References

1. Álvarez, J.: Congreso: Perspectivas y Tendencias del consumo de alimentos en Latinoamérica. IPSOS PERÚ. Obtenido de (2018). https://www.gcca.org/sites/default/files/2%20Perspectiva%20y%20Tendencias%20del%20Consumo%20de%20Alimentos%20en%20Latinaom%C3%A9rica.pdf
2. Cock, R.D., Bruneel, J., Bobelyn, A.: Making the lean start-up method work: the role of prior market knowledge. J. Small Bus. Manag. 1–21 (2019)
3. Clement, T.: Excelsior-Henderson motorcycles: could lean start-up methodology have averted a $100-million Mistake? Int. J. Entrep. Innov. 144–156 (2019)
4. Shabalina S., Shabalin E., Kurbanova A., Shigapova A., Vanickova R.: Social media marketing as a digital economy tool of the services market for the population of the Republic of Tatarstan. In: Digital Transformation of the Economy: Challenges, Trends and New Opportunities. Advances in Intelligent Systems and Computing, vol. 908. Springer, Cham (2020)
5. Shabalina, S., Shabalin, E., Kurbanova, A., Shigapova, A., Vanickova, R.: Social media marketing as a digital economy tool of the services market for the population of the Republic of Tatarstan. Adv. Intell. Syst. Comput. 356–367 (2018)
6. Alexander, T., Caldwell, J., Gonzalez, M., Harvey, D., Nye, B., Rodgers, C., Washer, A.: The Role of Emotional Intelligence (EI) in Project Management Over the Next Five Years. Organization Acceleration (2010)
7. PMBOK: sexta edición (2017)
8. Paluta, L., Kaiser, M. L., Huber, S., Wheeler, J.. Evaluating the impact of a healthy corner store initiative on food access domains. Eval. Prog. Plan. 24–32 (2019)
9. Rajat Kumar, B., Angappa, G., Shivam, G., Shampy, K., Pradip Kumar, B.. Personalized digital marketing recommender engine. J. Retail. Consum. Serv. (2019)

Lean Manufacturing Model Adapted for Waste Reduction in Peruvian Condiment Production Agri-Businesses

Axel Mesones-Guillén, Lizeth Palacios-Jesús, Edgardo Carvallo-Munar, Rosa Salas-Castro, and Luis Cardenas-Rengifo

Abstract The Peruvian export supply is primarily based on the segment of dressings and condiments. This paper seeks to adapt lean manufacturing tools using Poka-Yoke techniques and process standardization, which may align with existing processes to prevent production errors. This combination of tools is expected to reduce the percentage of waste generated throughout the condiment production process. Subsequent to an initial evaluation of the current method combined with the application of both the tools, 11.4% waste reduction was ultimately reported.

Keywords Lean manufacturing · Standard of procedure (SOP) · Scrap reduction · Poka-Yoke · Condiment production

1 Introduction

Perú ranks 16th among the suppliers of condiments and dressings to the USA. According to the Commission for the Promotion of Peruvian Exports and Tourism (PROMPERU) [1], mayonnaise exporters reported a sales of 2.5 million dollars in 2018. Therefore, waste levels must be reduced to maintain operating profitability. However, as evidenced in a survey conducted by the National Society of Industries

A. Mesones-Guillén · L. Palacios-Jesús · E. Carvallo-Munar · R. Salas-Castro ·
L. Cardenas-Rengifo (✉)
Universidad Peruana de Ciencias Aplicadas, Lima 15023, Peru
e-mail: pcsilcar@upc.edu.pe

A. Mesones-Guillén
e-mail: u201312737@upc.edu.pe

L. Palacios-Jesús
e-mail: u201221553@upc.edu.pe

E. Carvallo-Munar
e-mail: pcinecar@upc.edu.pe

R. Salas-Castro
e-mail: pcinrsal@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_35

(SIN), this production is affected by the high manufacturing costs of the product. The study focuses on a small condiment producer with the waste rates of 40.39% in the mayonnaise production process (dosimetry, liquefaction, and packaging process), representing 12% of their monthly turnover.

The problem is evidenced by an indicator provided by the Food and Agriculture Organization of the United Nations (FAO) [2], which defines average production losses of approximately 28% for agro-industrial businesses, implying a gap of approximately 12.39%. The key purpose of implementing an adapted lean manufacturing model is to use the combination of tools for the case study. On the other hand, every process requires a preliminary study to reduce waste through the optimal use of all resources [3]. Herein, the combination of tools is adapted to reduce waste in the liquefaction process, as it reports 14.99% of waste, thereby generating the greatest impact on losses in the different processes.

2 State of the Art

Implementing lean manufacturing with continuous improvements in the agribusiness sector generates greater process quality, safety, and productivity. Therefore, it is often used for managing costs and increasing the quality of the production system [4].

Lean manufacturing is based on the principles aimed at eliminating waste and maintaining coordination between processes [5], thereby operating with limited resources in a production chain where costs not only can be reduced but can also be improved [6]. In developing countries, even the implementation of lean is limited to several factors that intervene in the success of the changes proposed [7]. These barriers are present in companies, such as small and medium enterprises (SME), where the lack of communication, leadership, commitment, and resources results in long-term improvement proposals being usually discarded [5].

Nallusamy and Punna, in their small-scale manufacturing industry study, addressed the problem of 25% capacity reduction due to process losses using lean tools with root cause analysis [8]. Cauchick et al. provide a lean approach wherein the project focuses on a Brazilian food processing company, obtaining an overall productivity improvement of 17.5% [7]. Chutima and Ritprasertsri aimed their study at reducing production times in a condiment line using lean manufacturing to boost competitiveness and meet production goals. Consequently, total process times were reduced by 47.76%, using an effort approach based on issue prioritization with relevant causes.

2.1 Lean Tools

For the development of the proposed solution, the following tools are used in this work:

Poka-Yoke. Shigeo Shingo, who innovated this term, introduced simple mechanisms to prevent errors and defects in production. These errors are classified based on physical devices, functionality, and symbolism. There is also segregation of processes for correcting the error on the spot (reagents) and for preventing errors (proactive) [9].

Throughout the literature review, the possible correlation between human errors and the potential problems that may occur in the process is studied to suggest a process improvement approach. Therefore, to obtain good results in the implementation, it is necessary to assess the importance of the types of human errors in terms of the solution based on their frequency and criticality [10].

As activities that systematically exclude the error are eliminated, the expected error reduction for a Poka-Yoke mechanism is close to 100% because the problem is detected at an early stage [11]. For a Poka-Yoke to appropriately work, it must fulfill the following: have technical support for its implementation and changes in working methods; maintain quality and safety in the process; maintain process-related attributes, and maintain consistency for fault prevention [9].

Dhingra et al. [12] improved the process by reducing the impact of waste that was preventing the demand from being met. At that time, waste was estimated to cost 2,35,008 Indian rupees. Subsequent to implementing the Poka-Yoke tool, the incidence of errors was reduced by 100%. Vinod et al. [9], while reviewing the Poka-Yoke method, observed a relationship between this lean tool and the DMAIC methodology; therefore, they encouraged further research and studies to accomplish goals pertaining to productivity and the reduction of variability and incidents. Ahmad et al. [11] implemented a Poka-Yoke that not only guaranteed operator safety but also generated a 55% decrease in hazardous incidents for employees as well as for the process itself.

Standard of Procedure (SOP). This technique identifies areas or processes wherein methods or sequences could be improved to reduce process waste. For operators to understand the SOP registration document, the SOP document must be supplemented with instructions that facilitate understanding and help the resources adapt changes in a quicker manner [13].

Process standardization is defined as a set of instructions that clearly show the user how to sustainably achieve the results desired by the organization under standard conditions. These documents are only suitable for repetitive tasks. On the other hand, these are dynamic documents that reflect changes in work practices. If they are not modified on a timely basis, they can quickly become irrelevant or produce potential risks [14]. The standardization flow required is based on continuous improvement, which should govern the future proposal, along with any corresponding corrections [15].

In another study, Nallusamy and Saravanan were able to reduce lead times by 35% in a small business through standardization, thereby increasing customer satisfaction and productivity (owing to less waste generated). Costa and Bragaça stated the advantages of process standardization while observing the working of these processes, as well as their maximum capacity, and how collaborators can be assisted so that they can maintain certain activities while performing their process sequence.

Naravanan et al. [6] further demonstrate that using lean manufacturing through the standardization generated a 28% increase in the process efficiency at an agro-industry company. Mor et al. [13] also performed a process standardization in a manufacturing company wherein a 65% increase was reported in the total productivity, thereby evidencing how this lean tool may reduce activities that do not add any value to the processes and negatively impact the economic aspects.

2.2 *Pre-assessment Tool*

Systematic Interview Technique. To determine which part of the process requires modification, the system uses three phases (based on item prioritization). This assessment tool implements a hierarchical change in a systematic way, wherein a current situation can give way to a proposal that adds value to the selected process.

Miyagawa et al. [16] through statistical process monitoring concluded that while separating water and oil during the mayonnaise production process, a very important factor that contributes to destabilization is the exposure of the product to room temperature. In addition, the use of certain ingredients (egg yolk) can reduce this process affectation. As there is a factor that influences the production of mayonnaise itself, it is relevant to perform a preliminary evaluation to obtain better results prior to introducing process modifications.

3 Contribution

Based on the literature, lean tools are frequently directly used in processes that require improvement. Furthermore, the existing standardization trends generate process modifications based on their current situation. However, the International Labor Organization (ILO) recommends conducting a previous assessment of the structure used for current activities so that the tasks can implement a relevant improvement to the problem.

Therefore, an adapted model is used based on the models proposed in the literature review coupled with the previous assessment to mitigate the impact of the changes proposed.

This comprises three phases as shown in Fig. 1: (I) initial phase, (II) middle phase, and (III) final phase. The first phase is the evaluation of the current situation to determine the points of improvement for the process assessed. The second phase corresponds to the application of techniques to the process that will be modified with the new components, as well as the changes applied to the existing process. The final phase involves the preparation of a document wherein the processes are standardized and can be replicated in the future, contributing to the continuous improvement of the condiment production process.

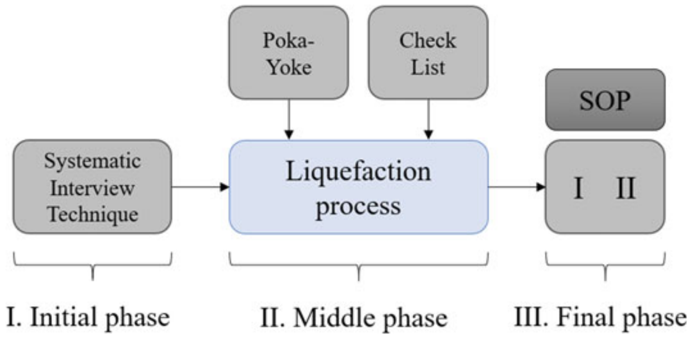


Fig. 1 Adapted proposed model. *Note* The gray squares are new components; whereas, the blue square is the modified component

3.1 Components

Systematic Interview Technique. During the initial phase I, the systematic evaluation is conducted as the different possibilities are required to be prioritized in a predetermined order. In this way, study methods may be conducted based on the basic conditions suggested by the ILO to achieve good results. To proceed with the subsequent stage where modifications to the current process are proposed, it is important to reduce the complexity of process activities.

The deliverable for this component includes changes to the current method being assessed, which may be evidenced using a flowchart.

Poka-Yoke. The reason for using Poka-Yoke is the need to grant operators less control over the activities conducted as well as facilitating the vision of the order and sequence corresponding to the production in the process. Therefore, subsequent to the changes provided in the initial phase "I", operators may be able to adopt the new process guidelines during the middle phase "II". Furthermore, the Poka-Yoke tool provides visual support that prompts the operators of critical activities and those that have become critical due to modifications.

The Poka-Yoke introduced in this component will be visual, and it must reduce process variability subsequent to implementing the proposed changes. To measure the results, an error incidence indicator is used as follows:

$$\text{Error incidence} = \text{Number of occurrences of the error to be prevented using Poka-Yoke} \quad (1)$$

Checklist. The checklist is another tool that complements the middle phase "II". With the help of this tool, the critical activities performed by the operator may be reviewed for compliance with process parameters. In addition, Poka-Yoke may also be consolidated as a change in the current process. As a result of the complementation of the tools, expected actions are achieved by the operators in their work.

SOP. In the final phase "III", process modifications and aggregate values are consolidated through an SOP document for the new procedure adapted to the proposed tools. At this stage, all key process players would already be acquainted with the SOP document, as well as documenting all activities for future replication (e.g., to address staff changes or to introduce modifications to the model proposed).

In addition, this phase also defines objectives, process owners, scoping, activities, and all documentation required and generated during the process. In other words, this phase provides an overview of everything this tool implies.

3.2 Case Study: The ABC Condiment Production Company

ABC is a Peruvian company that produces condiments, salad dressings, dried mash potatoes, pasta, and beverages targeted to socioeconomic sectors B, C, and D. Over the last two years, it obtained new contracts that allowed it to increase its capital and invest in plant improvements to meet production levels.

The condiment production line at ABC accounts for approximately 90% of the total sales. Based on the data from 2018, within the condiment varieties produced, mayonnaise generates 66% of the average monthly revenue, which is why this study focuses on this product.

3.3 Current Problem Indicators

ABC's current mayonnaise process exceeds the waste limit established by FAO (28%) as it produces 40.39% waste. Among the three primary production processes, the liquefaction process contributes to 20.15% of the average waste.

Taking this into consideration, the causes that considerably affected the process in terms of economic matters were determined, thereby concluding that 37.44% of monetary losses were due to sequence errors when placing the ingredients in the mixer (liquid ingredients must be placed before solid ingredients).

The current values that determine the foundation for subsequent validation are as follows:

- Average weekly losses due to sequence errors: 20.21%.
- Process losses associated with sequence errors: 14.99%.
- Incidence of sequence errors: 11.33 times per batch.

4 Validation

In agribusiness companies, the processes contemplate unique product characteristics, which must also be achieved when consumed under regulated standards. For condiments such as mayonnaise, an oil–water emulsion contains a large percentage of the oil as the primary element. The separation of these elements easily occurs during the preparation of the product; therefore, special care must be taken when adding ingredients to the mixer, especially those ingredients acting as emulsifying agents, as they allow the mayonnaise mixture to remain stable.

Therefore, the scenario shown focuses on the proper use of emulsifiers in the raw material liquefaction stage to obtain the final product in the sequence established by the company.

4.1 Scenario Presented by the ABC Company

The proposed model was validated through a pilot test in the liquefaction process during the two-week period when more than 50% of the monthly production is accomplished in the ABC company. For 30 batches, an order of 18,524.5 kg was arranged.

In this test, a change was introduced to the order of the ingredients, grouping activities to reduce the number of steps to perform the liquefaction process. In this way, during the middle phase "II", the implementation of the visual Poka-Yoke was favored by adapting the process to the proposals.

Furthermore, the reduction in complexity of the process favorably impacted the error incidence indicator in the sequence because the operators had not only the Poka-Yoke mechanism but also a checklist and a counter to connect each step of the liquefaction process with the corresponding ingredients.

Finally, the procedure was documented based on the results from the final phase "III", wherein both the method changes and the tools were added to the mixing activities performed. Moreover, the records that were to be created, now that the wrong sequence indicator was known, were also considered.

4.2 Results

Comparing the current process without modifications against the literature review, a 55% decrease in error occurrence and losses is expected.

Table 1 shows that the study was unable to achieve the desired indicator value for weekly average losses and process waste; however, a 59.84% decrease was reported in sequence errors.

Table 1 Obtained results

Indicator	Obtained results		
	Current	Expected	Validation
Weekly average waste (%)	20.21	9.20	11.40
Decrease in the liquefaction process (by sequence errors) (%)	14.99	6.67	8.77
Sequence errors (batches)	11.33	5.16	4.55

Note The weekly average is based on the weekly data obtained according to the results for the period of time, and process waste is given by the relationship between the process losses and the total number of kilograms entered during the period

Regarding the weekly average waste reduction due to sequence errors, a 43.59% decrease was reported, implying an 8.81% reduction from the current total of 40.39%. Therefore, process waste now accounts for 31.58%, i.e., only 3.58% above the average limit established by the FAO.

5 Conclusions

According to the data obtained when validating the adapted model, it is possible to reduce waste during the liquefaction process by reducing sequence errors in the activities performed by the operator through lean and systematic interview techniques. Although the average process waste percentage was not reduced to a value lower than the limit established by the FAO, the possibility of assessing new scenarios would lower the waste generated to values closer or below this limit (28%). The efforts could be focused on other mayonnaise production processes that were not included in the validation performed. The standardization process itself coupled with new activity components generates resistance to change.

Therefore, certain controls must be established to ensure that the new parameters are met and, in addition, to evaluate the implementation of periodic training for both current and future operators. These aspects should be taken into account in future scientific research work on this subject.

References

1. "PROMPERUSTAT - Estadísticas de Exportaciones Peruanas." [Online]. Available: http://www.siicex.gob.pe/promperustat/frmPaises_x_Partida.aspx. Accessed 23 Jun 2019
2. "Líneas de acción contra mermas alimenticias -." [Online]. Available: https://www.mundohvacr.com.mx/2014/12/lineas-de-accion-contra-mermas-alimenticias/. Accessed 23 Jun 2019
3. Prasanta, K.D., Chrisovalantis, M., Debashree, D., Soumyadeb, C., Fouad, B.A. Could lean practices and process innovation enhance supply chain sustainability of small and medium-sized enterprises? Bus. Strategy Environ. 1–17 (2019)

4. Desai, D.A., Kotadiya, P., Makwana, N., Patel, S.: Curbing variations in packaging process through six sigma way in a large-scale food-processing industry. J. Ind. Eng. Int. (2015)
5. Dadashnejad, A.A., Valmohammadi, C.: Investigating the effect of value stream mapping on operational losses: a case study. J. Eng. Des. Technol. **16**(3), 478–500 (2018)
6. Rao, G.V.P., Nallusamy, S., Rajaram Narayanan, M.: Augmentation of production level using different lean approaches in medium scale manufacturing industries. Int. J. Mech. Eng. Technol. **8**(12), 360–372 (2017)
7. Scheller, A.C., Sousa-Zomer, T.T., Cauchick-Miguel, P.A.: Lean Six Sigma in developing countries: evidence from a large Brazilian manufacturing firm. Int. J. Lean Six Sigma (2018)
8. Punna Rao, G.V., Nallusamy, S.: Execution of lean and ZF production system in manufacturing line for productivity enrichment—a case study. Int. J. Eng. Res. Africa **40**, 171–183 (2018)
9. Vinod, M., Devadasan, S.R., Sunil, D.T.V., Thilak, M.M.: Six Sigma through Poka-Yoke: a navigation through literature arena. Int. J. Adv. Manuf. Technol. **81**(1–4), 315–327 (2015)
10. Rigolin Ferreira Lopes, M.E., Quartucci Forster, C.H. Application of human error theories for the process improvement of requirements engineering. Inf. Sci. **250**, 142–161 (2016)
11. Ahmad, A.A., Abd, A., Freddawati, R., Wong, R., Iqbal, M.: Worker safety improvement at paper pleating production line using Poka-Yoke concept—a case study in automotive industry. J. Mech. Eng. Vol SI **4**(5), 183–196 (2017)
12. Kumar, S., Dhingra, A.K., Singh, B.: Process improvement through Lean-Kaizen using value stream map: a case study in India. Int. J. Adv. Manuf. Technol. **96**(5–8), 2687–2698 (2018)
13. Mor, R.S., Bhardwaj A., Singh, S., Sachdeva, A.: Productivity gains through standardization-of-work in a manufacturing company. J. Manuf. Technol. Manag. (2018)
14. Wilson, R., Lavarack, B. The development of standard operating procedures for boilers at Mackay Sugar Limited. Int. Sugar J. **120** (2018)
15. Kang, N., Zhao, C., Li, J., Horst, J.A.: A hierarchical structure of key performance indicators for operation management and continuous improvement in production systems, Int. J. Prod. Res. **7543**(February) (2016)
16. Miyagawa, Y., Ogawa, T., Adachi, S.: An accelerated method for estimating the dispersion stability of mayonnaise by oil–water separation under reduced pressure. JAOCS, J. Am. Oil Chem. Soc. 95(11), 1367–1372 (2018)

Production Model Under Lean Manufacturing and Change Awareness Approaches to Reduce Order Delays at Small and Medium-Sized Enterprises from the Clothing Sector in Peru

Sheyene León-Guizado, Anthony Castro-Hucharo, Pedro Chavez-Soriano, and Carlos Raymundo

Abstract This study proposes a production model that increases the manufacturing capacity in a small and medium-sized enterprise (SME) of garments with an aim to reduce the nonfulfillment of order deliveries. An assessment has been done and waiting times between production processes have been identified, along with defective products and inefficient work methods. This study proposes the design of a lean manufacturing model under the change management approach, whose methodology comprises five phases. In phase 0, awareness and training sessions are conducted (change management). Then, phase 1 reorganizes the work area (plant layout redistribution and 5S) and phase 2 seeks better workload balances (line balance and Heijunka implementation). Later, phase 3 standardizes work methods (standardization). Finally, the proposed model will be validated to determine whether the selected operating tools are supported by the awareness that contributes to increasing production.

Keywords Production model · Lean manufacturing · Change management · Efficiency · Awareness

S. León-Guizado · A. Castro-Hucharo · P. Chavez-Soriano · C. Raymundo (✉)
Universidad Peruana de Ciencias Aplicadas, Lima 15023, Peru
e-mail: carlos.raymundo@upc.edu.pe

S. León-Guizado
e-mail: u201223444@upc.edu.pe

A. Castro-Hucharo
e-mail: u201212779@upc.edu.pe

P. Chavez-Soriano
e-mail: pedro.chavez@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_36

1 Introduction

Globally, the Asian and Central American countries, which are considered as the main competitors of Peru in the textile sector, have strengthened their market positions owing to their lower labor costs. Another key aspect is that China is currently the main garment supplier in Peru, representing 70.3% of the total incoming imports for the country, because of high nonwage labor costs and labor market rigidities, which depend mainly on the workforce, as suggested by ADEX [1]. One of the main constraints is the several tariffs that the domestic producers are required to pay to import their raw materials; these tariffs are not properly returned (drawback). However, incoming of finished products from China is exempted from tariffs. This situation hinders the growth of the Peruvian textile industry and has become an area of concern for small and medium-sized enterprises (SMEs); in addition to tax burdens, SMEs also face significant labor and environmental cost overruns. The latter takes place due to subsidies, implementation of promotional mechanisms, more flexible labor regimes, and the creation of free trade zones. INFOTRADE [2] reports on the main products exported by the textile and clothing sector of the country have revealed a decrease in exports in comparison to that in 2012.

The basic economic pillars of Peru focus on the mining, agribusiness, and textile sector [3], thereby increasing the country's international presence and global competitiveness. However, these sectors are still weak from the viewpoint of productivity. According to the Ministry of Production (PRODUCE) [4]. Hence, it can be concluded that Peruvian clothing companies must take action to successfully address the decrease in export and prevent them from affecting their economic situation. On the contrary, the main export destination for Peruvian clothing is China, which currently represents 22% of all clothing-related exports [5].

It should also be noted that Peru has lost export presence globally. However, there are indications of a slight economic recovery. Starting at the end of 2017 and to 2018, Peruvian clothing exports approximately totaled US$435 million, reporting an increase of 11.9% during these months, as well as the USA and Ecuador as their main export destinations [6].

2 State of the Art

2.1 Production Model in SMEs in the Manufacturing Sector

SMEs play an important role in developing the economy due to the large percentage of employment they generate and GDP they represent. Nevertheless, the production models often used SMEs do not consider innovation, quality, or learning changes, which places them at a great disadvantage against large companies. Within this context, the incorporation of new technologies will significantly help in overcoming these issues. In fact, all SMEs would benefit from implementing quality-oriented management changes and innovations as central strategies [7–11].

2.2 Lean Manufacturing Tools in SMEs Operating Within the Manufacturing Sector

According to the literature, SMEs should implement lean manufacturing techniques to successfully deal with problems such as low production capacity, nonfulfillment of order deliveries, a high percentage of defective products because these tools may help them increase their productivity and sustainability. For these purposes, the manufacturing industry recommends using the 5S methodology, which is a lean tool focused on improving operative performance. In fact, worldwide manufacturing studies reveal positive results regarding its contribution to the improvement in quality, production, safety and work environment processes, yielding fast returns at low cost [12–14].

Since a quick response capacity is always required, lean practices seek to meet high quality and service demands from customers.

3 Contribution

3.1 Proposed Model

The model has been structured in three blocks according to the PDCA cycle (Fig. 1). The upper block includes variables grouped under the plan. The main process flow, which allows production to take place, is shown along the central part in the do block, covering the different activities that production management must perform to generate the product. The third block corresponds to the check/act actions, which include production management monitoring and assessment against the objectives and plans, and corrective and preventive actions.

The proposed method aims to reduce order delays for SMEs, reducing bottlenecks in cycle times and thereby increasing production capacity. When compared against other models, this model includes an initial change awareness stage to avoid any lean tool implementation issues that may arise. As can be seen in the model, production management begins and ends with the customer, represented by marketing and sales management. Therefore, the customer is displayed both on the right and left sides of the chart. Next, the chart details the requirements of the customer. This becomes the input for the production management system (represented in the diagram inside a rectangle). This input enters through the plan top block. The requirements from the customer form one of the foundations both for product design and development and for planning the supply of demand.

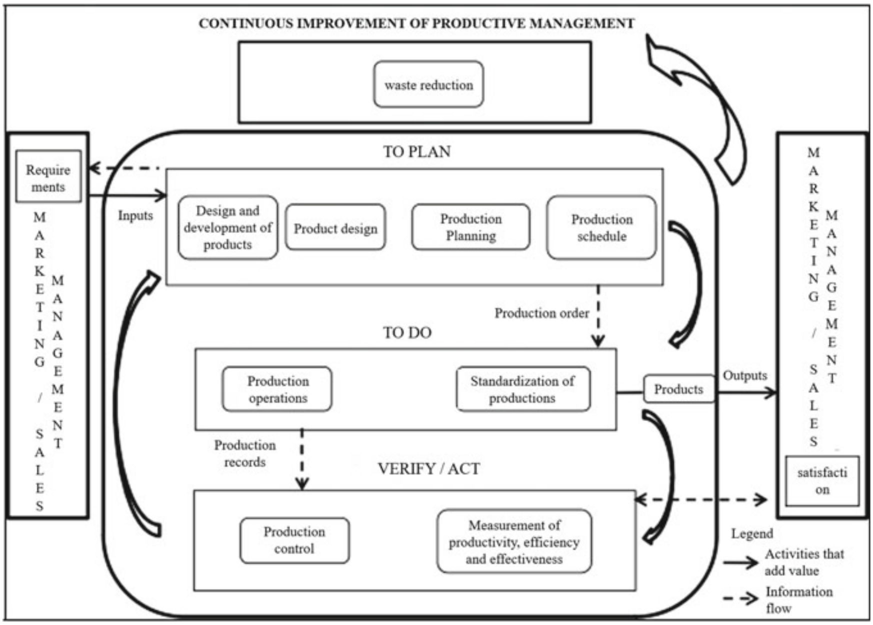


Fig. 1 Production management model

The production management process starts in the plan block. The first step includes product design and development that seeks to obtain simple and robust designs, wherein the most appropriate parameters are optimized and controlled, and the highest quality is achieved at minimum cost (Fig. 2).

Planning Process: Absence of new product designs. Plant layout distribution modifications have been implemented without considering the process flow. There is an obvious disregard for production scheduling (last minute “urgent” orders are introduced). In addition, there is no effective communication between production management and marketing/sales management; therefore, the wrong products are often manufactured or the company receives complaints from the customer because specifications were not followed as requested.

Production Process: Clear presence of bottlenecks. The company keeps a high inventory of in-process products to cover “emergencies.” There is little automatic inspection. In addition, no operation standards have been defined; therefore, operations are conducted based on experience or routine (operators work as they consider suitable). There is a significant number of delayed production orders. There is no task standardization.

Control Process: There are no idle time records and process control systems are virtually nonexistent. SMEs are focused on assessing performance based on units produced per each hour worked (a paradigm based on the notion that greater production leads to greater profitability). This approach to productivity implies that all management efforts are aimed exclusively at meeting production goals.

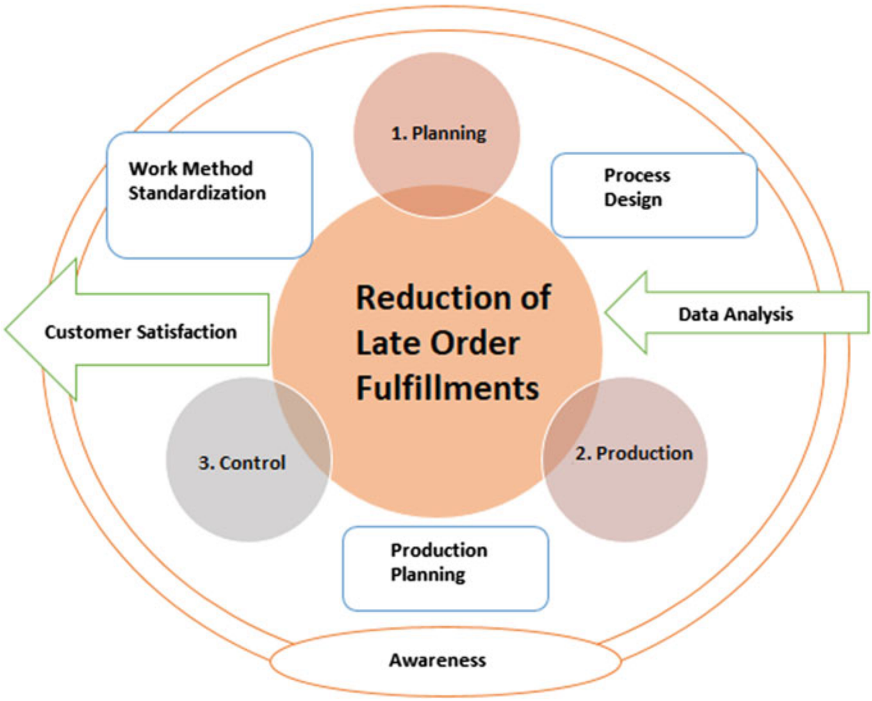


Fig. 2 Proposed model

Continuous Improvement: There is no motivation for making improvements to the production process. No awareness of analysis and improvement tools. In general, the concept of waste is not considered.

3.2 Proposed Method

To ensure guarantee a successful and ideal method implementation, a step-by-step implementation guide was prepared (Fig. 3).

3.3 Indicators

Using these formulas, a proper performance analysis will be conducted to assess the effectiveness and impact of the design described above so that the sector companies may determine the efficiency of the implementation of this design.

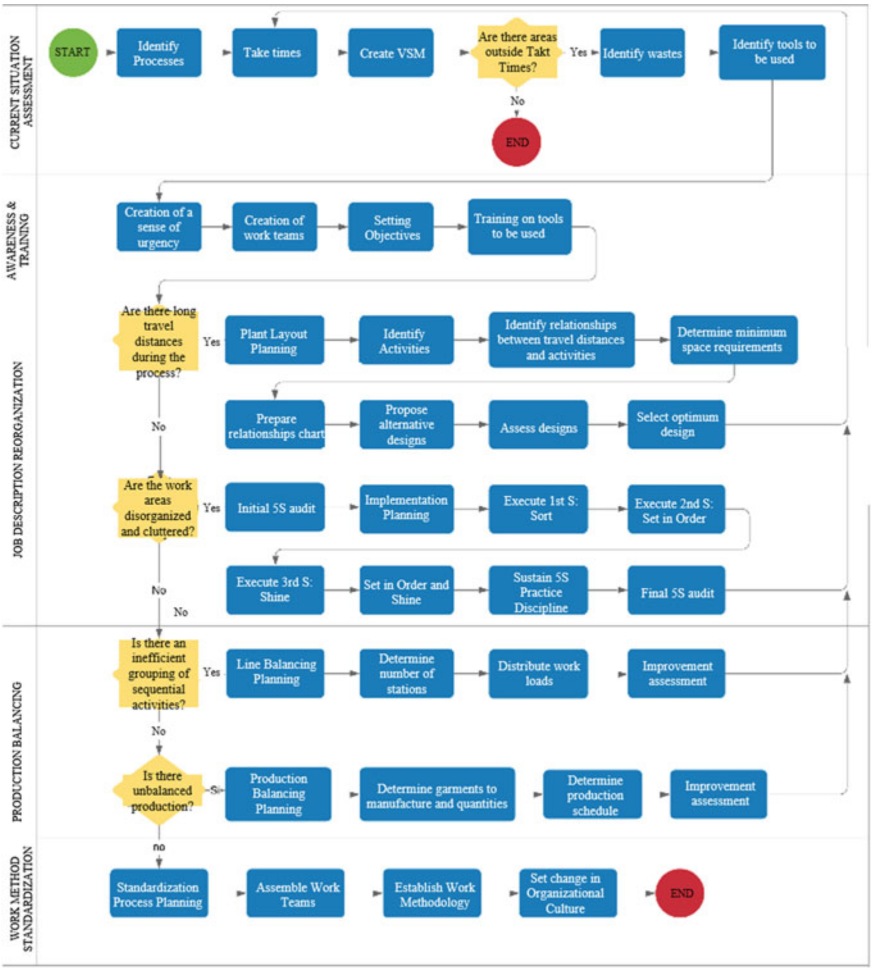


Fig. 3 Proposed method

Attendance percentage. This indicator measures meeting and training session attendance percentages.

$$\frac{\text{Number of Sessions Attended}}{\text{Total Number of Sessions}} \tag{1}$$

Commitment compliance percentage. Percentage of orders completely fulfilled over the scheduled time.

$$\frac{\text{Scheduled Time}}{\text{Actual Average Time}} \tag{2}$$

Percentage of On-Time Deliveries. Percentage of orders delivered on time over total deliveries.

$$\frac{\text{Number of Orders Fulfilled on Time}}{\text{Total Number of Scheduled Orders}} \tag{3}$$

Efficiency. This indicator shows cycle time efficiency in terms of costs.

$$\frac{\text{Results Reached}}{\text{Total Costs}} \times \text{Invested Time} \tag{4}$$

4 Validation

To validation of the study regarding the lean manufacturing model, it was possible due to granted the appropriate permissions from responsible authorities of the Peruvian textile SME.

The case study focuses on a Peruvian textile SME that manufactures baby garments, such as onesies, bodysuits, polo shirts, and pants. To better understand the business and evaluate sales behavior, this assessment focused on the product which records the highest demand and, in turn, represents the highest revenue for the company. The bestselling product is one-piece rompers or onesies.

The implementation is conducted in three phases: job description reorganization, a grouping of sequential activities and production balancing, and work method standardization. In addition, there is a phase 0, i.e., awareness and training.

4.1 PHASE 0: Awareness and Training

This phase is common to all phases since change resistance needs to be reduced and employees need to get involved with the project to successfully implement any changes or establish new things within the organization.

For these purposes, the following five application steps were followed: (1) Needs or Training Requirements Detection, (2) Training Program Design, (3) Resources, (4) Implementation, (5) Final Evaluation.

4.2 PHASE 1: Job Description Reorganization

There are some engineering tools that help us reduce high displacements and movements that do not add value to the product.

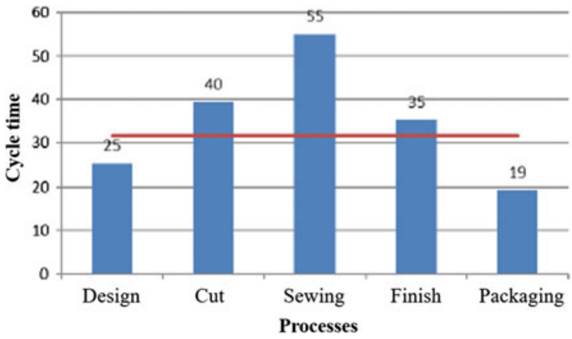


Fig. 4 Initial takt time

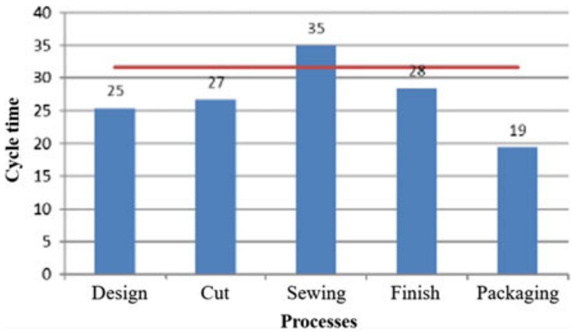


Fig. 5 Takt times after phase 1

Next, the study seeks to detail the development, execution, and monitoring of improvement activities, according to the scheduled dates, with evidence and indicators assessed throughout the process. Current processes were evaluated to identify which areas are above takt times. Then, improvement opportunities were proposed and training sessions on plant layout distribution and 5S implementation were provided (Fig. 4). Based on the research project features, classification, order, cleaning, standardization, and discipline activities were implemented, which reduced takt times in the cutting and finishing areas (Fig. 5).

4.3 PHASE 2: Grouping of Sequential Activities and Production Balancing

In this phase, production line workloads were assessed. Using the Arena software, the authors conducted a simulation of the proposed production model to balance production lines and obtain a more efficient production. In this manner, this study demonstrated that manufacturing area times may be reduced after the implementation of the proposed model (Fig. 6).

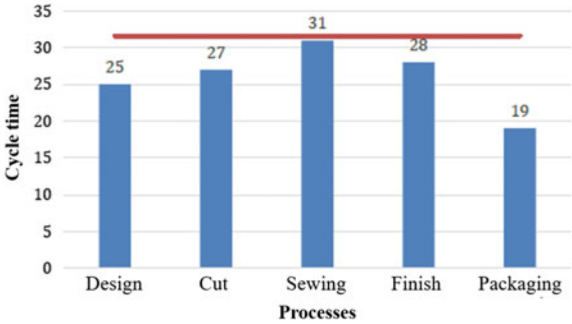


Fig. 6 Takt times after phase 2

4.4 PHASE 3: Work Method Standardization

This phase seeks to sustain the implementation developed as production variability can be reduced through work standardization. For these purposes, work instructions were developed for different process areas. These forms will be implemented during the manufacturing process of the sample product.

5 Conclusions

Two main causes were identified for nonfulfillment of orders delivery, namely high manufacturing times and a high number of reprocesses. However, after implementing the proposed model, 86% of the problems were solved. The resistance to change exhibited by operators is a critical factor that can affect the viability of the model since it prevents 100% of project completion. The proposed model reduced operation times and determined the number of operators required at each workstation to reduce cycle times.

References

1. Ministerio de Producción – PRODUCE: (2018). https://www.gob.pe/produce
2. Sociedad Nacional de Industrias-SNI: Estadísticas de Exportación del Sector Textil y Confecciones (2017). http://comitetextilperu.com/BoletinTextil/docs/expor.pdf
3. Asociación de exportadores – ADEX: ADEX Data Trade (2018). http://www.adexdatatrade.com
4. Sociedad nacional de industrias – SIN: (2018). http://www.sni.org.pe/reportes-economicos/
5. The observatory of economic complexity – OEC: (2018). https://atlas.media.mit.edu/es/
6. Proinversion–Agenciadelapromociónprivada Perú: Estadísticas Comerciales -SECTOR EXPORTADOR (2018)

7. Bourke, J., Roper, S.: Innovation, quality management and learning: short-term and longer-term effects. Res. Policy **46**(8), 1505–1518 (2017)
8. MacBryde, J., Paton, S., Clegg, B.: Understanding high-value manufacturing in Scottish SMEs. Int. J. Oper. Prod. Manag. **33**(11/12), 1579–1598 (2013)
9. Ibarra, M., González, L., Demuner, M.: Competitividad empresarial de las pequeñas y medianas empresas manufactureras de Baja California. Estudios fronterizos **18**(35), 107–130 (2017)
10. Womack, J.P., Jones, D.T.: Lean Thinking, Primera edn. Grupo Planeta, Barcelona, España (2012)
11. Sayid, A., Nure, A., Uddin, K.: Court shoe production line: improvement of process cycle efficiency by using lean tools. Revista de Pielărie Încălţăminte **17**, 3 (2017)
12. Moeuf, A., Tamayo, S., Lamouri, S. R., Pellerin, A.: Strengths and weaknesses of small and medium sized enterprises regarding the implementation of lean manufacturing (2016)
13. Duque, J.L.S.: Los procesos de cambio organizacional y la generación de valor. Estudios Gerenciales **30**(131), 162–171 (2014)
14. Xuechang, Z., Yu, L.: Does lean manufacturing improve firm value? (2017)

A Strategic Lean Procurement Model Based on Supplier Approval to Reduce Unplanned Downtime in a Textile Small and Medium-Sized Enterprises

Maricarmen Ramirez-Mitma, Jose Rojas-Garcia, Carlos Torres-Sifuentes, and Carlos Raymundo

Abstract In Peru, companies within the textile sector encounter problems such as a decrease in the contribution to an industrial gross domestic product from 10.6 to 7.5%, a decline in financial credits by about 0.7% p.a., and the direct competition of Central American countries across the textile chain. The finishing of fabrics shows the worst performance, with losses of a 38.8% variation in industrial production and a 7.5% growth rate per textile industrial sector. In this scenario, a strategic lean procurement model based on 5S and supplier approval was designed. The lean objective is to streamline workflow in the raw material warehouse that generates excessive lead time to production and leads to unplanned downtime. In addition, supplier approval boosts procurement efficiency and provides strategic value in assessing and planning raw materials to ensure efficient supply. The model is validated through a case study involving a textile small and medium-sized enterprise (SME), and drastic improvements are made. These improvements address the problem of unplanned downtime by reducing its incidence and making economic contributions to the company.

Keywords Supply chain management · Supplier · Kraljic · Lean

M. Ramirez-Mitma · J. Rojas-Garcia · C. Torres-Sifuentes · C. Raymundo (✉)
Universidad Peruana de Ciencias Aplicadas, Lima 15023, Peru
e-mail: carlos.raymundo@upc.edu.pe

M. Ramirez-Mitma
e-mail: u201514680@upc.edu.pe

J. Rojas-Garcia
e-mail: jose.rojas@upc.pe

C. Torres-Sifuentes
e-mail: carlos.torres@upc.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_37

1 Introduction

Across the world, textile and apparel industries have a high turnover rate and low industrial margins, and their competitiveness depends on reliable logistics, transportation, and delivery services [1]. Textile companies in Peru face a major problem in raw material supply because they are forced to import most resources, thereby raising logistics costs [2].

To support the competitiveness of Peruvian textile companies, the Peruvian government has implemented several customs systems to improve commercial promotion in new export destinations [2]. The first successful case is related to a standardized procurement system of textile exporters, where just-in-time manufacturing and procurement standardization processes are used. As a result, the system reduced five types of waste and expanded exports by 15% [3]. The second case is the application of the six sigma methodology to the internal processes of a clothing manufacturing company, which delivered 50% of the savings of former losses [4]. According to Jorge Luis Vidal, dry cleaners' success is based on the use of management skills to control the textile chain because the flaws that were not fixed in previous stages are exposed in this phase of the process [5].

The purpose of this study is to reduce textile SMEs' unplanned downtime causing economic losses, which could be used to foster development and growth. Faced with this problem, a strategic lean procurement model is introduced based on supplier approval and the adoption of 5S in warehouses.

2 State of the Art

2.1 Supplier Approval in Procurement Models in the Textile Sector

After reviewing the literature concerning the first research topic on supplier approval, we found that the problem is addressed through different solution approaches. Considering the evaluation and performance of textile suppliers, several authors have proposed multi-criteria decision-making models based on TOPSIS, Best Worst Method (BWM), the Fuzzy Interference System [6–8]. Considering the flexibility required for supplier approval in procurement areas, some authors determine that the implementation of fuzzy tools streamline the decision-making process [6, 8]. However, the BWM adopts a more structured method for evaluating and comparing sets to produce more reliable results [7]. Moreover, another author notes that a two-dimensional or Kraljic matrix based on uncertain criteria produces a model to solve the problems involved in sorting strategic suppliers [9].

2.2 *Application of the Lean Approach to Procurement Models in the Textile Industry*

We analyzed scientific papers concerning the lean approach and found that the goals are similar. Considering warehouse efficiency and textile supply, some authors propose 5S, hybrid 5S, or Kaizen [10, 11]. Another author proposes value stream mapping for textile management in the areas of production and planning; however, the leaner approach in this sector will have both positive and negative effects because it will reduce lead time and waste but will cause stress to employees [4]. In principle, these concepts reduced waste and costs [4, 10, 11]. Based on the feedback for a successful model, one author proposes that the model be aligned to the Deming cycle (PDCA) to share knowledge with the involved parties and thereby enhance procurement performance autonomously [12].

3 Contribution

3.1 *Overview*

Model proposed. The model proposed is shown in Fig. 1 and is developed from the state of the art analysis to propose a solution model based on the problems faced by the textile sector and their distinctive features.

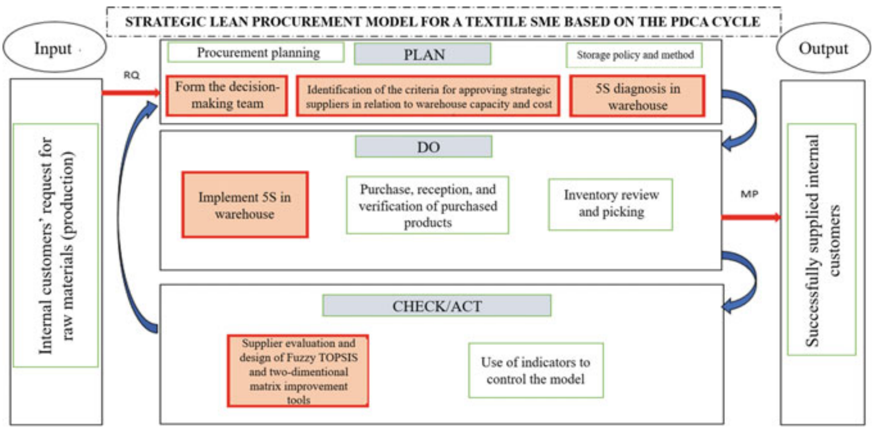


Fig. 1 Strategic lean procurement model based on supplier approval

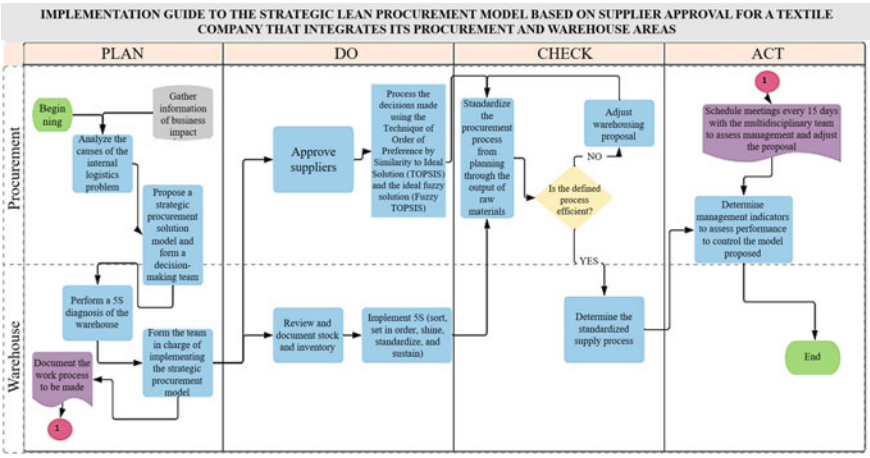


Fig. 2 Implementation guide to the model proposed

3.2 Process View

Figure 2 shows the implementation of the model proposed, which may be applied to any case study that meets the distinctive features mentioned in the introduction.

3.3 Indicators View

To control the functionality of the model in any case study, the indicators should examine the description, periodicity, area in charge, impact, evaluation limit (traffic light), and calculation method and scale so that any employee may obtain efficient information.

The indicators proposed will have a direct impact on the decrease in unplanned downtime caused by raw material supply (Figs. 3 and 4).

Description	Control the changes in the consumption of raw material purchased with respect to sales					
Frequency	Monthly			Area	Procurement	
Impact	Negotiation with suppliers				Target	35%
Limit	Critical	45% or more	Precaution	Between 35% and 45%	Ideal	35% or less
Indicator calculation method						
Calculation formula	$\frac{\text{Value of the purchases of raw materials}}{\text{Total sales}} \times 100$				Scale	Percentage

Fig. 3 Indicator No. 1: volume of purchased raw materials

<table>
<tr><td>Description</td><td colspan="6">Inspect the quality and timeliness of the purchases and services generated by suppliers</td></tr>
<tr><td>Frequency</td><td colspan="3">Every 15 days</td><td>Area</td><td colspan="2">Procurement</td></tr>
<tr><td>Impact</td><td colspan="4">Effort made by procurement to solve problems, losses of sales, and unplanned downtime due to supply shortage</td><td>Target</td><td>80%</td></tr>
<tr><td>Limit</td><td>Critical</td><td>60% or more</td><td>Precaution</td><td>Between 60% and 80%</td><td>Ideal</td><td>80% or less</td></tr>
<tr><td colspan="7">Indicator calculation method</td></tr>
<tr><td>Calculation formula</td><td colspan="4">$\frac{\text{Purchase orders successfully issued}}{\text{Total purchase orders issued}} \times 100$</td><td>Scale</td><td>Percentage</td></tr>
</table>

Fig. 4 Indicator No. 2: quality of the purchase orders issued

4 Validation

4.1 Case Study

To validation of the study regarding the Lean Procurement, it was possible due to granted the appropriate permissions from responsible authorities of the Peruvian company Corporación Textil del Sur (COTEXSUR).

The Peruvian company Corporación Textil del Sur (COTEXSUR) has successfully operated in the textile industry for over a decade. This company offers textile dyeing and finishing services at national and international levels. It also has cutting-edge machinery and a laboratory to develop laboratory dips to meet each customer's specific needs and quality in agreement with the colors approved.

4.2 Initial Diagnosis

The model is implemented at the textile company Corporación Textil del Sur S.A.C, and the areas involved in the validation process are logistics-procurement and warehouses. The initial symptom of production downtime is shown in the 2018 Pareto chart (Fig. 5). The lack of raw material loading and reprocessing of raw materials in poor condition constitute unplanned downtime.

We then diagnosed the areas involved in the main causes of unplanned downtime and determined, based on the indicators included in the model proposal, that the volume of purchased raw materials stood at 39%, the quality of purchase orders issued amounted to 77%, downtime for stock-out totaled 32 min, and inventory turnover stood at 3.15%.

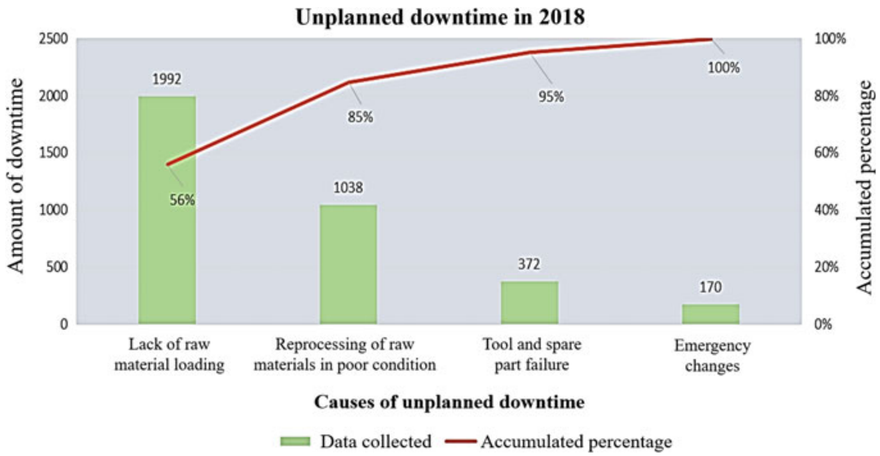


Fig. 5 Pareto chart for downtime

4.3 Application of the Model Proposed

4.3.1 Plan Phase

Decision-making team. The decision-making team is formed to define the supplier's approval criteria.

Identify supplier approval criteria. The criteria are identified based on predetermined Fuzzy TOPSIS evaluation criteria.

Procurement planning. Procurement is planned based on the meeting with the Warehouse Manager and consideration of three types of products: frequent (weekly), low turnover (bimonthly), and inactive (when there is no more stock). A table including all chemical products is compiled and will be filled based on the usage frequency of the type of product.

5S diagnosis in the warehouse. According to the evaluation made of warehouse areas, we considered implementing the 5S methodology to reduce the downtime and lead time caused by the shortage of supplies.

4.3.2 Do Phase

5S implementation (Figs. 6, 7 and 8) (Table 1).

- Seiri (sort)
- Seiton (straighten)
- Seiso (shine)
- Seiketsu (Standardize).

Fig. 6 Before and after Seiri photos

Fig. 7 Before and after Seiton photos

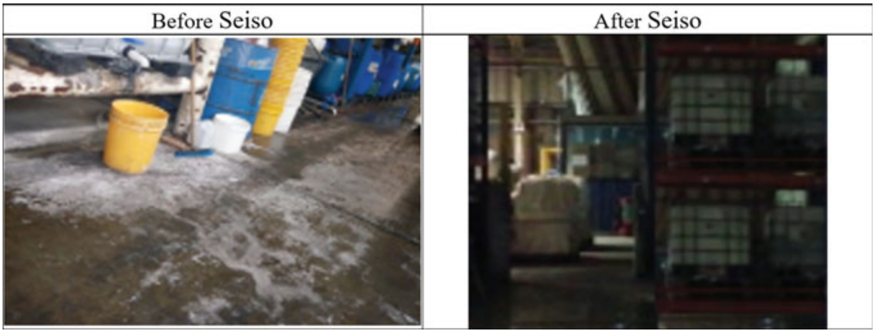


Fig. 8 Before and after Seiso photos

Table 1 The self-discipline of the 5S proposed

Self-discipline
• Perform 5-minute shine tasks per work shift (day and night)
• Assess the 5S audit on a weekly basis, and once a six-point score is reached, perform it on a monthly basis
• Return the type of raw material used to the related shelf based on whether it is a chemical input, dye, or industrial salt
• Provide evidence to show the improvements achieved

The initial score obtained in the audit on the status before 5S was 20 points, a critical score. After the implementation, the audit scored 49, a significant improvement in the area.

- Shitsuke (Sustain).

Check-act phase

The Fuzzy TOPSIS [13] arithmetic method is adopted according to the criteria defined during the planning phase. Standardization, distance, and ranking are calculated for both cost and delivery models (Fig. 9). These coordinates are subsequently expressed in a two-dimensional matrix explained by an author [13].

This matrix is considered the action plan in supply and procurement policies (Table 2).

We then assessed the impact of the improvement in the indicators calculated concerning unplanned downtime (Table 3).

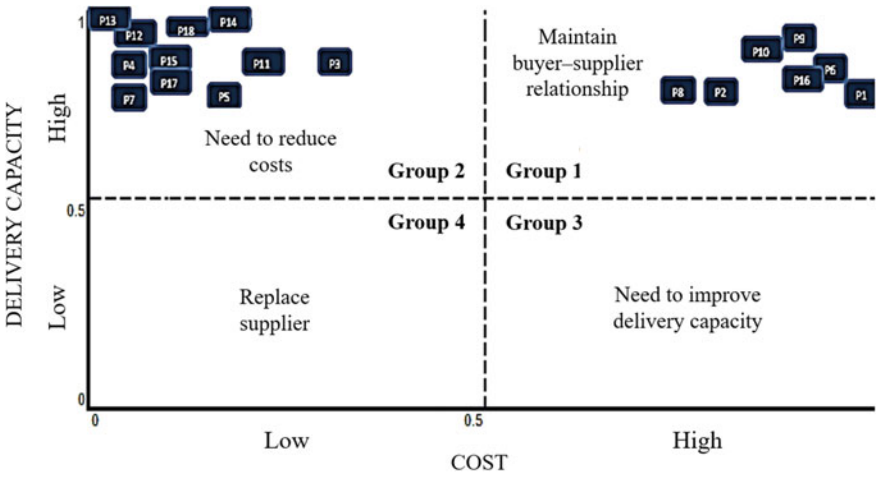


Fig. 9 COTEXSUR cost and delivery two-dimensional matrix

Table 2 Model indicators before and after implementation

Indicators	Results before implementation	Results after implementation
The volume of purchased raw materials (%)	39	34
Number of purchase orders issued (%)	77	80
Downtime caused by stock-out (min)	32	20
Raw materials turnover (times)	3.15	4

Table 3 Impact on unplanned downtime

Month	# monthly cases of downtime before implementation 2018	Month	# monthly cases of downtime after implementation 2019
August-18	329	August-19	320
September-18	342	September-19	308
October-18	397	October-19	293

5 Conclusions

Productivity in logistics areas was improved because these areas were the top issues that caused 80% of downtime according to the Pareto chart. Lead time for raw material was reduced from 32 to 20 min owing to the efforts made through the implementation of 5S and the agreed-upon warehousing processes. The decline in downtime for the first, second, and third months stood at 9, 12, and 15, respectively. This paper encouraged innovation in supplier management because a Fuzzy TOPSIS tool is used to replace obsolete forms that assessed two or three ambiguous criteria applied by the company studied.

References

1. Oré y Ramos: Propuesta de mejora en el proceso de compras de las pymes exportadoras del sector textil de confecciones de prendas de vestir de tejido de punto de algodón, aplicando herramientas de Lean Manufacturing. (2018) https://repositorioacademico.upc.edu.pe/handle/10757/62512215
2. DEMI: Estudio de investigación del sector textil y confecciones (2015)
3. Iturrizaga: Propuesta de mejora para el proceso de abastecimiento interno y producción en una empresa que confecciona prendas de vestir (2015). https://repositorioacademico.upc.edu.pe/bitstream/handle/10757/625166/Iturrizaga_VJ.pdf?sequence=4&isAllowed=y16
4. Hasan, M.M., Shohag, M.A.S, Azeem, A., Paul, S.K.: Selección de proveedores de criterios múltiples: un enfoque difuso. Revista Internacional de Sistemas y Gestión Logística (2015)
5. APTT: La clave del éxito en la gestión de la tintorería (2018)
6. Lima-Junior, F.R., Carpinetti, L.C.R.: Combining SCOR® model and fuzzy TOPSIS for supplier evaluation and management. Int. J. Prod. Econ. (2016)
7. Manfredsson, P.: Gestión textil habilitada por el pensamiento lean: un estudio de caso de pymes textiles. Planificación y control de la producción (2016)
8. Rezaei, J., Nispeling, T., Sarkis, J., Tavasszy, L.: Un enfoque de ciclo de vida de selección de proveedores que integra criterios tradicionales y ambientales utilizando el mejor y peor método. Journal of Cleaner Production (2016)
9. Garzón, F.S., Enjolras, M., Camargo, M., Morel, L.: Una metodología de adquisición ecológica basada en Kraljic Matrix para la evaluación y selección de proveedores: un estudio de caso del sector químico. Supply Chain Forum: An Int. J. 1–17 (2019). https://doi.org/10.1080/16258312.2019.1622446
10. Kumar, S., Dhingra, A. K., Singh, B.: Process improvement through Lean-Kaizen using value stream map: a case study in India (2018)
11. Neyra, J., Muñoz, J., Eyzaguirre, J.: 5S Hybrid management model for increasing productivity in a textile company in Lima. Hum. Interaction Emerg. Technol. (2019)

12. Garzón, FS, Enjolras, M., Camargo, M., Morel, L.: Una metodología de adquisición ecológica basada en Kraljic Matrix para la evaluación y selección de proveedores: un estudio de caso del sector químico. Supply Chain Forum: Int. J. 1–17 (2019). https://doi.org/10.1080/16258312.2019.1622446
13. Amindoust, A., Saghafinia, A.: Selección de proveedores de textiles en la cadena de suministro sostenible utilizando un modelo modular de sistema de inferencia difusa. J. Text. Inst. (2016)

Digital Transformation Model with a Focus on Total Quality Management and Lean Manufacturing to Increase Online Sales in Textile SMEs

Erika Briones-Castañeda, Gianmarco Carlos-Ramon, Carlos Torres-Sifuentes, Jose Rojas-García, and Carlos Raymundo-Ibañez

Abstract This work proposes a model that seeks to increase small and medium enterprises (SMEs) productivity through digital transformation (DT), using its three fundamental pillars, beginning with a change in the organization's mindset. This process alters the manner in which business is conducted, from production to delivery of products and services, to offer a better customer experience, which changes through the use of technology. The business model may change as a result of the application, which aims to increase the reach through data collection, thus getting to know the customer base and utilizing the trends generated by the information analyzed. This model consists of three stages. In the first stage, the inventory stocks are made known; then, the demand is observed; and a database is created or added, depending on the case. The second stage comprises storage of customer information, order confirmation, online sale, order fulfillment, stock update, and distribution. In the third stage, the product is delivered, and customer satisfaction is assessed. The database of this process allowed the stock control time to decrease by 91%, likewise, the time of income of products to the inventory is reduced by 93%.

Keywords Total quality management · Lean · E-commerce · Digital transformation · Textile · SMEs

E. Briones-Castañeda · G. Carlos-Ramon · C. Torres-Sifuentes · J. Rojas-García · C. Raymundo-Ibañez (✉)
Universidad Peruana de Ciencias Aplicadas (UPC), Lima 15023, Peru
e-mail: carlos.raymundo@upc.edu.pe

E. Briones-Castañeda
e-mail: u201416367@upc.edu.pe

G. Carlos-Ramon
e-mail: u201212189@upc.edu.pe

C. Torres-Sifuentes
e-mail: carlos.torres@upc.edu.pe

J. Rojas-García
e-mail: jose.rojas@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_38

1 Introduction

Since 2012, e-commerce sales have tripled in the Latin American region, bringing in about 40 billion dollars. Moreover, Perú shows a higher growth compared to other countries in the same region. According to a consultancy, this growth could annually reach 30.2%. Furthermore, it is known that 51.7% of Peruvians frequently use the Internet, which drives companies to consider new alternatives to increase their sales.

In the last three years, companies that belong to the traditional commerce sector have evolved to adapt to new market trends, thereby managing to launch e-commerce. According to the National Institute of Statistics and Computer Science, around 1000 companies join new forms of marketing per year, while about 70,000 were created only in the last quarter of 2018, i.e., a small percentage is present in e-commerce.

The ways in which consumers purchase have changed, and now, around 6 million users prefer to make purchases online, whether due to accessibility, speed, or convenience. According to the Lima Chamber of Commerce (CCL), they represent 31% of the total number of potential customers. However, a problem prevails within the remaining 69%, as they exhibit distrust while making online purchases; such purchases do not materialize on many occasions due to the fear immersed in customers relating to product compliance, lack of communication, distrust about the authenticity of the product data, etc., in addition to lack of knowledge about ways to use an online payment method [1]. Hence, new methodologies that resolve these problems must be implemented. According to the CCL, Peru is experiencing a period of growth of 30% of the national market; the country's e-commerce is likely to grow at a constant value with an annual growth rate of 12% until 2020 [2].

SMEs are challenged by resource constraints and lack of guidance on the benefits realization of digital transformation (DT), which is important, as digital innovation based on data is employed to improve existing processes and change the business model or reinvent it [3]. The lack of emphasis on digitalization in Peru is concerning, as, in several of the most digitalized European countries, 33% of the inhabitants are somewhat resistant to change due to lack of knowledge [4]. In Peru, companies establish 0.7% of their income in DT, which directly seems to affect productivity [5].

Given the loss of customers and the decline in sales, the aim is to capitalize on these tools through a model based on DT to increase productivity and achieve greater economic benefits, along with total quality management (TQM) and lean methodology to support management to facilitate better services and to optimally utilize marketing channels.

2 State of the Art

2.1 Digital Transformation in SMEs

In the modern digital economy, SMEs are constantly being challenged by digital enterprises that continue to alter the manner in which business is conducted [6]. Hence, there prevails a need for a structured framework and a tool to support the SMEs to adjust their organizational strategies to environmental changes in the digital age. As SMEs, particularly, face several challenges in transforming their business, a structured and easy-to-use approach needs to be developed that considers the challenges and opportunities of an SME such as limited resources or high flexibility, enabling a higher potential for success [7]. As a result of the implementation, the tools were extended to include all the relevant business aspects of a company, such as marketing, processes, change management, and innovation; this enabled an innovative thinking approach, along with a focus on the impacts of the digital era [8]. In turn, an approach involving the underlying one was developed because the complexity of DT projects becomes manageable when using action elements that are understood and guided by support units adapted to SMEs, as it begins with the basics, builds standardized, practical tools for SMEs in a specific industry, incorporates participatory formats for a better scale, and considers a neutral position when selecting a solution [9]. Conversely, another case study proposes a service-oriented portal dimension, which consists of portal maintenance service, B2B function, and cloud computing, which significantly influence organizational performance and company productivity [10].

2.2 Total Quality Management

The concept of total quality has been studied and applied in different companies to improve quality performance. TQM's goal is to achieve superior quality through the participation of all organizational areas. A 2019 study indicates that TQM practices can positively improve competitive performance, as well as the quality, cost, and delivery performance of a product [11]. According to many researchers, TQM generates value through a variety of benefits. In this way, it is possible to achieve a better understanding of customer needs, enhance problem-solving, improve customer satisfaction, facilitate better internal communication, develop a stronger relationship with suppliers, generate less waste, facilitate greater employee commitment and motivation, and make fewer errors [12].

The success of TQM depends not only on the adoption of its attributes but also on complementary factors that are apparently not related to TQM; thus, it can be deduced that it requires a culture that is receptive to change, leadership qualities, and motivation to make further improvements; it is difficult to imitate complementary resources without these intangible factors [13]. The core of TQM's ability to create value is its power to efficiently utilize specific valuable knowledge at all organizational levels. This contribution to knowledge creation leads to improvements in some distinctive company skills such as management skills, employee knowledge, external cooperation skills, development of collective consciousness, organizational commitment, stimulation of the organizational learning process, and speed and flexibility in the new product development [14].

Previous research shows that the choice of human resource management systems can have an economically significant impact on a company's performance. Sensible changes in a human resource management system can affect the likelihood of a company's survival and may have intermediate results (impacting productivity, profits, product quality, sales, customer satisfaction, etc.). In summary, human resources can make a significant difference to a company's bottom line.

3 Contribution

3.1 Proposed Model

A DT model for SMEs based on lean service and TQM tools is proposed to obtain customer satisfaction, product delivery compliance, and generate a base with management attributes as shown in Fig. 1.

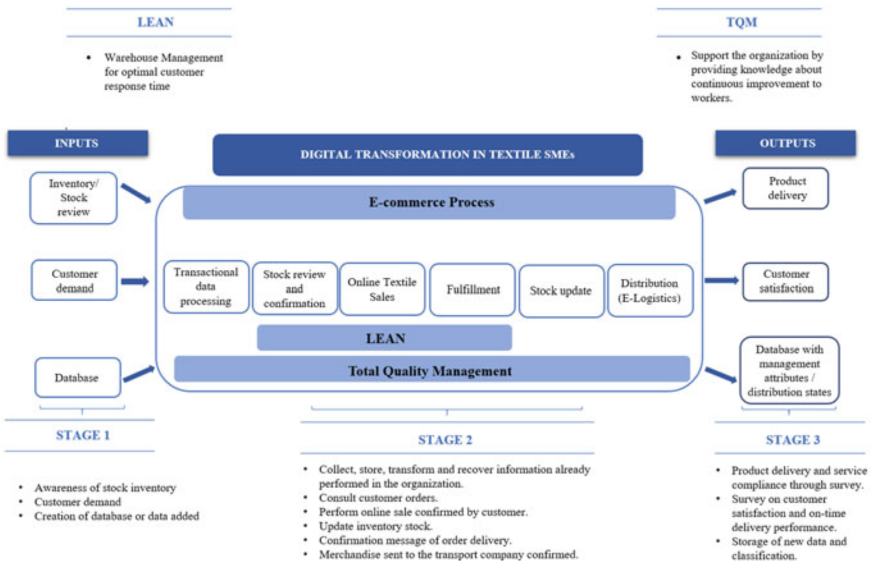


Fig. 1 DT model in textile SMEs

The components in question originate from the analysis of the textile sector's current situation and the search for a solution based on low growth in comparison with that of the market.

Further, each one is briefly explained.

Inputs

- Customer requirements: This component is generated by analyzing the company's current situation where customer needs are identified.
- Database creation: A database is generated by providing space for the relevant information that will be utilized for contacting customers.
- Stocktaking: This component involves counting the organization's on-hand inventory; in the absence of this component, the company would fall into erroneous sales and order cancellations.

E-commerce Operational Processes

- Transactional data processing: Data collection and placement are conducted according to its importance. In this way, operational tasks are automated, such as the recording of information enabling savings in personnel costs.
- Offer review: Here, customer orders are made known.
- Online textile sales: This involves confirmation by the client to make the payment and validate the payment method.
- Stock control: Once the sale is confirmed, the stock is updated so that problems do not occur in subsequent orders due to a lack of garments.
- Contactability: The purchase is confirmed, and a message with the purchase data is sent to the customer.
- Distribution: The organization in charge of the delivery is contacted, and the order collection and shipment proceed.

Outputs

- Product delivered: It is confirmed that product delivery is correct.
- Customer satisfaction: A survey is conducted to find out the customer satisfaction level.
- Database with management attributes: New information about the customer is obtained, which will be used to predict future purchases.

3.2 Proposed Method

Further, the flowchart of the proposed method is shown in Fig. 2.

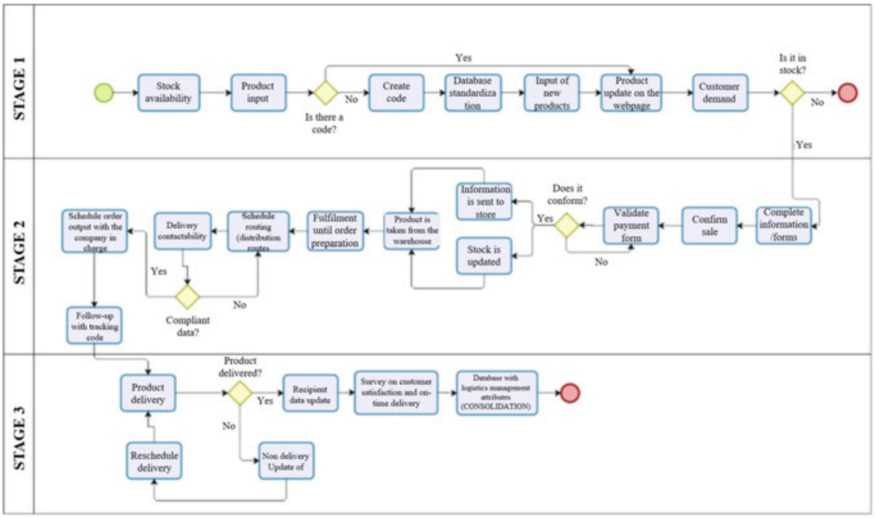


Fig. 2 Process flowchart

3.3 Indicators

Percentage of returns. This factor provides information on the number of deliveries that have not been appropriately conducted (due to issues that may be related to quantity, quality, and delays in delivery) during a specific period, without specifying the cause of return.

$$\text{Percentage of returns} = \frac{\text{Number of orders returned}}{\text{Number of fulfilled orders}} \times 100 \tag{1}$$

To achieve optimum speed with the proposed model, this must be at 20%.

Properly attended sales. It means that the order is fulfilled correctly, and the items and quantities supplied are exactly the same items in the quantities requested by the customer.

$$\text{Properly attended sales} = \frac{\text{Number of properly fulfilled orders}}{\text{Number of fulfilled orders}} \times 100 \tag{2}$$

To achieve a properly attended sale, the KPI should be greater than or equal to 50%.

Customer satisfaction. It refers to a good shopping experience of customers who in the long run purchase the product again. Furthermore, it is a value with which customer satisfaction can be measured with respect to the company and its performance.

Table 1 Initial indicators versus indicators after implementation

Indicator	Initial result (%)	Final score (%)	Variation (%)
Percentage of returns	50	29	21
Properly attended sales	25	66	41
Customer satisfaction	32	65	33

$$\text{Percentage index of customer satisfaction} = \frac{\text{Percentage index of customer satisfaction} =}{\text{Total evaluations achieved}} \times 100 \tag{3}$$

Acceptable values of greater than or equal to 50% are considered.

4 Validation

4.1 Description of the Case Study

The waste management model was implemented in '*Xiaomi Confecciones Textiles,*' a company dedicated to the sale of polo shirts, pants, and shorts; it specializes in corporate and urban clothing retail and wholesale.

4.2 Current Situation

The company's main problem is that online sales have been incurring losses over the years. The monetary loss for the year 2018 amounted to PEN 185700, which is perceived to be a high amount for a textile SME, due to different reasons such as the percentage of returns, improperly attended sales, and the level of customer dissatisfaction. Table 1 displays the initial values of the indicators compared to the case study.

4.3 Initial Indicators Versus Indicators After Implementation

Further, the variation in the indicator values after the implementation of the digital transformation model is shown.

As can be seen, the indicators have improved remarkably; the percentage of returns decreased to 29%, sales attended increased to 66%, and customer satisfaction increased considerably.

As observed in Table 1, the percentage of returns was reduced by 21%, customer satisfaction increased by 33%, and properly attended sales increased by 41% due to the newly implemented database and warehouse management.

5 Conclusions

The proposed method applied in the e-commerce process reduces the stock control time by 91%. In addition, the time from product input to inventory was reduced by 93%, indicating an improvement in the managing and updating of warehouse products. Lean service facilitated the implementation of a digital transformation model, and there was a 38% increase in the amount of packaging, allowing a quick response to the requests requested by customers. Due to database improvement, the percentage of returns decreased by 21%, either for the improvement of records by 32% or by the decrease in incomplete records. In addition, it was observed that the attended sales increased by 41% because the company received training in the warehouse and format management to improve the database. This aspect further allowed the company to increase its contact with the customer and ascertain customer satisfaction level, which in this case was 65% after implementation, thereby enabling it to create a digital culture and achieve better results.

References

1. Wong, J.K.W., Ge, J., He, S.X.: Digitisation in facilities management: a literature review and future research directions. Autom. Constr. **92**, 312–326 (2018). https://doi.org/10.1016/j.autcon.2018.04.006
2. El Peruano Diario Oficial del Bicentenario: https://elperuano.pe/noticia-ecommerce-crecio-47-el-peru-70898.asp. Online accessed: 18 de enero de 2019
3. Salem, R., Musharavati, F., Hamouda, A.M., Al-Khalifa, K.N.: An empirical study on lean awareness and potential for lean implementations in qatar industries. Int. J. Adv. Manuf. Technol. **82**(9–12), 1607–1625 (2016). https://doi.org/10.1007/s00170-015-7421-7
4. Instituto Nacional de Estadistica e Informatica. Peru: Tecnologia de Información y Comunicación en las Empresas. Lima. Recuperado (2016). https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1406/libro.pdf
5. Ismail, A., Majid, A.H.A., Rahman, M.A., Jamaluddin, N.A., Susantiy, A.I., Setiawati, C.I.: Aligning Malaysian SMEs with the Megatrends: the roles of HPWPs and employee creativity in enhancing Malaysian SME performance (2018)
6. Bouwman, H., Nikou, S., de Reuver, M.: Digitalization, business models, and SMEs: how do business model innovation practices improve performance of digitalizing SMEs? (2019)
7. Gallaugher, P.A.J.M.: Factors affecting the adoption of an Internet-based sales presence for small businesses. Inf. Soc. **13**(1), 55–74 (1997)
8. Chen, Y.-Y.K., Jaw, Y.-L., Wu, B.-L.: Effect of digital transformation on organisational performance of SMEs. Internet Res. **26**(1), 186–212 (1997)
9. Phan, A.C., Nguyen, H.T., Nguyen, H.A., Matsui, Y.: Effect of total quality management practices and JIT production practices on flexibility performance: empirical evidence from

international manufacturing plants. Sustainability (Switzerland) **11**(11), Article Number 3093 (2019). https://doi.org/10.3390/su11113093
10. Ruiz, E., De Pablo, J.D.S., Muñoz, R.M., Peña, I.: Do high performance work systems enhance business performance? Examining the mediating influence of total quality management. Zbornik Radova Ekonomskog Fakultet au Rijeci **37**(1), 235–258 (2019). https://doi.org/10.18045/zbefri.2019.1.235
11. Oliveira, G.S., Corrêa, J.E., Balestrassi, P., Martins, R., Turrioni, J.: Investigation of TQM implementation: empirical study in Brazilian ISO 9001-registered SMEs. Total Qual. Manag. Bus. Excell. **30**(5–6), 641–659 (2019). https://doi.org/10.1080/14783363.2017.1328273
12. Haroon, U., Shariff, M., Yusof, M.: Exploring the relationship between innovation capability, total quality management and SMEs performance in electric fan industry of Pakistan: moderating effect of relational learning. Int. J. Financial Res. **10**(3), 353–359 (2019). https://doi.org/10.5430/ijfr.v10n3p353
13. Manfredsson, P.: Textile management enabled by lean thinking: a case study of textile SMEs. Prod. Plan. Control **27**(7–8), 541–549 (2016) https://doi.org/10.1080/09537287.2016.1165299
14. Lins, M., Zotes, L., Caiado, R.: Critical factors for lean and innovation in services: from a systematic review to an empirical investigation (2019). https://doi.org/10.1080/14783363.2019.1624518

Poka-Yoke Model for Controlling Unit Entering and Fall Reduction in the Transportation of Toilets

Luis Roca-Ramos, Piero Vargas-Zamalloa, Edgardo Carvallo-Munar, Rosa Salas-Castro, and Luis Cardenas-Rengifo

Abstract Small and medium toilet manufacturing businesses in Peru face significant problems such as low productivity because of a high rate of production losses. To address this issue, the present study proposes the application of a lean manufacturing technique, such as the Poka-Yoke model, which improves the production flow within the company by implementing a transportation cart with a safety system to help reduce material falls during the manufacturing process. The results showed a 24% increase in productivity and reduction in problems of damaged products that did not meet quality standards, thus preventing their subsequent reprocessing.

Keywords Productivity · Poka-Yoke · Lean manufacturing · Toilets

1 Introduction

The production of toilets has always presented different problems over time. These problems are caused by the material with which it is manufactured, so special care must be taken when handling or transporting. Some of these problems are breakage, cracking, or impurity of materials. Peruvian population is constantly growing, which demands greater production in different sectors. Even when the country has been

L. Roca-Ramos · P. Vargas-Zamalloa · E. Carvallo-Munar · R. Salas-Castro ·
L. Cardenas-Rengifo (✉)
Universidad Peruana de Ciencias Aplicadas (UPC), Lima 15023, Peru
e-mail: pcsilcar@upc.edu.pe

L. Roca-Ramos
e-mail: u201212484@upc.edu.pe

P. Vargas-Zamalloa
e-mail: u201217221@upc.edu.pe

E. Carvallo-Munar
e-mail: pcinecar@upc.edu.pe

R. Salas-Castro
e-mail: rosa.salas@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_39

sustainably growing at approximately 2.5% per year, the growth of this sector is indirectly proportional to the company's productivity owing to the high index of losses and cost overruns reported [1], thus hindering the sector from achieving better positioning at an international level. As this problem directly affects the objectives of the company, any issues that do not contribute to the improvement within the company under study must be addressed and solved; this drives companies to seek methods to improve their productivity using various tools and engineering strategies that provide such benefit to the organization [2]. Productivity issues in companies are very frequent, so the application of a simple method to help them address these problems and, in turn, provide an overall improvement to the company is essential for achieving better competitiveness and increasing customer satisfaction [3]. The Poka-Yoke model is a key tool for preventing production errors, implementing better controls, and improving the production flow. Therefore, this study aims to disseminate knowledge regarding the importance of applying the Poka-Yoke methodology to improve the production flow and overall productivity of the company.

2 State of the Art

Manufacturing companies are directly related to the generation of production losses. In some cases, losses can be recovered; however, in other cases, they become waste, generating losses that can be significant depending on the number of products involved. For this reason, companies must do everything within their power to minimize defective products or products that suffer damages during the production process.

A considerable number of authors have researched and discussed the details of using a model or tool to prevent production failures or errors. For example, the Poka-Yoke model helps reduce the process and labor time, which usually leads to higher costs owing to the aforementioned problems [4]. In general, these devices or improvements are easily applicable because they are based on implementing considerable improvements that render a stable production process. Moreover, using a Poka-Yoke-based device can solve the problems negatively affecting a company because this methodology also helps companies improve their efficiency while performing their work [5].

As part of the conclusions for the case study in this paper, the tool was able to generate significant success in the company. By implementing a system that regulates production activities and decreases production accidents, a new line design consisting of only three workstations is designed to help increase productivity and decrease the losses incurred [6].

Furthermore, the implementations allowed us to determine the estimated time for which improvements can be observed. This lean manufacturing improvement, combined with other tools, can unveil areas of opportunity for waste reduction and productivity improvement within the company. In particular, production time can be reduced because production delays due to errors within the production process are eliminated. [7].

For any company, it is feasible to generate the highest profits with the lowest possible investment; this tool is based on a concept that a simple idea can generate significant impact [8].

3 Contribution

3.1 Overview

It was observed that the transport and control problems affecting units entering the kiln and dryer generated a 33% reduction in productivity. To solve this problem, a Poka-Yoke model was implemented to help the safe transportation of units between the areas and have a limited capacity that can serve as a measuring tool when transferring units to the kiln and dryer.

3.2 Specific View

3.2.1 Poka-Yoke

- As a first step, it is necessary to assemble a work team comprising a production manager and a head of occupational safety and health. Their functions include planning, programming, execution, and supervision of process modifications.
- Second, a new transportation cart that will reduce the number of losses was designed. All the changes implemented must be coordinated by the work team to the technical and operational team. This also includes the placement of safety bars at the bottom of the two cart levels, which allows units to be loaded onto the cart but prevents them from slipping.
- During the implementation, the study proposes conducting training sessions to help the operator adapt and learn the new processes. For this purpose, external subject matter experts are required to provide knowledge based on good practices implemented in other companies and replicate them in the toilet manufacturing company.
- Finally, the satisfaction indicators for the implemented process must verify whether they are yielding favorable results.

The implementation will apply to all 100 transportation carts in operation within the company. Only 20 units can be loaded into each cart, ten on the top level, and the rest on the second level. The transportation process requires two operators (a senior and a junior operator); the senior operator will be in charge of validating the units entering the kiln and dryer. The kiln and dryer must host ten full carts at all times to reduce the stress of entering fewer units and ensure that no pieces come out wet.

Fig. 1 Prototype of the transportation cart

The success indicators used to measure the effectiveness of the implementation are the percentages of failures and falls. Failures can occur inside the kiln or dryer because of incompliance with the input units limit (units to be entered into the oven). Moreover, during the transportation process, products can also be mishandled, such as falling on the ground or be dented. Any defects generated are counted in the unit of time for comparison purposes between previous and current results.

Sidebars are attached to the lower cart level to reduce the number of daily falls because of the lack of protection. The design, as shown in Fig. 1, also includes two bracers for ease of movement or anchoring the cart. The proposed model was based on a previous design prepared by for similar purposes [9].

3.3 Process View

For implementing the change management together with the planned maintenance, the following steps are considered as reference. Figure 2 presents these steps, according to the researched successful cases, and are considered as theoretical steps for the implementation of each methodology.

The process generally starts through the directors and managers acknowledging the importance of the issue. Next, they strive to communicate and build work teams to address the identified low productivity problem. For this, prior training is required to ensure that the techniques will be safely and timely applied.

Fig. 2 Process view

3.4 Indicators

The following indicators were presented to monitor the production process; these indicators measure the methodology and its contribution to reducing low productivity:

- Any defective units found at the end of the baking and drying processes indicate that the units were not properly placed inside the kiln. Therefore, the level of incidence on defect reduction is measured using the transportation carts with safety bars as a unit of measure.

$$\%\text{Failures} = \frac{\text{Total Defective Products}}{\text{Total Processed Products}} \times 100 \tag{1}$$

- The number of falls will assess the implementation of safety bars by counting the average number of pieces that fall during daily transportation.

$$\%\,\text{Falls} = \frac{\text{Total Fallen Products}}{\text{Total Trasported Products}} \times 100 \tag{2}$$

4 Validation

4.1 Validation of Poka-Yoke Load Protection

A prototype simulating the new cart design was created for the validation of the transportation cart, and the number of daily units that could be lost in falls was calculated. In this case study, the daily falls for the lack of safety bars in the transportation units had previously reached 24 units a day. With the support of the proposed prototype, this number is expected to be reduced to four units a day, going from 8595 lost units per year to just 1440 units, which accounts for an 84.6% reduction. The variation in daily falls is described in the chart below (Table 1).

The final results denote a 24% increase in the company's productivity. Considering the contribution of the prototype, which incurred a total cost of PEN 919, it is necessary to simulate the transportation of 15 kg objects at 3–5 km/h to record the number of times the units make contact with the floor after including the safety bars in lower cart levels (Table 2).

Table 1 Results per indicator

Indicator	Before	After	Improvement %
Number of daily falls	26 units/day	4 units/day	85
Failures (%)	55	25	30

Table 2 Implementation cost

Costs	Poka-Yoke	
Cost of training and opportunity		
1. Salary of participants	PEN	110.80
2. Salary of substitutes	PEN	110.51
3. Salary of internal trainer	PEN	0.00
4. Salary of external trainer	PEN	2,000.00
5. Materials	PEN	60.00
6. Equipment and supports	PEN	0.00
7. Rent	PEN	0.00
8. Catering	PEN	0.00
9. Cost of opportunity	PEN	1,364.69
Implementation cost		
1. Process owner salaries (managers)	PEN	48,000.00
2. Materials (cart)	PEN	10,000.00
3. Equipment (scales)	PEN	8,415.00
Total cost	PEN	70,061.00

Implementing the safety bars, training the operators, and executing the system to control the units would cost PEN 70,061 for the 100 carts the company currently owns.

5 Conclusions

Validations were carried out with products from other production lines. This showed that applying Poka-Yoke models to reduce falls and failures is replicable in other ceramic products as well since the results obtained were favorable in all of these with an average of 18.23% improvement. The products that participated in the experimentation were laundry, toilet tank, toilet cups, and toilets, which showed an improvement of 20%, 15%, 14%, and 24%, respectively. The study demonstrated that the placement of safety bars in the transportation carts produces an 85% reduction in the number of daily falls. By using the carts as measure units for the baking and drying processes, the average number of failures improves by 30%. The productivity improvement objective was reached by achieving a 24% increase attributed to the implementation of a control system for placing units inside the kiln and dryer along with the attachment of safety bars to the transportation carts, which reduce falls. With the help of the prototype made to simulate the transport of toilets within the company, it is possible to demonstrate that the reduction of falls of toilets is possible by using safety bars that restrict the movement of the toilets inside the transport carts, and the improvement represents 85% compared to its previous version. By using the new transport carts as a unit of measure to enter the ovens and dryers, it is possible to identify the improvement of 30%. This indicates that the order they impose when organizing toilets for defined quantities helps reduce the number of defective products by not respecting the optimal production units. This study aims to increase toilet manufacturing productivity by reducing the falls and failures in the baking and drying processes. However, an assessment of the incidence of non-productive times and reprocesses on production must also be conducted in the future.

References

1. Instituto Nacional de Estadística e Informática: Producción Nacional Enero 2019 (2019). Available: http://www.inei.gob.pe/media/MenuRecursivo/boletines/01-informe-tecnico-n01_produccion-nacional-nov2018.pdf
2. Kotter, J.: Leading the change. Harv. Bus. Rev. (2012)
3. Quirant, A., Ortega, A.: El cambio organizacional: la importancia del factor humano para lograr el éxito del proceso de cambio: Revista de Empresa (2006)
4. Wan Saidin, W.A.N., Mohd Idris, A.Z., Ravi, S., Ahmad Zaidi, A.M., Kasim, N.I.: Detection of nut welding using Poka-Yoke roller coaster jig. Appl. Mech. Mater. **761**, 170–174 (2015). https://doi.org/10.4028/www.scientific.net/amm.761.170
5. Helmi, S.A., Nordin, N.N., Hisjam, M.: Errors prevention in manufacturing process through integration of Poka Yoke and TRIZ (2017). https://doi.org/10.1063/1.5010642
6. Mckone, K.E., Weiss, E.N.: TPM: planned and autonomous maintenance: bridging the gap between practice and research. Prod. Oper. Manag. **7**, 335–351 (1998). https://doi.org/10.1111/j.1937-5956.1998.tb00128.x
7. Akter, S., Yasmin, R.: Implementation of kaizen for continuous improvement of productivity in implementation of kaizen for continuous improvement of productivity in garment industry in Bangladesh (2017)
8. Choomlucksana, J., Ongsaranakorn, M., Suksabai, P.: Improving the productivity of sheet metal stamping subassembly area using the application of lean manufacturing principles. Procedia Manuf. **2**(February), 102–107 (2015). https://doi.org/10.1016/j.promfg.2015.07.090
9. Jiménez, M.: Carretilla multifunción. ESP Patent 1,068,334, October 1, 2008

Lean Six Sigma Fleet Management Model for the Optimization of Ore Transportation in Mechanized Underground Mines in Peru

Jorge Huaira-Perez, Orlando Llerena-Vargas, Humberto Pehóvaz-Alvarez, Hugo Solis-Sarmiento, Vidal Aramburu-Rojas, and Carlos Raymundo

Abstract Mining activities around the world are undergoing constant change and modernization owing to technological and scientific advancements. Consequently, there are frequent proposals to streamline and enhance processes in mining operations. This study deals with ore transportation in mechanized mining units and aims to optimize fleet management using the Lean Six Sigma methodology to obtain a model in this specific process. The proposed method was implemented using a Lean Six Sigma instrument known as DMAIC (Define, Measure, Analyze, Improve, and Control). The case study was applied to an underground mine located in the Huancavelica region, Peru. The simulation showed that 24% of the time in the ore transport cycle is un-productive time and the improvement potential time represents 53% of the transportation process time.

Keywords Fleet management · Lean Six Sigma · Ores transportation · Mechanized underground mining

J. Huaira-Perez · O. Llerena-Vargas · H. Pehóvaz-Alvarez · H. Solis-Sarmiento · V. Aramburu-Rojas · C. Raymundo (✉)
Universidad Peruana de Ciencias Aplicadas, Lima 15023, Perú
e-mail: carlos.raymundo@upc.edu.pe

J. Huaira-Perez
e-mail: u201518696@upc.edu.pe

O. Llerena-Vargas
e-mail: u201523751@upc.edu.pe

H. Pehóvaz-Alvarez
e-mail: pcgmhpeh@upc.edu.pe

H. Solis-Sarmiento
e-mail: pcgmhsol@upc.edu.pe

V. Aramburu-Rojas
e-mail: pcgmvara@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_40

1 Introduction

Currently, Peru's mining sector is a major contributor to the country's economy and development, as according to the 2018 yearbook published by the Ministry of Energy and Mines, this sector represented 10% of Peru's total gross domestic product and around 61% of the total value of Peruvian exports [1]. Peru has fostered regular management system innovations because it seeks to focus on the optimization of mining processes through mechanized equipment. Upon examining underground mining, Peruvian mines do not have integrated automated fleet management systems because the country's geography and climatology poses a great challenge to develop these systems within mining units.

Similarly, in general mining, the management of load, haul, and dump machinery (LHD equipment) represents the main component of operating costs in mining production, which can amount to 50 and 60% of the mine operating costs [2]. The main objective of ore transportation and hauling is to optimize operating times; however, these are the least controllable variables in ore transportation. Many organizations are aware that the existence of downtime in operational processes is a problem that affects their efficiency [3]. Therefore, the execution of a Lean Six Sigma model to ensure the quality of ore transportation in mechanized underground mines has been considered the best alternative. Lean Six Sigma identifies errors in a process and suggests alternatives for improvement [4].

2 State of the Art

2.1 Fleet Management in Mining Transportation

Ore transportation from the mine to the plant is a relevant factor in mining operations, because it is linked to costs and productivity, as is the case of LHD machinery, which represents a big operating cost in mining activities [5, 6]. The mining activity experiences multiple disadvantages and excessive resource consumption, poor and/or dangerous operating environments, low process efficiency, serious safety risks, high production costs, and severe pollution; these factors are the most significant but least controlled [7, 8]. Operational or incidental problems can be found in fleet management in mining, such as the mining plan, truck availability, production areas, transportation routes, vehicle disturbance, tonnage capacity, loss of downtime, or incidental cases, such as rockfalls, flat tires, and roads in poor conditions [9].

Over the years, the goal is to develop these processes with a high degree of quality, to guarantee such process efficiency [5, 6, 8]. Currently, owing to the great variety of management systems, the search for quality is no longer interpreted as an expense but rather as an investment for the future since quality increase as a resource for controlling processes and the company's profitability is considered an unlimited, continuous, and endless process. Therefore, it is appropriate to implement the Six Sigma methodology for fleet management in mechanized underground mining, as favorable results were observed regarding the fleet management systems currently used in mines [7, 9].

2.2 Use of the Six Sigma Methodology in the Transportation Management System

Six Sigma serves as an alternative support for fleet management in mining transportation. This tool is a process improvement methodology that seeks to boost productivity. It is based on a set of instruments aimed at improving processes by reducing their variability, resulting in the reduction and elimination of defects in such process or operation [10–13]. The Six Sigma methodology bases its implementation in the statistics field, taking standard deviation as the main statistical measure, which is used to analyze the process data that has been noted previously [12, 14]. These data are collected to find the main failures or delays that exist in the process development.

Hence, the implementation of the Six Sigma methodology for the optimization of ore transportation for an underground mining company is important because it reduces and eliminates defects in the production cycle [10, 13]. The entire purpose of optimizing fleet management is to encourage production and reduce operating costs. This methodology was implemented in a copper surface mine, where processes were improved in the Tintaya mining unit of Glencore by enhancing equipment performance [11, 12, 14].

3 Contribution

3.1 Proposed Method

The proposed method is disclosed in Fig. 1, where we focus on the development of three important processes.

We will seek to describe the process, structure, and the KPIs of the study and diagnose the behavior of the operations to be studied. Conversely, we should be able to visualize the operational process flow, activity synchronization, and study process capacity and to describe the functions performed by each party involved.

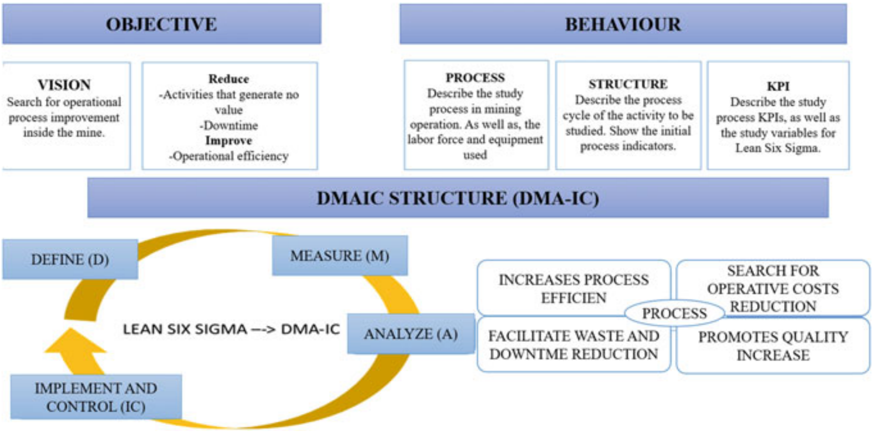


Fig. 1 Lean Six Sigma operational evaluation methodology

3.2 *Proposed Method*

Once the operational evaluation method of the process has been defined, through a flowchart, we will explain the development of the Lean Six Sigma methodology and its DMAIC method for improving the process quality of the process. The application of the methodology, through the flowchart, seeks to provide mining companies with the possibility of improving any operational process in the mine. Figure 2 shows the methodology diagram and its implementation within the working areas of mining operations.

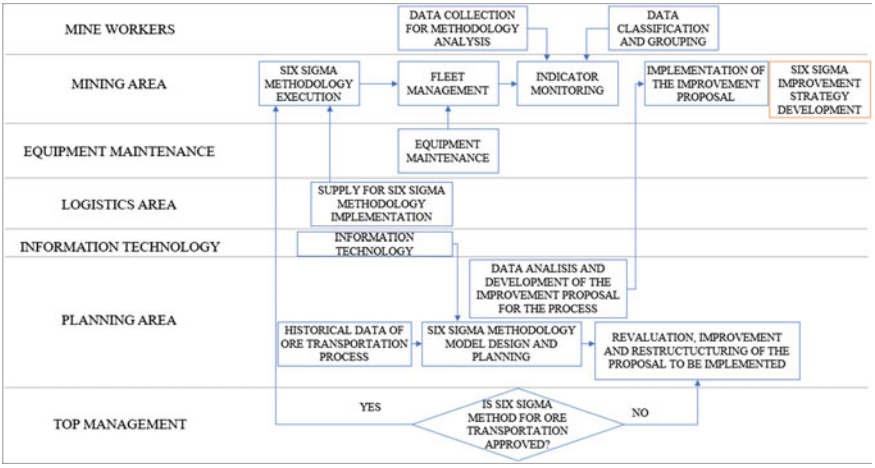


Fig. 2 Lean Six Sigma methodology diagram in mining operation

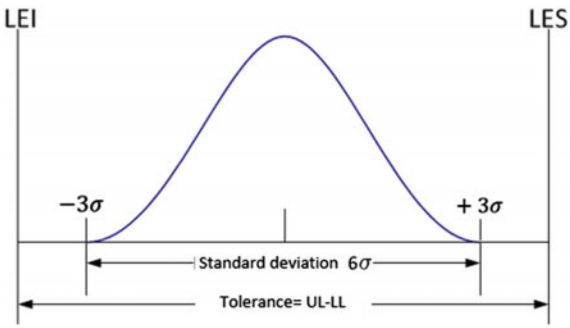


Fig. 3 Calculation of Cp and Cpk indicators

3.3 Indicators to Take into Consideration

3.3.1 Capability Index

For the calculation of the indicators, as shown in Fig. 3, the graph of capability index and Formulas (1), (2), (3) will be considered.

$$Cp = \frac{\text{LES} - \text{LEI}}{6\sigma} \tag{1}$$

$$Cpu = \frac{\text{LES} - \mu}{3\sigma} \tag{2}$$

$$Cpl = \frac{\mu - \text{LEI}}{3\sigma} \tag{3}$$

Cpk = Lesser value between Cpu and Cpl
σ = Standard deviation
μ = Quality characteristic average.

3.3.2 Performance

This indicator measures the operational performance of the low-profile tipper for indoor transportation within the mine (30 Tn tipper truck). Moreover, through the performance Formulas (4), (5), we will try to achieve a reliable measurement of the performance of the low-profile tipper for the operation.

$$\text{OP(ton)} = \frac{60(\text{min/hr}) \times \text{Nc}}{\text{Ct}} \tag{4}$$

$$\text{OP}\left(\text{m}^3\right) = \frac{60\left(\frac{\text{min}}{\text{hr}}\right) \times \text{Nc}}{\text{Ct} \times \text{Sf} \times \text{Pt}} \tag{5}$$

where

OP: Operational performance (in this case of the low-profile tipper of 30 Tn)
Ct: Transportation cycle time
Nc: Nominal capacity of the equipment (ton)
Sf: Swell factor (fraction)
Pt: Wet bulk density (ton/m^3).

3.3.3 Efficiency

This indicator is used to measure the efficiency of ore transportation operations inside the mine. To this end, we will use the different resulting times that were obtained through the value diagram, which was developed in the analyze stage of the sDMAIC process.

$$\text{Operational Efficiency} = \frac{\text{Executiontime}}{\text{Executiontime} + \text{Downtime}} \tag{6}$$

4 Validation

4.1 The Location Used to Validate the Methodology

To validate the study regarding the Lean Six Sigma, it was possible due to grant the appropriate permissions from responsible authorities of mechanized mine.

The validation was conducted in a mechanized mine in Peru located in the department of Huancavelica at about 2600 m.a.s.l. The mine is a polymetallic deposit, and its main metal is copper, in addition to other byproducts such as lead, zinc, and silver.

a. **Description of the Proposed Scenario**

For the development of the study methodology, we visited the mining unit, specifically Zones V and III, which are the locations where the company is running its operations. In these zones, the operational processes for ore extraction are drilling, blasting, hauling, and transporting ore. The present research will focus on the implementation of Lean Six Sigma during the ore transportation process inside the mine through low-profile tippers that can carry a maximum of 30 tons per trip. Conversely, the operational process for transporting ore inside the mine through a low-profile tipper must be described in detail. This cycle consists of five stages, mentioned and developed as follows: (1) load ore to the tipper, (2) haul ore to the pocket, (3) dump ore to the pocket, (4) return the empty equipment back to work, (5) wait to load on tipper.

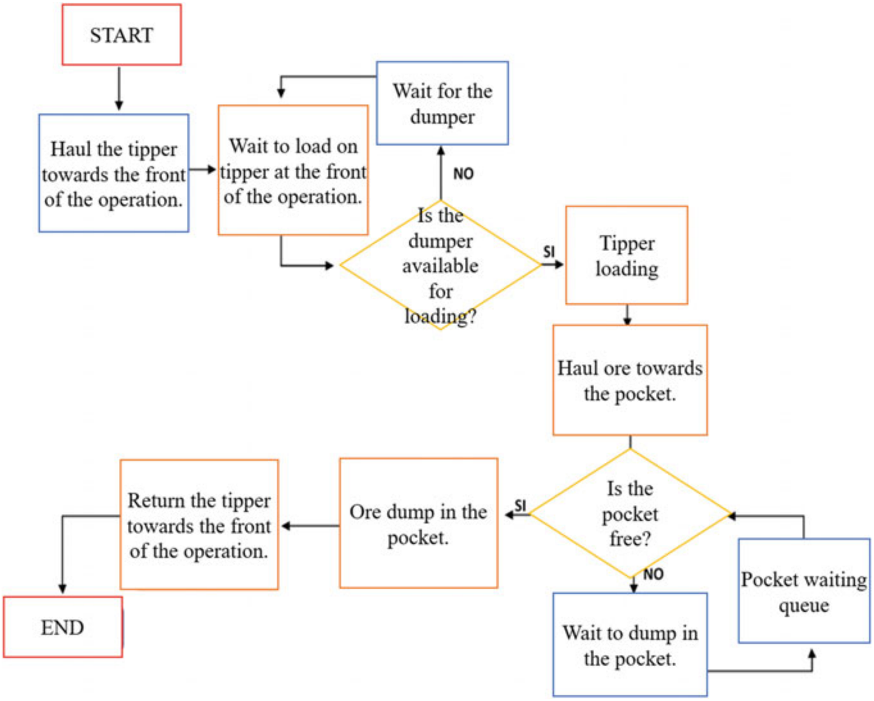


Fig. 4 Mapping of the ore transportation process inside the mine

b. **Implementation of the Lean Six Sigma Methodology proposal**

4.1.1 Problem Definition (D)

To start the DMAIC process, we need to define the variables to be used for the case study to reduce failures, delays, and downtimes during the ore transportation process inside the mine. For this case study, we could determine the variables found: machine hour, effective hour, available hours, operational delays, percentage of mechanical availability, covered kilometers.

The process mapping is presented in Fig. 4, as shown below.

4.1.2 Variables and/or Data Measurement (M)

In the second phase, time was measured considering the activities related to the effective hours variable for mining transportation; each data considered for the current study is stated in seconds. We divided the effective hours variable into five sub-variables. Furthermore, data measurements were made throughout January 2018 with one equipment (tipper No. E-24) that performs operational tasks inside the mine areas. From the data collected, in addition to the data that were collected daily and monthly, we obtained the following statistical indicators, as shown in Fig. 5.

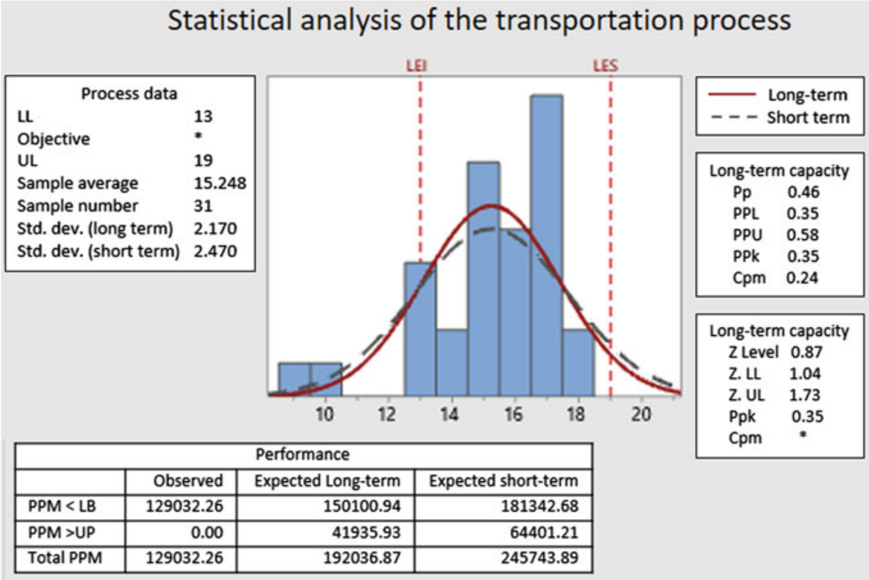


Fig. 5 Statistical analysis of the transportation process

c. Scenario and Results

Through the data measured in Minitab, we obtained the process results. The resulting method from the studied operation is related to the ore transportation process inside the underground mine under study (Table 1).

As shown, the results obtained at a 95% confidence level in the scenario with the predicted results are positive; thus, the implementation of the methodology tends to optimize the process under study.

d. Data Analysis

As shown (Table 2), the results obtained at a 95% confidence level in the scenario with the predicted results are positive; thus, the implementation of the methodology tends to optimize the process under study.

Table 1 Indicators

INDICATORS (operation average)	
Process capability index (*C*p)	0.40
Process capability index (*C*pk)	0.30
Sigma level (*Z*)	0.80
Operational efficiency (%)	76

Table 2 Indicators after implementing

Indicators (average of the three operations)		Study indicators	Indicators after implementing the improvement implementation
Process capability index (Cp)	0.40	0.46	1.02
Process capability index (Cpk)	0.30	0.35	0.86
Sigma level (Z)	0.80	0.87	2.45
Operational efficiency (%)	76	76	87

5 Conclusions

From the Lean Six Sigma analysis, it can be concluded that the ore transportation process inside the mine with a process capability of Pp 0.46 and Ppk 0.35 does not reach the minimum value to prove the capability or stability of the process (greater than 1.33). Moreover, the Six Sigma level of 0.87 is way below the expected level (the Sigma value for a capable process ranges between 3 and 4). Therefore, ore transportation inside the mine has great potential for operational improvement and fleet management. The value diagram showed that 24% of the time in the ore transport cycle is unproductive time or downtime. It also indicates that the improvement potential time represents 53% of transportation process time, which shows that the ore transportation process should be improved through fleet management to boost tipper productivity and increase operational performance and efficiency in terms of tonnage ore transportation per hour. In conclusion, if the activities that cause downtime or delays during the ore transportation process could be eliminated or reduced, the performance and efficiency of the low-profile tippers would increase.

References

1. MINEM. Anuario Minero 2018 (2018)
2. Matamoros, M.E.V., Dimitrakopoulos, R.: Stochastic short-term mine production schedule accounting for fleet allocation, operational considerations and blending restrictions. Eur. J. Oper. Res. **255**(3), 911–921 (2016)
3. Haviland, D., Marshall, J.: Fundamental behaviours of production traffic in underground mine haulage ramps. Int. J. Min. Sci. Technol. **25**(1), 7–14 (2015)
4. Bogacz, P., Migza, M.: Zastosowanie Lean Six Sigma w doskonaleniu procesów produkcyjnych w przemyśle wydobywczym. Inz. Miner. **17**(2), 23–29 (2016)
5. Greberg, J., Salama, A., Gustafson, A., Skawina, B.: Alternative process flow for underground mining operations: analysis of conceptual transport methods using discrete event simulation. Minerals **6**(3), 65 (2016)

6. Guo Li, J., Zhan, K.: Intelligent mining technology for an Underground metal mine based on unmanned equipment. Engineering **4**(3), 381–391 (2018)
7. Mixtaj, L., Weiss, E., Weiss, R., Labant, S.: The economic importance of tracking expenses related to the quality of the process management in an environment of mining company. Tech. Univ. Košice, 673–680 (2013)
8. VanHilst, M., Garg, P.K., Lo, C.: Repository mining and Six Sigma for process improvement. ACM SIGSOFT Softw. Eng. Notes **30**(4), 1 (2005)
9. Song, Z., Schunnesson, H., Rinne, M., Sturgul, J.: Intelligent scheduling for underground mobile mining equipment. PLoS ONE **10**(6), 1–22 (2015)
10. Ras, E., Visser, J.K.: A model for continuous improvement at a South African minerals. South African J. Ind. Eng. **26**, 191–206 (2007)
11. Nikolaevna, S.N.: The Introduction of the conception « Lean 6 Sigma » As a way to sustainable positions of transport enterprise. Econ. Process. Manag. (2015)
12. Weiss, E., et al.: Proposal of a six sigma project economic effectiveness evaluation model, **XIX**(5), 34–38 (2014)
13. Tovar, O.: Optimización de operaciones mediante aplicación de Six Sigma mediante aplicación de Six-Sigma (2016)
14. Lévano Crespo, L.A.: Incremento de productividad mina mediante herramientas six sigma, p. 239 (2012)

Production Control Model Using Lean Manufacturing Tools and Kanban/CONWIP Systems to Improve Productivity in the Process of Sand Casting in a Heavy Metalworking SME

Luz Prada-Echevarría, Jorge Chinchay-Grados, Fernando Maradiegue-Tuesta, and Carlos Raymundo

Abstract The contribution of the metalworking sector to the gross domestic product is decreased by 8.6% in the 2017–2018 period because of problems such as high rejection rates, lead times, and raw material losses. Consequently, the sector's production was reduced by PEN 1,200,000 while demand remained flat. Thus, this article proposes a production control model using 5S, *Single-Minute Exchange of Dies,* and tools such as Kanban and Constant Work-in-Progress (CONWIP) to increase the productivity of a sand casting line in a Peruvian metalworking enterprise that manufactures parts for mining on request. The improvement proposal involves the implementation of lean manufacturing tools and analyzes low productivity effects. The problems identified are typical of the Peruvian metalworking sector; thus, this study may contribute to finding possible solutions for issues faced by other enterprises.

Keywords Sand casting · Lean manufacturing · Make to order (MTO) · Kanban/CONWIP · SMED

L. Prada-Echevarría · J. Chinchay-Grados · F. Maradiegue-Tuesta · C. Raymundo (✉)
Universidad Peruana de Ciencias Aplicadas, Lima 15023, Perú
e-mail: carlos.raymundo@upc.edu.pe

L. Prada-Echevarría
e-mail: u201416030@upc.edu.pe

J. Chinchay-Grados
e-mail: u201416860@upc.edu.pe

F. Maradiegue-Tuesta
e-mail: fmaradie@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_41

1 Introduction

The December Inflation Report of the Peruvian Economy Institute (PBI, by its acronym in Spanish) presented a gross domestic product (GDP) growth forecast of 4% for both 2018 and 2019. Further, President Martín Vizcarra highlighted that domestic demand will remain the main growth driver in 2020 and the GDP will grow by 3.9%, which is lower than the 4.4% forecast for 2019.

The following section will present the state of the art, comprising the research and analysis of 20 articles on the methodologies used for waste reduction in metalworking enterprises. Next, we analyze the contribution of this article to the metalworking sector, followed by validation of the methodology; finally, we discuss the study results and conclusions.

2 State of the Art

Various metalworking enterprises have attempted to implement manufacturing strategies to improve production control. A Colombian enterprise developed a manufacturing strategy system to define the objectives, competitive priorities focused on cost, quality, delivery, flexibility, service, and innovation. Research findings show a percentage increase of 35% in priority indicators [1–3].

Other research findings indicate that Kanban and CONWIP systems are used to enhance production control, which is affected by chaotic behavior and unpredictable demand [4, 5]. The cases in these studies used Kanban and CONWIP cards, thus implementing a smart information system to control production processes and to improve the variable system according to common variable demand.

Several studies compare and verify the performance of manufacturing systems when using the theory of constraints (TOC) or just-in-time method or the Kanban and CONWIP systems [6–8]. The transition from the practices of highly accepted management philosophies has generated considerable interest in the TOC.

3 Contribution

3.1 Basic Framework

The use of lean manufacturing tools such as work standardization and SMED reduced lead time (33%), non-value-added time (19%), and set-up time (55%) in an auto parts manufacturing industrial enterprise located in Sakthi, India. The enterprise was dealing with large work-in-progress (WIP) inventories and a high rejection rate resulting from failures related to airlocks, and more broken cores. Therefore, the management decided to implement lean principles in the foundry shop floor [9, 10].

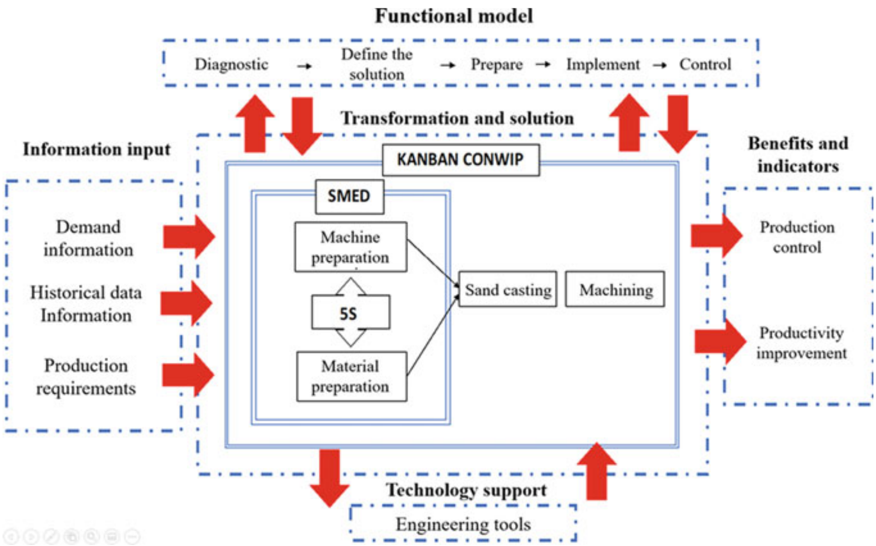


Fig. 1 Functional model overview

3.2 Overview

Figure 1 shows the proposed model.

3.3 Detail View

The detail view consists of four phases (Fig. 2).

3.4 Process View

Next, the graphic design of the proposal to be developed is shown (Fig. 3).

3.5 Indicator View

See Table 1.

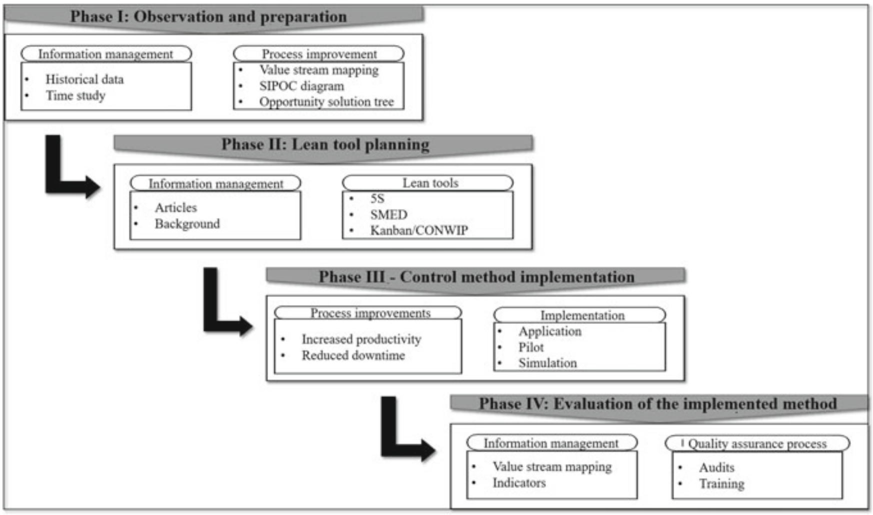


Fig. 2 Tool graphic design

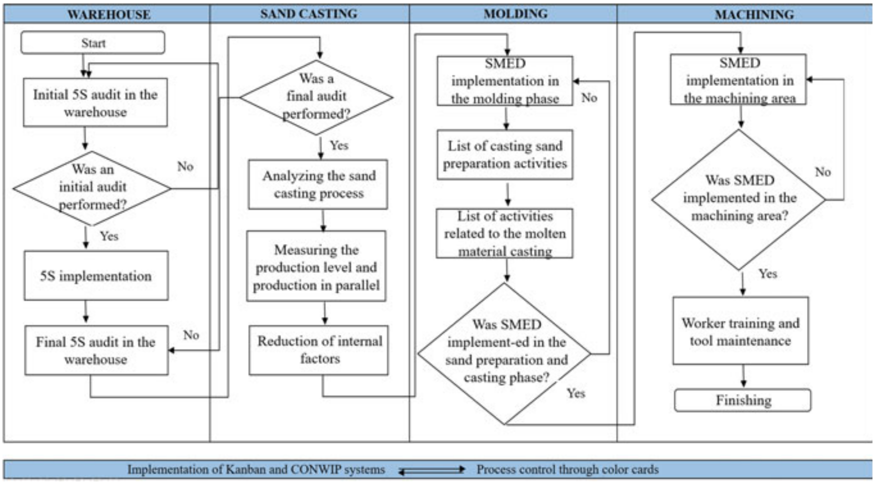


Fig. 3 Tool graphic design

Table 1 General indicator view

Indicator	Concept	Formula	Interpretation
Work-in-progress (WIP) inventory level	Indicates the amount of raw materials used in the production process	*ACTUAL INVENTORY LEVEL (KG)/SCHEDULED INVENTORY LEVEL (KG)*	The number of kilograms that has been used as scheduled
On-time delivery	Indicates the number of parts or items delivered to the client in the established time	*NUMBER OF ITEMS DELIVERED ON TIME/NUMBER OF ITEMS REQUESTED IN THE PURCHASE ORDERS (POs)*	A high number indicates that the parts were delivered on time
Total productivity	Indicates the number of tons produced during hours worked	*TOTAL QUANTITY OF TONS PRODUCED/TOTAL HOURS (MACHINE HOURS + LABOR HOURS)*	A high number indicates improved productivity
Machine downtime level	Indicates the percentage of machinery that was unavailable during the working time	*NUMBER OF HOURS OF MACHINE DOWNTIME/NUMBER OF HOURS OF MACHINE AVAILABILITY*	A low number indicates downtime reduction

4 Validation

4.1 Scenario

We developed the model through a study (Table 1) of the enterprise 'INDUSTRIAL MINING AND SOLUTION (IMS) SAC,' a metalworking SME located in Peru. IMS SAC is engaged in the production of mining sludge parts, which is prepared according to customer specifications. It also provides comprehensive technical solutions for various systems in mining, such as slurry water pumps.

4.2 Initial Diagnostic Values

We assessed the current state of the enterprise through an analysis of the sand casting production during an 8-year interval. For this purpose, we collected enterprise data from 2010 to 2018.

Figure 4 shows that production has remained relatively constant. However,

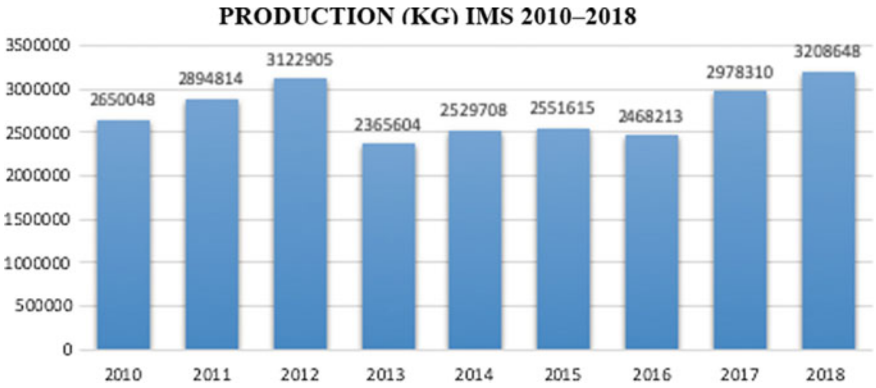


Fig. 4 Production (2010–2018)

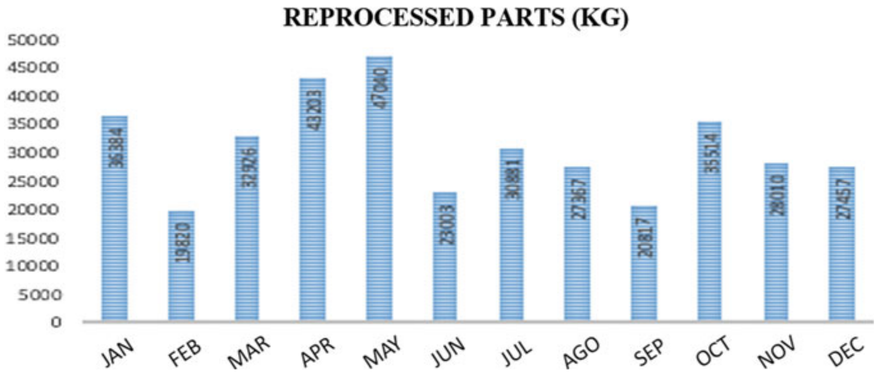


Fig. 5 Reprocessed parts (kg)

production recorded a trend of improvement with variations of **20%** in the 2016–2017 period and of 10% in the 2017–2018 period. This analysis shows that production improvement was a result of various factors, such as increased sales (domestic and/or international market).

Figure 5 shows a variable trend of reprocessed pieces; however, during the reported period a total of 372.42 kg of reprocessed pieces was registered.

4.3 Implementation of the Proposed Method

As mentioned above, in order to develop our proposal, we created a current state value stream map for the enterprise under study (Fig. 6).

After identifying problems in various areas, we conducted research on similar proposals to apply them to our case study (Fig. 7).

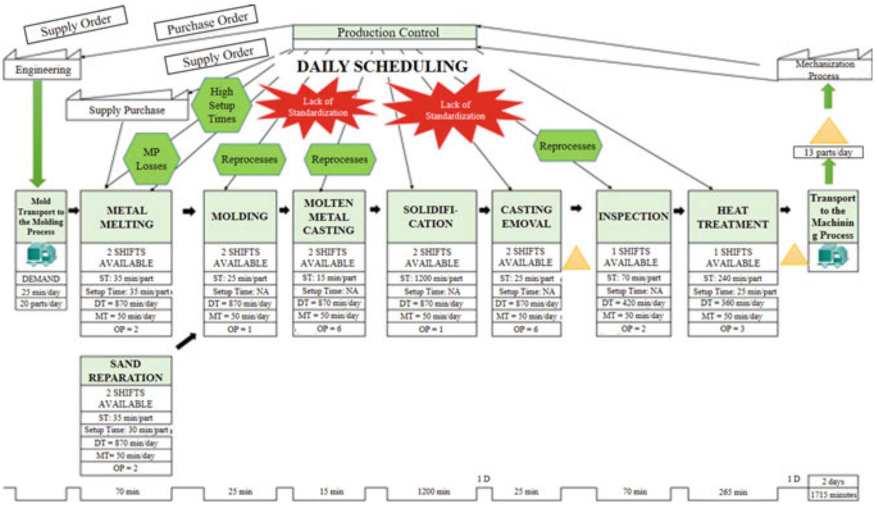


Fig. 6 Value stream mapping

Fig. 7 Left (before) Right (after)

Finally, we proposed that Kanban and CONWIP systems be used to control the entire production area because of their card information system, which is applied for each process (Fig. 8).

Then, we validated the model using Arena simulation software [11] based on the enterprise's time study. We, thus, validated the proposed improvements and compared current state indicators with improved state indicators (Table 2).

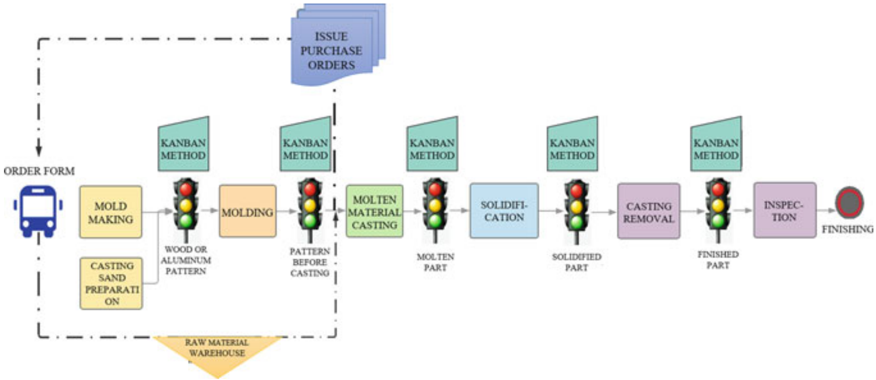


Fig. 8 Kanban/CONWIP model

Table 2 Current state versus improved state indicators

Indicator	Diagnostic (%)	Result (%)	Variation (%)
Work-in-progress inventory level	35	28	25
On-time delivery	72	78	10
Total productivity	74	80	8
Machine downtime level	45	39	15

5 Conclusions

The Kanban/CONWIP methodology allows the production and assembly stages to be synchronized in manufacturing plants, through a change in the way of performing and organizing the work. Fundamentally, the implementation of this methodology is the training of personnel. The implementation of the SMED tool allowed to reduce the number of machines stops, for maintenance or stops, since the company did not have a base model to carry out the implementation. In terms of production control, the use of the above tools resulted in an increase in productivity (20%), reduction in the WIP inventory level (50%), and improvement of on-time deliveries (40%).

References

1. Sociedad nacional de industrias (SNI).: Reporte Sectorial- Industria metalmecánica. https://www.sni.org.pe/octubre-2018-reporte-sectorial-metalmecanica/ (2018)
2. Reporte sectorial del Instituto de Estudios Económicos y Sociales (IEES).: Reporte Sectorial-Industria metalmecánica. https://www.sni.org.pe/octubre-2018-reporte-sectorial-metalmecanica/ (2018)

3. Saldarriaga, J.: Metalmecánica prevé tres años de vacas flacas en la minera, El Comercio. https://elcomercio.pe/economia/peru/metalmecanica-preve-tres-anos-vacas-flacas-mineria195491/ (2015)
4. Ramachandran, G.M., Neelakrishnan, S.: Un enfoque para mejorar la atención al cliente la entrega a tiempo con la fecha promesa original. Sudáfrica J. Ind. Eng. **28**, 109–119 (2017)
5. Dhiravidamani, P., Ramkumar, A.S., Ponnambalam, S.G.: Nachiappan Subramanian Implementación del sistema de manufactura esbelta y auditoría magra en una industria de fabricación de autopartes: un estudio de caso industrial. Int. J. Comput. Integ. Manuf. (2017)
6. Kumar, S., Dhingra, A., Singh, B.: Process improvement through Lean-Kaizen using value stream map: a case study in India. Int. J. Adv. Manuf. Technol. **96** (2018). https://doi.org/10.1007/s00170-018-1684-8
7. Kukla, S.: Quality and Safety assurance of iron casts and manufacturing processes. Arch. Foundry Eng. **16** (2016). https://doi.org/10.1515/afe-2016-0019
8. Júnior, R., Nunes, A., Luana, C., Silva, D.: Creating value with less impact: lean, green and eco-efficiency in a metalworking industry towards a cleaner production. J. Clean. Product. **196** (2018). https://doi.org/10.1016/j.jclepro.2018.06.064
9. Lukina, S., Korshunova, E., Dorozhkin, I.: Methods of automated control over composition and structure of metalworking equipment. MATEC Web of Conferences, 224, 01095 (2018). https://doi.org/10.1051/matecconf/201822401095
10. Carlos Lemos, M., Pereira, T., Pinto Ferreira, L., Silva, F.J.G.: A codification system roadmap: case study in a metalworking company. Proc. Manuf. **17**(2018), 688–695 (2018)
11. Arena simulator, extracted from https://www.arenasimulation.com/

Lean Manufacturing Application in the Laminating Machine Manufacturing Process in a Metalworking Company

Katherine Emilia Camacho Sanchez, Jenner Jose Saavedra Rosales, Yeimy Salvatierra Garcia, and Grimaldo Wilfredo Quispe Santivañez

Abstract This study describes an improvement implementation in the manufacturing process of a laminating machine using lean manufacturing tools. After identifying the activities that do not add value to the process, the Ishikawa diagram and Pareto diagram were applied to determine the potential causes of it. Finally, the visual board, the 5 S, and the poka-yoke were successfully implemented, achieving a 25-day reduction in the manufacturing process time, which represents an efficiency of 36.57%.

Keywords Lean manufacturing · Metalworking company · Visual board · Poka-yoke

1 Introduction

The current situation demands an extremely competitive market wherein customers seek lower costs and expect quick product deliveries [1]; therefore, the progress and technological accessibility have made companies more competitive [2]. The metalworking sector covers a wide range of productive activities from melting to transformation. In this regard, this study aims to design an improvement proposal based on the lean manufacturing methodology, as it is a company's productivity and efficiency improvement philosophy [3] because the ultimate goal is the continuous

K. E. Camacho Sanchez · J. J. Saavedra Rosales · Y. Salvatierra Garcia
Universidad Tecnológica del Perú, Lima, Perú
e-mail: 1130221@utp.edu.pe

J. J. Saavedra Rosales
e-mail: 1420414@utp.edu.pe

Y. Salvatierra Garcia
e-mail: ysalvatier@utp.edu.pe

G. W. Quispe Santivañez (✉)
Universidad Nacional Autónoma Altoandina de Tarma, Tarma 12651, Peru
e-mail: gquispe@unaat.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_42

improvement and customer satisfaction, for which it is very important to work on the waste elimination. This philosophy is speculated by Eiji Toyoda and Taichi Ohno. The term "Lean" is coined in the book "The machine that changed the world" written by Womack, wherein they finally name it as lean manufacturing [4]. The Lean philosophy strongly believes in human talent and not utilizing it properly is considered as its wastage [5].

This article presents the implementation of lean manufacturing using three tools: 5S, visual board, and poka-yoke. In this regard, the used methodology is explained further and finally, the results, conclusions, and recommendations are presented.

2 Methods

The following tools will be used for the following investigation: Visual Management or Visual Board, 5 S, Poka-Yoke, Value Stream Mapping (VSM), and Genchi Genbutsu.

2.1 Value Stream Mapping

It is a tool in which the sequence and flow of the material and information of the value chain are shown using icons and graphics. It is a guide to map the company's current situation [6].

2.2 Suppliers Inputs Process Outputs Customers (SIPOC)

It is a tool that describes the inputs, outputs, customers, and other attributes that are involved in a process's inherent activities [7]. Prior to taking any action toward a process, it must be thoroughly analyzed [8].

2.3 Thread Diagram

It is a graphic tool that starts from a scale plan or model where the path of workers, materials, or equipment is measured with a thread during a given succession of events [9].

2.4 Genchi Gembutsu

It refers to being present at the place of origin and visualizing the events that occur to comprehend the real situation of the problem [10].

2.5 Ishikawa Diagram

It serves as a vehicle to help work teams to have a consensus regarding a complex problem, with all the elements and relationships clearly visible at any level of detail required [11]. The participants contribute ideas about the causes that can produce the effects and these are recorded and illustrated in the diagram, and subsequently, the causes are rearranged in a hierarchical manner and the repeated ones are then eliminated [12].

2.6 Pareto Diagram

The Pareto diagram is used to graphically separate the significant aspects of a problem from the trivial ones so that a team knows where to direct their efforts for improvement. The Pareto law or the 80/20 rule states that 80% of the result is generated by 20% of the causes [13].

2.7 5 S

The 5S allows establishing the foundations for continuous improvement. This tool goes beyond cleaning; it must be considered more like a work philosophy than one of life [14].

2.8 Visual Board

It is defined as the delivery of quick and understandable visual information of all operations to raise awareness and see how it can be improved; however, the participation of all collaborators is required because improvement is obtained through the sequential suggestion of ideas [15].

2.9 Poka-Yoke

This tool is used to avoid errors within the processes and functions of the organization. They could be simple errors; however, they can generate waste, and bottlenecks, consequently leading to customer dissatisfaction [16].

3 Study Case

The target company of this study manufactures machines for the graphics industry. It has been in the Peruvian market for 33 years and has participated in national and international graphic fairs. In that regard, the primary problem is the delayed delivery of machines to customers. Therefore, the requirement to implement the lean manufacturing philosophy arises.

3.1 Current VSM Development

Figure 1 shows the current VSM of the laminating machine. From this, it is concluded manufacturing the laminating machine requires 67 days.

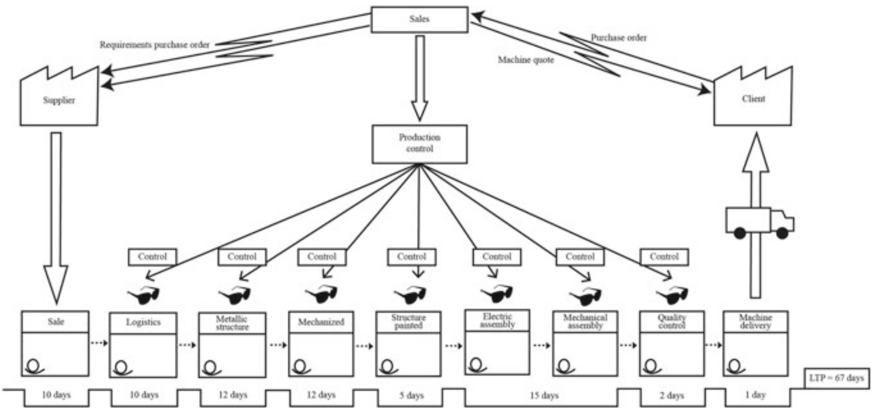


Fig. 1 Initial VSM of the laminating machine

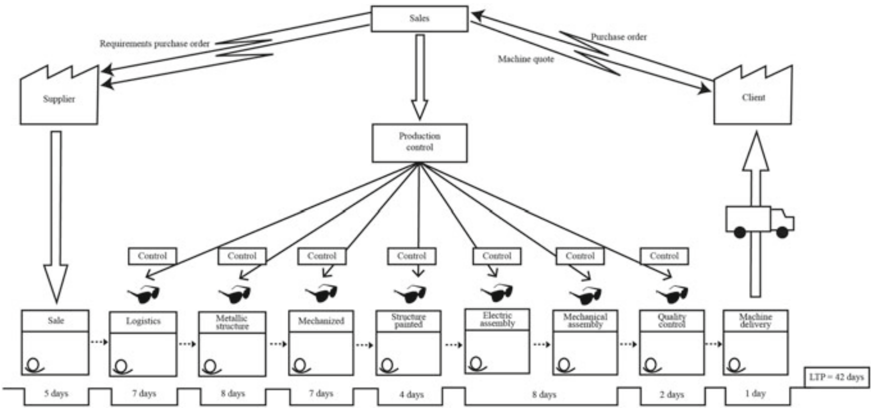


Fig. 2 VSM based on the improvement proposal

3.2 Future VSM Development

As a result of the improvement proposal, the time to manufacture the laminating machine will be 42 days, thereby achieving a reduction of 25 days. Figure 2 shows that the processes maintain the order in a similar manner as in the current VSM because standard machines cannot be altered.

Efficiency has been obtained owing to the decrease in manufacturing days. Currently, the efficiency is 98% as the manufacturing days have decreased from 67 to 42 days, thereby generating an efficiency increase of 36.57%, which allows the workers to work on other machines or advance more in the second laminating machine to be manufactured.

3.3 Visual Board Implementation

The implemented visual board is divided into nine areas and its purpose is that collaborators in all areas have a transversal knowledge of the machine manufacturing processes. Under this premise, order and planning with respect time and working method are sought as they will be able to know what the manufacturing progress is with respect to other areas.

3.4 Poka-Yoke Implementation

The checklist was used as a tool for the Poka-Yoke. This list details all the necessary components and tools to perform the correct assembly of the laminating machine.

Figure 3 shows the checklist of the electrical area, which comprises 76 materials and allows the electrical and logistics area to work in a more organized manner, and Fig. 4 shows the checklist of the mechanical area, which comprises 53 materials that must be supplied by the logistics area as soon as possible.

The area prior to applying the 5 S is shown in Fig. 5. Figure 6 shows the application of the 5 S.

	SEMIAUTOMATIC LAMINATING MACHINE CHECKLIST						
Machine						DATE:	
Client							
Country						July 10th, 2019	
Purchase order date		General aspects		Three-Phase 220 VAC 11KVA RESISTANCE		CHECKED BY:	
	LOGISTICS - ELECTRICAL AREA						
ITEM	Material Description	Qty.	Brand	Model	Material delivery date	Observations	✓
1	40×60 mm wireway	4 m	FOTEK	--------		2×45.8 2×77.5 1×40.7 1×55.3	
2	8X1/2" self-tapping + 1/2" flat washer	38	--------	--------		24:wireway 10:channel 4:solid state roll base	
3	6×1/2" self-tapping + 1/2" flat washer	4	--------	--------		For solid state relay base fan	
4	5×3/4" self-tapping + 1/2" flat washer	4	--------	--------		For speed variator support	
5	5/12×1 1/2" bolt + 5/12" flat washers + 5/12" nut + 5/32" pressure washers	4	--------	--------		To hold three-phase power wrench	

Fig. 3 Checklist of the electrical area

AUTOMATIC LAMINATING MACHINE CHECKLIST						
Machine					DATE:	
Client						
Country					July 10th, 2019	
Purchase order date		General aspects	Three-Phase 220 VAC 11KVA RESISTANCE		CHECKED BY:	
LOGISTICS - NATIONAL PRODUCTS - MECHANICAL AREA						
ITEM	Material Description	Qty.	Material	Material delivery date	Observations	✓
1	Chain guide	4	Fe			
2	Stop speaker	1	Fe			
3	Main Silicone Roller Rail	1	Fe			
4	Pressure arm system	2	Fe			
5	Segmented entry silicone roller	1	Fe			
6	Chrome-plated pressure roller	1	Fe			
7	Silicone pressure roller	1	Fe			

Fig. 4 Checklist of the mechanical area

Fig. 5 Mechanical and electrical area prior to using 5 S

Fig. 6 5 S application in the mechanical and electrical area

4 Results and Discussion

Table 1 shows the results obtained from the improvement implemented in the study. The total time for the manufacturing process was reduced by 25 days, which represents an increase in production of 0.6 more machines due to the availability of additional days to manufacture more machines. In terms of efficiency, this represents an increase of 36.57% and consequently, an increase of 27,636 in overall sales is observed.

Table 1 Global results

Process indicators	Original value	New value	Improvement
Total time	67 days	42 days	25 days
Production	1 machine	1.6 machines	0.6 machines
Efficiency	61.43%	98%	36.57%
Sales revenue	PEN 46,060.00	PEN 73,696.00	PEN 27,636.00

5 Conclusions

The analysis allowed identifying the main causes of the problematic situation, and these are disorder in workspaces, lack of procedures, and lack of knowledge of the process as a whole. Lean manufacturing tools allowed to reduce the total time of the manufacturing process of the rolling machine from 67 to 42 days.

As future work, the results will be expanded with a methodology and simulations to provide the best possible alternatives in the manufacturing process of rolling machines to obtain better results.

References

1. Antonio, G., Bedolla, J., Chiabert, P.: A novel methodology to integrate manufacturing execution systems with the lean manufacturing approach. Proc. Manuf. **11**(2017), 2243–2251 (2017)
2. Alkhoraif, A., Rashid, H., McLaughlin, P.: Lean implementation in small and medium enterprises: literature review. Oper. Res. Perspect. **6**(2019), 1–19 (2019)
3. Marulanda, N., González, H.: Objectives and operational strategic decisions as support for lean manufacturing. Dimensión Empresarial **16**(1), 29–46 (2017)
4. León, G., Marulanda, N., Gonzales, H.: Factores claves de éxito en la implementación de Lean Manufacturing en algunas empresas con sede en Colombia. Revista de la Facultad de Ciencias Económicas y Administrativas **18**(1), 85–100 (2017)
5. Rojas, A., Gisbert, V.: Lean manufacturing: Herramienta para mejorar la productividad en las Empresas. 3C Empresa, 116–124 (2017)
6. Tejeda, A.: Mejoras de lean manufacturing en los sistemas productivos. Ciencia y tecnología, **45**(2), 276–310 (2011)
7. Felizzola, H., Luna, C.: Lean Six Sigma in small and medium enterprises: a methodological approach. Revista Chilena De Ingenieria, **22**(2), 263–277 (2014). https://doi.org/10.4067/S0718-33052014000200012
8. Torres, C.A.: Orientaciones para implementar una gestión basada en procesos. Ingeniería Industrial **32**(2), 159–171 (2014)
9. Puma, G.: Propuesta de redistribución de planta y mejoramiento de la producción para la empresa. 150 (2011). Retrieved from https://www.mendeley.com/library/
10. Ibarra, V., Ballesteros, L.: Manufactura Esbelta. Ciencia y Tecnológica. **53**, 54–58. Manufacturing, L. Lean manufacturing : implantación 5 S. 5(Edición 20), 16–26 (2016). Medianas, P.Y., Colombianas, E., Carlos, J., Lora, T.: D, (32), 41–77 (2012)
11. Romero Bermúdez, E., Camacho, JD.: El uso del diagrama causa-efecto en el análisis de casos, **40**(3–4), 127–142. Retrieved from https://www.redalyc.org/pdf/270/27018888005.pdf (2010)

12. Ruiz-Falcó Rojas, A.: Herramientas de Calidad, Universidad Pontificia Comillas ICAI, 1–70 (2009)
13. González, J.: Ley de Pareto: 80/20. 1–25. (2017). Retrieved from www.jggomez.eu
14. Ramírez, M., Gisbert, V.: Lean manufacturing: Implantacion 5 S. 3C Tecnologia **20**(5), 16–26 (2016)
15. Sarria, M., Fonseca, G., Bocanegra.: Modelo metodológico de implementación de Lean Manufacturing. Revista EAN **83**, 51–71 (2017)
16. Ibarra, V., Ballesteros, L.: Manufactura Esbelta. Ciencia y Tecnológica **53**, 54–58. Manufacturing, L.: Lean manufacturing: implantación 5 S. 5(Edición 20) (2016)

The Perspectives of Integration Between Lean Manufacturing and Industry 4.0

Luana Spósito Valamede and Alessandra Cristina Santos Akkari

Abstract Lean manufacturing (LM) has become the leading approach to develope efficient processes in the industry since the early 1990s. Nowadays, LM is widely recognized and accepted in organizational configuration and refers to the integration of employees in the manufacturing process, which is focused on continuous improvement and value-adding activities to avoid waste. However, a new paradigm called Industry 4.0 has recently emerged in the manufacturing sector, based on intelligent systems and integrations between machines, products, information, and individuals across the value chain. In this context, this work aimed to study how these two approaches can coexist and support each other in the digital era. Through the systematic literature review, data standardization was developed to identify how LM can integrate with a productive environment based on smart factories, cyber-physical systems, IoT, and IoS. To study the principles, technologies, and attributes from Industry 4.0 and LM, it was used the academic databases Web of Science, Scopus, and ScienceDirect, with a time restriction for articles publication from 2012 to 2018. Among the results, it was observed that three relationship perspectives were identified, consisting of (i) LM as an Industry 4.0 facilitator, (ii) the potentiation of Lean Manufacturing by smart factories, (iii) and the mutual correlation between Industry 4.0 and LM. Based on the interpretation of these perspectives, it was concluded that while smart factories will facilitate and encourage the promotion of Lean principles, LM will be the base to implement and manage the changes which came with the Fourth Industrial Revolution.

Keywords Industry 4.0 · Lean manufacturing · Smart factories

L. S. Valamede · A. C. S. Akkari (✉)
Mackenzie Presbyterian University, Campinas, Brazil
e-mail: alessandra.akkari@mackenzie.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_43

1 Introduction

Artificial intelligence, high-performance computing, robotics, and virtual simulations are some technological advances brought by the Fourth-Generation Industrial Revolution. Based on concepts Smart Factory, Cyber-physical Systems (CPS), Internet of Things (IoT) and Internet of Services (IoS), the so-called Industry 4.0 will afford the process optimization, increased incomes, and consequently, companies' profits [1].

The benefits and innovations resulting from these changes will provide many impacts [2, 3]. With these changes, will current production systems be able to maintain themselves in this new production environment and get integration with new technologies?

Among these production systems, there is lean manufacturing (LM), originated from the Toyota Production System (TPS). The LM principles are widely used in various productive sectors, such as financial, automotive, and electromechanical and aim to promote the continuous improvement of organizational activities, focusing on the elimination of waste, on the value stream of the entire production chain and on getting excellence through customer perspectives [2].

When it is compared to LM and Industry 4.0, the divergent points between the two approaches are highlighted, once these are fundamentally different. While the first keep processes as simple as possible and reduce system variability by focusing on mass production; the second one promotes the use of complex Information Technologies (IT) solutions for machines, people and processes, and is characterized by flexibility and mass customization. However, the objective of constantly adding value to production systems is equivalent in both ideologies.

In this sense, this paper aims to study the possibility of these two concepts coexisting and the contributions that they can provide to each other, through the identification and standardization of information of Industry 4.0 and LM principles and attributes.

2 Theoretical Framework

Wyrwicka and Mrugalska [4] define LM as a dynamic process guided by a system of principles and practices, aiming to achieve continuous improvement through the elimination of waste.

In the beginning, LM was only focused on reducing the number of unproductive practices. However, when this approach has popularized, it has started to encompass many other aspects of manufacturing from the life cycle stages of products, such as product development, procurement, and distribution management [4].

LM is known for simplicity and high efficiency, and it is considered a popular production system since the 1990s. Also, LM still has been seeing as a foundation of productive organizations, and it is applied across a range of business domains.

This integrated management practice is able to relate quality processes, work teams, customer, and supplier networks in an integrated manner, which becomes one of the most important concepts that help companies gain competitive advantages in the world market [5].

In order to put into practice the characteristic philosophy of LM, it is necessary a joint effort of all the employees to implement lean actions in the organizational environment. Also, the organization must work in an integrated way to establish the LM system, using tools such as 5S, Kanban, Poka-Yoke, Value Stream Mapping (VSM), Kaizen, Total Productive Maintenance (TPM), Single-Minute Exchange of Die (SMED), and Andon.

Still in the context of manufacturing, nowadays, the Fourth Industrial Revolution has been highlighted, which provides digitization and its incorporation in industrial systems, incurring the optimization of the shop floor, revenue increase, and maximization of profitability. This transformation, which impacts on competitiveness, society, and economy, has been named as Industry 4.0 by members of the German government at a press conference during the Hannover Trade Fair event in 2011 [2].

According to the newly formed “Industrie 4.0 Working Group,” organizations will be able to establish global networks, which incorporate their machines, storage systems, and production facilities in a CPS form. In fact, CPS acts along with the intelligent factories, IoT, IoS, which constitute the four pillars that sustain the Fourth Industrial Revolution [2].

CPS consists of the collaboration of computational entities, which are in intensive connection with the surrounding physical world and its ongoing processes. These entities provide and use, at the same time, the access and data processing services available on the Internet [6].

Due to the new production logic, the smart factory consists of vertical integration of various components in order to deploy a flexible and reconfigurable manufacturing system. Its structure incorporates industrial network control terminals that oversee the intelligent shop floor objects, which enable the CPS to monitor the entire production chain, identify defects and make corrections through autonomous decisions and cooperation between agents to achieve a high level of efficiency and quality [7].

IoT refers to global and dynamic network infrastructure. There are several self-configuring elements, which operate in an integrated manner and are dependent on sensory communication, networks, and information processing technologies. In this system, physical and virtual “objects” have identities, physical attributes, and virtual personalities, and use intelligent interfaces [8, 9].

Finally, IoS is able to build and deliver a large number of new types of services, combining those that are available separately on the Web to form value-added and customer-perceivable services [1].

3 Methodology

This study performed a standardization of data through a systematic literature review, which involved the search for current and relevant articles in the field of LM and Industry 4.0. It was used to bibliographic research the academic databases Web of Science, Scopus, and ScienceDirect. These databases were used due to their popularity and availability of a large number of papers.

All identified and selected articles have been transferred to the Mendeley, a management software used for their control and classification to avoid possible duplication of articles that were published in more than one searched database.

The time restriction from 2012 to 2018 was applied. The search fields used to choose the articles were limited to the abstract, title, and keywords, which resulted in the following search string: Title-Summary-Keyword ("Industry 4.0" or "Fourth Industrial Revolution" or "Smart Factory") and ("Lean Production" or "Lean Manufacturing" or "Lean Management"). The keywords used to conduct this research were broad enough to not restrict the search, and they aimed to approach the different objectives established by this work. Only articles in English were considered to compose this study.

Once introduced to the LM concept and its relation with TPS, it was explained the importance of Lean approach use on product systems, and studied the most notorious Lean tools for this work (5S, Kanban, Poka-Yoke, VSM, Kaizen, TPM, SMED, and Andon).

About the Fourth Industrial Revolution, it was evidenced by the changes caused by Industry 4.0 in the context of manufacturing. The four pillars that consist the basis 4.0 were explained, consisting of smart factories, CPS, IoT, and IoS.

The exclusion criterion selected articles that adopted an Industry 4.0 reference model and/or its concepts, framework, and approach. Therefore, articles that had a superficial or partial approach to Industry 4.0 were excluded.

Finally, the perspectives of integration between LM and Industry 4.0 were discussed, based on the characterization of each approach.

4 Results and Discussion

It was observed that the combination between LM and Industry 4.0 allows three different interpretations of this subject. While the first describes LM as the basis or one prerequisite to implement technologies 4.0, the second one envisions Industry 4.0 as a conclusion of LM or as a potential for increasing LM approach efficiency. The third perspective treats about the mutual correlation between Industry 4.0 and LM, which generates positive synergies and is influenced by the two previous interpretations [10, 11]. These three perspectives will be discussed during the next subsections of this study.

In fact, the third interpretation presents the most complete analysis of the three associations between these two concepts, once it points out the contributions that one can add to the other. Methods and principles of LM emphasize the simplicity and standardization of organizational processes by aiming at processes control and continuous improvement. This facilitates the entry of new technologies and solutions into production flows. One the other hand, technologies 4.0 will boost lean practices, because they are able to collect and process varied and voluminous data with high speed and security.

4.1 Lean Manufacturing as an Industry 4.0 Facilitator

LM acts to introduce processes with efficiency, waste-free, defined standards, and high customer focus. Through the implementation of lean methods, planning of the entire production flow can be performed, which contributes directly to the identification, classification, and organization of all constituent processes. The introduction of new technologies is facilitated by the management of lean methods and continuous improvement activities. This contributes to the insertion of innovations 4.0.

In order to be consistent with future production systems, which already have the emergent pillars and technologies from the 4th Industrial Revolution, organizations should define their production processes in an efficient way. Therefore, once this step is completed, these institutions will be able to automate their processes. Then, to develop the smart factories, it is necessary the definition of processes, suppliers, customers, tasks, and times. Similarly, standardized, transparent, and reproducible processes are fundamental for the introduction of Industry 4.0 [2, 11].

The lean approach supports the installation of innovations by reducing product and process complexity, which enables the cost-effective use of technologies 4.0. Moreover, by contributing in a mutual way, the implementation of horizontal and vertical networks by Industry 4.0 provides better integration of customers and suppliers in the value-added process [2, 11].

Furthermore, lean does not exclude automation. Taiichi Ohno, the TPS pioneer, stated that production processes should be automated and supervised by workers, and he called this principle as Jidoka or autonomous, considered the second pillar of TPS. Jidoka corresponds to principles of the 4th Industrial Revolution, whereby human beings, with emerging and innovative technologies, will assume the same role [12].

By analyzing the relationship between LM and Industry 4.0 and the influence of this association on an operational performance level of organizations, a study made with 110 companies indicated that those have low degrees of maturity in a production system and of interaction with Industry 4.0. In addition, once the lean systems' maturity level provides greater awareness of underlying practices and principles, it can be concluded that LM is an important variable in an association process with the Fourth Industrial Revolution. Therefore, having a solid lean system, organizations will benefit from Industry 4.0′s technology implementation through increased operational performance [13].

4.2 Industry 4.0 Potentializes Lean Manufacturing

According to this perspective, it is considered that the application of new information and communication technologies introduced by the Fourth Industrial Revolution in existing LM systems can improve the performance of LM, which stabilizes and refines production and logistics processes more efficiently.

The influence of Industry 4.0 on LM can also be seen as a factor to generate new business models and as a great impact on operational factors. Basing on successful cases of Industry 4.0 implementation in multinational organizations, such as Audi, Volkswagen, ThyssenKrupp, and Wolfkran, the insertion of solutions using emerging technologies, such as big data, enabled significant improvements in Just-in-Time system, waste reduction and in quality improvement [14].

At ThyssenKrupp, one of the world leaders in the elevator segment, a fusion between the physical world and data networks has been identified, which created a CPS during the production of elevator parts. Using CPS, the company reports sustainable gains in production management, maintenance, logistics processes, reduction of nonconformities, and energy savings [14].

According to the lean approach, when the flow strengthens and non-value added activities are minimized, it causes cost savings. Therefore, any effort made to reduce these wastes is compensated in terms of reducing operating costs. In the current context, this effort comes from digitizing and integrating resources, which is done by Industry 4.0. In addition to transforming factories in smart and integrated production systems, the financial benefits provided by the Fourth Industrial Revolution are also generated by reducing or eliminating waste. With advanced information and communication systems in place, along with a lean operating structure, the industry must have the potential to expand to new horizons.

As an example, with real-time data collection, that improves the transparency and quality of information, smart factories will be able to attend the fluctuating market demand, an advantage that exceeds the traditional LM’s stable production. As a result, flexibility increased by Industry 4.0 will provide these organizations to work with highly complex production systems [11].

4.3 Mutual Correlation Between Industry 4.0 and Lean Manufacturing

As a result of the combination between the two previous perspectives, the correlation of Industry 4.0 and LM provides mutual contribution and can be attributed to similarities regarding objectives, central pillars, and principles of the lean system and Fourth Industrial Revolution.

Based on the possibility that Industry 4.0 and LM can coexist and support each other, some terms have been developed to refer to the combination of these two approaches such as Lean 4.0, Lean Automation, Smart Lean Manufacturing, and Lean Industry 4.0 [11].

With markets growing due to the virtualization of the purchasing process, the guidance focused on customer requests, one of the LM principles, will remain essential during the 4$^{\text{th}}$ Industrial Revolution. At the same time, the influence of customers will increase, due to the introduction of emerging technologies, which causes more complex demands (such as shorter lead times and higher product quality) that cannot be attended just with manual lean methods. Thus, these methods must be complemented by technologies 4.0 to produce for these new demands [10].

In addition to customer focus, process improvements through the elimination of waste continuous improvement and creation of fluid and pulled production lines will be indispensable for the smart factory's implementation. Emerging technologies, in turn, must support and improve the Lean methods already used for process improvement. Also, to analyze and optimize the stages of current production systems, the performance of professionals with knowledge of these new technologies and their functions and possibilities, as well as the lean tools already employed, becomes crucial [10].

About the socio-technical factors resulting from the interaction between LM and Industry 4.0, it is considered the systems that involve a complex interaction between people, machines, and the environmental aspects of organizational systems. While the conventional relationship of a management system predominantly controls workers, the association between LM and Industry 4.0 will lead to active employee participation, providing a mutual transfer of knowledge between management and operational levels. Thus, management decisions will be optimized based on the operators' shared knowledge. Factory floor workers, in turn, will no longer be passive agents who perform their tasks without any reference. The role of this new type of knowledge worker will be multifaceted and will include ideas from each distinct discipline to maintain alignment within the heterogeneous Industry 4.0 network. The Fourth Industrial Revolution is characterized by providing greater freedom in the technical–social relationship, which was not conceivable in the context of conventional technologies [15].

5 Conclusion

The present work aimed to study the possibility of integration between the concepts of Industry 4.0 and lean manufacturing. Through bibliographic research for data standardization, the principles, attributes, tools, and practices that LM and Industry 4.0 were identified and interpreted to analyze the potential for integration between these two concepts. Then, there were identified three integration possibilities, consisting of (i) LM as a facilitator for Industry 4.0, (ii) the potentiation of lean manufacturing by smart factories, (iii) and the mutual correlation between Industry 4.0 and LM.

It can be concluded that both concepts facilitate decentralized structures over large and heterogeneous systems, sharing the common goal of providing small production modules, which have easy integration and low complexity.

By focusing on the simplicity of machines and workstations to eliminate waste and unnecessary activities in production processes, LM speeds up production flow, facilitates the digitization of manufacturing steps, and emphasizes visual and transparent control, making it easier for the identification of failures. These changes, brought by Lean principles, suggest that the implementation of LM contributes significantly to the development of Industry 4.0, as advanced manufacturing technologies will be underpinned by Lean practices to maximize production performance.

On the other hand, the introduction of smart factories helps increase the maturity of LM. It is noteworthy that for the success of the Just-in-Time pillar, the information generated must be accurate, up-to-date, and shared every moment. Through IoT and IoS, the supply chain can be digitized to provide real-time inventory levels, equipment status, and product location data.

However, it is necessary to promote further research to a deeper understanding of improvements and adaptations in the evolution of technologies 4.0 on LM influence. In addition, Industry 4.0 academic studies used in this paper were purely theory-driven, not readily adaptable to an application.

References

1. Thuemmler, C., Bai, C.: Health 4.0: How Virtualization and Big Data are Revolutionizing Healthcare. Springer International Publishing, Switzerland, pp. 23–37 (2017)
2. Dombrowski, U., Richter, T., Krenkel, P.: Interdependencies of industrie 4.0 & Lean production systems: a use cases analysis. Proc. Manuf. **11**, 1061–1068 (2017)
3. Wagner, T., Herrmann, C., Thiede, S.: Industry 4.0 impacts on lean production systems. Proc. CIRP **63**, 125–131 (2017)
4. Wyrwicka, M.K., Mrugalska, B.: Mirages of lean manufacturing in practice. Proc. Eng. **182**, 780–785 (2017)
5. Kolberg, D., Zühlke, D.: Lean automation enabled by industry 4.0 technologies. IFAC-Papers OnLine, **28**(3), 1870–1875 (2015)
6. Navickas, V., Kuznetsova, S.A., Gruzauskas, V.: Cyber-physical systems expression in industry 4.0 context. Financ. Credit Act. Theory Pract. **2**(23), 188–197 (2017)

7. Wang, S., Li, D., Zhang, C.: Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination, Comput. Netw. **101**, 158–168 (2016)
8. Kagermann, H.: Change through digitization–value creation in the age of industry 4.0. In: Albach, H., Meffert, H., Pinkwart, A., Ralf, R. (eds.) Management of Permanent Change. Springer Fachmedien Wiesbaden, Berlin, pp. 23–32 (2015)
9. Da Xu, L., He, W., Li, S.: Internet of things in industries: a survey. IEEE Trans. Ind. Inf. **10**(4), 2233–2243 (2014)
10. Lock, C., Reinhart, G., Bauer, H., Brandl, F., Lock, C., Reinhart, G.: Integration of industrie 4.0 in lean manufacturing learning factories. Proc. Manuf. **23**, 147–152 (2018)
11. Mayr, A., Weigelt, M., Kühl, A., Grimm, S., Erll, A., Potzel, M., Franke, J.: Lean 4.0-a conceptual conjunction of lean management and Industry 4.0. Proc. CIRP **72**, 622–628 (2018)
12. Buer, S., Strandhagen, J., Chan, F.T.S.: The link between industry 4. 0 and lean manufacturing: mapping current research and establishing a research agenda. Int. J. Prod. Res. **56**(8), 2924–2940 (2018)
13. Prinz, C., Kreggenfeld, N., Kuhlenkötter, B.: Lean meets industry 4.0–a practical approach to interlink the and method world and cyber-physical world. Proc. Manuf. **23**, 21–26 (2018)
14. Fettermann, D.C., Cavalcante, C.G.S., de Almeida, T.D., Tortorella, G.L.: How does industry 4.0 contribute to operations management? J. Ind. Prod. Eng. **35**(4), 255–268 (2018)
15. Davies, R., Coole, T., Smith, A.: Review of socio-technical considerations to ensure successful implementation of industry 4.0. Proc. Manuf. **11**, 1288–1295 (2017)

Agile Project Implementation Methodology for Introducing New Technologies in Industry 4.0

William F. Franco and Frank Herman Behrens

Abstract The Industry 4.0 era promises to position companies at a new competitive level by changing production scales and being able to connect factories to consumer needs. To this end, it is necessary to incorporate new technologies into the production process, which enable and transform the way of producing and managing, seeking productivity and generating qualified jobs. The main problem encountered by managers and professionals involved in industry modernization activities is the difficulty of implementing projects for the adoption of these new technologies, for example, where to start introducing these technologies, how to make the connection between new technologies and existing processes, and, finally, how to convince their peers and their leadership in choosing the best projects for companies where they work. This paper aims to demonstrate how agile methodology can be used in project management and execution for the introduction of new technologies, with innovation and quick response, proving its effectiveness and applicability in the context of industrial modernization required by Industry 4.0. As a practical example, this work describes and analyzes a real case study, a project for the introduction of a collaborative robot in a production line for the product packaging step.

Keywords Industry 4.0 · Agile methodology · Project management · Collaborative robot

1 Introduction

The industry has changed dramatically in the last decade thanks to the exponential advancement and the maturation of various disruptive technologies applicable to the industrial environment. New sensors, wireless communication, automation,

W. F. Franco · F. H. Behrens (✉)
Pontifical Catholic University of Campinas, Campinas, SP 13087-571, Brazil
e-mail: frank@puc-campinas.edu.br

W. F. Franco
e-mail: william.ff@puccampinas.edu.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_44

cyber-physical systems, task robotization, real-time data analysis, cloud computing, artificial intelligence, etc., [1–3] are the new tools at industry disposal that positively impact productivity. Additionally, the rational use of resources and energy, and the economic and environmental sustainability of production processes make the industry more efficient and more profitable, positioning it in a new level of competitiveness.

Many experts call this industry-level shift "The Fourth Industrial Revolution" known worldwide as Industry 4.0 or as advanced manufacturing in the USA [4]. Making Industry 4.0 a reality implies the use of enabling technologies, which are based on the digitization of information, the direct communication between machines, products, and people, and the creation of automatic computational models for production management based on data that integrate horizontally throughout the production chain [4, 5]. This work demonstrates how the agile methodology can be used in project management and execution as a tool to introduce new technologies into the industrial environment without impact on the ongoing production, proving its effectiveness and applicability in the context of the Industry 4.0. The example of installing a collaborative robot in a production line is described and analyzed.

2 Background

For the introduction of new enabling technologies, innovation in industries is a key factor. Employees working in this new industry need to have in mind which new and disruptive technologies form the backbone of Industry 4.0 [6].

2.1 The Importance of Industry 4.0

The term Industry 4.0 was first used in 2011 at the Hannover fair in Germany. In essence, it aims to integrate the real world with the virtual one, incorporating in its context data science, the Internet, and information technology. In other words, Industry 4.0 can be defined as factories that become intelligent by adopting new enabling technologies and bringing together different technology fields to be incorporated into this new industry.

Countries such as Germany and the USA have major companies that have focused their efforts on developing and advancing smart factories in the context of Industry 4.0. The Boston Consulting Group (BCG) has researched this topic involving 617 companies, with revenues over \$50 million. The result of this study revealed that almost 50% of German companies have already developed initial Industry 4.0 projects and that American companies are also working hard on this subject. When asked what the plans would be for the next 1 or 2 years, 40% of US companies said they have projects on new technologies and about 60% of German companies have the same strategy [7]. Industry 4.0 will bring productivity and cost savings over a short time. Long-term advantages will be better demand visibility, end-to-end automation of the supply and production chains, resource and energy optimization, waste reduction, and less environmental impact.

2.2 The Enabling Technologies for Industry 4.0

In BCG research, nine enabling technologies for Industry 4.0 were identified [5, 7].

- **Advanced Manufacturing Solutions**. Implementation of advanced, highly flexible, high performance, interconnected and modular production systems. Among the main systems in use is the robotics applied to the concept of collaborative robots.
- **Additive Manufacturing**. The widespread use of 3D printing for prototyping, maintenance parts, and devices for continuous improvement, coupled with digital development software tools.
- **Augmented Reality**. Visualization systems to support production processes, creating new human–machine interfaces to display production parameters and data.
- **Simulation**. Use of software tools connected to physical systems to simulate applications and virtually test solutions for increased productivity and optimization of operations.
- **Integrated Systems**. Data and information integration across all areas of the supply and production chains, from suppliers to end-users.
- **Industrial Internet of Things**. Multidirectional communication between production elements, sensor devices, and actuators, both internal and external to the machines, through data communication via the internet.
- **Cloud Computing**. Implementation of computing solutions and open systems data management techniques hosted in the cloud.
- **Cyber Security**. New security rules to protect data against unauthorized access from the various internal and external system connections.
- **Big Data Analytics**. Analysis of large amounts of data from all elements of the production chain, aiming to optimize production processes.

2.3 Industry Participation in Country Indicators

First world countries are at the forefront of using innovative technologies and methodologies in their respective industries. Also, the industrial parks of these countries are more developed compared to the Brazilian scenario. This situation distances Brazil from the application of Industry 4.0 technologies and, consequently, increases the challenge of the Brazilian industry to remain competitive, as first world countries are at higher levels of Industry 4.0 maturity.

Industry 4.0′s new paradigm will make industries more productive. Thus, companies that execute a larger number of internal projects aiming at the incorporation of new technologies in their production processes will, consequently, have more competitive costs. Therefore, Brazilian industries need to react quickly toward adopting the new world paradigm to remain competitive at the world level.

Considering that the industrial sector in 2017 was responsible for 21% of the Brazilian Gross Domestic Product (GDP) [8], the effect of not adopting new technologies and the lack of investments in Industry 4.0 projects will compromise our economy.

It should also be taken into account that Brazil, as a member of the group of countries called BRICS, had a prominent position in the world economic scenario until 2014. However, in recent years, the country has lost importance in the ranking of competitiveness in relation to the BRICS members and other more competitive countries, ranking the 80th position in the 2017–2018 report [9], partly due to the economic crisis that began in 2014, but also due to the lack of investment in adequately modernizing its industry.

Thus, the adoption of the concepts and technologies used in Industry 4.0 may reverse this scenario, restoring the relevance of the industrial sector and the importance to the country in the world economy. The above context indicates the industrial sector needs to setup aggressive plans and speed up the execution of projects to implement new technologies in the industry to return to the global competitiveness level. Agile project methodologies can be a fundamental tool to ensure assertiveness and productivity in the management of innovation projects toward the Industry 4.0 level.

2.4 Agile Project Methodology

Agile methodology has the principle of working on an adaptive approach to solve complex problems. It focuses on productivity through communication and planning, which gives teams the freedom to discover different ways to design solutions. It also provides a more efficient decision-making process in case a design change is required. Additional advantages include reduced cost due to constant communication and increased quality, ensuring that all teams are aware of problems and changes.

One of the most widely used agile methods for managing and developing software projects is Scrum [10]. In this method, the project is divided into parts or blocks, called Sprints, which in turn have a series of activities for planning, construction, testing, and review, maximizing deliverables for project progress.

At the end of each Sprint, a project step is produced, a portion of the project with some useful functionality is delivered to the customer (making a small delivery), and the team meets to review, provide feedback, and plan new activities. At these meetings, the effectiveness and the quality of activity delivery, as well as the performance of the tests, are evaluated. Feedback serves to measure customer satisfaction with small deliveries, which allows adjustments to solutions and deliveries at the end of each Sprint.

3 Development

To investigate if the application of the agile methodology is effective in the execution of a project to introduce new technology in the production process, a case study was proposed aiming at the application of a collaborative robot for packing products in a production line. The objective is to verify that the Scrum method meets the requirements for industrial project development, as it is widely used in software development but not widely known among production, automation, and mechanical engineering professionals.

3.1 The Case Study

The stage of the industrial process chosen for application of the agile methodology was the packaging of products in the last step of a production line. In the original packaging process shown in Fig. 1, an employee manually places the finished products in the shipping box.

Scrum method was used to provide an automated solution to the process where a set of product units are positioned in the shipping box, replacing manual human action with mechanized action performed by the collaborative robot. At the factory, there are several similar manual packaging jobs and, once this prototype has been proven successful, the solution can easily be replicated in other production steps.

Fig. 1 Manual packing process

3.2 The Productivity Indicator

Among the various performance indicators of a production line, the productivity indicator is fundamental for the industry to remain competitive. Productivity can be measured by the ratio between the number of parts produced and the number of employees working in this line, according to Eq. (1).

$$\text{Productivity} = \frac{\text{Number of parts produced}}{\text{Number of employees}} \tag{1}$$

The production line has 33 employees working in three production shifts, 24 h a day. As the line produces 40 pieces per minute, 57,600 parts are produced in one day. Applying Eq. (1), daily productivity with manual operation (Fig. 1) is 1746 pieces per employee.

3.3 The Sprint Example in the Case Study

An example of applying the Scrum method to illustrate the application of the agile methodology was the Sprint for the definition of the gripper tool for the collaborative robot model UR3e (Universal Robots). A one-week-long Sprint was designed to deliver the definition of the best collaborative robot gripper for the application in the selected case study. The gripper is a tool to be developed and installed on the robot arm, responsible for taking some product units that arrive by the conveyor belt and properly accommodating them inside the shipping box.

One of the challenges of this Sprint was to assess whether the gripper should be designed as a sophisticated mechanism for grasping and moving a set of products, mechanically adjusting them to its final position. Alternatively, they could be positioned close and simply slide to the shipping box by gravity.

The validated solution at the end of Sprint was the evidence that a unit package wrapped by plastic cellophane slides easily by gravity without friction at the time of transfer, as shown in Fig. 2.

Validation examples like the one above allowed the executing team to choose the best solution for the project. Also, agile methodology gave them the freedom to create and innovate, allowing certain functionalities to be quickly tested and validated in advance, and then to choose the best way to move forward.

The result of this Sprint was fundamental for understanding the possibilities for project definition. It has been found that the speed of collaborative robot cycles is low compared to the production line speed, requiring an optimized gripper layout with a shorter working course and simultaneous feeding of two shipping boxes to achieve the proper project cycle speed. The production line produces 40 pieces per minute while the robot's speed is 14 s to perform one cycle handling 5 units. If we feed only one shipping box, the robot would reach a speed of 22 pieces per minute. Doubling the gripper tool to handle 10 units each robot cycle, we can feed two shipping boxes simultaneously and reach 44 pieces per minute, meeting the speed of production line.

Fig. 2 Testing the gravity sliding of a set of product units to the shipping box

4 Results

The speed of implementation of new technology introduction projects for Industry 4.0, using the agile methodology, was proven in the experiment of deploying a collaborative robot in the product packaging stage, and the results were satisfactory.

With testing and deliveries of Sprints tasks, the project was taking shape and the team felt engaged, following the methodology, exceeding the expectation of innovation, agility, and project deliverables.

All machine elements of the gripper tool incorporated into the collaborative robot arm were modeled in 3D, with the aid of specific computer design software, to automatically obtain all parts orders for machining, 3D printing commands, and the final assembly drawings. Figure 3a shows the final design aspect of the gripper tool.

Also, to avoid design errors, Sprints were performed to simulate the solutions, aiming to understand the assemblies and analyze the interactions between the parts. Figure 3b shows a view of the gripper tool attached to the robot arm, to study the possible movements of this set considering the requirements of the product packaging application.

Finally, the 3D model was extended to incorporate the design of the robot gripper assembly installation into the actual factory layout, as shown in Fig. 4a, to visualize any kind of interference the designed solution could cause in the production space. Figure 4b shows a view of the collaborative robot installed on the production line.

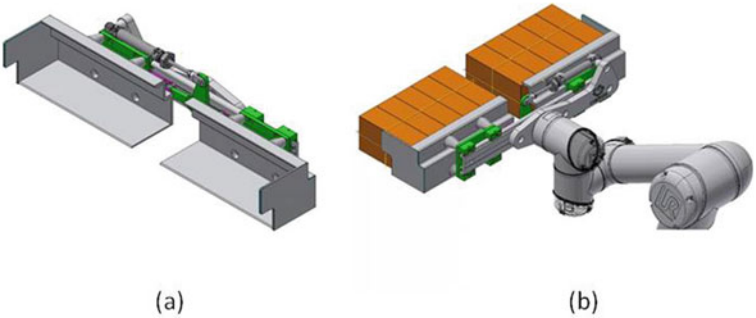


Fig. 3. 3D modeling: **a** front view of the robot gripper tool and **b** the gripper assembled in the robot arm

Fig. 4 Final implementation: **a** deployment 3D simulation in the factory layout and **b** the robot finally installed on the production line

All of the tasks performed on the various Scrum Sprints helped the team to be more productive in anticipating errors and delivering the project in a much shorter timeframe of 53 days, rather than the 120 days in traditional design methodology.

The collaborative robot implemented in the production line achieved daily productivity of 1920 pieces per employee. With the automated process, the line now has 30 employees, reducing three people from the original team, which means a 9% increase in productivity over manual packaging.

5 Conclusions

This work allowed us to evaluate the use of agile methodology in the introduction of new Industry 4.0 technologies in factories. As a case study, we designed the implementation of a collaborative robot for product packaging, aiming to increase productivity. The project execution time was reduced by about 55% when compared to the traditional methodology. Besides, the use of agile methodology has enhanced innovation in the tasks performed by the project team and has helped to change the mindset required for the use of new technologies in traditional manufacturing.

The speed in the execution of a project introducing a new technology in the industrial environment, such as the case study analyzed, is a clear demonstration of the benefits obtained with the adoption of the agile methodology. Examples like this can be replicated in several other practical situations, to drive the implementation of Industry 4.0 by Brazilian industries.

Therefore, it is important that industry professionals qualify for new design methodologies, as they are valuable and necessary tools for the resumption of growth and modernization of the Brazilian industry.

References

1. Lu, Y.: Industry 4.0: a survey on technologies, applications and open research issues. J. Indus. Inf. Integ. **6**, 1–10 (2017). https://doi.org/10.1016/j.jii.2017.04.005
2. Yan, J., Meng, Y., Lu L., Li, L.: Industrial big data on an Industry 4.0 environment: challenges, schemes and applications for predictive maintenance. IEEE Access **5**, 23484–23491 (2017). https://doi.org/10.1109/ACCESS.2017.2765544
3. Jazdi, N.: Cyber physical systems in the context of Industry 4.0. In: 2014 IEEE International Conference on Automation, Quality and Testing, Robotics, pp. 1–4. IEEE Press, Cluj-Napoca, Romania (2014). DOI: https://doi.org/10.1109/AQTR.2014.6857843
4. Lasi, H., Fettke, P., Kemper, H. G., Feld T., Hoffmann, M.: Industry 4.0, Business & Information Systems Engineering, **6**(4), 239–242 (2014). https://doi.org/10.1007/s12599-014-0334-4
5. Wan, J., Cai, H., Zhou, K.: Industrie 4.0: enabling technologies. In: Proceedings of 2015 International Conference on Intelligent Computing and Internet of Things, pp. 135–145. IEEE Press (2015). https://doi.org/10.1109/ICAIOT.2015.7111555
6. Kagermann, H., Wahlster, W., Helbig, J.: Recommendations for implementing the strategic initiative INDUSTRIE 4.0—securing the future of German manufacturing industry. Final report of the Industrie 4.0 Working Group. Forschungsunion (2013)
7. The Boston Consulting Group. https://www.bcg.com/capabilities/operations/embracing-industry-4.0-rediscovering-growth.aspx
8. CNI, Brazilian Industry in Figures. https://www.portaldaindustria.com.br/estatisticas/industry-in-figures/
9. BRICS—Brazil, Russia, India, China. South Africa. https://www.itamaraty.gov.br/pt-BR/politica-externa/mecanismos-inter-regionais/3672-brics
10. Srivastava, A., Bhardwaj, S., Saraswat, S.: SCRUM model for agile methodology. In: International Conference on Computing, Communication and Automation (ICCCA), IEEE Press, Great Noida, India (2017). https://doi.org/10.1109/CCAA.2017.8229928

Factors Limiting the Industrial Development of Marine Macroalgae in the District of Marcona, Perú

Santos Edgard Sebastián Bernuy, Karina Silvera Sanchez, and Grimaldo Wilfredo Quispe Santivañez

Abstract The aim of the present research was to examine the factors that limit the industrial development of marine macroalgae in the district of Marcona, Peru. To achieve this, we considered several variables that characterize its industrial development, such as exploitable biomass volume, extraction methods, transformation processes, aquaculture, and government regulation, correlating each variable under the Pearson's coefficient. Hence, it was determined that the factors most associated with the industrial development of macroalgae in Marcona, Peru are extraction and transformation processes, because they exhibit a higher degree of correlation.

Keywords Marine macroalgae · Industrial development · Pearson's correlation

1 Introduction

Most species of algae, a worldwide scale, are exploited taking advantage of both the biomass naturally carried to the beach and directly harvested. In Peru, algae are exported and used as raw material, in the alginate, carrageenan and agar industry; and to a lesser extent, consumed as human food. During the last decade, the growing economic importance of the resource has led to increasing levels of exploitation. According to the Institute of the Sea of Peru (IMARPE) [1]. In Peru, brown algae (mainly *aracanto*) are being used as raw material for alginate extraction. The extraction is mainly supported by two species: *Lessonia nigrescens* and *L. trabeculata.*

S. E. S. Bernuy
Universidad Tecnológica del Perú, Lima, Perú
e-mail: 1330577@utp.edu.pe

K. S. Sanchez · G. W. Quispe Santivañez (✉)
Facultad de Ingenieria, Universidad Continental, Huancayo, Perú
e-mail: gquispe@continental.edu.pe

K. S. Sanchez
e-mail: 1410805@utp.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_45

Brown algae as a raw material for alginate extraction has historically been supported by the collection of individuals resulting from natural mortality.

According to the United Nations Food and Agriculture Organization (FAO) [2], the worldwide seaweed industry offers a variety of products, the total annual production value is estimated between the US $5500 and the US $6000 million, with a volume of 11.3 million tons. The most commonly used hydrocolloids are alginate, from brown algae, for the pharmaceutical industry and textile production; Agar–agar, from red algae; the gelling agent that is used for crops and in the food industry, and finally Carrageen, of red algae, which is used for dairy production and meat production. The Latin American algal industry plays an important role worldwide, and approximately 17% of the industrialized algae is obtained from Chile, which, besides Peru, is the only country in the region where seaweed culture is a highly developed activity. Unlike Argentina, Brazil, and Mexico, where algae processing plants exist, algae are exploited and processed only in Chile [3]. The largest volumes of marine macroalgae production in the department of Ica, Peru, takes place in the cities at 'San Juan de Marcona' and 'Pisco'.

A few years ago, the authorities from Perú have regulated the marine macroalgae fishery, unlike other countries, such as Chile, Spain, and China, according to the National Fishing Society (SNP) [4]. Furthermore, there is no government regulation that encourages state investment to industrialize this resource. Compared with other countries within the region, Peru has a low level of development in macroalgae aquaculture.

According to FAO [5], as 96% of the production of seaweed is obtained from aquaculture, the macroalgae extraction/harvest should be associated with crops in order to ensure that this resource is sustainable and become a successful industry.

This work seeks to identify the factors limiting the industrial development of marine macroalgae in the district of Marcona, Ica, Peru. To address it, the analysis of the growth of Peruvian investment regard this resource serves to perform a comparative study between the local scenario and other countries of the world.

2 Problem Statement

This study was conducted between April and July 2008, and it found that the main commercially extracted macroalgae species are *Macrocystis pyrifera*, *L. trabeculata*, and *L. nigrescens* and that they are mainly distributed to the south of 'Punta San Juan', located at the city of the 'San Juan de Marcona', Ica, Perú [6]. Furthermore, it was determined that the size of commercially extracted plants showed a high diameter, beyond the 20 cm, whose diameter allows the efficient extraction of this resource. Based on these results, government regulations were suggested for the extraction of brown macroalgae owing to their economic importance in the center and southern coastal areas.

The Regional Directorate for Production of the department of Ica (DIREPRO-Ica) states that 18.8% of formally registered artisanal fishermen worked directly in the extraction and commercialization of brown macroalgae in 2013 [7]. In 2008, the ministry of production established a closed season to protect the biodiversity of the Peruvian sea, and the general fisheries regulation was modified the next year in order to pave the way for legal framework regards production, harvest techniques and species sustainability in the activity of extraction. However, despite the established restrictions and regulations, there have been several changes during this period that have modified the direct and indirect stakeholders involved in the brown macroalgae production chain. According to the consulting firm ICON-INSTITUTE [8] in responding to the growing demand for the resource, the IMARPE regional headquarters-Pisco Ica developed strategies with scientific support, and pretend to lay the foundations to administer this kind of fishery, safeguard its conservation and the economic efficiency of its exploitation. For this reason, as 'San Juan de Marcona' city has characterized by their fishing community [9], where associations are involved in artisanal activities for harvesting fish, shellfish, and seaweed, this work analyzes the problematic situation regards the industrial development of marine macroalgae and serves as a reference by regulators for improving the extraction process of this place.

3 Methodology

The methodology used in this research is based on a search for data from entities related to the algae exploitation sector available on government entities such as IMARPE[1] (Sea Institute of Perú) and SINIA[2] (National System of Environmental Information). Also, open-access alga production data from exporting companies in the sector in the period 2005 until 2009 is considered thanks to Veritrade's[3] information. Likewise, to validate the research question the analysis of the degree of correlation of the selected variables into the industrial development of marine macroalgae is employed by means of the structural equation model [10].

In this work, the structural model is used to represent a network of relationships in the form of a set of linear equations that link endogenous latent variables with exogenous latent variables. It involves the transfer of the measurement of such constructs to a set of linear equations that define the relationships between the constructs (structural model), via the surface variables (measurement model), and the possible correlations between the different variables, As shown in Fig. 1, the proposed structural model indicates latent variables endogenous to the factors of production and exploitation of algae and exogenous variables to factors that limit industrial development such as technology and manufacturing activity and variables such as investment by private companies and the sustainable environment.

[1]The oficial site of the governamental entity is: https://www.imarpe.gob.pe/.

[2]See Footnote 1.

[3]The oficial site of this enterprise is: https://www.veritradecorp.com/.

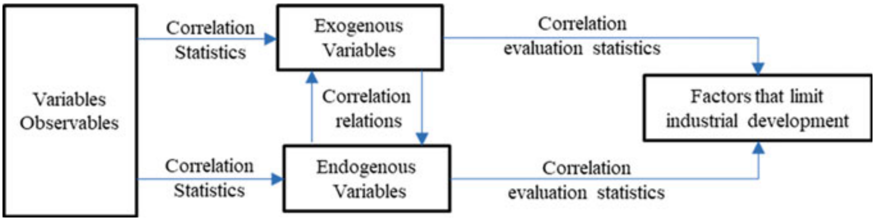


Fig. 1 Proposed causal model

Table 1 Tables of dependent factor

Macroalgae production volumes in Marcona (ton/year)					
Year	2005	2006	2007	2008	2009
Volume 'metric tons'	402.94	26.05	749.99	348.4	0
Marine macroalgae extraction methods					
Year	2006	2007	2008		
Weighting coefficient	0.4	0.5	0.5		
Raw material transformation					
Year	2005	2006	2007	2008	2009
Weighting coefficient	0.8	0.8	0.85	0.85	0.85
The volume of macroalgae produced through aquaculture					
Year	2005	2006	2007	2008	2009
Volume 'metric tons'	500	3118	9213	12,107	3874
Government regulations in Peru					
Year	2005	2006	2007	2008	2009
Weighting coefficient	0.1	0.1	0.1	0.1	0.6

Table 1 presents the values along the years about the chosen dependent factors. As part of the proposed structural model, the respective relationships between are described below.

3.1 Exogenous Surface Variables

*C*1: Investment of private companies, *C*2: Environmental Sustainability.

3.2 Endogenous Surface Variables

$X1$: Volumes of exploitable biomass, $X2$: Extraction methods, $X3$: Transformation processes, $X4$: Aquaculture, $X5$: Regulations.

3.3 Latent Exogenous Variable

$Y1$: Limited industrial development.

3.4 Endogenous Latent Variables

$F1$: Technology, $F2$: Manufacturing activity.

3.5 Exogenous Latent Variable Relationship with Endogenous Latent Variable

$H1$: Limited industrial development with respect to technology, $H2$: Limited industrial development with respect to manufacturing activity.

3.6 Endogenous Latent Variable Relationship with Endogenous Latent Variable

$H3$: Technology with respect to manufacturing activity.

3.7 Relations of the Exogenous Surface Variables with the Exogenous Latent Variable

$M1$: Investment of private companies with respect to industrial development, $M2$: Environmental sustainability with respect to industrial development.

3.8 Relationships of Endogenous Surface Variables with Endogenous Latent Variables

$A1$: Volumes of exploitable biomass with respect to technology, $A2$: Extraction methods with respect to technology, $A3$: Transformation processes with respect to manufacturing activity, $A4$: Aquaculture with respect to manufacturing activity, $A5$: Regulations with respect to manufacturing activity.

The mathematical relationship between them are described below:

$$F1 = A1 - A2 \tag{1}$$

$$F1 = A3 - A4 - A5 \tag{2}$$

$$H1 = Y1 - F1 \tag{3}$$

$$H2 = Y1 - F2 \tag{4}$$

$$M1 = Y1 - C1 \tag{5}$$

$$M2 = Y1 - C2 \tag{6}$$

4 Results

4.1 Results Obtained

Table 2 shows the factors most associated with the industrial development of macroalgae in Marcona, Perú. These are the production volume, extraction methods, transformation processes, and government regulations because they exhibited a higher degree of correlation according to Pearson's correlation coefficient [11].

Table 2 Correlation of the variables using Pearson's correlation coefficient

	Technology	The intervention of private companies/year-records	PT manufacturing activity/year
Production volumes	− 0.35	− 0.25	0.83
Extraction	0.50	0.33	1.00
Transformation	0.76	0.33	1.00
Aquaculture	0.61	0.61	− 0.15
Regulations	0.53	− 0.04	0.93

4.2 Interpretation of the Results of the Structural Model

The variables and their relationships are described in Table 3 according to the data obtained in Table 1. Also, Table 4 presents the estimation and evaluation of exogenous variables.

The following results were obtained after applying Eqs. 1, 2, 3, 4, 5 and 6.

- $M1$: The factor that limits the industry is related to the investment of private companies that impact the result of $M1$ with a value of 42.79
- $M2$: The factor that limits the industry is related to the environmental sustainability that impacts the result of $M2$ with a value of 3685.08

Therefore, the variables that limit most of the industrial development are aquaculture, technology, and regulations, this is because the results obtained in the structural equation are greater than the volume and extraction methods. Finally, Fig. 2. Shows the resulted causal model.

Table 3 Estimation and evaluation of the model

	Dependent variables		Mean	Variance	Typical deviation
Technology	$A1$	Production volumes	154.84	60,724.97	246.42
	$A2$	Method of extraction	2.40	3.85	1.96
Manufacturing Activity	$A3$	Transformation processes	567.08	325,083.51	570.16
	$A4$	Aquaculture	4,164.00	15,489,681.25	3,935.69
	$A5$	Regulations	625.11	414,234.41	643.61
Technology	$H3$	Manufacturing activity	627.25	411,567.69	641.54

Table 4 Estimation and evaluation of exogenous surface variables

Exogenous surface variables		Mean	Variance	Typical deviation
$C1$	Investment of private companies	276.50	48.75	6.98
$C2$	Environmental Sustainability	8,146.00	14,036,264.67	3,746.50

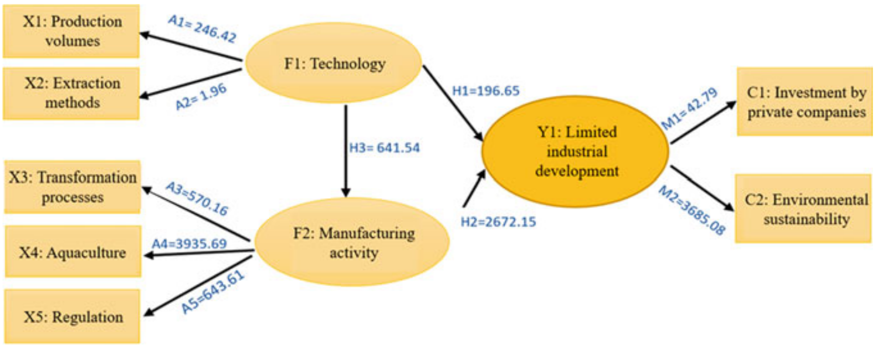


Fig. 2 Causal model

5 Conclusions

Macroalgae are crucial for manufacturing industrial products, and currently, they are considered a highly profitable resource worldwide. We concluded that the following factors limit the industrial development of marine macroalgae are volume, extraction methods, transformation processes, and regulations; furthermore, extraction methods and transformation processes exhibited a stronger correlation, according to the Pearson's coefficient. With respect to the structural model, the highest standard deviation was observed in the aquaculture variable, which indicates that macroalgae aquaculture is important for sustainable industrial development in the future, leading to furtive work on seaweed production hatcheries.

References

1. Instituto del Mar del Perú IMARPE.: Estudios sobre macroalgas pardas en el sur del Perú. 2011–2015 (ISSN 0378-7702) (2012) https://biblioimarpe.imarpe.gob.pe:8080/handle/123456789/21817. Last accessed 07 June 2019
2. FAO.: Global aquaculture outlook in the next decades: an analysis of national aquaculture production forecasts to 2030. FAO Fisheries Circular No. C1001, Roma (2004) https://www.fao.org/3/a-y5648e.pdf. Last accessed 07 June 2019
3. Moya Ortiz, J.: Centro de investigación y desarrollo tecnológico en algas. Disponible en (2012). https://repositorio.uchile.cl/handle/2250/100374
4. Sociedad Nacional de Pesca SNP. Aprueban reglamento de ordenamiento pesquero de las macroalgas marinas y modifican reglamento de de la ley general de pesca aprobado por decreto supremo nº 012–2001-pe y el reglamento de inspecciones y sanciones pesqueras y acuicolas aprobado por decreto supremo nº 016–2007-PRODUCE (2009). https://snp.org.pe/docs/9%20ROP%20MACROALGAS.pdf. Last accessed 07 June 2019
5. FAO.: El estado mundial de la pesca y la acuicultura 2018. Cumplir los objetivos de desarrollo sostenible, Roma (2018). https://www.fao.org/3/I9540es/i9540es.pdf. Last accessed 07 June 2019
6. Salavarría Palma, E.A.: Análisis de la variabilidad genética de Macrocystis spp. (Laminariales) en la costa centro sur del Perú, empleando marcadores mitocondriales (2014)

7. "Gobierno Regional de Ica", Regionica.gob.pe (2020). https://www.regionica.gob.pe/pdf/transparencia_2018/resoluciones/direpro/rd_642_direpro.pdf. Last accessed: 06 May 2019
8. ICON-INSTITUT.: Estudio de investigación de poblaciones y de las condiciones de viabilidad ecológica de las actividades extractivas de algas pardas e invertebrados en la zona costera sur, en apoyo a la investigación y desarrollo del Instituto del mar del Perú (IMARPE), (2009). https://www.mincetur.gob.pe/wp-content/uploads/documentos/comercio_exterior/Sites/ueperu/licitacion/pdfs/Informes/156.pdf. Last accessed: 10 May 2019
9. Contreras, C., Tostes.: La Gestión de interesados en la recolección y comercialización de Macroalgas Pardas en el distrito de Marcona, Nazca, Perú (2018)
10. Portillo, E., Hernández Gómez, M.T., Jesús, A., Ortega, E.V., Martínez, M.G.: Modelos de ecuaciones estructurales: características, fases, construcción, aplicación y resultados. Ciencia Trabajo **18**(55), 16–22 (2016). https://doi.org/10.4067/S0718-24492016000100004
11. Restrepo, L.F., González, J.: From pearson to Spearman. Revista Colombiana De Ciencias Pecuarias **20**(2), 183–192 (2007)

Hardware Modeling Challenges Regarding Application-Focused PCB Designs in Industry 4.0 and IoT Conceptual Environments

Edson Camilo, Yuzo Iano, Gabriel Gomes de Oliveira, Diego Pajuelo, Ana Carolina Borges Monteiro, and Reinaldo Padilha França

Abstract Industry 4.0 brings with it a lot of technology in the areas of communication, artificial intelligence, Internet of Things (IoT), and others. One of the great challenges for all these activities is in the design of the hardware including not only the electronic electron project but also the printed circuit board (PCB) design. Working with signals in gigahertz all the tracks of signals, paths, layers, place of components, and processes involved affecting the final product that the Industry 4.0 so longs for. This research presents a detailed discussion along with a description modeling of the main topics that are faced in relation to the challenges of PCB design and manufacturing techniques.

Keywords PCB · Layout · Signal integrity · CISPR 22 · Communication · Signals

E. Camilo (✉) · Y. Iano · G. Gomes de Oliveira · D. Pajuelo · A. C. Borges Monteiro · R. Padilha França
School of Electrical and Computer Engineering (FEEC), University of Campinas, UNICAMP, Av. Albert Einstein, 400, Barão Geraldo, Campinas, SP, Brazil
e-mail: edscamilo@gmail.com

Y. Iano
e-mail: yuzo@decom.fee.unicamp.com

G. Gomes de Oliveira
e-mail: oliveiragomesgabriel@live.com

D. Pajuelo
e-mail: diego.pajuelo.castro@gmail.com

A. C. Borges Monteiro
e-mail: monteiro@decom.fee.unicamp.com

R. Padilha França
e-mail: padilha@decom.fee.unicamp.com

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_46

1 Introduction

Industry 4.0 is a recent term that explains the application of new technologies in major industrial processes. Among the most striking features of this concept are task automation and data and information control. This new concept proposed a paradigm shift in the way industrial processes would be modified to include intelligent and interconnected devices along the entire production chain. It encompasses all the technological innovations created in recent years to facilitate industrial processes and generate more profit for companies, and the term came from a project of technology-focused German government strategies and was first used at the Hannover Fair in 2011 [1, 2].

The basic concept of the Fourth Industrial Revolution understands that by connecting machines, systems, and assets, companies can create intelligent networks throughout their production network and thereby control various actions autonomously. That is, smart factories will have the capacity and autonomy necessary to schedule maintenance, predict process failures, and adapt to unplanned requirements and changes in production [1–3].

For this transformation to come, true systems and technologies had to be developed in the context of automation, digitization, and communication between systems, machines, products, people and the Internet of Things (IoT) become reality. Industry 4.0 has covered a variety of research, from those involving IoT, autonomous cars, low-abstraction, low memory-consuming telecommunications systems [4], powerful and easy-to-use computer systems [5], medical diagnostic methodologies through digital processing techniques for low cost, high precision medical diagnostics [6–8].

The digitization of data from machines, processes, and devices complement the operational layer of an industrial plant. IoT, as it is known, is the technique that allows you to connect general information from devices to the Internet. This makes it possible, within Industry 4.0, to interconnect data and systems, allowing us to form the cyber ecosystem, where we can achieve full and complete interoperability of the industrial plant, where we can call it the digital plant. Therefore, knowing all the processes in Industry 4.0 and the Internet of Things (IoT) is essential for the company to achieve competitive conditions in high quality. The flexible and modular production of this system makes advanced manufacturing so important for the industrial sector because it can produce according to the proposed demand, optimizing inventory and resources [9–11].

In the area of applications and solutions via the Internet, the real race for new business strategies, new forms of commerce, information exchange, and information was available to everyone at the click of a finger on the mobile screen. In times where hardware frequencies were kHz and MHz are facing challenges, due to this era of Industry 4.0 where there are technologies of artificial intelligence, wireless communications, networked electronics, images, sounds, and sensors, making the equipment more and more decisive, the speed of operation in MHz is no longer enough [1, 11].

To work with so much information and in an increasingly short time, our hardware also needs to be more efficient, fast, and robust. The convenience of pushing a button and like a magic pass the gate opens, in a stall of the fingers the light laughs, video cameras connected to the network provide people with real-time images of the office or residence; yet through the network, they speak through video and voice; and in the same sense, the automotive industry already works with automated cars where there is no need for a human driver and just as through the union of IoT and Industry 4.0, the possibilities of the drones with regard to surveillance, delivery of placards among other possibilities, will be increasingly present in people's daily lives [12, 13].

Therefore, this paper aims to provide a discussion on the topic hardware for Industry 4.0 and IoT approaching its particularities, categorizing, and synthesizing potential points of technologies.

2 Methodology

This survey carried out a discussion and description modeling on the main topics that are faced in relation to the challenges of PCB design and manufacturing techniques.

3 Results and Discussion

Nowadays, electronic circuits are everywhere; they are found on the wrist; through smartwatch, in the pocket; they hide in the car; and they are even present in space, since they provide a huge number of applications in electronics by realizing ideas using the right knowledge and tools to do the most extraordinary things imaginable [14].

Printed circuit board (PCBs) are boards dedicated to a special circuit, the biggest advantage of these boards is that they are extremely reliable, can be installed perfectly in chassis, chassis, among other places of a specific design, are smaller electronic components that can shorten your circuit beyond allowing the use of double-sided plates. Nowadays, the circuits are made using computer-aided design (CAD) software, with many programs to choose from, which can be arranged between component terminals and defining the shape of the board in general as a rectangle, but with PCBs, many shapes are possible, such as circles, triangles, rings, stars, among others [15, 16].

Because PCBs are typically designed for different purposes, from project prototyping to the production of rugged, high-tech equipment, this is an important study material for both engineers and companies, serving in the initial scope of the manufacturing process. PCBs are the assumptions of execution, integration, composition, placement, and dimension orientations.

The challenges start soon when the appliance is plugged into the power supply as long as it is desired to have the convenience of being able to connect at 110 or 220 VAC; but for this, a power source has to be prepared for it. The design generally implies a switched source that enables it to be connected within a voltage range, not just one. This technology demands its development because this type of circuitry is an excellent noise generator as well. The problems faced relating to meeting EMC/EMI standards. A layout without due consideration of track spacing, creepage, and clearance the choice of PCB laminate related to CTI, TG parameters leads us to failures by the IEC when it comes to safety.

Failure to consider track widths due to currents, current loops, correct place of components, always minimizing track lengths in this region will cause unwanted noise and will certainly cause the product to fail in the EMI-CISPR 22 tests, electromagnetic interference (EMI) and in the same way as CISPR 22 alludes to an analog standard with respect to the European standard EN 55022, which is generally mentioned in all European electromagnetic compatibility standards, covering ranges in the range from 30 MHz to 1 GHz.

Communication signals such as USB3.0, PCI Express, and signal require impedance parameters for maximum signal transfer from its source to its destination. When we speak of impedance parameters, we are talking about PCB construction processes involving thicknesses between layers, parameters related to the dielectric coefficient of the laminate materials and the welding mask. It was also considering thicknesses of the tracks and their spacings between the signals of the differential pairs, which are in the order of magnitude 6 mils (0.15 mm).

Other no less important constraints should also be considered for critical signals such as keeping the minimum distance between them and all the rest. Other no less important constraints should also be considered for critical signals such as keeping the minimum distance between them and all the rest.

These considerations are valid for USB, PCI Express, TDMI, Serial SATA, LVDS, and others.

All these high-speed signals must also be routed considering the shortest distance between the connection points, every layer change through the roads shakes these signals, because for each row we are introducing another equivalent circuit, some have path limits as, for example USB signals that support at most 2-way (USB3.0). Cross talk is another problem that can appear, and signals considered aggressors must be isolated from the victim signals (Fig. 1).

Usually, the GND planes are used as shielding to protect these signals in both internal planes and planes involving the signals on the TOP and BOTTOM, and signal integrity (SI) considerations are becoming increasingly necessary in PCB projects. Since it needs more and more processing speed, for the memory data signals, the DDR 2, 3, 4, and 5 memories require timing control, which means to route all the signals with controlled track length, without a path, and in a specific layer. USB signals, PCI-Express, HDMI, Serial SATA, SuperSpeed, TDMS, LVDS, command signals, data strobes, data mask, clocks need to be routed with impedance controls, spacings, track widths, number of track lengths.

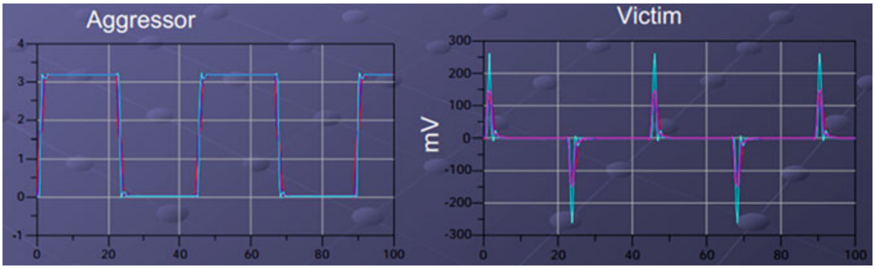


Fig. 1 Cross talk

In this context, differential pairs must be routed according to specific rules of width and spacing of tracks on specific layers with impedance matching. The length of their tracks is also controlled to have a maximum of 5–10 mils of difference between them. This becomes a simple task to do when there is a project with countless signals being processed with this technique. Finding the paths to all of them requires a lot of experience and routing techniques (Fig. 2).

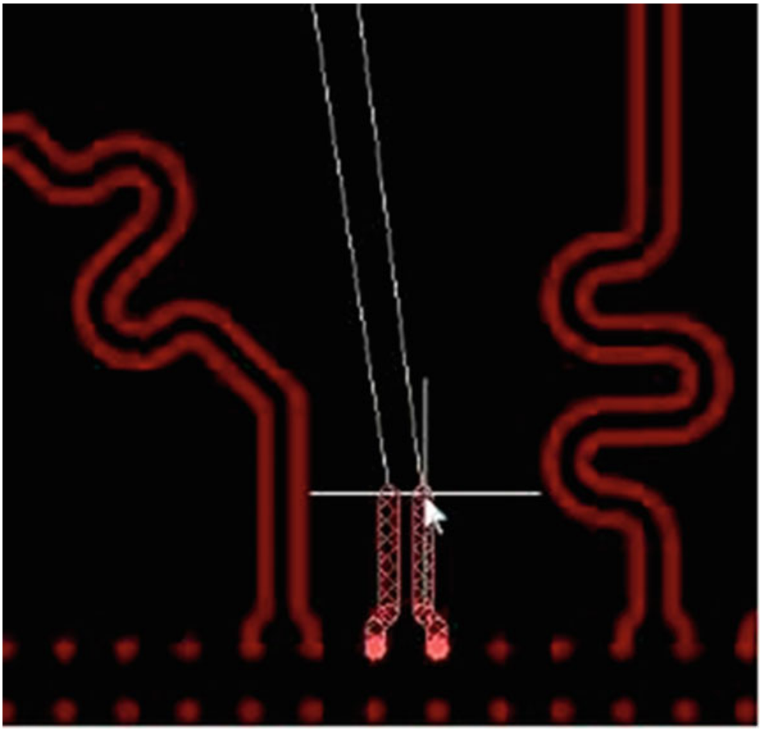


Fig. 2 Differential pair of routing strategy

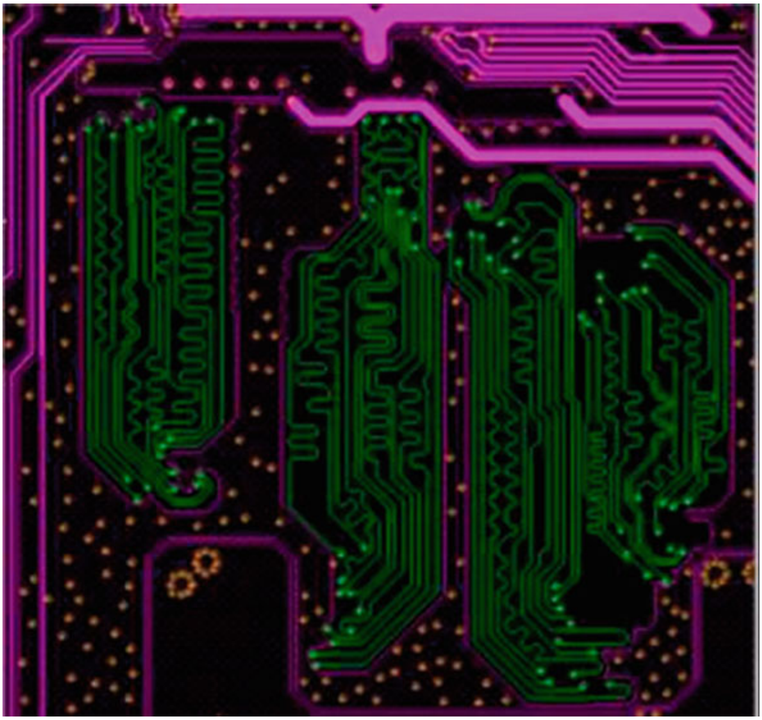


Fig. 3 DDR tracks with match length

Power integrity is part of the challenges, and memories are working with voltages of 1.5 V for DDR3, 1.2 V for DDR4; this means having well-defined the tracing of these feeds in order to avoid losses. After everything is routed and checked, it is possible to perform simulations that are fundamental due to the complexity and costs involved in the manufacture of PCB (Fig. 3).

Trafficking data and communication signals at gigahertz speeds are becoming increasingly critical for correct data transfer either in input or output requirements or in the interpretation or generation of this data.

These signals traveling on trails through the layout PCB undergo deformations in their structure. Signal levels that should represent 0 or 1 end up not having these levels well-defined which causes errors in their interpretation. Such signal deviations and deformations are part of the phenomena that occur in the PCB environment as controlled impedance, whose origin is stacking up and every signal that travels from layer top to layer bottom trough a via will also have a slight deformation at either the rising or falling edge, any signal that travels without reference to a GND plane is also affected, as well as cross-talk effects.

The more defined the rising and falling edges of the signal mean the better interpretation of the data. The better the routing of the tracks observing the critical items related to signal integrity, the better the expected results (Fig. 4).

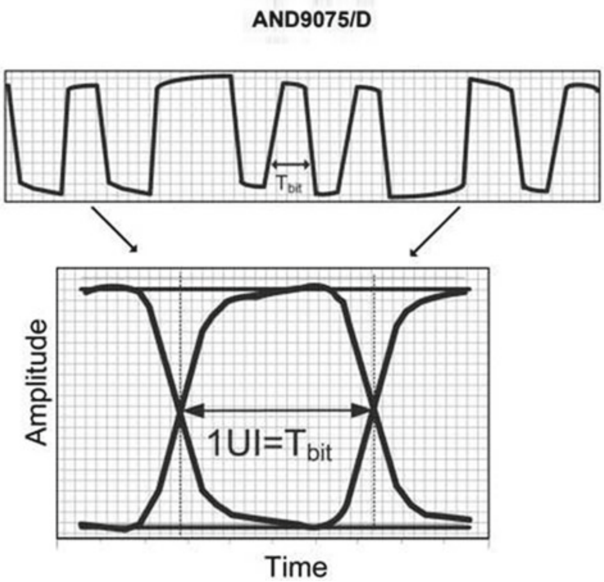


Fig. 4 Typical high-speed signal with eye diagram

It is already known to companies that the technology of PCB fabrication with boring buried and blind routes; however, the industry tries to avoid this technology due to the high costs involved, but when doing the design of electronic circuits with gigahertz frequencies for communication signals and given the use of this technology, it is necessary to keep in mind that there will be parasitic capacitances acting between the stubs of the pathways.

The parasitic capacitance causes signal deformation, impairing the correct interpretation of the data. Through-hole paths as shown in figure should be avoided for high frequency (GHz) (Fig. 5).

It is still the project of this PCB related to manufacturing processes because of these concerns regarding the tracks of 4 mils, on average (0.1 mm), also routes, in this order of magnitude, weld pads geometries, weld mask openings, micro-vias with 4 mils diameter, stencil, test points, positioning of components according to the welding processes.

Remembering that for mass production processes, this means more or fewer costs. The vast majority of products that have wireless communication are mandatory by ANATEL (National Telecommunications Agency—Brazil), if they have passed the electromagnetic interference (EMC)/electromagnetic compatibility (EMI) tests and again face other challenges, gigantic noises are injected through power cables (conducted noise) and others emitted via antennas (irradiated noise); the product also cannot emit conducted or irradiated noise and many are blocked by CISPR 22, and it is impossible to have a product approved by them and other standards International

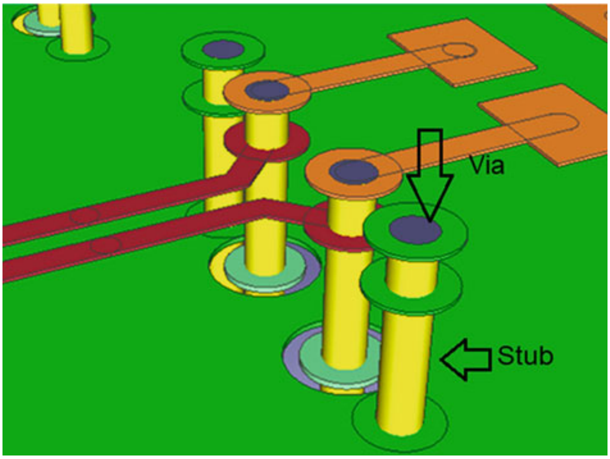


Fig. 5 Stub vias through-hole technology

Electrotechnical Commission (IEC), Underwriters Laboratories Inc. (UL) without a rigorous and meticulous PCB design.

The difference between these standards is that IEC describes minimum specifications on device safety requirements while UL has comprehensive specifications regarding technical details with respect to product safety and application (Fig. 6).

The error detection through CISPR22 is clear but the biggest challenge is to find its cause and propose solutions ranging from changes to the electrical scheme by adding filters but the problem may also be coming from a poor layout of routing, decoupling, and the tracks themselves may be radiating these strong signals. Finding out where they come from is the great challenge of the hardware engineers (Fig. 7).

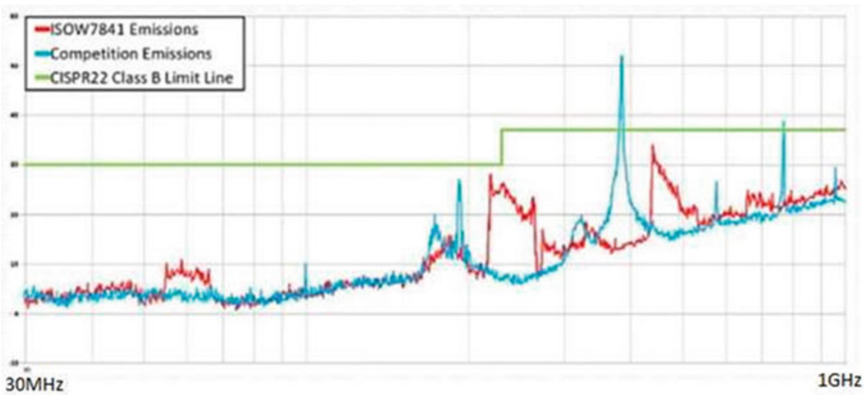


Fig. 6 CISPR 22 test fail-EMI

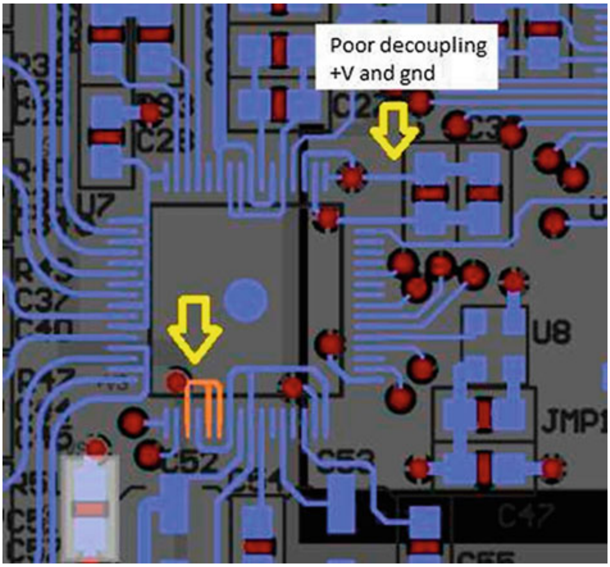


Fig. 7 Poor decoupling layout strategy

4 Conclusions

With cost-effective and highly reliable advantages, providing economical for high volume production, modeling software is color-coded for different connections and is therefore easy to install while mentioning the device rating that aids in its maintenance.

While having drawbacks with regard to layout design, the higher initial cost compared to point-to-point built-in circuits or coiled wires, etching of circuit boards generates chemicals that are hazardous to the environment, yet considering that it cannot be updated once it is printed.

As well as the hardware challenges for Industry 4.0 go beyond tracing the paths and securing their connections and should be considered considerations related to time, controlled impedance, materials, routing strategies, cross talk, PCB manufacturing techniques, EMC/EMI-related techniques, power and signal integrity are becoming increasingly mandatory, as well as Electronic CADs have made a great contribution which facilitating these works strengthening the intellectual knowledge of the HDW designer is the biggest differential.

Thus, good modeling practices and organization criteria for electronic components should be performed to avoid short circuit and noise, or a waste of time, as the board design process is geared to meet the need for specific designs aimed at the quality of a well-developed layout.

References

1. Lasi, H., et al.: Industry 4.0. Business & information systems engineering **6**(4), 239–242 (2014)
2. Vogel-Heuser, B., Dieter, H.: Guest editorial industry 4.0–prerequisites and visions. IEEE Trans. Autom. Sci. Eng. **13**(2), 411–413 (2016)
3. Devezas, T., Sarygulov, A.: Industry 4.0. Springer (2017)
4. Padilha, R., Iano, Y., Monteiro, A.C.B., Arthur, R., Estrela, V.V.: Betterment proposal to multipath fading channels potential to MIMO systems. In: Brazilian Technology Symposium pp. 115–130. Springer, Cham (2018)
5. França, R.P., Peluso, M., Monteiro, A.C.B., Iano, Y., Arthur, R., Estrela, V.V.: Development of a kernel: a deeper look at the architecture of an operating system. In: Brazilian Technology Symposium, pp. 103–114. Springer, Cham (2018)
6. Monteiro, A.C.B., Iano, Y., França, R.P., Arthur, R., Estrela, V.V.: A comparative study between methodologies based on the hough transform and watershed transform on the blood cell count. In: Brazilian Technology Symposium, pp. 65–78. Springer, Cham (2018)
7. Monteiro, A.C.B., Iano, Y., França, R.P., Arthur, R.: Methodology of high accuracy, sensitivity and specificity in the counts of erythrocytes and leukocytes in blood smear images. In: Brazilian Technology Symposium, pp. 79–90. Springer, Cham (2018)
8. Monteiro, A.C.B., Iano, Y., França, R.P., Arthur, R.: toxoplasmosis gondii: from discovery to advances in image processing. In: Brazilian Technology Symposium, pp. 91–101. Springer, Cham (2018)
9. Saha, H.N., Mandal, A., Sinha, A.: Recent trends in the internet of things. In: 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), pp. 1–4. IEEE (2017)
10. Gilchrist, A.: Industry 4.0: The Industrial Internet of Things. Apress (2016)
11. Shafiq, S.I., Sanin, C., Szczerbicki, E., Toro, C.: Virtual engineering object/virtual engineering process: a specialized form of cyber physical system for Industrie 4.0. Procedia Comput. Sci. **60**, 1146–1155 (2015)
12. Roblek, V., Meško, M., Krapež, A.: A complex view of industry 4.0. Sage Open, **6**(2), 2158244016653987 (2016)
13. Okano, M.T.: IOT and industry 4.0: the industrial new revolution. In: International Conference on Management and Information Systems September, vol. 25, p. 26 (2017)
14. Zebulum, R.S., Pacheco, M.A., Vellasco, M.M.B.: Evolutionary Electronics: Automatic Design of Electronic Circuits and Systems by Genetic Algorithms. CRC Press (2018)
15. Coombs, C.F.: Printed circuits handbook, vol. 1. In: Holden, H.T. (ed.) McGraw-Hill, New York (2001)
16. Clark, R.H.: Handbook of Printed Circuit Manufacturing. Springer Science & Business Media (2012)

Smart Contracts and Blockchain: An Application Model for Traceability in the Beef Supply Chain

Anderson Domeneguette Felippe and Antonio Carlos Demanboro

Abstract While demand for food increases as a result of the world population growth, supply chains face bigger challenges concerning the quality and traceability of goods. Thus, this paper focuses on the beef supply chain and its current traceability tools, which have shown several flaws as they fail to provide reliable means to track products back to their source. Since blockchain and smart contracts technology have emerged providing a decentralized and secure platform for information exchange, it has changed how people and a wide range of institutions handle assets transactions such as currency and data. Therefore, this paper demonstrates an application model using blockchain and a smart contract algorithm as a viable and trustworthy traceability tool for the beef supply chain.

Keywords Smart contracts · Blockchain · Traceability · Beef supply chain

1 Introduction

World population prospects presented by [2] states that, by 2050, there will be more than 9.7 billion people on earth. As that number grows every year, food demand will be one of the core challenges for humankind [3] and, on the table of most of the families, meat and mainly beef will keep playing a central role as the main source of animal protein [4]. Though vegetarian and vegan diets, as well as alternative sources of animal protein such as insects, have been largely studied and used in the last years [5] in order to reduce meat consumption, an estimation showed that beef consumption itself will increase up to 16% in the next six years [4]. As the world's largest exporter of beef, Brazil may experience up to 40% growth in this demand over the same period [4]; hence, the Brazilian beef market is highlighted in this work. One of the most

A. D. Felippe (✉) · A. C. Demanboro
Pontifical Catholic University of Campinas, Campinas, Brazil
e-mail: adfelippe@gmail.com

A. C. Demanboro
e-mail: ademanboro@gmail.com

© Springer Nature Switzerland AG 2021
Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_47

frequent issues faced by the beef supply chain is series of scandals that have been disclosed, such as Tesco's horsemeat scandal or the latest Brazilian meat scandal in 2017, which severely affected beef supply on a global scale [6]. This scenario will bring great challenges for the beef supply chain ranging from environmental footprint to product quality and traceability.

There are many methods for traceability and some will be explored in this paper, and, although they have been extensively used in the beef supply chain [7], all of them have failed to integrate the actors throughout the process while featuring an immutable, easy-to-obtain, verifiable, and secure source of data for tracking cattle from birth to retail sales [8]. The goal of this paper is to provide a traceability model based on blockchain and smart contracts that can be used to ensure cattle are securely tracked throughout their life cycle and their data (birth, origin, vaccines, diseases, etc.) could be verified by any actor in the beef supply chain.

This paper is organized as follows. In Sect. 2, we review the supply chain present and future challenges, as well as main traceability tools currently in use and reasons why they have failed to deliver value to the process. In Sect. 3, we explain how blockchain and smart contracts work, their current use in untrusted environments such as finance and also present links to supply chains. We show and analyze, in Sect. 4, our blockchain application with a smart contract algorithm to serve as a traceability tool in the beef supply chain and its benefits to the market. In Sect. 5, we present our conclusions.

2 Beef Supply Chain Review

2.1 Current Scenario

The most common method for tracking cattle is RFID ear tags. Federal agencies around the world have implemented regulations that force farmers to use some sort of traceability method [6], such as the Brazilian Agriculture Ministry (MAPA). MAPA regulates the use of a tracking system called SISBOV, using ear tags and a nationwide database for beef exportation, though it has certain flaws in the process that cannot guarantee reliable traceability throughout the supply chain [7].

An opinion review performed with customers, governmental agencies, and suppliers regarding current beef traceability methods concluded that every actor in the supply chain has a different idea about their real value and final consumer reckons that a traceability tool providing reliable information about the product could change their quality perception and would help on their purchasing decisions [6, 9].

2.2 Future Challenges

Brazil is currently responsible for 19.29% of the beef exportation market share [10] and considering the previously mentioned 40% growth in the exportation figures [4], supply will certainly face many challenges as demand increases. Since there is no common widely adopted standard for beef traceability and each producer can use its own solution [11], this may cause in the future a variety of issues such as untracked products (in case of scandals and public health issues), unreliable quality, and a great difficulty to be able to supply beef in a liable and timely manner.

3 Blockchain Technology

3.1 Blockchain

The concept of a secure chain of blocks was born in the 90s, but only later in 2008, it took the shape we know nowadays, since it was created to serve as Bitcoin's underlying protocol [12]. Blockchain is basically a distributed ledger (or database) on a decentralized peer-to-peer (P2P) network for information exchange (Fig. 1).

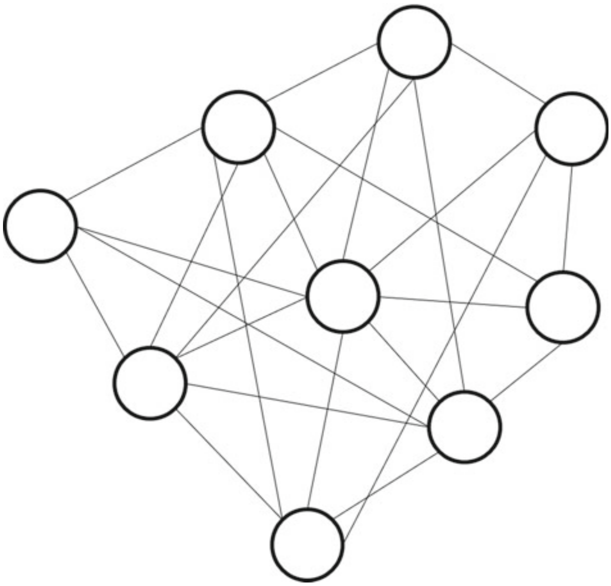


Fig. 1 A decentralized network of nodes without a central authority. Each node has a light or full copy of the entire blockchain and participates in the "competition" for mining new blocks individually. When a new block is mined, the winning node spreads it to the network

Fig. 2 A Blockchain example

The structure consists of a chain of blocks (Fig. 2) containing a dataset formed by its own cryptographic signature, a header, the signature of the previous block, and the log of all validated transactions (what and how much is sent to whom) in that time frame. Blocks are linked, similarly to a linked list (therefore forming a chain), and every new block is always written after the last one [13]. There are, however, strict rules to accept a new block in the blockchain, which involves a very hard computational effort known as mining. Mining is a process of continuously compute the candidate block's header data changing an integer number known as a nonce in order to find a hash that is smaller than a pre-defined target and, once this hash is found, the block is finally written to the blockchain [13].

Hash functions have such an important role in blockchain implementation due to the way they work. For a data set input of any size, a hash function will always produce an output with a fixed size [14]. Using Bitcoin as an example, its hash function (SHA-256 function applied twice in the block header) has a fixed size of 160 bits or 2^{160} possible combinations [1], and thus, finding the target cannot be done by other means but trying every possible combination while varying the nonce until it is found, a work that will expend a lot of time and power even on machines with high computational capacity. Finding this target is essentially performed by a trial and error algorithm called proof of work (POW) [12]. This algorithm is called POW since it will, in fact, prove that a lot of work has been done in order to validate the information involved in that transaction. POW is the algorithm used in many cryptocurrencies today such as Bitcoin and Ethereum [13, 15]. There are other mining methods, such as proof of stake (POS), and even an indirect consensus approval method used in IOTA's Tangle [16] (another form of distributed ledger technology that differs from the blockchain), though we focus on POW as it is widely used by the main crypto-assets in the market and is also immune to some forms of attack that can jeopardize POS-based networks [17].

The block mining rate, in Bitcoin, for example, must be kept at a 10-min average. As more miners enter the network, and therefore, more computational capacity becomes available; the nonce difficulty is increased to keep new blocks mining rate at the same 10-min window. If that capacity decreases—meaning that fewer nodes are mining—difficulty will also decrease over time [18]. Additionally, as every block carries the hash of the previous block, if any block is somehow changed, all subsequent blocks will have to be mined again, demanding a computational power that is unavailable in practical terms. In a matter of a few weeks, it would be impossible

```
contract MyCoin {

      function MyCoin() {
            this_balance[origin] = 1000;

      function sendCoin(address receiver, uint amount) {
            balance[receiver] = amount;
      }

      function getBalance() {
            return this_balance[origin];
      }
}
```

Fig. 3 A simple smart contract example

to alter the blockchain with only a few nodes running in the network [13]. All these security layers have proven to be one of the most appealing features for the adoption of blockchain, mainly in environments with an untrustworthy nature [19].

3.2 Smart Contracts

Smart contracts were originally created to act as a digital entity to assure the execution of contractual terms on networks, aiming to be embedded in hardware and software applications [20]. The concept has now evolved to be a computer program or script that executes specific tasks or payments as some pre-established agreements and rules are met in the digital contract. In a blockchain environment, a smart contract has its own hash address and accepts transactions that will be verified by its own rules and written to the blockchain if the contract terms are executed. Those contract terms can be as simple as programming logic, such as "if this, then that" (Fig. 3) [15].

3.3 Standout Use Cases

Due to its technical feature of being able to provide means for trustless environments to exchange assets such as currency, blockchain has been extensively tested in the financial market. Several banks such as JPMorgan Chase, Citi, and Bank of America have invested and implemented some variance of a blockchain system to enforce their security on transactions and data reliability [21, 22]. On supply chain's side, blockchain has been used as a tracking tool combined with RFID presenting very good results and many logistics players are willing to embrace the technology, which can provide much more than just traceability as some authors point out that companies' return on investment may increase with the adoption of such reliable techniques [23, 24]. IBM and Walmart also teamed up in order to use blockchain to track pork in China and the USA to improve traceability, avoid facing health issues and add value to their products [25]. Many players in the supply chain business have adopted blockchain traceability systems such as Waltonchain, a blockchain designed to be used with RFID in logistics [26], and VeChain, which also has Walmart by their side [27], supporting the increasing adoption of the technology in their market.

4 A Blockchain-Based Traceability Tool

As pointed out in this work, traceability will be one of the main issues faced by the beef supply chain, and current methods and tools have failed to provide a viable solution for it. In order to face this problem, the proposed methodology is an application model based on blockchain smart contracts that can provide a traceability register to the cattle, which can be checked upon at any time during their life cycle. In order to achieve that level of traceability, a traceable entity must be created during breeding, the animal's first stages of life. It is at this point that a smart contract comes around. Farmers must have a set of hardware and software tools to create a first reliable register of a single animal as a new smart contract in the blockchain. Those tools involve the need for a hardware platform such as a PC or a smartphone that is capable of running a decentralized application (DApp) to interact with the contract acting as a node in the blockchain network. This contract that is being proposed would have only two types of functions available: a getter and a setter as well as data validation (Fig. 4).

Getter functions could also be used for parametric search using known cattle attributes. At the contract creation, the individual information of the animal must be inserted through a setter function, logging it permanently in the blockchain. This piece of individual information can vary from body marks, tattoos, or even best a DNA fingerprint, which can assure a non-interchangeable form of identification [28]. From breeding to retail sales, a variety of data is collected from the animal such as weight, RFID, location, vaccines, diseases, transportation, and others. All that information can be added up on the smart contract through its setter function at any time; however, these data have to be valid and common to cattle traceability to be accepted in the contract, which will all be done in software.

Smart contracts can, and preferably should, be programmed to have a lifetime. In this model, a smart contract should be self-destroyed after N months from its creation. N can be calculated as in (1).

$$N = S_{\mathrm{t}} + P_{\mathrm{t}} + C_{\mathrm{t}} \tag{1}$$

```
contract CattleTracer {
        struct cattle_data;
        bool is_contract_new;

        function CattleTracker() {
                is_contract_new = true;
                new cattle_data this_animal;
        }

        function InsertDataInt(uint value, type data_type) {
                if (value.validate(value, data_type) == true)
                        this_animal.data_type = value;
                else
                        return ERROR;
                return OK;

        }

        function InsertDataFloat(float value, type data_type) {
                if (value.validate(value, data_type) == true)
                        this_animal.data_type = value;
                else
                        return ERROR;
                return OK;

        }

        function InsertDataString(string str, type data_type) {
                if (value.validate(str, data_type) == true)
                        this_animal.data_type = value;
                else
                        return ERROR;
                return OK;

        }

        function getData() {
                return this_animal;
        }
}
```

Fig. 4 Pseudocode for a smart contract for beef traceability

S_t is the time until slaughter, P_t is time to process, and C_t is time until consumption. All periods should be considered as long as possible to avoid that the smart contract is halted earlier than expected, though values can vary depending on cattle breed, feeding, fattening method, etc. S_t can range from 18 to 26 months [29], P_t should not be longer than 1 month, and C_t can be as long as 8 months [30] if the meat is processed and kept frozen, thus N could range from 19 to 35 months. It is important to note that once a smart contract is destroyed, it will only cease to allow access to its functions, though its contents can be read and verified at any time as they are written in the blockchain. During its lifetime, the getter function will return all fields and their data in an organized way. After its lifetime has ended, only the raw data written in the blockchain can be retrieved using the contract's hash address [15]. The result of having a smart contract in the blockchain is an immutable register containing cattle tracking data from their birth to the consumer's hands.

A blockchain application will bring various advantages to the beef traceability in the supply chain:

- Secure and safe environment.
- Easy implementation once software infrastructure is developed.
- Possible low implementation cost (considering current cloud infrastructure costs).
- Immutable registers.
- Fast access to data online.
- Highly traceable records.
- Scalable.
- Decentralized (no need for central authority).

5 Conclusion

This work reviews the upcoming challenges that the beef supply chain will face with the growth of the world population and evaluates the main issues involving traceability in the sector. Blockchain and smart contracts technology are also discussed and are linked to markets that are willing to adopt them as a trustworthy means for assets exchange. An application model is proposed using smart contracts in the blockchain that can provide a very secure and immutable tracking record for cattle, from birth to consumers, creating a powerful tool that is capable of filling the technical and operational gaps left by current traceability methods. It is assumed that the implementation cost can be supported by the players in the beef chain, because, in doing so, it could aggregate value to the products to be sold in the market. We suggest that further studies need to explore the costs involved and that any implementation, at first stages, is done on a well-established blockchain platform such as Ethereum or HyperLedger, for example. Since blockchain is still considered a technology under development, the acceptance could be low, hampering the implementation on a large scale; however, it has been widely tested and used in similar markets with positive results. Early adopters can become a reference for future use of the technology, and the beef supply chain can certainly benefit from it since it will offer a standard approach for cattle traceability anywhere, surpassing current challenges shared by most of the actors involved in it all over the world.

References

1. Kahn, M.A., Salah, K.: IoT security: review, blockchain solutions, and open challenges. Future Gener. Comput. Syst. **82**, 395–411 (2018)
2. United Nations, World Population Prospects: The 2017 Revision, Key Findings and Advance (2017). https://population.un.org/wpp/Publications/Files/WPP2017_DataBooklet.pdf
3. de Bakker, E., Dagevos, H.J.: Reducing meat consumption in today's consumer society: questioning the citizen-consumer gap. Agric. Environ. Ethics **25**(6), 877–894 (2012)
4. Saath, K.C. de O., Fachinello, A.L.: Crescimento da demanda mundial de alimentos e restrições do fator terra no Brasil. Rev. Econ. Sociol. Rural **56**(2), 195–212 (2018)
5. de Castro, R.J.S., Ohara, A., Aguilar, J.G. dos S., Domingues, M.A.F.: Nutritional, functional and biological properties of insect proteins: Processes for obtaining, consumption and future challenges. Trends Food Sci. Technol. **76**, 82–89 (2018)
6. Sander, F., Semeijn, J., Mahr, D.: The acceptance of blockchain technology in meat traceability and transparency. British Food J. **120**(9), 2066–2079 (2018)
7. Lopes, M.A., dos Santos, G.: Custo da implantação da rastreabilidade em bovinocultura utilizando os diferentes métodos de identificação permitidos pelo SISBOV. Ciência Ani. Brasileira **8**(4), 657–664 (2007)
8. Grande, E.T.G., Vieira, S.L.: Beef traceability by radio frequency identification system in the production process of a slaughterhouse. J. Inf. Syst. Technol. Manage. **10**(1), 99–118 (2013)
9. Galvez, J.F., Mejuto, J.C., Simal-Gandara, J.: Future challenges on the use of blockchain for food traceability analysis. TrAC Trends Anal. Chem. **107**, 222–232 (2018)
10. Formigoni, I.: Principais exportadores de carne bovina em 2018 (2018). http://www.farmnews.com.br/mercado/principais-exportadores-de-carne-bovina-2. Last accessed 27 Apr 2019
11. Scarvada, A.J., Batalha, M.O., Ribeiro, P.C.C.: Implementando RFID na cadeia de negócios: tecnologia a serviço da excelência. In: Hessel, F., Villar, R.S.G., Baladei, S.D.P. (org.) Implementando RFID na Cadeia de Negócios. 3rd ed., pp. 261–270. EDIPUCRS, Porto Alegre (2014)
12. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. (2008). https://bitcoin.org/bitcoin.pdf. Last accessed 28 Apr 2019
13. Antonopoulos, A.M.: Mastering Bitcoin, 2nd edn. O'Reilly Media Inc., Sebastopol (2017)
14. Dworkin, M.J.: SHA-3 Standard: permutation-based hash and extendable-output functions. Federal Inf. Process. Stds. (NIST FIPS) 202 (2015). https://doi.org/10.6028/NIST.FIPS.202. Last accessed 01 May 2019
15. Antonopoulos, A.M., Wood, G.: Mastering Ethereum. (2019). https://github.com/ethereumbook/ethereumbook. Last accessed 06 Sep 2019
16. Popov, S.: The tangle (2018). https://iota.org/IOTA_Whitepaper.pdf. Last accessed 19 Sep 2019
17. Deirmentzoglou, E., Papakyriakopoulos, G., Patsakis, C.: A survey on long-range attacks for proof of stake protocols. IEEE Access **7**, 28712–28725 (2019)
18. Oliver, C.G., Ricottone, A., Philippopoulos, P.: Proposal for a fully decentralized blockchain and proof-of-work algorithm for solving NP-complete problems. Distrib. Paral. Cluster Comput. (2017). https://arxiv.org/abs/1708.09419. Last accessed 01 May 2019
19. Tian, F.: An agri-food supply chain traceability system for China based on RFID & blockchain technology. In: 13th International Conference on Service Systems and Service Management (ICSSSM). Kunming, China (2016). https://doi.org/10.1109/ICSSSM.2016.7538424. Last accessed 01 May 2019
20. Szabo, N.: Smart Contracts (1994). http://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html. Last accessed 01 May 2019

21. Kelly, J.: Forty big banks test blockchain-based bond trading system. Reuters. (2016). https://www.reuters.com/article/banking-blockchain-bonds/forty-big-banks-test-blockchain-based-bond-trading-system-idUSL8N16A30H. Last accessed 04 June 2019
22. Power, B.: JPMorgan and BofA lead race for blockchain banking supremacy. BREAKERMAG (2019). https://breakermag.com/jpmorgan-and-bofa-lead-race-for-blockchain-banking-supremacy. Last accessed 04 June 2019
23. Marr, B.: How blockchain will transform the supply chain and logistics industry. Forbes (2018). https://www.forbes.com/sites/bernardmarr/2018/03/23/how-blockchain-will-transform-the-supply-chain-and-logistics-industry/#789f15e85fec. Last accessed 04 May 2019
24. Deloitte, Breaking blockchain open—Deloitte's 2018 global blockchain survey (2018). https://www2.deloitte.com/content/dam/Deloitte/cz/Documents/financial-services/cz-2018-deloitte-global-blockchain-survey.pdf. Last accessed 04 June 2019
25. Hackett, R.: Walmart and IBM are partnering to put chinese pork on a blockchain. Fortune (2016). https://fortune.com/2016/10/19/walmart-ibm-blockchain-china-pork. Last accessed 06 June 2019
26. Waltonchain White Paper. (2019). https://www.waltonchain.org/en/Uploads/2019-04-25/5cc171763aebb.pdf. last accessed 19 Sep 2019
27. Vechainworld, Walmart China Takes on Food Safety with VeChainThor Blockchain Technology. (2019). https://bbs.vechainworld.io/topic/208/walmart-china-takes-on-food-safety-with-vechainthor-blockchain-technology. Last accessed 19 Sep 2019
28. Haberfeld, A., Hillel, J.: Development of DNA fingerprint probes: an approach and its application. Anim. Biotechnol. **2**(1), 61–73 (1991)
29. Bortoluzzo, A.B., Pedrinola, P.D., Martins, S.R.: Tempo Ideal Para Abate de Gado de Corte Via Maximização do Lucro. Insper Working Paper (2011). https://www.insper.edu.br/en/wp-content/uploads/2013/12/2011_wpe239.pdf. Last accessed 06 June 2019
30. Minerva Foods, Como conservar a carne bovina. (2015). https://portal.minervafoods.com/blog/post/como-conservar-carne-bovina. Last accessed 04 May 2019

Twenty Years Survey of Big Data: Definition, Concepts, and Applications in Engineering

Massaki de O. Igarashi, Paulo E. Sartorelli, and Mariana Z. T. de Lima

Abstract In the last decade, there was an exponential growth in data generation from different sources especially due to advances in information and communication technology. Thus, organizations have seen the potential to gain competitive edges from the analyses of this data, changing it in the information that, without Big Data tools, could not be obtained. In this context, this work brings a survey about Big Data and explains this concept has changed during the years. Moreover, this paper aims to elucidate the last twenty years of Big Data and its applications in different areas of engineering: civil, electrical, manufacturing, mechanical, materials, chemical, and software engineering.

Keywords Big Data · Engineering · Areas · Data generation

1 Introduction

In the last two decades, the Internet has shown significant growth, which was especially responsible for the amount and the speed of data generation [1]. Currently, most of the actions performed on the Internet generating data, have been identified in order to analyze customer preferences, behavior patterns, evaluate trends, and even detect potential crises and fraud [2]. The importance of data generation and its applications increases considering the estimative of data volume growth for 2020 around 40 zettabytes [2, 3]. Among the possibilities to extract value from this large amount of data (structured or not), one of the outstanding ways it is to identify existing patterns in databases through the most frequently used information. Another way is that companies can create and store data and get detailed information across a range of areas, such as inventory forecasts, demand prospects over the coming months, and then use that information to make better decisions and improve organizational performance [4].

M. de O. Igarashi (✉) · P. E. Sartorelli · M. Z. T. de Lima
Mackenzie Presbyterian University, Campinas, SP, Brazil
e-mail: massaki.igarashi@mackenzie.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_48

Table 1 Different areas of Big Data applications

Precision agriculture	Games and social networks	Precision medicine
Banks and credit	Government and public management	Health
Insurance	Financial sector	Decision making
Education	Logistics and supply chain	Transportation
Judicial	Marketing	

The term Big Data is used to define a large and complex data set whose traditional database techniques, tools, and software are no longer efficient. Therefore, scaling this data, the diversity and complexity require new techniques, architectures, and algorithms for its management and analysis, allowing, an easier extraction of value and knowledge.

Big Data tools and techniques help to extract value and useful information for better decision making in the most diverse areas and possibilities (Table 1).

1.1 Theoretical Framework

According to Chen [5], the use of data as the information is not recent. Since the '50s, the administration and information technology areas have already used this concept and information systems. The first research was reported in the '70s, giving rise to data processing methods. In the '90s, software's has emerged and has been used to analyze the amount of data [6].

Big Data started to be used more frequently in the 2000s, with the expansion of new software's, processing techniques, storage, and data transmission [6]. Chen [5], Davenport [7], and his collaborators emphasized the emergence of new technologies and the spread of the Internet and global e-commerce was defined as the beginning of a new era of data generation that is transmitted over this worldwide network. The authors mentioned examples of actions taken from the Internet that began to drive a steady and growing stream of data, such as patterns of web browsing, clicks, online commerce, content that users themselves generate on social media, Web sites, blogs, and platforms. Thus, the term Big Data is not related to a whole new concept, but to the updating of several technologies, tools, and techniques that have already existed; this union provides a clearer understanding of Big Data today [6].

A. The 5 V's of Big Data

There is not a well-established consensus on the definition of the term Big Data, although it is known as a large amount of data that is daily generated daily. The first attempts are defined by three pillars: volume, velocity, and variety [8]. But other authors, when defining Big Data, use five pillars (the 5 V's of Big Data): volume, velocity, veracity, variety, and value.

Volume: It is related to a large amount of data that is generated daily around the world, coming from different sources [9].

Velocity: The speed in which data is generated [10].

Variety: data diversity by the many sources, structured or unstructured data [11].

Veracity: Data authenticity is essential for organizations [12]. It is related to the quality of the information available for decision making [9].

Value: The value that influences decision making based on data analytics [9].

Trying to systematize these V's that define Big Data, a timeline was created to relate publication dates, definitions, and their number of letters V's (Fig. 1).

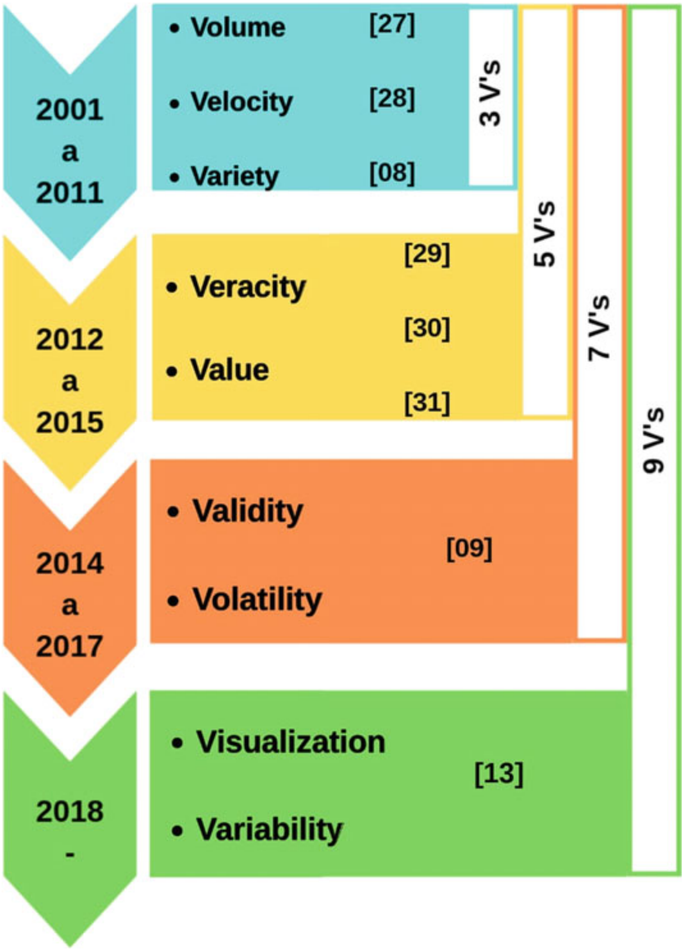


Fig. 1 Big Data evolution timeline [27–31]

Although most authors define only the five pillars (volume, velocity, variety, veracity, and value), there are other authors who add more pillars. For example, Ali-ud-din Khan [9] added two other pillars: validity and volatility. Sahoo [13] and collaborators highlight two others: visualization and variability. There are even some current blogging attempts to define the term through 10 V's, but without academic recognition.

2 Methodology

This research has a qualitative approach (which objective is to analyze Big Data definitions in time), with explanatory purpose about the potential of Big Data application in different areas of engineering and to describe its definition along the time.

A literature review was conducted from academic publications from national and international literature that was searched and used from academic bases Google Scholar, ScienceDirect, and ISI Web of Science.

The search was performed based on the keywords: "Big Data" + "engineering area," obtaining publication numbers from year to year, from the 2000s to the present date.

3 Analysis and Discussion

3.1 Big Data Applications in Engineering Areas

This study presents some Big Data investigation and applications in seven engineering areas (civil, electrical, manufacturing, mechanical, materials, chemistry, and software).

A. **Civil Engineering**

Liang Wang [14] reports a new vehicle detection and tracking system based on image data collected by unmanned aerial vehicles (UAV). This system uses consecutively captured frames to generate dynamic vehicle information: position and speed.

This system can be used by "smart cities" in traffic management centers, for the location of accidents, in the real monitoring of a highway, among others, as area of construction, search and rescue applications, structural inspection, and health monitoring; and with some modifications, it can also be used in jungle animal tracking.

Rathore [15] presents an IoT-based system for smart city development and urban planning using Big Data analytics. The current system uses different types of real-time sensors (weather sensors, water sensors, and intelligent parking), surveillance cameras, emergency buttons on the streets, among other devices. Its implementation is structured in the collection, data filtering, classification, preprocessing, processing, and decision making. The sensors generate large amounts of data at high speeds, which are processed by the Hadoop and MapReduce frameworks and guarantee system scalability and efficiency. It cannot only benefits citizens, but also authorities by providing information that makes decision making faster and more efficient.

B. **Electrical Engineering**

Chou [16] reports a Big Data analysis framework for smart grids and components of an energy-saving decision support system. The structure analyzes, in real-time, the data of electricity consumption, identifies consumption patterns, and predicts energy consumption, in addition to provide optimal operation schedules for appliances. The measurement infrastructure was installed in a residential building in the experimental simulations.

But, Lee [17] reports the development of smart charging infrastructure for electric cars using Big Data tools. The structure is being implemented in Jeju Province, South Korea, where the authors present the distribution of chargers and the number of chargers in operation. The goal is to combine various statistical and machine learning techniques to identify the charging demand pattern of electric vehicles and integrate renewable energy to power charging networks.

C. **Manufacturing Engineering**

Zhang [18] presents a proposal for a general Big Data-based analysis architecture for the product lifecycle management (PLM). This architecture integrates Big Data analytics and service-driven standards that aided decision making, coordination, and optimization of the Cleaner Production (CP) process. By utilizing the technology and the IoT concept at every stage of the PLM, develop an intelligent manufacturing and maintenance environment.

On the other hand, Lee [19] exposes the possibilities and trends of integrating advanced data analytics into manufacturing, products, and services, linking Big Data, the advancement of Information and Communication Technologies (ICT), and the fact that cyber-physical systems (CPS) facilitate the systematic transformation of massive data into information and assist in decision making. In addition, the group also features a systematic architecture called 5C that encompasses the steps required to fully integrate CPS into manufacturing.

D. **Mechanical Engineering**

Bumblauskas [20] reports a smart maintenance decision support system (SMDSS) based on a company's corporate data. The system is able to provide the user with recommendations for improving asset life cycles. The authors illustrate how Big Data analytical tools can be used to prioritize equipment maintenance and define the architecture in which a physical asset owner can enter usage parameters for interconnected equipment and receive a comprehensive proposal for the service to meet the recommendations generated by the analytical model.

Fernandes [21] presents the results of a study in a metallurgy company where data analysis and resource selection methods were employed from continuous equipment monitoring. Machine learning and data mining techniques were used to extract information and assist in decision making. The information gained will assist in the development of adaptive learning models capable of handling complex information that can be applied across a complete line of industrial products and equipment.

E. **Materials Engineering**

Lu [22] provides an overview and gives examples of research progress toward the discovery of new materials and classifies this research area as Materials 4.0. The work also highlights the development of machine learning protocols and speaks of quantum material property simulation software, digitized material data, intelligent machine learning algorithms, among others. The author emphasized the considerable reduction in time between concept and commercialization of new products and also speaks of the recovery after the conclusion of the product life cycle. The challenge is still aggregating data from multiple sources, as well as managing and analyzing unstructured data to enable 4.0 materials, and further recommends that web-based material research platforms need to be developed to explore opportunities and identify gaps.

F. **Chemical Engineering**

May [23] presents the possibilities of advanced multidimensional separations in mass spectrometry (MS) using Big Data tools, where hybrid analytical instrumentation based on MS has been a powerful technique for meeting the challenges in science and medicine, including helping in drug discovery and synthetic biology. The paper highlights the possibility of large-scale measurements to obtain information and also highlights the challenges of Big Data in chemical analysis: from the enumeration of chemical isomers to this field of multidimensional analysis based on MS.

Chiang [24] shows recent advances in some areas such as the chemical and pharmaceutical industry, among others. The objectives of this paper are to educate the chemical engineering community about the capacity of Big Data and to improve reliability and operational efficiency in various industry sectors, including directing future research. The study also points out that 88% of chemical executives recognize that data analysis will be crucial to maintaining a competitive advantage over a 5-year interval.

G. **Software Engineering**

Arndt [25] explores how software engineering (SE) technology can support the development of Big Data projects, and how Big Data techniques can be used to develop new processes and evolve SE techniques. The paper points out that a lot of research has been done in recent years to improve the production of Big Data systems and to ensure fast and elastic scaling whenever and wherever needed. One of the main researches focuses on applying SE methods to the production of Big Data systems is in the area of software architectures development.

According to Casale [26], software engineering as a discipline must be consolidated, because despite presenting impressive achievements in the field, it is one of the latest scientific disciplines. Currently, researchers in software engineering have been distinguished by the high quality of software and the adoption of controlled practices in development and operation.

3.2 Publications with the Term Big Data by Area

The search for the term "Big Data" + "engineering area" studied at this work in the last 20 years, also held on 27th September 2019 in Google Scholar from the year 2000 until 2019 enabled the drawing of the graph of Fig. 2.

The graph shows the trend of publications in each engineering area studied in this work for the same period. The chart shows a greater number of publications for the terms "Big Data" and "software engineering," with 826,000 publications in the period. The lowest number of publications among those surveyed was "Big Data and" "civil engineering" with 86,000 publications.

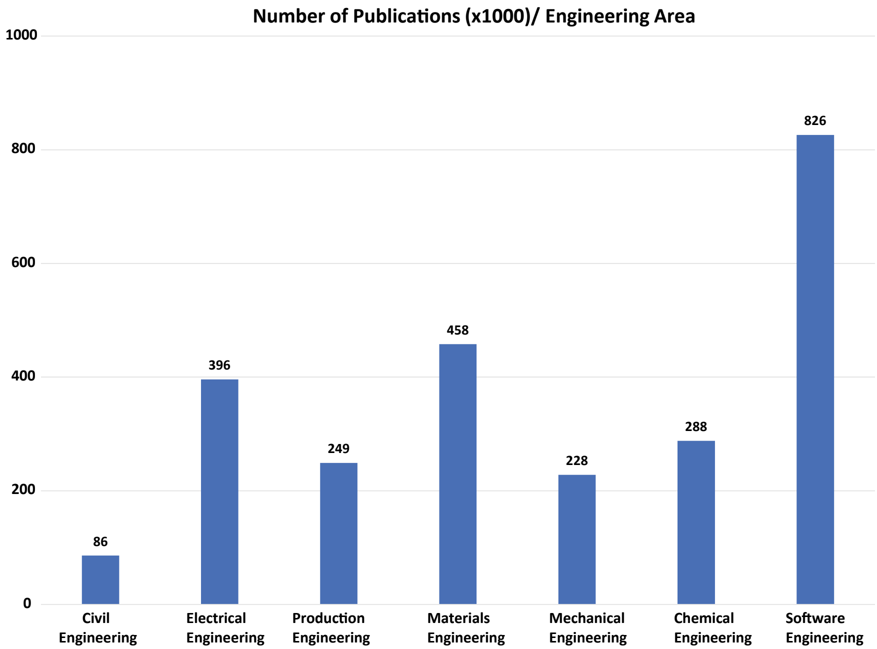


Fig. 2 Number of publications (×1000) by engineering area

4 Conclusions

This research presented a brief overview that, in addition to the evolution of publications over a period of 20 years, highlighted some applications in civil, electrical, manufacturing, mechanical, materials, chemistry, and software engineering.

Most of the case approaches to exemplify Big Data applications in engineering highlight: IoT, decision making, or intelligent sensing.

The results of this research demonstrate that some engineering areas, as expected, are of greater interest. But in any case, the results highlight applications that may help in future engineering research.

References

1. Ge, M. et al.: Big data for internet of things: a survey. Future Gener. Comput. Syst. (2018)
2. John, J.G., David, D.R.: The digital universe in 2020: big data, bigger digital shadows, and biggest growth in the far east. Idc iview: idc analyze the future, 2012. Available: https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf. Last Accessed on 28 Aug 2019
3. We Are Social. Digital in 2018: world's internet users pass the 4 billion mark. New York: We Are Social Ltd, 30 Jan. 2018. Available: https://wearesocial.com/blog/2018/01/global-digital-report-2018. Last Accessed on 31 July 2019
4. Manyika, J.: Big data: the next frontier for innovation, competition, and productivity, 2011. Available: http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation. Last Accessed on Sept 23, 2018
5. Chen, H., et al.: Business intelligence and analytics: from big data to big impact. Mis. Q. (2012)
6. Park, H.W., Leydesdorff, L.: Decomposing social and semantic networks in emerging "big data" research. J. Inf. (2013)
7. Davenport, T.H., et al.: How'big data'is different. Mit Sloan Manage. Rev. (2012)
8. Mcafee, A., et al.: Big data: the management revolution. Harvard Bus. Rev. (2012)
9. Khan, et al.: Seven v's of big data understanding big data to extract value. In: Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education. IEEE (2014)
10. Barton, D., Court, D.: Get started with big data: tie strategy to performance. Harvard business review, out. 2012. Available: https://hbr.org/2012/10/getting-started-on-a-big-data. Last Accessed on 15 Aug 2018
11. Tabuena, J.: What every internal auditor should know about big data. Compliance Week (2012)
12. Taurion, C.: Você realmente sabe o que é big data. Blog da ibm, 2012. Available: https://www.ibm.com/developerworks/community/blogs/ctaurion/entry/voce_realmente_sabe_o_que_e_big_data?lang = en. Last Accessed on 18 July 2018
13. Sahoo, M., et al.: The emergence of big data: a survey. Int. J. Comput. Sci. Mob. Appl. India (2018)
14. Wang, L., et al.: Detecting and tracking vehicles in traffic by unmanned aerial vehicles. Autom. Constr. (2016)
15. Rathore, M., et al.: Urban planning and building smart cities based on the internet of things using big data analytics. Comput. Networks (2016)
16. Chou, J., Ngo, N.: Smart grid data analytics framework for increasing energy savings in residential buildings. Autom. Constr. (2016)
17. Lee, J., et al.: Big data analysis for an electric vehicle charging infrastructure using open data and software. In: Proceedings of the Eighth International Conference on Future Energy Systems. Acm (2017)

18. Zhang, Y., et al.: A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. J. Cleaner Produc. (2017)
19. Lee, J., et al.: Industrial big data analytics and cyber-physical systems for future maintenance & service innovation. Procedia cirp (2015)
20. Bumblauskas, D., et al.: Smart maintenance decision support systems (smdss) based on corporate big data analytics. Expert Syst. Appl. (2017)
21. Fernandes, M., et al.: Data analysis and feature selection for predictive maintenance: a case-study in the metallurgic industry. Int. J. Inf. Manage. (2019)
22. Lu, W., et al.: Data mining-aided materials discovery and optimization. J. Materiomics (2017)
23. May, J.C., Mclean, J.A.: Advanced multidimensional separations in mass spectrometry: navigating the big data deluge. Ann. Rev. Anal. Chem. (2016)
24. Chiang, L., et al.: Big data analytics in chemical engineering. Ann. Rev. Chem. Biomol. Eng. (2017)
25. Arndt, T.: Big data and software engineering: prospects for mutual enrichment. Iran J. Comput. Sci. (2018)
26. Casale, G., et al.: Current and future challenges of software engineering for services and applications. Procedia Comput. Sci. (2016)
27. Laney, D.: 3d data management: controlling data volume, velocity and variety. Meta Group Research Note (2001)
28. Russom, P., et al.: Big data analytics. Tdwi best practices report, fourth quarter, v.19(4), 1–34 (2011)
29. Feldman, B., et al.: Big data in healthcare hype and hope. Dr. Bonnie (2012)
30. Ularu, E.G., et al.: Perspectives on big data and big data analytics. Database Syst. J. (2012)
31. Gandomi, A., Haider, M.: Beyond the hype: big data concepts, methods, and analytics. Int. J. Inf. Manage. (2015)

IoT Security: A Simplified Analysis of the Literature

Tiago Guimarães Monteiro, Marcos Almeida do Amaral, Lia Toledo Moreira Mota, and Yuzo Iano

Abstract This article has the goal of analyzing scientific studies about the Internet of Everything (IoE) and, under the security information concepts, understanding the relationship between then. With this objective, this article identifies researches, models, processing types, and tools about related issues of this theme. This research uses a qualitative methodology besides a systematic review of the literature to check all related keywords, such as OSI-ISO model, IoT, IoE, and Security, used to develop technological applications for smart buildings, houses, and cities. For this article, some databases were used: Scopus, Science Direct, IEEE, and ISO. This article provides detailed information on the present-day scientific articles focusing on security issues in IoT applications. This paper could serve as a basis for researchers seeking information for further investigation. The findings provide valuable contributions to scholars researching on the subject, and public managers considering implementing them in practice in their cities.

Keywords OSI · IoE · IoT · Security · Smart cities · Public management

T. G. Monteiro (✉) · L. T. M. Mota
Pontifical Catholic University of Campinas, Campinas, Brazil
e-mail: tiago.monteiro@gmail.com

L. T. M. Mota
e-mail: lia.mota@puc-campinas.edu.br

M. A. do Amaral
Mackenzie Presbiterian University, Campinas, Brazil
e-mail: marcos.amaral@mackenzie.com.br

Y. Iano
University of Campinas, Campinas, Brazil
e-mail: yuzo@decom.fee.unicamp.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_49

1 Introduction

With the considerable increase in the number of devices connected to the Internet, it has been tried to ensure that all information generated and exchanged by these networked devices has guarantees of privacy and confidentiality [1].

Within the message exchange of data, audio, video, and any other means of communication, either through applications or by the device itself, there is a need for information security. This growing mass of data called "Big Data" is the biggest proof that every day more devices are generating information exchange and lacking security [2].

Recently in Brazil, in 2014, Law No. 12965/2014 was approved, being known as the "New Civil Mark of the Internet" (free translation for Novo Marco Civil da Internet). This new law highlights various aspects of information exchange, the use of third-party data, and other important items to ensure the privacy of users which has been assured and regulated.

This new law, at least theoretically, guarantees privacy protection and attempts to impose greater and better control on the user data by Web sites and applications [3].

According to Palanza (2016): "The Internet of Things, Big Data, and Privacy are the triad of our technological future," so it is so important that they are studied together [4].

The most important and well-known aspects of information security are aimed primarily at ensuring three aspects: confidentiality, integrity, and availability—CIA security triad [5].

These aspects are important not only for securing device connections, but also for other important security aspects such as accountability, transparency, auditability, reliability, non-rejection, and privacy of connections [6].

When talking about connecting devices and buildings, it is vital to guarantee the security of all devices and buildings participating in secure networks. One weak point will be enough to open the system to strangers.

For these reasons, it is important to make a literature review in some databases in order to prove and certify the path pointed by the academic community, scholars, and practitioners of IoT.

Based on the suggested path, we will be able to analyze the quality of known information security tools to ensure proper management of the IoT network.

2 IoT versus IoE

In order to investigate the security in IoT (Internet of Things) devices, this definition must be separated from the IoE (Internet of Everything), since the first one deals with products and especially those in the very physical application are more susceptible to attacks [2].

In the past, communication was destined to be used by human beings to make the information exchange, and the devices were built around human decisions. Radios and computers had network access, but the decision upon the information was outside the network [7].

However, with the advance of computers and algorithms, the decision making could be done in the same devices that had the interface, being programmable decisions such as automation in a production line, or artificial intelligence analysis on a database [8].

This migration of the decision brought to the computers the capability to have heavy network traffic without a single human intervention. Also, as the devices have more technology available at lower prices and sizes, it became a trend to input communication applications in objects that were not designed with this objective.

Electro-mechanical machines had sensors in the third industrial revolution, but now in the fourth revolution, those sensors turned into a point of a network. This means that the machine is now sharing its behavior on a network that, eventually, will end on the Internet [8].

Products with radio frequency identification (RFID) tags can be located by interrogation devices, and their physical location can be traced back to the point where they were sensed and that information can also be shared online. This connection of the object with the Internet is IoT [7].

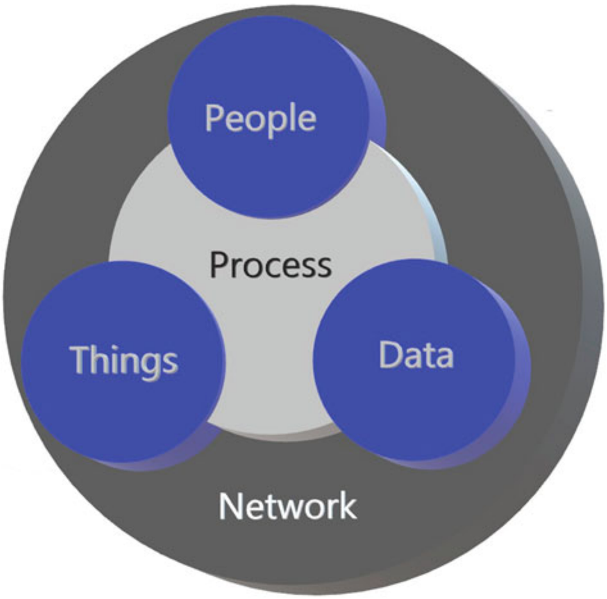


Fig. 1 Internet of Everything

Some manufacturers are now expanding this concept to generate an even broader term to define all communications that cover objects and people. If the first data link was supposed to connect people, newer ones connect things. Eventually, things are interlinked by processes and people, and now it can be defined as the Internet of Everything (IoE) [7].

IoE is the connection of four pillars: people, things, data, and process. Now, data is shared between things and people, based on processes. All integration can be viewed on Fig. 1.

3 OSI and IoT Model Layers

Since the beginning of computer networks, the adopted model has divided them into seven different layers, each with its own particularities and differences. This model was named Open System Interconnection (OSI), and it was developed by the International Organization for Standardization (ISO).

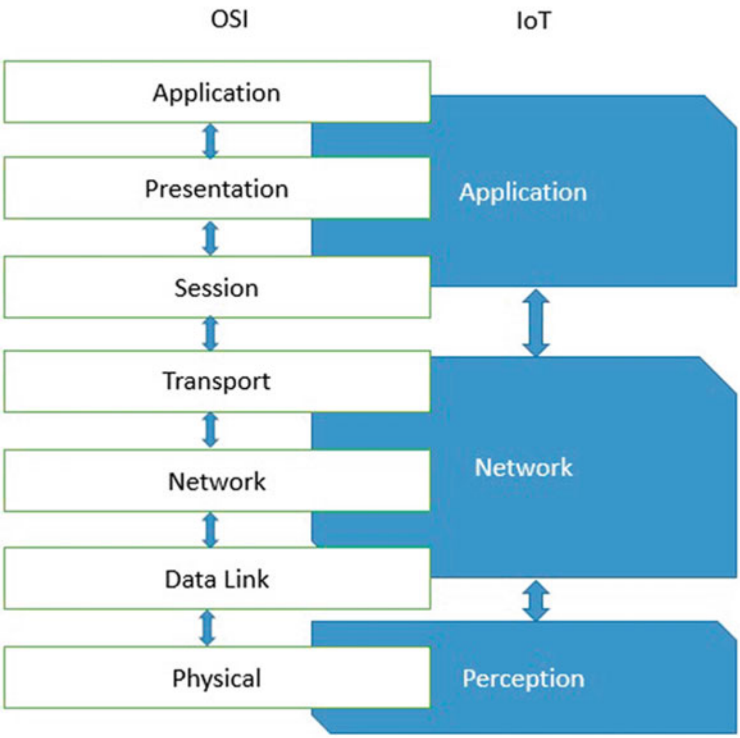


Fig. 2 OSI versus IoT layers

The OSI-ISO model is the best known and most widely used structural network model in the world. It ensures better inter-compatibility of network equipment. The OSI-ISO model has the following layers:

- Layer 1—Physical layer: Electrical signals and mechanical characteristics of bit transmission.
- Layer 2—Data link layer: Transmission of packets from one node to another according to the station address.
- Layer 3—Network layer: Sending data to different local area networks (LANs) and wide area networks(WANs).
- Layer 4—Transport layer: Ensures proper transmission and reception of data packets.
- Layer 5—Session layer: Starts and ends a transmission and controls the sequence of commands that ensure the full transmission of packets.
- Layer 6—Presentation layer: Ensures the security aspects with encryption and code conversions required such as American Standard Code for Information Interchange (ASCII) to Extended Binary Coded Decimal Interchange Code (EBCDIC)—8-bit character code (Fig. 2).
- Layer 7—Presentation layer: Determines the type of communication and sets standards for sending email messages, file transfers, and client/server connections.

In the IoT model, there are only three layers, and these encompass the same performance already seen in the OSI-ISO model. However, since IoT devices do not have the same power availability, processing power, and memory, it is easier to break their security [5].

The IoT model provides the following layers:

- Layer 1—Perception layer: Also known as wireless sensor network (WSN). It is the layer where data is acquired through the detection, collection, and processing of information.
- Layer 2—Network layer: Where the routing and data transmission between hubs and other IoT equipment is established. Communication is established through the latest technologies: WiFi, LTE, Bluetooth, 3G, Zigbee, among others.
- Layer 3—Application layer: This layer is where the triad of information security is established, thus ensuring the confidentiality, integrity, and availability of information.

4 Safety in both models

In the OSI model, security applications are spread in several layers, each with its own characteristics to meet a certain type of attack on that layer [9].

The biggest problems encountered in the OSI model are related to:

1. Masquerade
2. Illegal associations

3. Non-authorized access
4. Denial of service
5. Repudiation
6. Leakage of information content
7. Traffic analysis
8. Invalid message sequencing
9. Data modification or destruction
10. Information inference attacks
11. Illegal modification of programs [10].

Those threats are common. In order to eliminate them and protect the network, security management methods must be applied, such as:

1. Authentication management;
2. Access control management;
3. Key management;
4. Security audit trails and event handling;
5. Encipherment Algorithms [11].

All these management methods could be applied individually or in combination to get more specific results according to the threats on the network.

However, in the IoT model, security applications are mostly applied in the third tier. It is in this layer that there is an interaction between the user and the applications.

To ensure the security of IoT networks and devices, it is necessary to use different mechanisms such as:

1. Network software and devices must be original, authorized, and licensed;
2. When the IoT device is initialized, it must first authenticate to its network before collecting and sending data;
3. Because IoT devices have limited processing, memory, and power availability, there must be a device running firewall;
4. Security updates and patches should be installed on devices, but without generating excessive bandwidth consumption that could clog the network [12].

5 State of the Art

Nowadays, research on implementing the best type of configuration for an IoT network has some troubles with some of the most important features:

1. IoT networks are heterogeneous and dynamic.
2. Threats are constantly changed and optimized.
3. Possibility of integration between different manufacturers.
4. Use of the most varied protocols in the networks.
5. Network configuration runs away from traditional models [12].

These surveys show that the two main sources of content production on the subject, industry, and academia, have been addressing security issues primarily on the following subjects:

1. Attacks on the sensor layer;
2. Attacks on other IoT network devices;
3. Countermeasures that have already been taken to protect IoT networks;
4. More robust applications based on some, but not all, countermeasures already identified [6].

The research around the issues covered in this article highlights the concern of everyone involved in setting a security standard for IoT networks.

It is interesting to note that several articles already deal with IoT in the global context of IoE and thus make all the desired security definitions more complex and cumbersome.

6 Conclusion

The IoT reduced model brings the agility on the service, which is fundamental for this kind of application.

However, due to the heavy processing load used by all seven layers, the full-scale security systems bring to the circuits power consumption and sizes that are not compatible with the IoT philosophy, that is, to bring a way of connection even to the small devices.

Since the IoT networks are based on heterogeneous and dynamic applications, with a high diversity of manufacturers, the security application on the perception layer is the key point to the lack of security.

As this layer is installed within the consumer products and cannot afford more processing, for the time being, the possible solution is to insert a verification key within the interrogation device, since its capacity have more room for improvement.

References

1. BBVA Innovation Center.: BBVA Innovation Center Internet of Things (Innovation Trends Series) (2016). Available: https://amazon.com/kindle/ www.centrodeinnovacionbbva.com
2. Alcântara, L.K.: Big Data e IOT: Desafios da Privacidade e da Proteção de dados no Direito Digital. São Paulo (2017). Available: https://amazon.com/kindle
3. Marcacini, A.T.R.: Aspectos Fundamentais Do Marco Civil Da Internet: LEI Nº 12.965/2014. São Paulo (2016). Available: https://amazon.com/kindle
4. Palanza, S.: Internet of things, big data e privacy: la tríade del futuro (2016). Available: http://www.iai.it/en/pubblicazioni/internet-things-big-data-e-privacy
5. Etter, D.: Iot Security: Pratical Guide Book. São Paulo (2017). Available: https://amazon.com/kindle

6. Mosenia, A., Jha, N.K.: A comprehensive study of security of internet-of-things. In: IEEE Transactions on Emerging Topics in Computin, vol. 05, Nº 04. Out.–Dez. (2017)
7. Miraz, M.H., et al.: A review on internet of things (IoT), internet of everything (IoE) and internet of nano things (IoNT). In: 2015 Internet Technologies and Applications (ITA) (2015)
8. Kalyani, V.L., Sharma, D.: IoT: Machine to machine (M2M), device to device (D2D) internet of everything (IoE) and human to human (H2H): future of communication. J. Manage. Eng. Inf. Technol. (JMEIT) **2**(6) (2015)
9. Tanenbaum, A.S.: Redes de Computadores. São Paulo, 2003. Editora Elsevier, 946p (2003)
10. Patel, A.: Security management for OSI networks. Comput. Commun. **17**(7), 544–553 (1994)
11. Buffenoir, T.: Security in the OSI model. Comput. Stand. Interfaces **7**(1–2), 145–150 (1988)
12. Mahmoud, R., Yousuf, T., Aloul, F., Zualkernan, I.: Internet of things (IoT) security: current status, challenges and prospective measures (2016)

TRUMIT—Trust Manager for IoT

Nelson Bellincanta Filho and Luiz C. P. Albini

Abstract The main objective of this paper is to design a dynamic trust model to evaluate the reliability level of nodes in an IoT environment. To achieve this goal, TRUMIT—Trust Manager for IoT, a fully decentralized and distributed trust management model for IoT environments, is proposed. In this approach, each node self-organizes its trust network to obtain and recommend node reliability information. The proposed model demonstrated the ability of convergence and resilience to malicious nodes, thus ensuring reliability between the nodes.

Keywords IoT · Trust management · Direct trust · Recommendations

1 Introduction

Since its inception, the Internet has undergone many transformations, from the traditional bonding and sharing of computers and documents, to a platform for doing business and connecting people through social media. Now, a new paradigm called the Internet of Things allows billions or even trillions of heterogeneous objects called "things" to interact with each other over the Internet, anytime and anywhere. This interaction allows these objects to collect and share information, to coordinate decisions, to assist in the execution of everyday tasks [1–3].

However, the information collected and processed by these objects is prone to factors that make them unreliable. These factors may be related to technical or behavioral issues. Technical issues, relating to the conditions of the deployment environment and the compromise of information transmission, these problems focus on

N. B. Filho (✉) · L. C. P. Albini
Department of Informatics, Federal University of Parana (UFPR), Curitiba, Brazil
e-mail: nbfilho@inf.ufpr.br; nelson.filho@ifpr.edu.br

L. C. P. Albini
e-mail: albini@inf.ufpr.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_50

details such as verifying the authenticity of an object and determining what permissions the object has. To solve this problem, the techniques used include encryption, data hiding, digital signatures, authentication protocols, and access control methods. However, addressing object behavior problems is a defiant task due to the heterogeneous, large-scale and dynamic environment such as IoT [4–7].

In this way, creating reliable IoT environments mitigates unrecoverable and unexpected damage to develop intelligent, efficient, stable and flexible systems, providing the trust and belief that IoT objects and hence services perform as expected [8–11]. Given this context, this paper proposes TRUMIT, a fully decentralized trust management model distributed for IoT environments, capable of assessing the trust of IoT objects, considering the characteristics of this environment.

The remainder of this document is organized as follows, Sect. 2 presents state-of-the-art trust management systems for IoT. Then Sect. 3 introduces TRUMIT, describing its notation and its functionality. Section 4 presents the analysis of the results collected from the TRUMIT assessment. Finally, Sect. 5 contains the concluding remarks and suggestions for future work.

2 Related Work

Assessing trust in a dynamic network such as IoT is a challenging task and a topic that has received wide attention from the security community. This fact is due to the heterogeneity and the limited computational resources of the devices that make up the IoT environment. To overcome this challenge, several trust management solutions for IoT have been proposed, and each has significant gains for trust management for IoT.

In [12], a fuzzy mathematical model of trust is proposed, where each node employs a monitoring process for neighboring nodes to collect information about packet forwarding behavior. However, the network card must be in promiscuous mode, causing the device to receive all messages from its neighbors, even if they are not directed directly to itself. The [13–16] approaches are user-centric, so each device (node) has one owner and one owner can have many devices, each owner has a friend list that represents their social interactions. The authors do not consider the use of this model in environments with little or no user interaction.

The work [17] uses a Bayesian reputation system where each node calculates trust according to its direct observations. The direct trust is updated at each meeting or interaction activity, while the indirect trust is updated periodically. This approach does not consider updating direct and indirect trust at the same time. The work [18] considers a centralized trust manager, which stores all node reputation reports. Considering a centralized trust manager, this approach makes it difficult to implement in a distributed environment such as IoT.

[19] Proposes a trust score based on the following parameters: its own previous experience, acquired knowledge and other units recommendation. After calculating each of these parameters, trust is calculated using a fuzzy rule base. However, using various metrics for trust composition results in a longer time to calculate trust as well as higher consumption of computational resources. The work [20] addresses trust management theoretically and provides a general framework for developing trust management for IoT. According to the authors, it is difficult to establish a trust mechanism for the IoT as a whole, thus establishing trust mechanisms for the IoT in the form of layers.

The work [21] proposes a dynamic trust model to assess the reliability level of nodes in a social IoT environment. The proposed approach considers the spread of trust to be distributed and centralized; however, the proposal does not detail how the two means of spreading trust can be used together. The goal of [22] is to identify the behavior of malicious nodes and prevent possible on-off attacks in a multiservice IoT environment. This approach does not consider evaluating nodes that are not directly connected. In [23], the proposed system takes the context and environmental information from IoT sensor devices and delivers them to trusted agents who filter the information and send it to the cloud-deployed MAPE-K control loop. So, the trust operates with a cloud service.

3 TRUMIT

Trust Manager for IoT (TRUMIT) is a fully decentralized and distributed trust management system for IoT environments. TRUMIT does not need a centralized trust entity, so nodes are able to interact independently to evaluate, recommend, and store trust information. In this approach, each node self-organizes its trust network to obtain and recommend node reliability information. This information is obtained through direct and indirect experiences. Thus, the trust value calculation is performed locally, reducing the number of interactions needed between nodes to obtain the evidence of trust.

3.1 TRUMIT Features

For a better understanding of the proposed work, Table 1 presents the notations used in this article.

Table 1 Notation used

Notation	Description
N_{i}	Node identity *i*
G^i	Network node trust graph *i*
D_{ij}	Direct node trust *i* relative to the node *j*
D_{jk}	Node direct trust value *j* relative to node *k*
ΔT_p	Interval between information exchanges
IS	Successful interactions
IU	Unsuccessful interactions
R_{ijk}	Node recommendation *j* relative to the node *k*
T_{ij}	Node direct final trust value *i* relative to the node *j*
RT_{ik}	Node indirect final trust value *i* relative to the node *k*
α	Adjustment weight of direct trust value between [0, 1]

3.2 Trust Network Creation

Upon entering the system, each node creates its own trust network in the form of a static, connected, undirected trust graph $G^i = \{V^i, E^i\}$ in a self-organized way. Initially, nodes have information only about nodes with which they had direct trust relationships, and only this data is stored in the trusted network.

Thus, at predetermined time intervals ΔT_p, nodes exchange trust evidence stored in their trust networks with their neighbors. This evidence is used to populate the *node* trust table. This way the stored trust values can be propagated over the network.

3.3 Direct Trust

Direct trust is assessed based on the node's own experience with its neighbors through observations on whether interactions between nodes are successful or not. Thus, when a node interacts directly with a neighboring node, it evaluates the trust of the neighboring node according to the Eq. 1, where D_{ij} corresponds to the trust rating of node N_i relative to node N_j, *IS* is the number of successful interactions performed by node *i* with node *j* and *IU* is the number of unsuccessful interactions.

$$D_{ij} = \frac{IS}{IS + IU} \tag{1}$$

3.4 Indirect Trust

Indirect trust is evaluated based on the information the node receives from neighbor nodes, so when an interaction occurs between the directly connected nodes, the nodes involved in that interaction exchange the stored values in their trust tables. Thus, an N_k node that is not directly connected to a N_i node can be evaluated indirectly. This assessment is called a recommendation. Thus, the indirect trust of the N_k node based on the recommendation of the N_j node is calculated by the N_i node as follows:

$$R_{ijk} = \left(D_{ij} + D_{jk}\right) \tag{2}$$

Thus, R_{ijk} corresponds to the indirect trust assessment of node i relative to node k, D_{ij} indicates the direct trust value of node i in node j and D_{jk} the direct trust value of node j with respect to node k.

3.5 Calculation of Final Trust Value

The trust value is a real number in the range of [0, 1], where 1 indicates total trust and 0 distrust. Thus, the calculation of the final trust value of node i relative to node j is denoted by Eq. 3. Where α and $1 - \alpha$ are the weights to adjust the final direct trust value between 0 and 1. So $0 < \alpha < 1$, $0 < T_{ij} < 1$.

$$T_{ij} = \alpha D_{ij} + (1 - \alpha) \tag{3}$$

The calculation of the final trust value for our recommended nodes is denoted by Eq. 4. Where α and $1 - \alpha$ are the weights to adjust the final indirect trust value between 0 and 1. Since the direct trust values, D_{ij} and D_{jk} are in the range [0, 1], the indirect trust of node i to node k (based on node j recommendations) is always smaller than the values of D_{ij} or D_{jk}. If more than one node recommends node $_k$, node $_j$ considers the recommendation with the newest *timestamp.*

$$RT_{ijk} = R_{ijk} + (1 - \alpha) \tag{4}$$

4 TRUMIT Evaluation

Evaluations of the proposed model were performed considering the presence of honest and malicious nodes. Cooja Simulator was used to simulate the interactions between nodes in an IoT environment. Table 2 contains the parameters used in this evaluation.

Table 2 Evaluation parameters 1

Parameters	Scenario 1	Scenario 2	Scenario 3
No of the nodes	50	50	50
Communication technology	802.15.4	802.15.4	802.15.4
Transmission radius (m)	20	30	50
Simulation area (m^2)	10,000	10,000	10,000
Simulation duration (s)	240	240	240
Direct trust parameter (α)	0.0–0.5	0.0–0.5	0.0–0.5
Interaction interval (ΔTp)	10 s	10 s	10 s

Considering a scenario with no malicious nodes, TRUMIT was evaluated against the number of interactions required for a given node to have full knowledge of the network. As can be seen in Fig. 1, when the direct and indirect trust parameter (α) is zero, nodes need fewer interactions to obtain the evidence of trust from the entire network. This is because with $\alpha = 0.0$ the nodes accept evidence form all other nodes in the network. Thus, data is exchanged quickly. However, when the value of α increases, nodes need more interactions to get reliable evidence from the entire network.

In this evaluation, it was possible to verify that, for the scenario with a transmission radius of 20 m, nodes perform a higher number of interactions to obtain evidence of network reliability; however, in this scenario, there was no network convergence, nodes collect only 88% evidence of network trust. With respect to the other scenarios, it was found that with the 30 m transmission radius, the nodes required fewer interactions to obtain reliable network evidence compared to the other two scenarios.

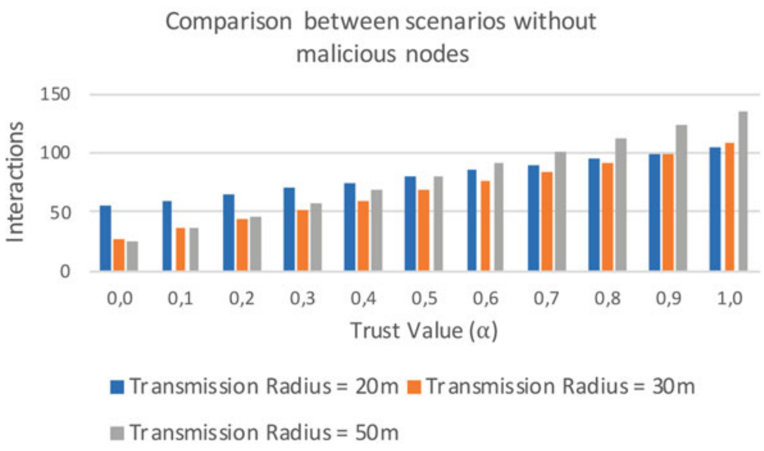


Fig. 1 Scenarios without malicious nodes

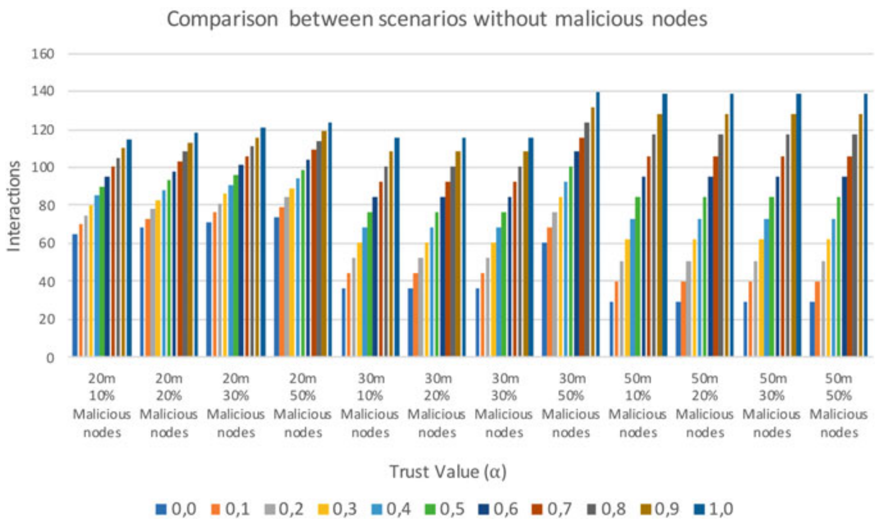


Fig. 2 Scenarios of malicious nodes

The evaluation considering malicious nodes was performed in scenarios with 10, 20 and 30% of malicious nodes in the network. In all evaluated scenarios, nodes behave maliciously after forming their trusted networks. The malicious nodes in this experiment change the node values of their confidence tables to 0.0, so these nodes will not be recommended to their neighbors.

Figure 2 shows a comparison of node interactions between scenarios with malicious nodes. Thus, it is possible to verify that in scenarios with higher density, there is no significant increase between the values of interactions performed by the node, proving the effectiveness and resilience of TRUMIT in environments with the presence of malicious nodes.

5 Conclusion

This paper proposes a new trust model called TRUMIT to manage trust relationships between devices in an IoT environment. The proposed model is fully decentralized and distributed, so devices can quickly identify which other devices are reliable or not. In this approach, each node self-organizes its trusted network to obtain, evaluate, and recommend device reliability information. Devices obtain this information through direct and indirect observations. Thus, a device calculates the trust value locally, decreasing the number of interactions between them. Over time, devices will gain trust information from other devices that are not directly connected through recommendations and add them to your trust network, so when there is the direct interaction between them in the future, the devices will have a certain level between them, without even interacting directly at another time. The proposed system proved resilient to bad-mouthing attacks. Thus, it can be concluded that the system has the functionality required for trust management in an IoT environment. Future work of TRUMIT includes implement a punishment scheme for nodes that do not match the metrics used for trust value composition, add new metrics for calculating trust value, such as energy consumption, cooperativity, and communities of interest and finally test TRUMIT in a real IoT environment.

References

1. Ngu, A.H., et al.: IoT middleware: a survey on issues and enabling technologies. IEEE Internet of Things J. **4**(1), 1–20 (2016)
2. Kumar, S.A., Vealey, T., Srivastava, H., Security in internet of things: challenges, solutions and future directions. In: 2016 49th Hawaii International Conference on System Sciences (HICSS). IEEE (2016)
3. Ben-Daya, M., Hassini, E., Bahroun, Z.: Internet of things and supply chain management: a literature review. Int J Prod Res **57**(15–16), 4719–4742 (2019)
4. Ali, B., Awad, A.: Cyber and physical security vulnerability assessment for IoT-based smart homes. Sensors **18**(3), 817 (2018)
5. Azzedin, F., Ghaleb, M.: Internet-of-things and information fusion: trust perspective survey. Sensors **19**(8), 1929 (2019)
6. Mendoza, C.V.L., Kleinschmidt, J.H.A.: distributed trust management mechanism for the Internet of things using a multiservice approach. Wireless Personal Commun. **103**(3), 2501–2513 (2018)
7. Chen, J., et al.: Trust architecture and reputation evaluation for internet of things. J. Ambient Intell. Humaniz. Comput. **10**(8), 3099–3107 (2019)
8. Guo, J., Chen, R., Tsai, J.J.P.: A survey of trust computation models for service management in internet of things systems. Comput. Commun. **97**, 1–14 (2017)
9. Palade, A., et al.: Middleware for internet of things: a quantitative evaluation in small scale. In: 2017 IEEE 18th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM). IEEE (2017)
10. Gu, L., Wang, J., Sun, B.: Trust management mechanism for internet of things. China Commun. **11**(2), 148–156 (2014)
11. Din, I.U., et al.: Trust management techniques for the internet of things: a survey. IEEE Access **7**, 29763–29787 (2018)
12. Chen, D., et al.: TRM-IoT: A trust management model based on fuzzy reputation for internet of things. Comput. Sci. Inf. Syst. **8**(4), 1207–1228 (2011)
13. Bao, F., Chen, I.-R.: Dynamic trust management for internet of things applications. In: Proceedings of the 2012 International Workshop on Self-aware Internet of Things. ACM (2012)
14. Chen, R., Bao, F., Guo, J.: Trust-based service management for social internet of things systems. IEEE Trans. Dependable Secure Comput. **13**(6), 684–696 (2015)
15. Chen, R., Guo, J., Bao, F.: Trust management for SOA-based IoT and its application to service composition. IEEE Trans. Serv. Comput. **9**(3), 482–495 (2014)
16. Chen, Z., et al.: A scheme of access service recommendation for the social internet of things. Int. J. Commun. Syst. **29**(4), 694–706 (2016)
17. Bao, F., Chen, R., Guo, J.: Scalable, adaptive and survivable trust management for community of interest based internet of things systems. In: 2013 IEEE Eleventh International Symposium on Autonomous Decentralized Systems (ISADS). IEEE, (2013)
18. Saied, Y.B., et al.: Trust management system design for the internet of things: a context-aware and multi-service approach. Comput. Secur. **39**, 351–365 (2013)

19. Mahalle, P.N., et al.: A fuzzy approach to trust based access control in internet of things. Wireless VITAE 2013. IEEE, (2013)
20. Wang, J.P., et al.: Distributed trust management mechanism for the internet of things. Appl. Mech. Mater. 347. (Trans Tech Publications, 2013)
21. Nitti, M., Girau, R., Atzori, L.: Trustworthiness management in the social internet of things. IEEE Trans. Knowl. Data Eng. **26**(5), 1253–1266 (2013)
22. Mendoza, C.V.L., Kleinschmidt, J.H.: Mitigating on-off attacks in the internet of things using a distributed trust management scheme. Int. J. Distrib. Sens. Networks **11**(11), 859731 (2015)
23. Namal, S., et al.: Autonomic trust management in cloud-based and highly dynamic IoT applications. 2015 ITU Kaleidoscope: Trust in the Information Society (K-2015). IEEE (2015)

Data Visualization Techniques for Monitoring Real-Time Information of Cold Chain

Jerson Rivas Tucto, Nora Castillo Talexio, and Pedro Shiguihara Juárez

Abstract Real-time monitoring of temperature is a critical factor in ensuring the integrity of food during the cold chain. In this work, we compare techniques related to real-time data visualization to contribute to more efficient monitoring of the cold chain. Three real-time data display attributes were evaluated, and we constructed a dataset based on the Frisbee database (CDD). In this paper, we proposed graphics containing different line and area techniques to be evaluated for a specialist. The proposed graphs contained the line and area techniques that, when performing the experiment, obtained a higher success rate compared to the auto-charting technique. However, it was evidenced that elements such as color facilitate the detection of anomalies and trends in temperature change due to its high percentage of effectiveness in the results.

Keywords Data visualization · Cold chain · Monitoring real-time information

1 Introduction

Companies in their daily operations generate data every day [1]. Visualization techniques are useful for data exploration, allowing them to understand complicated scientific findings to experts and inexperienced in the field [2]. In the cold chain, which handles a series of products highly sensitive to temperature and other environmental conditions, if it is not properly controlled, it can cause adverse effects on human health, product prices, and food availability. Products can be constantly

J. Rivas Tucto · N. Castillo Talexio (✉)
Department of Information System, Universidad Peruana de Ciencias Aplicadas, Lima, Peru
e-mail: u201411546@upc.edu.pe

J. Rivas Tucto
e-mail: u201412732@upc.edu.pe

P. Shiguihara Juárez
Department of Computer Science, Universidad Peruana de Ciencias Aplicadas, Lima, Peru
e-mail: pedro.shiguihara@upc.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_51

monitored through panels [3]. Three key problems were identified that limit the performance of the cold chain: insufficient capacity, lack of the latest technology or "optimal" equipment, and inadequate temperature monitoring and maintenance systems [4]. The cause of the first problem described is the poor visibility and understanding of the current state of the cold chain equipment. For this reason, the solutions addressed for these adequacy gaps are to develop an accurate graph of the capacity gaps of the cold chain based on current and future needs for the creation of an action plan and achieve the development of a resource mobilization plan for the implementation of actions and finally to monitor the implementation of the plan. On the other hand, the second problem addressed is the poor technology that affects the performance of the cold chain due to the lack of knowledge of the benefits, the new equipment has not to mention the costs incurred. Furthermore, the proposed solution for this problem is the implementation of pilot tests to identify better models that allow the collection of performance, profitability, and usability tests. Finally, the third problem covers inadequate temperature monitoring and maintenance systems, using temperature monitoring. The problem of data representation in visual media is identified, particularly in the visualization of real-time data [5]. Real-time sensor data visualization allows cold chain organizations to make better decisions during and after product distribution, such as adjusting refrigeration systems, minimizing temperature variability to ensure product quality, transport line optimization, and product tracking for withdrawals [3]. When users cannot decide which type of visualization is the most appropriate, they tend to interpret the data incorrectly and that is why visualization techniques are developed to help analyze data effectively [6]. We propose a comparison of techniques and elements of data visualization using time series for the cold chain to improve monitoring and decision making.

2 Related Work

In [7], the line technique is applied to demonstrate the real-time humidity of the stored rice bags. According to [8], this technique is related to the identification of trends and changes over a period of time. In [9], the auto-charting technique is used to visualize the seriousness of air pollution in real time in the city of Beijing. According to [10], auto-charting allows visualizing the variation of data regarding time through the bar graph. In [11], the area technique is represented through profitability and risk metrics that are updated in real time. Area graphs are similar to line graphs with areas of colors below them. In [12], time intervals in a view for real-time monitoring of the water temperature of the rivers are presented. In [13], it is mentioned that these should be placed close to the graphs to facilitate their understanding. In [14], the quality thresholds allow visualizing when the data exceeds the proposed limits. According to [7], in [11] the widget indicator is used to show the current value of some measurements in real time. Furthermore, concepts such as data visualization and dashboard are defined. In [15], the visualization of data is responsible for the design, development, and application of the graphs representation of the data obtained from different sources processed by the computer to facilitate decision making. According to [16], a dashboard allows visualizing the data for monitoring and easy understanding.

3 Method

For the construction of graphs, a time series from the cold chain is used. Data is collected and processed, and the visualization techniques to be used are selected.

3.1 Data Collection and Processing

In this paper, the process of monitoring of cold chain products is used. Figure 1 shows the workflow stages for the construction of data sets. The tool that will be used to extract data is named Frisbee Cold Chain Database of the European project.[1] CCD is freely accessible to all users. This database is contributed by scientists and those interested in the cold chain [17].

First, the following attributes were selected in CCD: stage of a cold chain, food storage temperature, characterization of food, and type of food. The values chosen were distribution warehouse, chilled, ready to eat, and milk products, respectively. Then the profile is assembled, the information is extracted in a flat-file format. The result obtained by the tool was 92 data sets that contain the field time (minutes) and temperature. To perform the analysis of anomalies that occurs during a cold chain path, we add three columns: maximum and minimum threshold of temperature, and state which indicates if there is an alert or not. Finally, we obtained resulting data sets for the experimentation. Data sets constructed are available on Kaggle repository[2] for the scientific community might access.

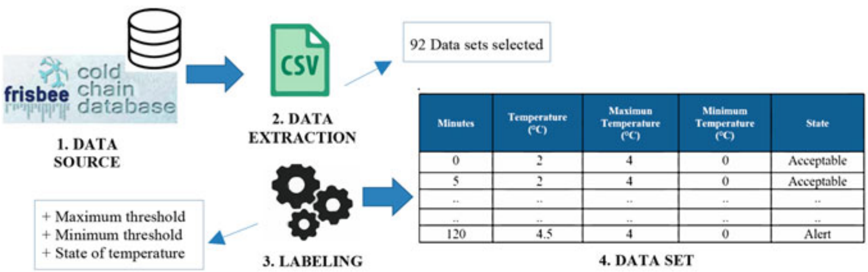


Fig. 1 Workflow for the construction of data sets about the cold chain of milk that is used as a data source for the case study

[1] http://www.frisbee-project.eu/coldchaindb.html.

[2] https://www.kaggle.com/u201411546/data-sets-cold-chain-milk.

3.2 Selection of Data Visualization Techniques and Elements Identify

For this work, we choose data visualization techniques and features described in the previous section. We selected three methods that allow us to show real-time data trends, changes, and anomalies. The techniques are the line chart, area chart, and auto-charting. The time is put on the X-axis and the measure on the Y-axis; in this way, changes are observed over time [1]. After having selected the techniques, we proceed to select data elements such as color, legend, quality threshold and others mentioned, for creating the design. Data visualization techniques are used, in conjunction with data elements. In Fig. 2, some of the charts constructed can be seen. Figure 2a shows an area chart that contains colored elements in a one-hour interval. Figure 2b shows the technique of auto-charting using the legend of data in the interval of a day. Figure 2c shows a line chart using quality thresholds. Figure 2d shows a line chart using four colors and its legend to represent the data state.

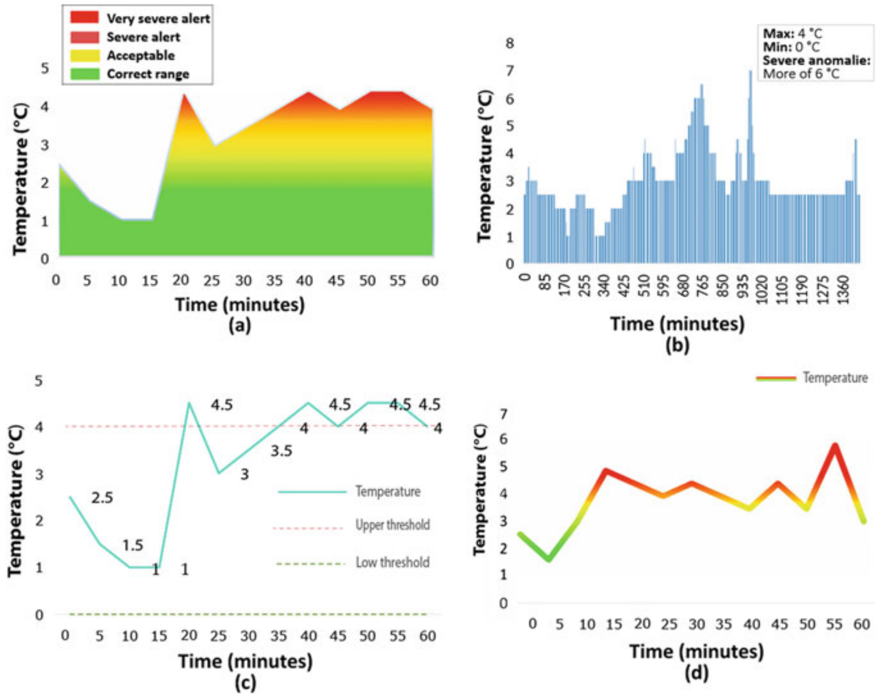


Fig. 2 Graphs related to the cold chain case

4 Experimental Settings

According to [11], to validate the data visualization techniques, specialists help to validate the graphs created. According to [10], data visualization should allow people to identify trends and anomalies. In [1], colors were used to represent data.

Also, as mentioned in previous sections, real-time dashboards should allow users to understand results easily to improve decision making.

5 Results

Table 1 shows the results obtained for the attributes of identification (which are related to identifying trends and anomalies, colors use to represents data and easy-understand form) and use of colors. An analysis was made on these aspects, which is presented in the subsequent table. The graphs that had the highest rate of success were graph 3 (50.00%). This proposal shows the data in a line chart using simple legends. For A2, Fig. 3 is still the one with the highest percentage of success in identifying the number of alerts (83.33%). Graph 3 in A2 also shows a line chart but it had the color element added. In addition, it can be observed that the graphs with a shorter time interval obtained a higher average. On the other hand, the failure rates of A1 are higher than those of A2 in all cases.

Table 1 Results of the identification and use of the color attribute. The following table describes the success percentages on anomalies detection

Analysis	Result	# Graph Showed		
		1 (%)	2 (%)	3 (%)
A1	Success	11.11	33.33	50.00
	Failure	88.90	66.70	50.00
A2	Success	27.78	61.11	83.33
	Failure	72.20	38.90	16.70

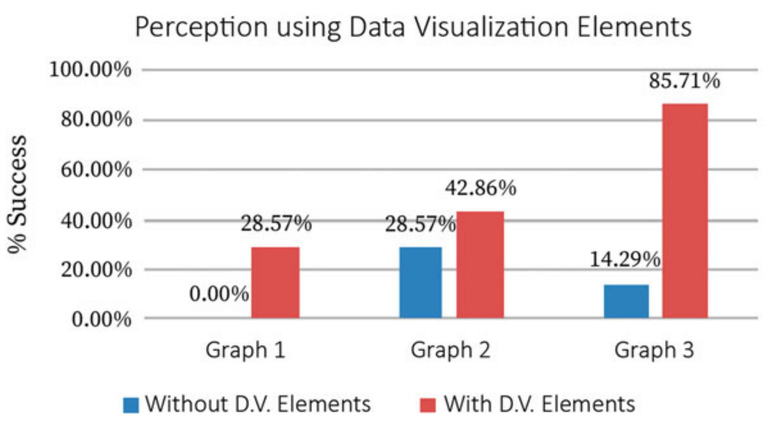


Fig. 3 Perception using data visualization elements

Table 2 Results of the easy-understand attribute with different instruction-level related to data visualization

Level	Graph 1 (%)	Graph 2 (%)	Graph 3 (%)
Novice	8.17	6.00	8.50
Intermediate	7.57	5.00	7.29
Advanced	9.20	5.60	8.00

According to Table 2, for all levels of knowledge on dashboards, the auto-charting technique obtained the lowest average. This dashboard contains a line chart with an interval of 30-min and quality thresholds. This proposed dashboard shows an area chart that contains colors to monitor temperature status. In addition, a gauge widget was added to show the current value of the temperature. It was found that the perception of the people who have some problems in the eyesight improves remarkably when adding data visualization elements (D.V Elements) in comparison with those when they observe simple graphs.

According to the results of the evaluation, line chart showed better performance when in 3 of the 4 defined attributes; as mentioned in [17], this technique allowed to visualize trends and changes in real time more effectively. Another technique mentioned in [17] with favorable results was the area chart. This technique is the best option for people with a novice level of knowledge in dashboards. The auto-charting technique showed the most relatively low results for the detection of anomalies and easy-understand attributes compared to the other mentioned techniques. Regarding data elements included on dashboards, the results showed the positive impact of using colors to detect anomalies in a cold chain scenario. As a consequence, this could allow users to make decisions at the right time, as mentioned in [9]. This element is represented in a legend to distinguish the meaning of each color level. Moreover, the graphs that showed the data in a shorter time interval (30 min), such as the graphs proposed in [12] were more efficient to be observed. In addition, studies show people had a better performance in the detection of anomalies with graphs that contained data elements in comparison with graphs without them. Which premise will be tested in the future with an analysis with individuals, where this will be regulated by the respective ethical committee, if estimated necessary. Results on the identification of anomalies are using time series through the use of data visualization elements in people who have some visual disability.

6 Conclusion

This work was aimed at determining the suitability of data visualization in the cold chain industry. A data set related to the milk cold chain was constructed. Attributes about trends and anomalies, color to represent data, and easy-understand form with the respective graphs were evaluated.

For the attribute of identification, the second and third graphs show the highest success percentages with 33.33% and 50%, respectively. These graphs were used line, and area techniques were without adding any data element. However, when evaluating the results of the second attribute mentioned, the third graph, which contained data elements such as color, obtained outstanding success percentages of 83.33% and only 16.70% failure. For the last evaluated attribute, the second dashboard that uses the auto-charting technique obtained the lowest punctuation for all levels of knowledge. In contrast, the first dashboard showed greater ease of understanding, obtaining a range from 7 to 10 and the third graph with an average of 8.50, this contained a greater number of elements than those shown in the other graphs. This research offers opportunities to improve the visualization of data from the perspective of the cold chain industry, and as future work, we will investigate how data visualization techniques can be adapted to different industries.

References

1. Sadiku, M., Shadare, A., Musa, S., Akujuobi, C.: Data visualization. Int. J. Eng. Res. Adv. Technol. **12**, 11–16 (2016)
2. Szabo, S., Johnston, K., Lloyd, A.: Advanced data visualization in health economics and outcomes research: opportunities and challenges. Appl. Health Econ. Health Policy **17**, 433–441 (2019). https://doi.org/10.1007/s40258-019-00476-5
3. Chaudhuri, A., Dukovska-Popovska, I., Subramanian, N., Chan, H., Bai, R.: Decision-making in cold chain logistics using data analytics: a literature review. Int. J. Logistics Manage. **29**, 839–861 (2018). https://doi.org/10.1108/ijlm-03-2017-0059
4. Ashok, A., Brison, M., Letallec, Y.: Improving cold chain systems: challenges and solutions. Vaccine **35**, 2217–2223 (2017). https://doi.org/10.1016/j.vaccine.2016.08.045
5. Holtz, S., Valle, G., Howard, J., Morreale, P.: Visualization and pattern identification in large scale time series data. In: IEEE Symposium on Large Data Analysis and Visualization, pp 129–130. IEE press, New York (2011) https://doi.org/10.1109/ldav.2011.6092333
6. Franciscani, G., Santos, R., Ottoni, R., Pesce, J.P., Meira, W., Melo-Minardi, R.: An annotation process for data visualization techniques. DATA ANALITYCS 2014: The Third International Conference on Data Analytics, pp. 24–32. IARIA XPS Press, Rome (2014)
7. Serikul, P., Nakpong, N.: Smart farm monitoring via the Blynk IoT platform. In: 16th International Conference on ICT and Knowledge Engineering (ICT&KE), pp. 1–6. IEEE Press, Bangkok (2018). https://doi.org/10.1109/ictke.2018.8612441
8. Data Visualization Techniques. https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/data-visualization-techniques-106006.pdf. Last accessed 2019/07/04
9. Chen, P.: Visualization of real-time monitoring datagraphic of urban environmental quality. EURASIP J. Image Video Process. **42**, 1–9 (2019). https://doi.org/10.1186/s13640-019-0443-6
10. Toasa, R., Maximiano, M., Reis, C., Guevara, D.: Data visualization techniques for real time information—a custom and dynamic dashboard for analyzing surveys results. In: 13th Iberian Conference on Information Systems and Technologies (CISTI), pp 1–7. IEEE Press, Caceres (2018). https://doi.org/10.23919/cisti.2018.8398641
11. Gürdür, D., Raizer, K., El-khoury, J.: Data visualization support for complex logistics operations and cyber-physical systems. In: 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Prague, pp 200–2011 (2019). https://doi.org/10.5220/0006569402000211

12. Ellison, J., Smethurst, P., Morrison, B., Keast, D., Almeida, A., Taylor, P., Bai, Q., Penton, D., Yu, H.: Real-time river monitoring supports community management of low-flow periods. J. Hydrol. **572**, 839–850 (2019). https://doi.org/10.1016/j.jhydrol.2019.03.035
13. Visual Analysis Besr Practices. http://www.dataplusscience.com/files/visual-analysis-guidebook.pdf. Last accessed 2019/07/04
14. Kozik, R., Choras, M., Puchalski, D., Renk, R.: Platform for software quality and dependability data analysis. In: Contemporary Complex Systems and Their Dependability: Proceedings of the Thirteenth International Conference on Dependability and Complex Systems DepCoS-RELCOMEX, pp 306–315. Springer, Brunow (2019). https://doi.org/10.1007/978-3-319-91446-6_29
15. Sosulski, K.: Data Visualization Made Simple: Insights into Becoming Visual, 1st edn, pp. 1–40. Routledge, New York (2018). https://doi.org/10.4324/9781315146096
16. Gogou, E., Katzaros, G., Derens, E., Alvarez, G., Taoukis, P.S.: Cold chain database development and application as a tool for the cold chain management and food quality evaluation. Int. J. Refrigeration **52**, 109–121 (2015). https://doi.org/10.1016/j.ijrefrig.2015.01.019
17. Young, M.: The Technical Writers Handbook: Writing with Style and Clarity. University Science Books, California (2003)

Bike-Sharing Management System Using IoT

Renato Mercado Luna, Gabriel Benavente Soto, Daniel Subauste Oliden, and Alfredo Barrientos Padilla

Abstract Nowadays, due to climate change and many other facts affecting daily life a trend to use eco-friendly transportation ways has arisen, and from them the one mostly used and with the highest acceptance is biking. Therefore, several companies have emerged offering bike-sharing systems, and those systems have been greatly accepted in the different metropolises around the world. The generalization of these systems has created a new need: to manage them efficiently. Consequently, in this work, we propose a software architecture and the implementation of a bike-sharing management system using the Internet of things (IoT).

Keywords Architecture · Management system · IoT · Cloud · Cloud architecture

1 Introduction

In recent years, the use of bike-sharing systems has increased. These systems allow many people to walk along with the city or to get around their work or study places while exercising, thus avoiding traffic and environmental pollution. The latter benefits had been determiners for the exponential growth experienced by this eco-friendly system. For that reason, several governments had decided to subsidize bike-sharing systems or to provide some support like infrastructure or bikes.

However, if these systems are not properly managed they create huge problems within the cities. Being one of these problems the overcrowding of bikes in delivery

R. M. Luna (✉) · G. B. Soto · D. S. Oliden · A. B. Padilla
Universidad Peruana de Ciencias Aplicadas (UPC), Lima 15023, Peru
e-mail: u201510738@upc.edu.pe

G. B. Soto
e-mail: u201314158@upc.edu.pe

D. S. Oliden
e-mail: daniel.subauste@upc.pe

A. B. Padilla
e-mail: pcsiabar@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_52

points. This usually happens because people do not have civic awareness regarding the use of these systems. This lack of control about delivery points has come to cause blockage in roads and pathways. In addition, the lack of information regarding the condition of bikes makes it difficult to use the system at the beginning. These problems had created the need to use efficient management systems.

It is because of this need that we suggest a cloud architecture for the efficient management of bike-sharing. This architecture seeks to have high availability and scalability, and for that reason we decided to use a serverless architecture. Regarding security, we chose to use Cognito, an AWS service for users' management and their correspondent sessions to keep the privacy of users' data [1]. Finally, the cloud service we chose was AWS due to the low latencies of its services and the opportunity to use AWS Lambda, a service used to create event-driven serverless platforms [2]. Besides, we took cloud-edge-beneath architecture [3] as a reference due to the high scalability and adaptability it possesses. For the IoT device, we selected a Raspberry Pi since that device is adequate for the needs of this project.

2 Literature Review

Bike-sharing systems had exponentially increased their use over the past few years. However, these systems tend to need state funding because of the high infrastructure expenses [4]. Likewise, these systems are likely to have a low use rate in relation to conventional transportation services, besides a high rate of losses [5]. For the difficulties already mentioned, these systems have a reduced use.

With those disadvantages on the scene, a few architectures have arisen seeking to provide improvements to the system. Each of those architectures has different elements that characterized them. One of them proposed as key benefit the environmental measurements within the cities; these measurements allow users to stay away from certain areas of the city in certain moments in order to avoid high temperatures, high pollution or excessive humidity. We need to highlight the fact that this architecture only works properly in intelligent cities [6]. Another architecture that we reviewed was cloud-edge-beneath, due to the high scalability and adaptability. On the other hand, we reviewed an architecture that uses big data for monitoring environment elements of a city by using sensors, such architecture proposes to capture local data just when a user is around a certain area, and besides it shares sensor readings to avoid unnecessary repetitions. Since this reading is massive, it has the opportunity to analyze information using big data and offer recommendations to users according to their behavior [7].

Other architectures reviewed recommend to add a router layer to improve server outputs and reduce latencies. Besides that, there are other architectures that advise adding repeaters to simplify the work of IoT devices as well as allowing them to send ongoing data. Similarly, we reviewed an architecture that proposes to use a SIM card to provide IoT device with autonomy, with the limitation of not being able to send enough information to carry out big data due to the balance limit on the SIM [8].

After a comprehensive review of these architectures, a new question arose: What are the communication protocols bike-sharing system modules use? First of all, some authors mentioned the speed benefits of MQTT protocol and proposed to combine it with HTTP protocol to provide high performance to cloud systems. They also highlighted that the MQTT protocol is advisable for IoT devices because it is lighter than others [9]. Besides, there are other authors that proposed [10] six architecture levels aiming to find the best efficiency of Internet of things applications. From these exhaustive studies, we determined that MQTT protocol is optimal for information massive sending since it works with a publisher–subscriber design pattern; even so, it does not ensure the security or integrity of information. Other authors suggested IoB-DTN protocol designed to nets that do not have ongoing connectivity. This protocol is focused on systems that only send information when the receptor is in range, and the biggest part of that information is stored in the device's cache, for example, a bike system connected among them that only transmits information when bikes are in the range [11].

After inquiring about communication protocols among system modules, our next step is to learn what are the usage modes of existing bike-sharing systems. The only modes currently used are systems that have docks and the ones that dispense them. The ones with docks are called third-generation bike-sharing systems, and the ones without them are fourth-generation systems. Third-generation systems are more expensive since they need docks but they allow people to maintain order within the city. The authors conclude that fourth-generation systems are viable for cities with a cycling culture and third-generation systems for cities without it [12].

Similarly, it is important to know the impact these systems have in the cities or areas that used them. Some authors [13] tried to identify the environmental benefits of bike-sharing systems and analyzed them from two sides: (a) substitution, which means the number of kilometers traveled by bicycle instead of using a conventional transport, and (b) complement, which means the difference in travel time of a person when using a bike to get to the bus or train station from home. The main result was that the use of bike-sharing systems dramatically reduces traffic and energy consumption; similarly, it reduces toxic gas emissions, improves public health and promotes economic growth. For one side, a few authors [14] studied bike-sharing systems in Asia through a simulation model to determine the impact those create; the key result of that study was that a free bike-sharing system can reduce environmental damage costs in about 16%, equivalent to 1, 5 million dollars per year. Likewise, the physical effort these systems suppose improves users' health and it can even prevent 22 early deaths per year. We need to mention that bike-sharing systems have low acceptance among the big motorbike population of that continent. Nevertheless, they concluded stating that bike-sharing systems can be a success in Asia, mostly in a big metropolis.

An additional element is to determine what system elements motivate users to use it. First, some authors [15] tried to determine which factors are decisive for users of bike-sharing systems, and to achieve this they surveyed 529 people. The result of the said survey showed three main factors: (a) health and environment, the benefits of these systems for health and environment; (b) social influence, the perception that society has against biking as a sustainable alternative; and finally, (c) cyclist lifestyle, the motivation for cycling since it is their lifestyle. Therefore, to improve the retention of bike-sharing systems it is essential to address these elements.

Finally, it is necessary to determine what benefits other bike-sharing systems have. On the one hand, a few authors [16] suggested improving users' retention by preventing them to run out of parking spots using a Petri net to determine the anxiety factor that regulates the probability that a station is full at a certain time. On the other hand, other authors [17] suggested to provide cyclists with the necessary information to be able to circulate, and they encouraged cyclists to use their phones to receive information from the accelerometer, gyroscope, elective compass and GPS to provide statistical data on travel, average distance, among others, highlighting the importance of statistical data for cyclists.

3 Architecture Development

Once the literature review is finished, it is necessary to plan an architecture for the system to be implemented. We took cloud-edge-beneath as a reference to propose a software architecture able to escalate and have high availability. The proposed architecture uses AWS service for the cloud and IoT Raspberry Pi for devices. Services and hardware used in the development of the system are outlined below.

3.1 Services Used

To implement the architecture, we used multiple AWS services: To store information created by the system, we used DynamoDB, a non-relational database; for the processing and interaction with the database, we used 12 lambda functions and an instance of an EC2 where we created an API and a Web application for the administrator users and system operators; for the interaction among bikes, users and bikes with the cloud we used an API gateway and AWS service to manage client requests made to the cloud through HTTP protocol; to run users' sessions, we used Cognito, since it allows to provide security to users' passwords and the creation of sessions within the application; to administer users' interactions, we created a mobile application in Android and IOS; to reach information from bikes, we used a microcontroller in the bike and for administrator users' interaction we created a Web application.

3.2 Hardware Used

- Raspberry Pi 3: Single-board computer used to control system hardware elements.

- Lithium-ion batteries: rechargeable batteries used to feed system engines and sirens.

3.3 Architecture Design

As we mentioned early, we chose cloud-edge-beneath as a reference architecture, and from that we adopted the three-layer distribution [3]. Regarding the architecture, a serverless solution was chosen, since this allows scalability to be obtained, which is crucial for the project due to the need for bike-sharing solutions to be able to provide good performance at peak times [18]. Besides, a non-relational database was used, this being considered a good practice by AWS for the creation of a serverless solution [19]. Finally, we defined three clients for the application: system bikes with their microcontrollers, users with mobile applications and system administrator users who use a Web platform for system administration and management. These clients interact with the elements that comprise the system such as bikes, parking places, among others; these operations are processed and stored in the cloud. In order to facilitate the understanding of our architecture proposal, we made the following graphic (Fig. 1) and explant it in the following paragraphs:

This architecture has three separate layers, each one with a specific purpose: (a) The first layer called the physical layer hosts the Raspberry Pi device, which sends and receives information from the cloud through a script. In addition, in this layer are the mobile clients and the Web graphic interface, which constantly communicate with the cloud; (b) the second layer or edge layer comprises the communication channels between the physical layer and the cloud. In the case of IoT devices, they communicate through an API gateway and the Cognito authentication service. The latter is only used for logins, and this service was chosen due to its high performance; (c) finally, the third layer or cloud layer contains all the processes, executed through Lambda functions, the Web application, contained in an EC2, and the database DynamoDB. This layered distribution favors scalability as it generates low cohesion between modules and high coupling.

Additionally, in this three-layer architecture, we adopted the serverless architecture which was applied in the cloud processes that were implemented as Lambda functions. Those functions are only instantiated when they are called by mobile devices or IoT devices and to perform a users' state review periodically in the system. The display of the system administrator users is shown in Fig. 2.

The image above presents how the system allows to visualize the bikes in movement, parking places and the incidents reported by the users, this data can be displayed in real-time. As explained in the system legend, the red markers represent incidents, the yellow ones are stations and the blue ones are bicycles in use. The unused bikes are not represented because they are contained in the stations, and this information can be viewed in the details of the station. With that information, the system can be effectively managed in order to provide a service that does not create the aforementioned problems associated with bike-sharing systems.

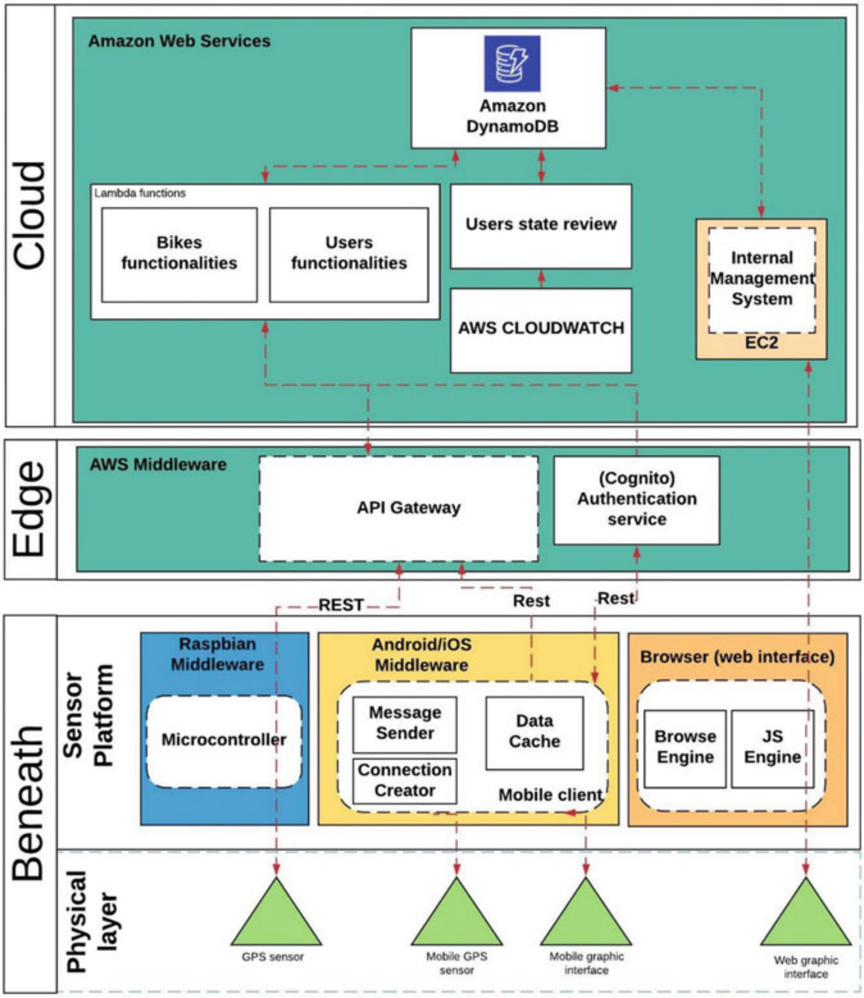


Fig. 1 Proposed architecture

4 Conclusions

This paper presents a three-layer cloud-edge-beneath architecture using a serverless architecture for cloud functions. After reviewing the submitted architectures and solutions, we implemented a bike-sharing management system using the designed architecture. Being the main advantage of this system are scalability and adjustability.

Finally, we identified that further investigation toward the analysis of the data generated by the usage of the system can be prosecuted with the usage of machine learning and analysis than can be done to the patters that can be found in the user's

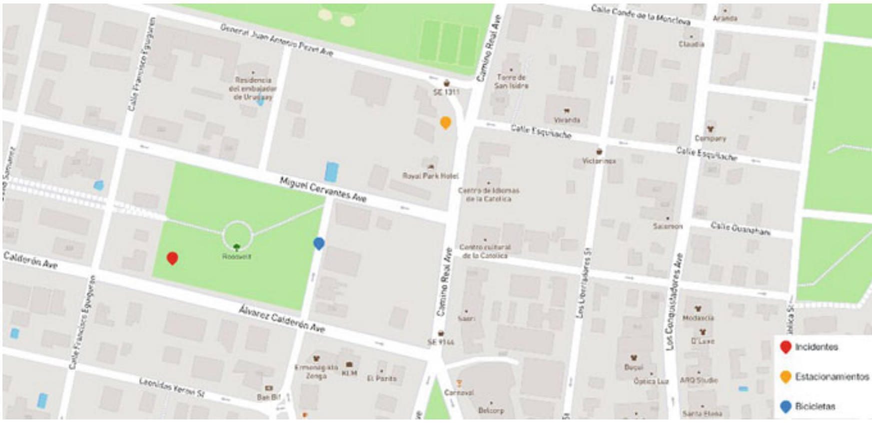


Fig. 2 Management system

behavior. Offering recommendations on real time to the users of the system would increase the usage of bike-sharing systems and enrich the experience for the user.

References

1. AWS| Gestión de identidades y autenticación de usuario en la nube (Online). Available: https://aws.amazon.com/es/cognito/. Last accessed: 2019/09/28
2. Grünberg, K., Schenck, W.: A case study on benchmarking IoT cloud services. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10967 LNCS, pp. 398–406 (2018)
3. Xu, Y., Helal, A.: Scalable cloud-sensor architecture for the Internet of Things. IEEE Internet Things J. **3**(3), 285–298 (2016)
4. Audikana, A., Ravalet, E., Baranger, V., Kaufmann, V.: Implementing bikesharing systems in small cities: evidence from the Swiss experience. Transp. Policy **55**, 18–28 (2017)
5. Peng, R., Zhang, M., Guo, C., Cui, J., Song, J.: Public bicycle operating system based on space-time security and the Internet of Things. Wuhan Univ. J. Nat. Sci. **23**(6), 541–548 (2018)
6. Corno, F., Montanaro, T., Migliore, C., Castrogiovanni, P.: SmartBike: an IoT crowd sensing platform for monitoring city air pollution. Int. J. Electr. Comput. Eng. **7**(6), 3602–3612 (2017)
7. Bellavista, P., Zanni, A.: Towards better scalability for IoT-cloud interactions via combined exploitation of MQTT and CoAP. In: 2016 IEEE 2nd International Forum on Research and Technologies for Society and Industry Leveraging a Better Tomorrow (2016)
8. Gusev, M., Dustdar, S.: Going back to the roots-the evolution of edge computing, an IoT perspective. IEEE Internet Comput. **22**(2), 5–15 (2018)
9. Hou, L., et al.: Internet of Things cloud: architecture and implementation. IEEE Commun. Mag. **54**(12), 32–39 (2016)
10. Prabhu Kumar, P.C., Geetha, G.: Web-cloud architecture levels and optimized MQTT and COAP protocol suites for web of things. Concurr. Comput. 1–8 (2018)
11. Zguira, Y., Rivano, H., Meddeb, A.: Internet of bikes: a DTN protocol with data aggregation for urban data collection. Sensors (Switzerland) **18**(9), 1–39 (2018)
12. Gu, T., Kim, I., Currie, G.: To be or not to be dockless: empirical analysis of dockless bikeshare development in China. Transp. Res. Part A Policy Pract. **119**, 122–147 (2019)

13. Qiu, L.Y., He, L.Y.: Bike sharing and the economy, the environment, and health-related externalities. Sustain **10**(4) (2018)
14. Lu, M., Hsu, S.C., Chen, P.C., Lee, W.Y.: Improving the sustainability of integrated transportation system with bike-sharing: a spatial agent-based approach. Sustain. Cities Soc. **41**, 44–51 (2018)
15. Cerutti, P.S., Martins, R.D., Macke, J., Sarate, J.A.R.: 'Green, but not as green as that': an analysis of a Brazilian bike-sharing system. J. Clean. Prod. **217**, 185–193 (2019)
16. Tsai, M.F., Chen, P., Hong, Y.J.: Enhancing the utilization of public bike sharing systems using return anxiety information. Futur. Gener. Comput. Syst. **92**, 961–971 (2019)
17. Zhao, Y., Su, Y.S., Chang, Y.: A real-time bicycle record system of ground conditions based on Internet of Things. IEEE Access **5**, 17525–17533 (2017)
18. AWS: Serverless Architectures with AWS Lambda Overview and Best Practices (2017)
19. AWS: Serverless Applications Lens AWS Well-Architected Framework (2018)

A Geometric Design of Traffic-Light Roads Crossing Using the Continuous Flow Intersections Methodology to Reduce Points of Primary Conflicts Caused by Left Turns

Betsi Chuco, Carlos Pérez, Manuel Silvera, and Fernando Campos

Abstract The continuous flow intersections (CFI) increase the operational capacity of road systems with congestion problems, by using shared lanes located at the access points before the intersection. However, the CFI presents road safety risks that increase the likelihood of vehicle collision. This article proposes a geometric design composed of raised delineators at a traffic-light intersection, whose objective is to eliminate the points of primary conflicts caused by the left turns of the vehicles. To do this, a study was conducted to identify the different maneuvers present in a wide cross-type intersection in a commercial area located in the city of Lima. A total of 3219 vehicles was collected, of which 561 vehicles turned left demonstrating the high density of this type of maneuver. The effectiveness of the proposed design is validated using a microsimulation in the Vissim program. The results show that it was reduced from 58 to 8 points of conflict, increasing operational capacity by 34.97%. Finally, the risks caused by CFI decreased by 83%.

Keywords Continuous flow intersection · Conflict points · Operational capacity · Traffic-light roads

---

B. Chuco (✉) · C. Pérez · M. Silvera · F. Campos
Universidad Peruana de Ciencias Aplicadas UPC, Av. Prolongación Primavera, 2390 Lima, Peru
e-mail: u201418532@upc.edu.pe

C. Pérez
e-mail: u201624447@upc.edu.pe

M. Silvera
e-mail: manuel.silvera@upc.edu.pe

F. Campos
e-mail: pccifcam@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_53

1 Introduction

In the present investigation, a wide cross-type intersection located in the city of Lima will be analyzed. The intersection has fifteen lanes distributed in its four accesses, generating a large number of conflict points that reduce its operational capacity. This is because drivers are free to perform different maneuvers such as left and right turns, permitted turns, prohibited turns, among others when finding in a wide space. One of the factors that affect the performance of intersections marked at the level is the presence of left-turning maneuvers [1]. Given this problem, the design of CFI is effective in increasing operational capacity with a high level of application flexibility, especially in conditions of traffic generated by left-turning maneuvers from drivers [2]. Consequently, this research aims to apply the CFI design with elevated eyeliners in order to reduce conflict points, improving the operational capacity and safety of the intersection. These delineators will allow separating the route of the vehicles in the direction of the mixed lanes of the CFI.

The structure of this research is presented below. In section two, the extension of the foundation of the CFI is discussed. The methodology of the present investigation will be developed in section three. In section four, the intersection proposal is validated based on the application of the CFI design with high delineators, through microsimulation in the Vissim software. This software allows the calibration and validation of microsimulation models [3]. Finally, the research findings are presented.

2 State of Art

Different investigations have implemented unconventional intersections in order to increase the capacity of traffic-light intersections, for example, a study carried out in Shanghai indicates that CFI is an unconventional design, which allows increasing the capacity of intersections with a high level of flexibility of application, especially in conditions where the traffic presents left-turning maneuvers. The results of this investigation indicate that the application of the CFI in different cities of the world presents an improvement in the operational capacity that goes between 48.7% and 79.5%. However, operational capacity will vary according to geometric configurations and traffic demand patterns [4]. On the other hand, a study conducted in the USA indicates that the CFI provides a unique solution, which combines the geometry and operational behavior of the intersection [5]. Another study carried out in Shanghai indicates that the geometric design of the CFI lies in the use of mixed lanes. Prior to entering the intersection, there is a secondary traffic light located at the beginning of a medium opening. On the other hand, the authors point out that the CFI affects the road safety of intersections since they generate certain risks such as red light violations in secondary traffic lights by 1.83%, violations when taking the wrong lanes to make the left turn at 11.07%, and reduced speeds by 18.75% of vehicles during the route through the mixed lanes [6]. Based on these investigations, the present study aims to develop a redesign proposal based on the geometric design of the CFI to improve the real situation of the intersection studied.

3 Methodology

This section describes the procedure developed for the base design of the wide cross-type intersection located in a commercial area of the city of Lima. Then, we propose a new intersection design with elevated eyeliners driven by the CFI principle. All these will be done following the guidelines of the Federal Highway Administration (FHWA) of the USA [3].

3.1 Collection of Field Data

For the collection of the field information, a Drone Mavic Pro with a video camera was used. With this equipment, it was possible to identify the conflict points generated at the intersection. To quantify the number of vehicles for each access, manual counters were used, taking into account the number of lanes in a 15-min interval. The field data collection was carried out on Wednesday, August 21, 2019, for which a period of time from 5:00 pm to 6:00 pm because, at that time, the intersection has the highest vehicular flow.

3.2 Analysis and Representation of Field Data

For the analysis of the field data, it was necessary to classify the vehicle composition into three groups in a period of one hour. Then, the vehicles were counted for each phase of the traffic-light cycle, taking into account the maneuvers of the left and right turns, permitted turns, prohibited turns, among others. Of the three phases analyzed, phase 1 has the highest vehicle flow, comprising 2242 vehicles. On the other hand, the left turns of phase 1 represent 10.16% of the total vehicles.

Next, Table 1 shows the results of the vehicular flow for each of the phases with their respective maneuvers. Also, Fig. 1 illustrates phase 1 of the traffic-light cycle of the intersection.

3.3 *Base Development Model*

To start the construction of the microsimulation model first, an image was entered into the Vissim software that shows the geometry of the intersection. Then, the geometry data and the physical characteristics of the study area were inserted, such as the number of lanes, lane widths, free flow velocities, curvatures. In addition, the proportions of the types of vehicles in the entire network and the volumes obtained in the field [7] were entered.

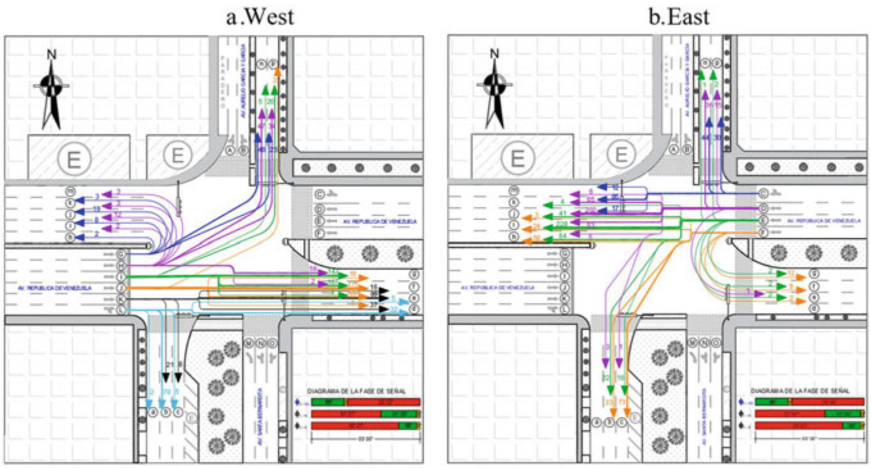


Fig. 1 Vehicular flowchart of phase 1 of the intersection

Table 1 Vehicular flow results for each phase of the intersection

Phase	Access	Movements				Vehicular flow
		Allowed			Not allowed	
		Right	Left	Head-on	U-turns	
Phase 1	West/east	234	327	1599	82	2242
Phase 2	South	41	128	226	0	395
Phase 3	North	229	106	247	0	582
Total		504	561	2072	82	3219

Table 2 Table of calibration results

Access west				
Data	Sample size (und)	Average (seg)	Two-tailed test with 95% reliability	Mean difference (seg)
Countryside	50	6.11	⟨ − 0.844 0.811⟩	0.43
Program Vissim	15	5.66		
Access south				
Data	Sample size (und)	Average (seg)	Two-tailed test with 95% reliability	Mean difference (seg)
Countryside	16	6.68	⟨ − 0.658 0.568⟩	−0.25
Program Vissim	15	6.94		

4 Calibration and Model Validation

4.1 Model Calibration

In any process of calibration of a model, an adjustment as close as possible between it and reality [8] is intended. First, the travel time of the vehicles was chosen as the efficiency parameter. A simulation was recorded with maximum speed to obtain the value of N (run number) where 15 runs were used assuming 95% reliability. The warm-up period chosen was 600 and 3600 s was used as the total microsimulation time. In addition, the three parameters of the Wiedemann 74 model were used, which changed depending on the results of the calibration [9].

To verify that the model is calibrated, statistical evaluation was carried out by means of a null hypothesis test of equality of means with a 95% confidence level. For this analysis, the vehicle travel times obtained in the field were compared with the results of the Vissim program. Table 2 shows the results of the studied lanes, both for the west and east access, respectively. Finally, it is concluded that the four lanes comply with the null hypothesis test of equality of means. Therefore, the calibration of the program with the Wiedemann parameters ax = 1, bx_add = 0.5 and bx_mult = 0.3 is demonstrated.

4.2 Model Validation

Finally, as the last step to analyze the simulation results, the model was validated with other input data. The registration of these data was carried out on Wednesday, September 11, 2019, for the period between 5:00 pm and 6:00 pm. Table 3 shows the results of the validation where the statistical test of mean differences was used with a 95% reliability using the vehicle travel time as an efficiency parameter. With these results, the model is validated using values of the Wiedemann parameters [9], ax = 1, bx_add = 0.5 and bx_mult = 0.3.

5 Proposed Method

After validating the microsimulation model, different design alternatives are developed in order to improve the real situation of the intersection studied. The next procedure is the coding of the improvement alternatives in a new microsimulation model, respecting the parameters of the validation model. For this new model shown in Fig. 2, a change in geometry was made based on the CFI. The first change that was made was the reduction of lane numbers in the main accesses, due to the imbalance between the supply and demand of vehicles. After proposing the necessary lanes in each access, the geometric design of the CFI is implemented. A length of 50 meters was considered for the four mixed lanes located at the west and east accesses of the intersection. The assigned distance will allow a minimum tour of the vehicles along the mixed lane, which will reduce the risks of the CFI. However, these risks are reduced by placing elevated eyeliners in the study area and using the correct signage.

Table 3 Validation results table

Access west				
Data	Sample size (und)	Average (seg)	Two-tailed test with 95% reliability	Mean difference (seg)
Countryside	50	5.53	⟨ − 0.916 0.908⟩	−0.13
Program Vissim	15	5.66		
Access south				
Data	Sample size (und)	Average (seg)	Two-tailed test with 95% reliability	Mean difference (seg)
Countryside	16	7.04	⟨ − 0.921 0.897⟩	0.04
Program Vissim	15	7		

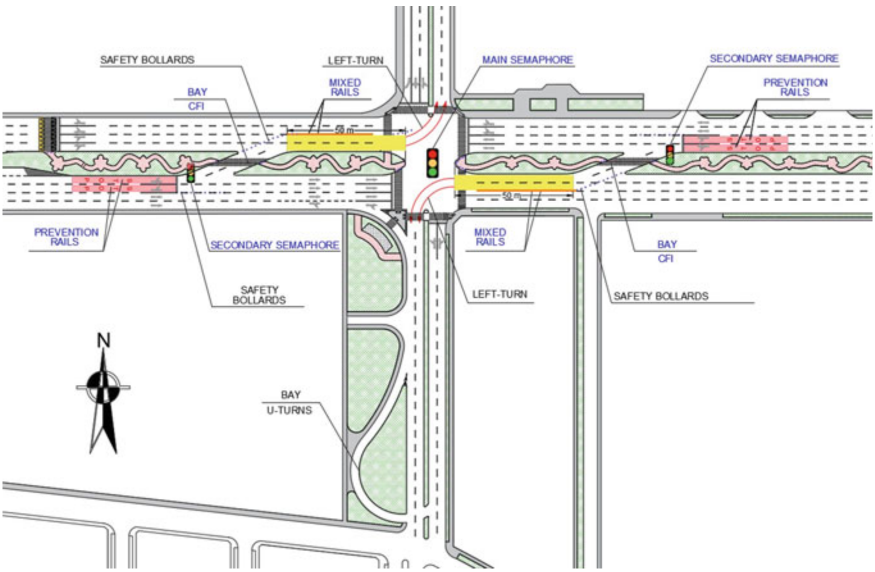


Fig. 2 Proposed design based on the geometric design of the CFI

6 Validation and Results

6.1 Number of Conflict Points

Figure 3 shows that a large number of conflict points occur for phase 1 at the intersection of analysis since drivers can perform different maneuvers by placing themselves at a wide intersection. The number of points of conflict type crossing at the intersection analyzed is 58 points in total, which only corresponds to the left-hand vehicular movement.

The application of the geometric design of the CFI makes it possible to reduce the points of cross-type conflict at the intersection that are generated through the left-turning movements. When applying the CFI at the intersection, phase 1 presents only 2 points of cross-type conflict. Consequently, with the new geometric design of the intersection, only eight points of conflict are generated in total because there are only four lanes in both accesses that allow users to make the left-turning movements. Figure 4 illustrates the number of points of conflict of each access of the intersection with the proposed design based on the CFI.

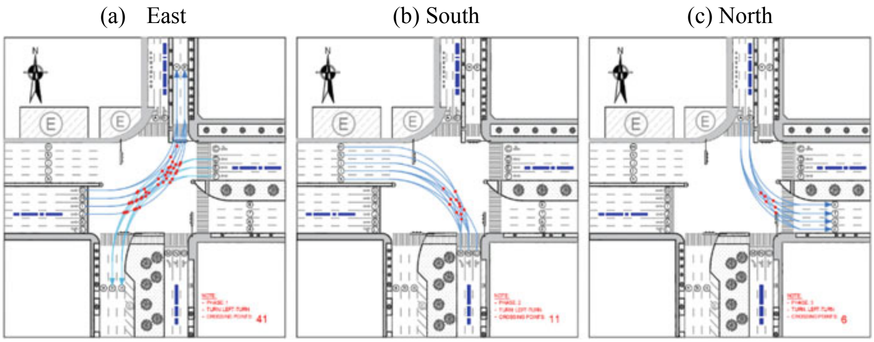


Fig. 3 Number of points of conflict type crossing for each access of the intersection studied

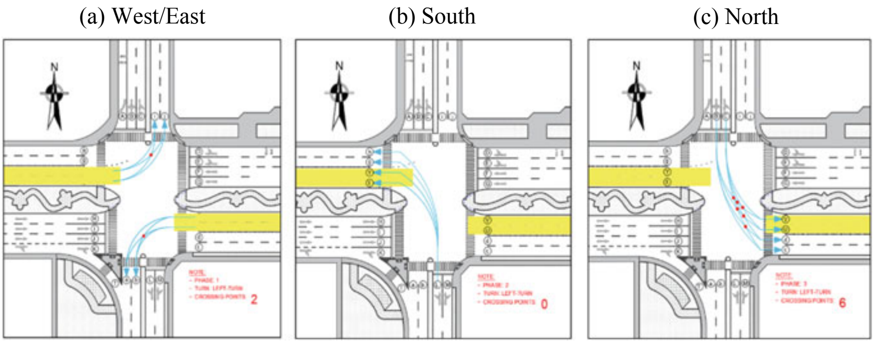


Fig. 4 Number of cross-type conflict points in the new geometric design based on the CFI

Table 4 Analysis of operational capacity

	The operational capacity of the intersection (Veh)
Real intersection	790
Intersection with new CFI geometric design composed with raised eyeliners	1215

6.2 Determination of Operational Capacity

The operational capacity of an intersection is the maximum number of vehicles that flow through a uniform point or section of a lane over a minimum interval of 15 min [10]. For the present study, the operational capacity was obtained under the conditions of the study area, vehicular traffic, and control signals. When making the new geometric design based on the CFI composed of high eyeliners, the operational capacity of the intersection increased by 34.97%. Table 4 shows the results of the operational capacity based on the number of vehicles registered in an interval of 15 min for each case.

6.3 Evaluation of Safety Results

The study area analyzed for being a wide cross-type intersection presents a large number of conflict points, caused by left turns. Consequently, these maneuvers threaten the road safety of the users, thus producing different safety risks. In the real intersection, 19 risks of frontal crashes and 27 risks of rear and lateral crashes were recorded. Figure 5a illustrates the types of risks at the current intersection. Regarding the design of the CFI, the main concern is the mixed lane, as users may not realize the new geometric design and the signals before reaching the intersection, resulting in two important risks that threaten the road safety and this would increase the likelihood of vehicle collision. These risks are frontal collisions and the risks of posterior and lateral collisions. From a total of 77 vehicles in the west access, 10 risks of a frontal collision and 8 risks of posterior and lateral collisions were recorded. Regarding the east access, of a total of 96 vehicles, 9 risks of a frontal collision and 7 risks of posterior and lateral collisions were recorded. Figure 5b illustrates the types of risks at the intersection with the new geometric design of the CFI. However, the risks of posterior and lateral collisions are reduced by the use of elevated eyeliners by 83%. This is illustrated in Fig. 5c.

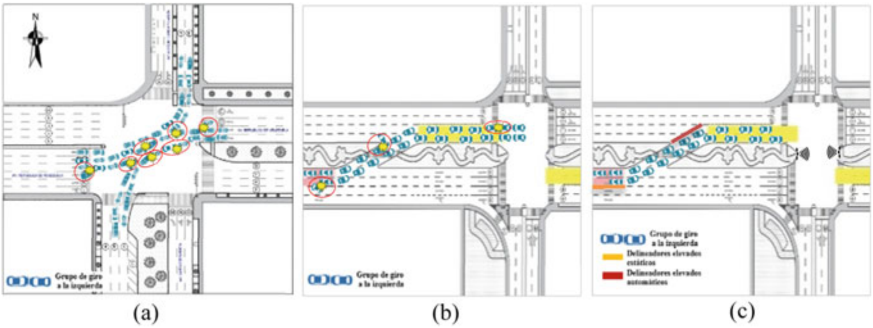


Fig. 5 Types of collisions at the intersection analyzed

7 Conclusions

In summary, the new geometric design based on the CFI improves the actual situation of the real intersection, since the number of cross-type conflict points caused by left vehicular movements was reduced, eliminating 50 points of cross-type conflicts of a total of 58 conflict points presented by the current intersection. Consequently, operating capacity increased by 34.97%, since the supply of vehicles increases by 425 vehicles. On the other hand, the real intersection presents 19 risks of frontal crashes and 27 risks of lateral and rear crashes. With respect to the CFI-based microsimulation model, 10 risks of a frontal collision and 8 risks of posterior and lateral collisions in the west access and 9 risks of a frontal collision and 7 risks of collisions in the east and lateral access were recorded. However, these risks are reduced by 83% with the implementation of elevated eyeliners. Finally, it is recommended to make an adjustment in the synchronization of the secondary traffic lights with the main traffic light to optimize the new geometric design of the intersection and ensure even more safety. In addition, it is proposed to install sensors in the central area of the intersection, so that they notice the total free space of the mixed lane after the signal change before the intersection and to sanction drivers who make unauthorized movements.

References

1. Tabernero, V., Sayed, T.: Upstream signalized crossover intersection: an unconventional intersection scheme. J. Transp. Eng. **132**(11), 907–911 (2006). https://doi.org/10.1061/(ASCE)0733-947X(2006)132:11(907)
2. Zhao, J., Ma, W., Zhang, H.M., Yang, X.: Increasing the capacity of signalized intersections with dynamic use of exit lanes for left-turn traffic. Transp. Res. Rec. J. Transp. Res. Board **2355**, 49–59 (2013)
3. Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software. Publication FHWA-HRT-04-040, U.S. Department of Transportation, 2010

4. Zhao, J., Ma, W., Head, L.: Optimal operation of displaced left-turn intersections: a lanebased approach. Transp. Res. Part C: Emerging Technol. **61**, 29–48 (2015). https://doi.org/10.1016/j.trc.2015.10.012
5. Carroll, D.H., Lahusen, D.: Operational effects of continuous flow intersection geometrics: a deterministic model. Transp. Res. Rec. **2348**(1), 1–11 (2013). https://doi.org/10.3141/2348-01
6. Zhao, J., Liu, Y.: Safety evaluation of intersections with dynamic use of exit-lanes for left-turn using field data. Accident Anal. Prevention **102**, 31–40 (2017). https://doi.org/10.1016/j.aap.2017.02.023
7. Fellendorf, M., Vortisch, P.: Validation of the Microscopic Traffic Flow Model VISSIM in Different Real-World Situations (2001)
8. Hollander, Y., Liu, R.: The principles of calibrating traffic microsimulation models. Transportation **35**(3), 347–362 (2008)
9. Wiedemann, R.: Simulation des Straßenverkehrsflusses, p. 8. Schriftenreihe des InstitutsfürVerkehrswesen der Universität Karlsruhe, Heft (1974)
10. HCM 2010: Highway Capacity Manual. Transportation Research Board, Washington, DC

Presence Sensors: A Comparison Between the Ultrasonic and the Infrared in the Detection of People and Vehicles

Rosana C. B. Rego and Rodrigo S. Semente

Abstract Sensors have become fundamental elements in automation and industrial processes, for updating in a precise and fast way. Among the various types of sensors, there are presence sensors. They are able to perform simple detection tasks more accurately and efficiently than people. This work presented the comparison between two types of sensors: the ultrasonic and the infrared. It is also proposed a method of estimating the length of the vehicle by presence detection. The preliminary results show that the technique is interesting and cheap to distinguish and detect vehicles types.

Keywords Presence sensors · Infrared · Ultrasonic · Vehicles

1 Introduction

The use of sensors is essential in the modern world. Whether it is to control industrial processes, monitor climatic and environmental conditions or simply facilitate everyday life procedures, we can find them in a variety of situations.

The purpose of a sensor is to respond to a stimulus and convert it into an electrical signal compatible with the circuits attached to it. The output signal of the sensor may be in the form of voltage, current, or charge and may be described in terms of amplitude, frequency, phase, or digital code [2, 5].

We can divide the sensors basically into two groups: analog sensors and digital sensors. The analog has as answer an analog signal, that is, assume numerous values. Unlike the digital ones that have a digital signal as its answer, it assumes the values

R. C. B. Rego (✉)
Federal Rural University of the Semi-Arid, Rio Grande do Norte, Brazil
e-mail: rosana.rego@ufersa.edu.br

R. S. Semente
Department of Engineering and Technology, Federal Rural University of the Semi-Arid,
Rio Grande do Norte, Brazil

© Springer Nature Switzerland AG 2021
Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_54

0 or 1 [6]. In this work, two types of presence sensors are presented. One has an analog signal on the output, and the other has a digital signal on the output.

An important group of sensors is the proximity sensors or presence sensors. They are used to detect the presence of people or any objects in a monitored area. The presence sensor produces a signal if an object is in its detection zone, whether this object is stationary or not [1, 2, 5]. There are several types of presence sensors, such as ultrasonic, infrared, optics, inductive (used to detect electromagnetic field) [7].

Presence sensors are commonly used in industrial processes, and recently an optimal application of presence sensors is at smart traffic lights. The sensors are typically installed at a mid-block location, which provides accurate measures of vehicle volumes and speeds. The information provided by system sensors can be used to support the following system functions: acquiring traffic flow information to compute signal timing, identifying critical intersection control (CIC), and selecting timing plans [3, 7].

Traffic managers all over the world use a camera or some presence sensor to monitor and manage traffic streams. Be it for monitoring motorists and pedestrians in urban areas, for detecting incidents on highways and in tunnels, or for traffic data collection purposes. However, the video detection by camera requires more data processing, consequently a higher cost. Thinking about this, this work carried out an analysis of two presence senors and how they behave in the detection of people or vehicles.

The aim of this work is to present the principle of operation of two presence sensors, as well as their characteristics, and how they behave in the detection of vehicles and people. It was made a comparison between an ultrasonic sensor and an infrared sensor. And a method for detecting the length of the vehicle is presented. This work is still an initial work in which tests are still being carried out to validate or not the technique.

2 Sensors

2.1 Ultrasonic

The ultrasonic sensor is a distance sensor. It emits sound pulse (transmitted wave) and calculates the time it takes for the pulse to be reflected back to the sensor. Fig. 1 shows the operation of an ultrasonic sensor.

A minimum distance from the sensor is required to promote a time delay, so that the echoes can be interpreted. In the case of the HCSR04 sensor model, which was used in this work, the minimum distance is 2 cm, and the maximum distance is 4 m. This ultrasonic is necessary only to supply a short 10 uS pulse to the trigger input to start the ranging, and then, the module will send out an 8 cycle burst of ultrasound at 40 kHz and raise its echo. An echo is a distance object that is pulse width and the range in proportion.

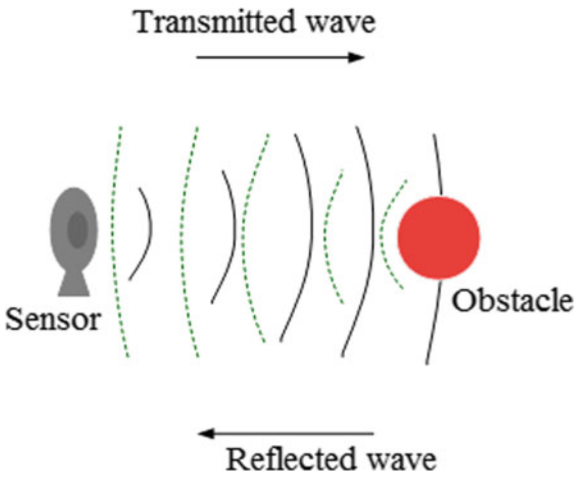


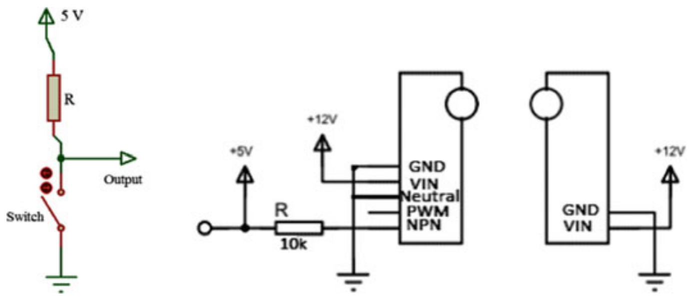
Fig. 1 Ultrasonic operation

Ultrasonic sensors are quite fast for most of the common applications. For example, it can be used for distance measurement of an object in the path of a person, equipment, or a vehicle. These sensors can also be used to detect the driving violations of slower vehicles, and especially large heavy trucks, traveling in an improper lane like the "passing lane," might cause serious negative effects on the highway traffic order, reduce highway traffic efficiency, and become a safety threat for other drivers who have to change lanes more frequently [4].

2.2 Infrared

Infrared sensors can be active or passive. The sensor used here is an active. It has two devices, one that emits (emitter) and another that receives (receiver) the infrared rays. The emitter and receiver form an infrared beam, and when this beam is interrupted, the sensor sends the detection signal to the controller. The sensors consist internally of a switch (transistor) in which they alternate between two levels: high or low. The high state is a voltage of 5 V, and the low state is a voltage of 0 V. This switch can be connected in two ways: with a pull-up resistor or pull-down resistor (Fig. 2a).

The switching of the sensor output is performed by a transistor. When no detection occurs, the sensor is not activated. The infrared output is an NPN output, where the load coupled to the sensor is connected between the positive terminal and the sensor output. When the sensor detects nothing, it remains at a high level, when it detects an object it goes to the low level. Fig. 2b shows the infrared sensor used. These sensors can be used in security applications such as alarms, perimeter protection, automatic lighting, garage doors, and others [6]. Here, it was used to detect vehicles or people.



(a) Switch connected to a resistor pull-up.

(b) The infrared sensor used.

Fig. 2 Infrared sensor

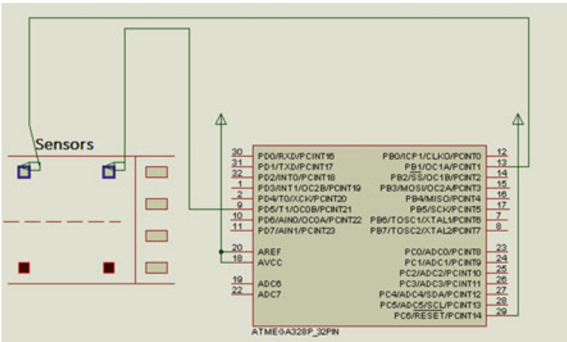


Fig. 3 Circuit diagram

3 Experiment

To get the data, an infrared barrier was created with the infrared receiver and the infrared emitter and the emitter and receiver of the ultrasonic remained in the same encapsulation. It was considered a distance of 300 cm between the infrared pairs. The receivers were at a distance of 400 cm from the emitters.

Figure 3 shows the scheme of the circuit with the microcontroller used. The controller used was an Atmega328. The system would cost approximately $70.00.

An analysis of the effect of one emitter on the other sensor was performed, varying the distance of the two, until the minimum distance to avoid interference between the two was found. The minimum distance found was 50 cm.

To detect the presence of something using the ultrasonic, a maximum distance (350 cm) was defined. If the sensor sends to the controller a distance smaller than the maximum defined, then it detected something.

The measured distance is calculated on the basis of travel time. As shown in Eq. 1.

$$\text{Distance} = \frac{(\text{travel time} \times \text{velocity of sound})}{2} \tag{1}$$

The vehicle speed can bee calculated by Eq. 2.

$$\text{speed} = \frac{\text{SDist}}{T_{i_2} - T_{i_1}}. \tag{2}$$

where SDist is the distance between the two sensors and T_{i_1} and T_{i_2} are, respectively, the initial time of detection of two sensors (Fig. 9). The length of the vehicle can be calculated by Eq. 3.

$$\text{Length} = \text{speed} \cdot \frac{(T_{f_1} - T_{i_1}) + (T_{f_2} - T_{i_2})}{2}. \tag{3}$$

where T_{f_1} and T_{f_2} are the final detection time sensors 1 and 2, respectively. In other words, Eq. 3 is the speed multiplied by the average between the detection time of sensor 1 and sensor 2. However, its measure is not precise, because there are measures imprecision at the detection time of the sensors and the vehicle shape. So, it is proposed that the vehicle signature value (VSV) can be estimated by the integrals of the received signals, S_1 and S_2, by the time, determined by the following Eq. 4,

$$\text{VSV} = \left(\int_{T_{f_1}}^{T_{i_1}} S_1 \mathrm{d}t + \int_{T_{f_2}}^{T_{i_2}} S_2 \mathrm{d}t \right) /2 \tag{4}$$

Two tests were performed in the presence of cars. The first test, a 4 m car (let us reference by car 1) passed through the sensors at speeds of 10, 20, 30 km/h. In the second test, with another car (let us reference by car 2) of 4.232 m passed the sensors at speeds of 14.4, 27 and 39 km/h. The obtained values and results are presented at Table 1. The data were obtained at a sampling rate of 0.2 s.

Table 1 Data obtained with the ultrasonic sensor response for the car 1 detection

	Detection 1	Detection 2	Detection 3
Speed (km/h)	10	20	30
Length (m)	3.34	3.34	4.16
VSV	8.78	4.36	6.58

4 Results and Analysis

In Fig. 4, it is possible to observe the behavior of the ultrasonic sensor in the detection of presence, in case of detection of people. And in Fig. 5, it is possible to observe the transient response of the infrared sensor (when detecting the person).

The ultrasonic sensor has a response rate slightly higher than the infrared sensor; however, the infrared is more stable than the ultrasonic. The model of the ultrasonic sensor used here has an accuracy of 2 cm. Because the ultrasonic output signal is an analog signal, it is more sensitive to noise as shown in Fig. 4. To improve the output signal, it is advisable to use a filter.

Fig. 6 shows the response of the infrared sensor in the presence of car 1. And Fig. 7 shows the response of the ultrasonic sensor for the same case. The car passed three times through the sensors at different speeds as you can see in Fig. 7. The response of both sensors was similar; however, the infrared sensor did not detect the car when it passed at a speed greater than 20 km/h. As can see in Fig. 6 in the time 112 s, it did not detect anything.

Comparing the two responses (infrared and ultrasonic), the use of infrared is suitable for presence detection of moving objects with low speed only. Since the ultrasonic, it detects the presence and still provides the distance, despite the oscillations in the response. Therefore, to detect only people, the infrared is recommended. However, if you want to detect the distance of the object or a very fast object, then the ultrasonic is recommended.

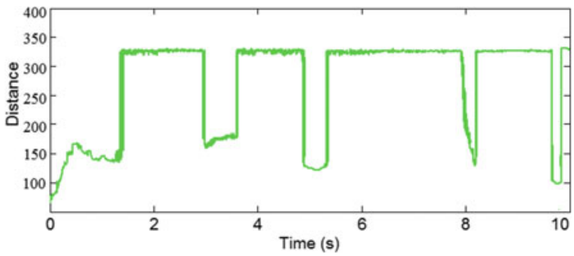


Fig. 4 Ultrasonic sensor curve

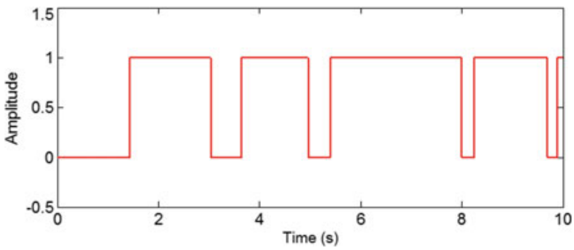


Fig. 5 Infrared sensor curve

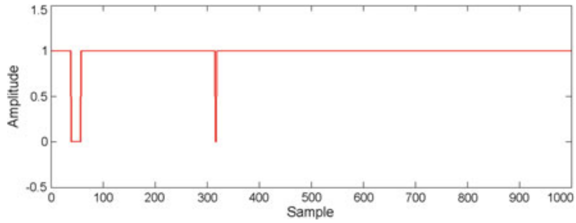


Fig. 6 Infrared sensor response for the car 1 detection

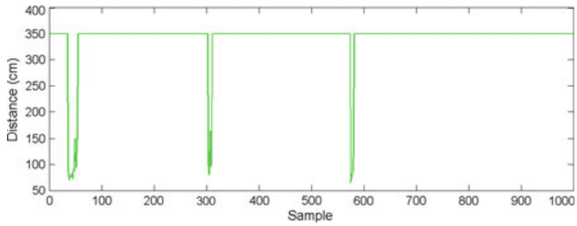


Fig. 7 Ultrasonic sensor response for the car 1 detection

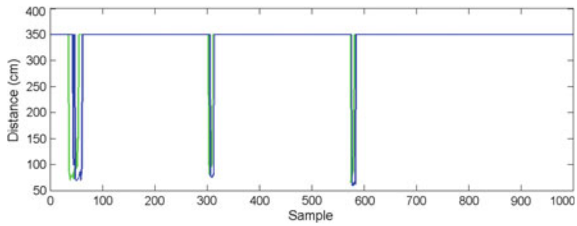


Fig. 8 Data obtained with the ultrasonics sensors for the car 1 detection

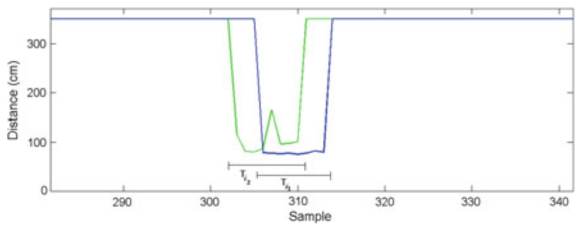


Fig. 9 Data obtained with the ultrasonic sensors in car 1 detection

The infrared sensor behaved properly when the vehicle passed at low speeds as shown in Fig. 6 on the first detection when the signal change from 1 to 0. Therefore, the analysis of the results obtained will be performed with the data coming from the ultrasonics sensors (Fig. 8).

Figure 8 shows the response of the sensors in the detection of the car 1.

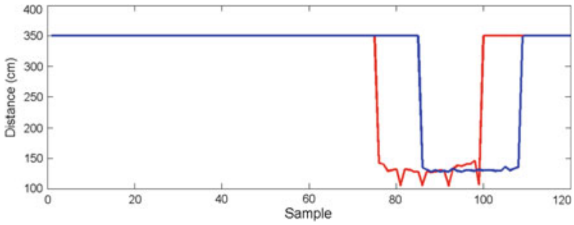


Fig. 10 Ultrasonic sensor response for the car 2 detection with a speed of 14.4 km/h

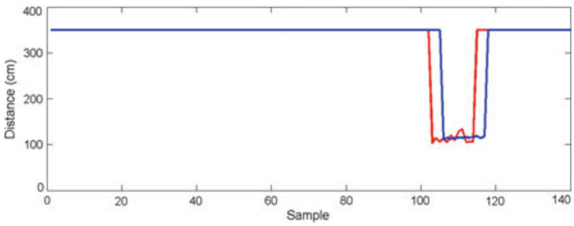


Fig. 11 Ultrasonic sensor response for the car 2 detection with a speed of 27 km/h

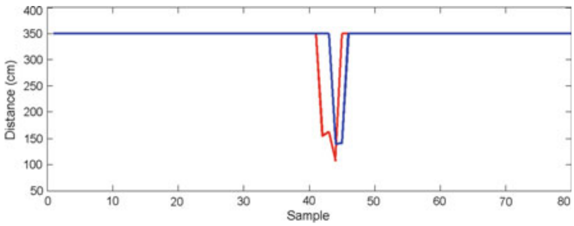


Fig. 12 Ultrasonic sensor response for the car 2 detection with a speed of 39 km/h

Table 2 Data obtained with the ultrasonic sensor response for the car 2 detection

	Detection 1	Detection 2	Detection 3
Speed (km/h)	14.4	27	39
Length (m)	4.9	4.875	4.873
VSV	6.41	7.03	8.34

Figures 10, 11, and 12 show the response of the ultrasonic sensors in the detection of the car 2 with different speeds. From the data obtained (Fig. 8), consider the following speeds 10, 20, and 30 km/h. The calculations of the estimated length of the vehicle are given in Table 1.

According to Table 1, the estimated average length was 3.5 m. The original length of the car 1 is 4 m. So there is an error of 12.50%.

According to Table 2, the estimated average length was 4.88 m. The original length of the car 2 is 4.23 m. So there is an error of 15.31%. It is necessary to calibrate the sensors beforehand, but these results suggest the feasibility of the technique used.

The VSV for Table 1 has a variation coefficient of 33.55%. And for Table 2, the variation coefficient is 13.85%

The VSV should be compared with others vehicles VSV to calibrate and used to identify the vehicle. It was observed that the energy of the system suffered little variation with the speed and could be used for the detection of the vehicle type. More testing with other vehicle types will be needed to validate the technique.

5 Conclusion

To choose the sensor to be used, you need to know what you want to detect. If it is only the presence of an object, the infrared is more appropriate. However, if it is necessary, know the distance that this object is, then the ultrasonic is more appropriate. Also, it is necessary to take into consideration the environmental factors, this is, the conditions of the environment in which the sensor will be installed are of extreme importance for your choice. For example, if it is an environment where it has a lot of light, the infrared sensor is no a good choice.

How it was seen, the infrared is more precise, but, it is necessary voltage power source (12 V), and it was not sensitive to the detection of cars.

As discussed previously to validate the technique of estimating the length of the vehicle, it is necessary to carry out more measurements with different types of vehicles with different sizes.

References

1. Diraco, G., Leone, A., Siciliano, P.: People occupancy detection and profiling with 3D depth sensors for building energy management. Energy Build. **92**, (2015)
2. Gunapala, S., et al.: High dynamic range infrared sensors for remote sensing applications. In: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 6342–6345 (2018)
3. Koonce, P., Rodegerdts, L., Lee, K., Quayle, S., Beaird, S., Braud, C., Bonneson, J.,Tarnoff, P., Urbanik, T.: Traffic signal timing manual (2008)
4. Liu, J., Han, J., Li, B., Lv, H.: An ultrasonic sensor system based on a two-dimensional state method for highway vehicle violation detection applications. Sensors **15**, 9000–9021 (2015)
5. Rego, R.C.B., Soares, R.S.: Presence sensor for people detection and reductionof pedestrian waiting time in traffic light. In: Brazilian Technology Symposium (2018)
6. Thomazini, D., Albuquerque, P.U.B.: Sensores industriais: fundamentos e aplicações, São Paulo, 8th ed. (2011)
7. Zhang, C., et al.: Coexistence-insensitive presence detection. U.S. Patent No. 10,305,766 (2019)

An Operational Approach for Enhancement Toll Collection in Trucks with Axis Suspended

Michell Miranda Borges Costa, Yuzo Iano, Gisele Maria Figueiredo Caruso, Diego Pajuelo, and Gabriel Gomes de Oliveira

Abstract With the emergence of highway concessions in Brazil, toll plazas were implemented. The collection rate is based on the number of axis truck which is in contact with the ground at the time of toll payment. The problem is when truckers decide to lift some loaded truck axis in order to pay a lower fare. The goal is to establish a Weigh-in-Motion (WIM) system for informing toll plaza whether trucks are normal loading conditions or not. If not, then truck would be charged according to the rule.

Keywords Roads · Moving weighing trucks · Shipping · Toll · Axis · and tariffs

1 Introduction

The highway concession program began in 1988 through Decree No. 94002, which authorized the National Department of Roads (DNER) to hire, through concession, construction, maintenance, and operation of highways. In 1993, through Ordinance No. 010, the Ministry of Transport creates a working group to implement the Program of Federal Highway Concessions (PROCROFE). Already in 1994, DNER was signed the first concession agreement with the Municipality of Rio de Janeiro and Government of Santa Catarina, uniting various regions of Florianópolis. In 1995 was signed

M. M. B. Costa (✉) · Y. Iano · G. M. F. Caruso · D. Pajuelo · G. Gomes de Oliveira
University of Campinas, Campinas, SP, Brazil
e-mail: michellmogi@gmail.com

Y. Iano
e-mail: yuzo@decom.fee.unicamp.br

G. M. F. Caruso
e-mail: giselecaruso787@gmail.com

D. Pajuelo
e-mail: diego.pajuelo.castro@gmail.com

G. Gomes de Oliveira
e-mail: oliveiragomesgabriel@ieee.org

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_55

the concession agreements of President Dutra Highway, Rio-Teresopolis, and Juiz de Fora-Rio, authorizing the union to delegate to the united highways and allows its concession. In 1998, they were signed nine concession agreements relating to stage 1 of the Highway Concession Program in São Paulo state [1].

In this sense, the nine dealers that first batch of the state of São Paulo awards had their fare collection model defined by the concession notice where there would charge tolls of motorcycle type vehicles, and also the suspended axis of trucks, i.e., those vehicles were only charged the shafts that were in contact with the ground (soil) at the time of passing through the toll plaza.

Highway tolls are priced through direct charges based on the distance traveled. These charges generally depend on vehicle type, number and configuration of vehicle axis, trailer types, and vehicle dimensions [2].

In July 2013, the governor of São Paulo announced that there would be no adjustment of toll rates, this adjustment provided for in the concession contracts, which tracks the index of annual inflation measured by the IPC-A (Broad Consumer Price Index). To allow the maintenance of economic and financial equilibrium of the concession contracts, among other measures, ARTESP, São Paulo state transport agency adopts the steps to recover the so-called suspended axle trucks. In-state tracks under concession, commercial vehicles had a suspended axis without contact with the ground, did not pay the toll. From July 1, 2013, all axes of trucks are billed, as well as what has happened in all federal highways. This toll collection methodology, charging the suspended shaft was revoked in May 2018 [3, 4].

With the cancelation of this collection, new problems come, i.e., loaded trucks suspend the shaft before reaching the toll plaza to pay a lower value, then cross the toll plaza, lower the shaft to the normal state.

The aim of this study is to propose a model of toll collection, which facilitates the collection of loaded truck tariffs in order to avoid fraud in highway toll plazas.

2 Development

2.1 Vehicle Weighing Process on Brazil

Whereas the loaded trucks overweight control is a global stamp of concern, especially coming from the constant increase in road transport cargo, it is necessary to permanently search for methods that preserve the pavement, especially in the preservation of the lives of road users.

The current model weighing is carried out in the so-called PGF—General Audit Office. Figure 1 shows this configuration.

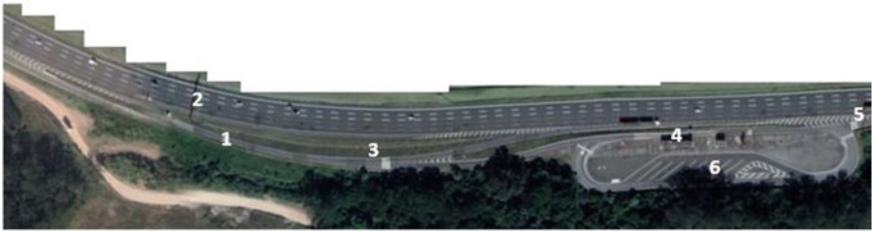


Fig. 1 PGF—General Inspection Post *Source* Google Maps, 2018

1. Selective Balance: Weighing the vehicles and classifies configuration with reference to the axis that is in contact with the ground, and the total gross weight adopted is the minimum category;
2. Control Evasion: Logs image of trucks does not step into for weighing in selective balance;
3. Control Evasion: Logs image of trucks flout the traffic lights and not step into for weighing in precision scale;
4. Precision Scale: Weigh the vehicle and classifies the configuration with reference to the axis in contact with the ground. A weighting agent, as determined by the manufacturer, limits the gross weight total adopted;
5. Control Evasion: Logs image of trucks flout the traffic lights and not step into the parking yard;
6. Parking Yard: Place where the trucks remain stranded waiting for weight or dimensions of regularization.

However, in this weighing model, besides the dependence of the presence of a public official to till the infraction notices also have other drawbacks such as weighing time, which prevents a massive weighing cargo vehicle.

Figure 2 demonstrates the low productivity according to the seven (07) PGF's in January 2019, where the enforcement weight average operated 21% of the month. The total (79%) of dead hours, 78% occurred due to lack of DER agent.

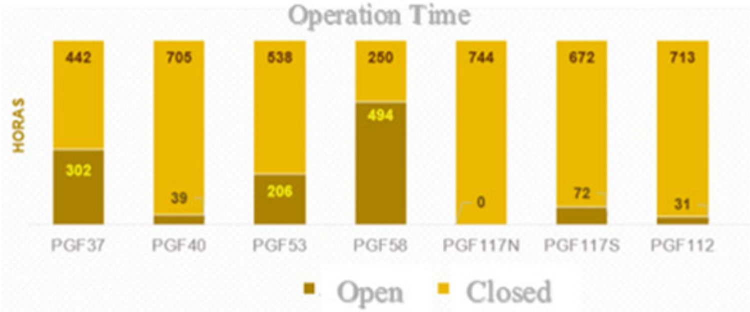


Fig. 2 Survey results of productivity of PGF—General Inspection Post

Table 1 Percentage of inspected trucks

Jan—2019							
Highway	SP- #1			SP- #2			
% Bus	11	9	6	7	7	3	3
PGF	#1	#2	#3	#4	#5	#6	#7
VDM Commercial	4916	4902	7917	9954	10,775	2102	2285
VDM Bus	521	461	435	657	711	61	64
VDM Truck	4395	4441	7482	9297	10,064	2041	2221
Total trucks no. SH	136.242	137.678	231.929	288.208	311.979	63.272	68.852
PGF selective inspected trucks	41.057	25.106	11.193	12.600	134.737	0	6.933
% Inspected trucks	30	18	5	4	43	0	10

Table 1 shows the percentage of trucks that were inspected on January 19, 2019, in which these PGF's were installed. This proves the low efficiency of the operating productivity of this surveillance system.

2.2 Billing Concepts Toll Rates

The collection of highway user's rate, which takes into account many items as the square location, type of highway, investment requirements, etc., corresponds to the number of axis of each vehicle, which is called category. The conventional car also called drive vehicle, receives the category 1 CAT1, and pays the corresponding toll rate to a single rate CAT 1. Category 2 or greater is designed for commercial vehicles and cargo and is determined according to the number of axis; considering the number of axis in contact with the ground and dual carriageways.

For a better understanding of this classification, Fig. 3 shows a truck with a single axis and dual carriageways, and Fig. 4 shows the time when the axis is suspended thanks to the pneumatic mechanism enabled on the truck.

2.3 Types of Commercial Vehicles

Considering the various manufacturers and types of loads, there is a wide range of varying types of commercial vehicles. For the same manufacturer, considering the same type of chassis and engine, it can still be a significant variation in the weight of the vehicle, considering only the type of body located therein.

According to the established by authorities, the cargo vehicles regard weight can be classified as shown in Fig. 5.

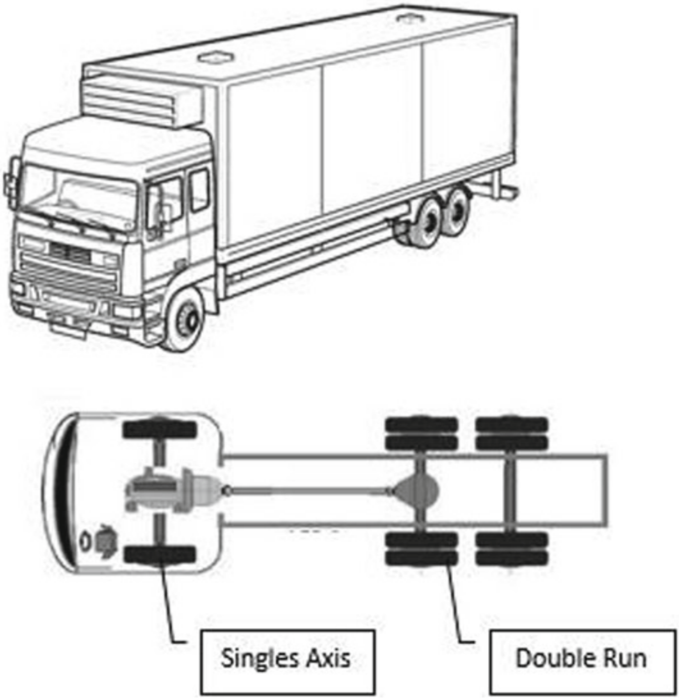


Fig. 3 Single-axis and dual carriageways

Fig. 4 Suspended axis

From the legal limit, 10% tolerance is allowed at the time of inspection and 12.5% tolerance regards the overflowing case. Also, the PBT/PBTC (Total Gross Weight) has a maximum of 5% tolerance.

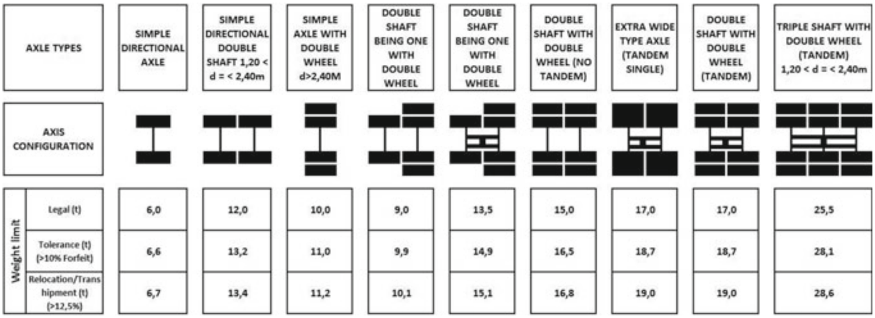


Fig. 5 Axle types

2.4 *Measures*

In 2019, we carried out measures about the weighing of empty trucks at the PGF stations. Table 2 shows the result of the weighing of these trucks, the percentage of variation between the minimum and maximum weight of the same truck-class (%) and the PBT/PBTC and all kinds of truck configuration.

With this survey, it was possible to demonstrate the wide variation in weight even when the truck is empty (off-loading), to the same class of vehicle.

3 Divergence of Same-Class Truck Weight Measures

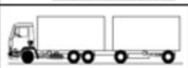
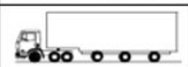
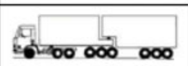
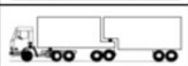
Currently, the weighing equipment does not identify the suspended axis of vehicles loaded or empty at the PGF point. The PGF has a truck weigh suspended axis on a selective scale and is found excessive; this truck is directed to the precision balance. If the excess occurs with the precision balance suspended axis, then the assessment is not made immediately; the driver is guided by the DER the lower the suspended axis and performs reweighing. If this does not occur excesses, then reweighing the vehicle is released [5]. Figure 6 demonstrates two identical, empty settings, but different PBT.

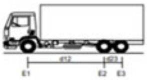
Through a data survey carried out at a fixed weighing point, where the weight of several empty trucks was found, it was possible to verify that, for the same type of truck, even when empty, the weight can be different, depending on the type of truck bodywork used. For instance, Fig. 7 presents the maximum weight for the truck double set in tandem with the legal limit of 17,000 kg in empty condition.

It is found that the maximum weight is 10,070 kg, that is, 59% of the legal limit weight, which is 17,000 kg.

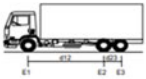
Table 2 Summary of measures regard the kind of trucks configuration

Weight Empty Truck

CAT	Class	Configuration	Maximum Weight Measured	Minimum Weight Measured	Var.%	Total Samples		Possible Axis Suspended	PBT/PBTC Regulated (ton)
5		2I1	16.090	10.940	47%	3	0,48%	2	43
4		2I2	18.250	16.050	14%	4	0,64%	1	36
5		2I3	24.700	16.220	52%	11	1,76%	2	45/46
4		2S2	20.060	12.620	59%	28	4,48%	1	33
5		2S3	27.710	9.830	182%	113	18,08%	2	41,5
3		3C	16.290	4.140	293%	227	36,32%	1	23
5		3C2	20.520	19.380	6%	2	0,32%	1	43
6		3C3	22.730	22.730	0%	1	0,16%	2	45/50
6		3I1	19.300	18.200	6%	4	0,64%	3	45/50
5		3I2	23.650	19.940	19%	3	0,48%	2	43
6		3I3	23.870	14.260	67%	43	6,88%	3	45/53
9		3M6	32.300	20.460	58%	24	3,84%	5	74
5		3S2	19.820	16.340	21%	3	0,48%	2	40
6		3S3	27.840	14.270	95%	99	15,84%	3	45/48,5
7		3T4	24.250	18.780	29%	23	3,68%	3	45/57
9		3T6	29.010	22.610	28%	4	0,64%	4	74
4		4CD	16.140	8.410	92%	33	5,28%	2	29
	Total Samples					625			



	E1	E2+E3	PBT E1+E2+E3
Empty	3.960	8.530	12.490
Capacity	2.040	8.470	10.510
Limit (axis/PBT)	6.000	17.000	23.000
Empty Parameter		50%	54%



	E1	E2+E3	PBT E1+E2+E3
Empty	3.860	4.980	8.840
Capacity	2.140	12.020	14.160
Limit (axis/PBT)	6.000	17.000	23.000
Empty Parameter		29%	38%

Fig. 6 Weight variation for the same type of truck

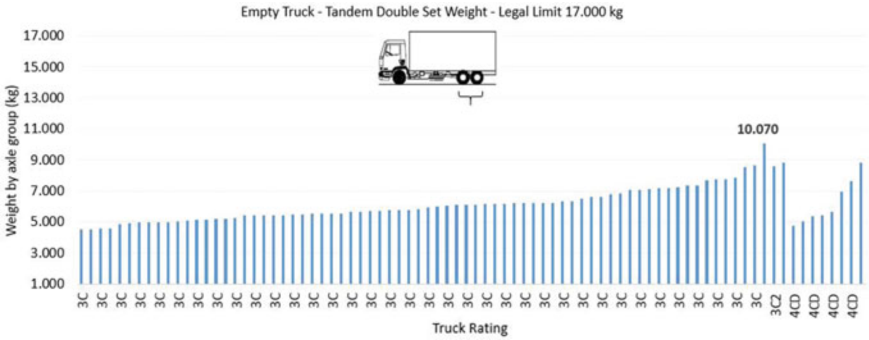


Fig. 7 Weight variation by axle group—truck

4 Technological Proposal and Discussion

Some prior works [6–9] addresses technologies on toll system which can aid to eradicate tariff payment evasion of trucks with axis suspended. Two proposals are being explored in this work.

Proposal 1 The equipment makes the current configuration of the reading (considering suspended axis) after considering a parameter for the cut (s) shaft (s)/set of suspension shafts (s). For example, a truck setting in tandem double set the limit to be considered would be 50% empty shaft legal limit—the legal limit = 17,000 kg; 50% of the legal limit = 8500 kg. However, many trucks, road equipment, and accessories contribute to a weight variation of the same empty truck configuration, parameterization and forcing it difficult to use a higher standard cut (perhaps 60 or 70%). This criterion will benefit some of the trucks, loaded partially, use the hanger shafts for not paying the toll.

Proposal 2 The equipment makes the current configuration of the reading (considering suspended axis) after considering a parameter for cutting PBT/PBTC according to the setting. For example, the weight of the empty truck limited to 50% of PBT the legal limit. Then, we also have a wide range of weight for the same vehicle, with different settings.

To do this, it is being considered at the checkpoint system based on Weigh-in-Motion (WIM)—Weighing in motion, with systematic mounted for operation as a high-speed weigh station, maintaining the operating speed of the track. The use of the WIM still needs to be standardized and regulated in order to function as a surveillance tool. Below, we present a layout of the proposal from the work (Figs. 8 and 9).

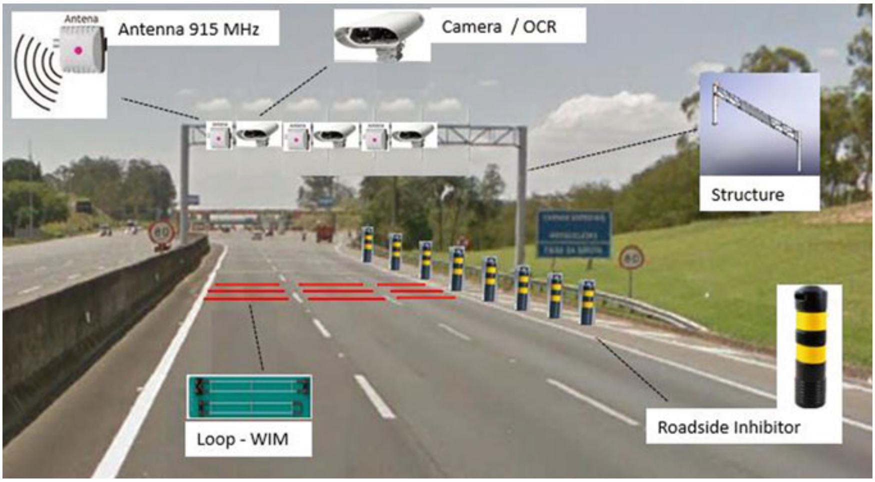


Fig. 8 Arrangement of equipment at the checkpoint

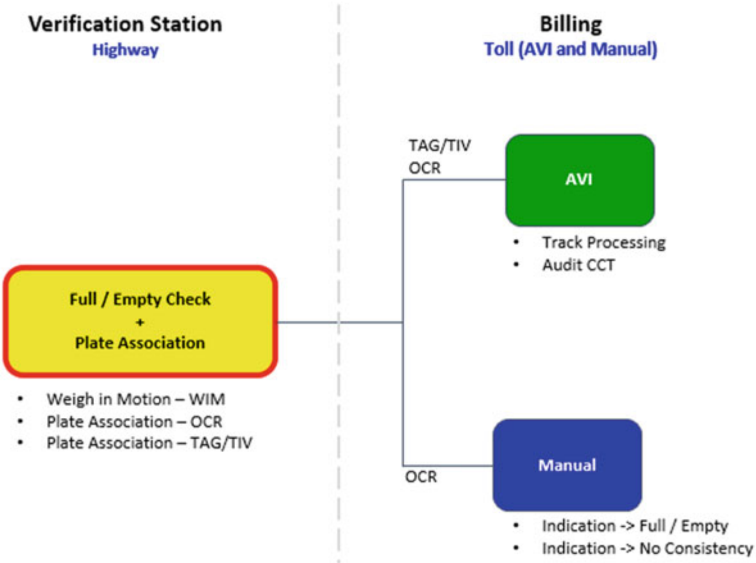


Fig. 9 Proposal layout

5 Discussion

In the current scenario, Brazil is discussing measures for specific subjects within a given area of activity related to planning, development, review, and implementation of metrological technical regulation. Considering the demand for metrological technical regulation of HS-WIM "High-Speed Weigh-in-Motion," a preliminary report was prepared in order to subsidize the decision on the required regulation.

Once the technical and legal issues have been overcome, the WIM can be used, beyond the purposes of our work, also for the inspection of excess load on highways, providing a longer useful life to the pavement and bringing more safety to road users.

References

1. ABCR: ABCR History. Available at: www.abcr.org.br. Accessed 05/31/2019
2. Milenkovic, M; Glavic D; Mlenadovic M.: Decision-support framework for selecting the optimal road toll collection system. J. Adv. Transp. **2018**. Article ID 4949565, 16 p (2018). https://doi.org/10.1155/2018/4949565. Accessed 03/09/2019
3. Senate News: Commission approves toll exemption for suspended axle trucks. Available at https://www12.senado.leg.br/noticias/materias/2018/06/19/comissao-aprova-isencao-de-pedagio-para-caminhoes-com-eixo-suspenso. Last accessed 2019/04/06
4. Brazil: Provisional Measure No. 833, of May 27, 2018. Amends Law No. 13,103 of March 2, 2015. It provides that in all national territory, empty cargo vehicles will be exempt from tolls on axles that keep suspended. Brasilia, issue 100-C, section 1-extra, p 1, May 2018
5. Machado, R.D.: Motion Weighing Systems: A Prototype for Data Analysis. (2016). Course Conclusion Paper—Federal University of Santa Catarina, Santa Catarina (2016)
6. Burnos, P., Gajda, J.: Thermal property analysis of axle load sensors for weighing vehicles in weigh-in-motion system. Sensors **16**(12) (2016). http://www.mdpi.com/1424-8220/16/12/2143
7. WEIGH-in-Motion of Road Vehicles: COST32, Final Report. Appendix 1: Eurpean WIM Specification (1999)
8. ISWIM: What is WIM and history. http://www.is-wim.org/index.php?nm=2&nsm=1&lg=en. Last accessed 2019/08/15
9. Conway, A., Walton, C.M.: Potential Application of ITS Technologies to Improve Commercial Vehicle Operations, Enforcement, and Monitoring. Center for Transportation Research, University of Texas, Austin, TX (2016)

A Look at the Evolution of Autonomous Cars and Its Impact on Society Along with Their Perspective on Future Mobility

Alysson G. Oliveira, Yuzo Iano, Diego Pajuelo, Ana Carolina Borges Monteiro, Reinaldo Padilha França, and Gabriel Gomes de Oliveira

Abstract Autonomous cars range from those with a slightly higher degree of driving assistance to those capable of driving without human interference, since the technologies are the same in both, further changing the way they are integrated into the on-board computer. Contributing from a number of components that work together to ensure that a self-driving car has the insight and guidance, it needs to drive safely, from cameras, sensors, and radars working like the eyes of the vehicle to detecting obstacles, road signs, traffic lights, pedestrians, relief among other aspects. The great advantage is that these devices do a real scan around the vehicle, allowing it to see in 360 degrees. In this way, he can perceive situations that the human eye could hardly catch, considerably increasing the safety of autonomous cars. In this sense, the fascination of humanity by car automation is the study of this paper. This research presents a detailed discussion along with historical review regarding the development of autonomous cars from the beginning to the present day and provides a perspective regarding the future of urban mobility.

Keywords Autonomous driving · Self-drive car · DARPA · Smart roads

---

A. G. Oliveira (✉) · Y. Iano · D. Pajuelo · A. C. B. Monteiro · R. P. França · G. G. de Oliveira
School of Electrical and Computer Engineering (FEEC), State University of Campinas—UNICAMP, Av. Albert Einstein 400, Barão Geraldo, Campinas, SP, Brazil
e-mail: superalysson@gmail.com

Y. Iano
e-mail: yuzo@decom.fee.unicamp.com

D. Pajuelo
e-mail: diego.pajuelo.castro@gmail.com

A. C. B. Monteiro
e-mail: monteiro@decom.fee.unicamp.com

R. P. França
e-mail: padilha@decom.fee.unicamp.com

G. G. de Oliveira
e-mail: oliveiragomesgabriel@live.com

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_56

1 Introduction

An autonomous car, also called a self-driving car or a stand-alone vehicle, is a robotic vehicle that is designed to travel between destinations without the driver having to make any effort. However, to qualify as fully autonomous, this vehicle must be able to navigate without human intervention to a predetermined destination by roads that have not been adapted for its use [1].

In the early twentieth century, mankind was amazed at technology fairs, very common at that time. These fairs have presented all kinds of technologies, inventions, and ideas that were being studied and could somehow get a day for daily use. The great attraction of these fairs was the vision of the future that had smart highways that drove cars "autonomous" they did not need the user to be in the vehicle control during the journey.

Several major racing line companies invested efforts and many resources to imagine a way to try to approach this utopian vision of the future. Many of these improvements were achieved with small advances throughout history [2].

Many of the initially designed cars simply could not drive more than a few kilometers without causing an accident, which at the time many of them fatal, so many projects were abandoned early on. It is possible to mention a few landmarks of development the autonomous car as that of Francis P. Houdina, an electrical engineer from the US Army and founder Houdina Radio Control, who had the first demonstration of a controlled car by radio control [3].

Another great exponent in the development of autonomous cars was the Ernst Dieter Dickmanns engineers who led his team in the Bundeswehr University and had several advances, which was possible to highlight mainly in the field of vision adapted cars, where he managed to create sensors and use in Mercedez doing so could have direction, acceleration, and brakes controlled by computers [4].

The autonomous car, although it seems recent, has been in the making for a long time, since by definition a driverless car has more control units, computing power, code lines, and wireless connections to the outside world than a normal car today. The so-called age of autonomous driving already has the efforts of giants such as General Motors, Hyundai, Google, Mercedes-Benz, Nissan, Volvo, Toyota, Honda, and Audi. They are all in this race with the goal which manages to put the first vehicle of its kind on the market [5].

The completion of such projects requires investments and development of powerful operating systems [6] and telecommunications methodologies that present easy usability, low level of abstraction, and low memory consumption of the devices involved [7]. In addition, digital image processing techniques related to detecting geometric shapes such as the Hough transform and watershed transform can assist in the detection of objects and present in the course of an autonomous car [8, 9].

In this sense, this paper aims to provide a historical review regarding the development of autonomous cars from the beginning to the present day and provides a perspective regarding the future of urban mobility.

2 Methodology

This survey carried out a historical review from the beginning to the present day relating the idealization to the development of autonomous cars enabling a look at the future of urban mobility.

3 Theoretical Background

At the fair New York 1939, the autonomous car idea gained wide public exposure at GM's Futurama exhibition at the World's Fair, was presented in its highways and Horizons pavilion, where General Motors showed how it would be the 20-year world in front, as shown in Figs. 1 and 2, this pavilion was presented the concept of intelligent roads, which allowed it to experiment with how electronics could be used to guide and maintain the proper tracking distance, developed by Bel engineer Geddes [10].

Following in the year 1958, a model with a coiled front end was tested that could "detect the alternating current of a road-embedded wire and adjust the steering wheel accordingly," which GM described the experiment as positive and claimed that the car drove along the road and stayed between the ends of the track without the driver's hands on the wheel, as shown in Fig. 3.

Fig. 1 General Motors pavilion at the 1939 World's Fair in New York [11]

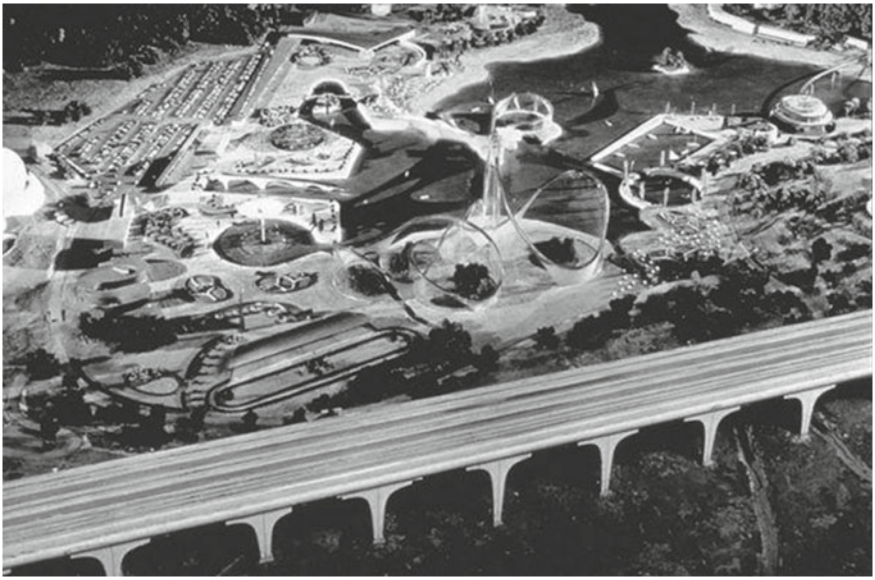


Fig. 2 Highways and Horizons pavilion [11]

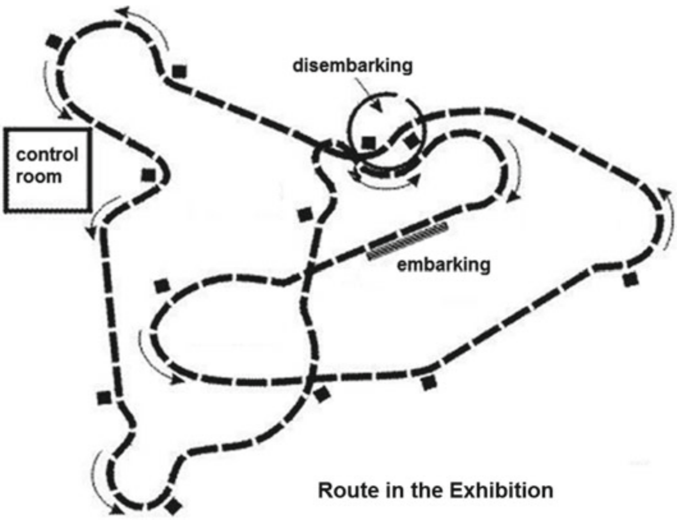


Fig. 3 Flag of the route model [11]

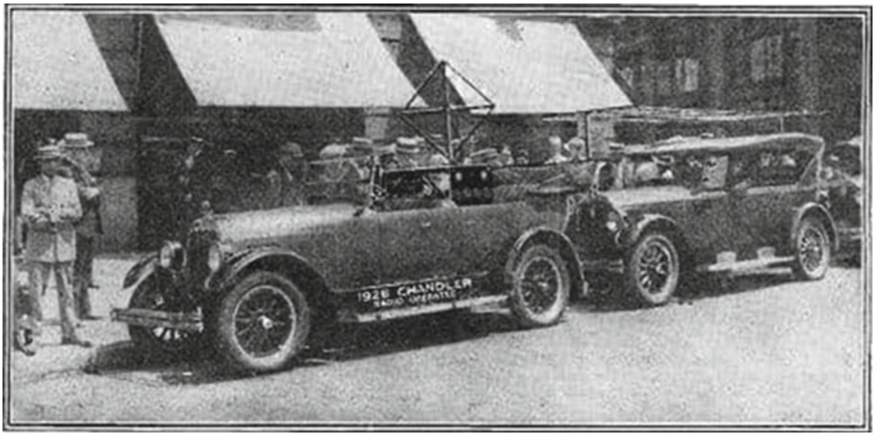


Fig. 4 A 1926 Chandler controlled by radio control in 1925 in NY [13]

It was at the Futurama exhibition at the World's Fair that there was the first big spotlight in interest in autonomous cars that had a great impact worldwide. However, some years earlier, around 1925, it was documented in the streets of New York, a controlled car radio control, originated by Francis P. Houdina which adapted to 1926 Chandler, as shown in Fig. 4, with a radio antenna and controlled with an oncoming car with another transmitting antenna that vehicle [12].

In the decade 50 was introduced by RCA Laboratories a car model that was guided by wires inserted into the ground. This idea has been enhanced by the traffic engineer Leland M. Hancock, who decided to apply the concept presented in real streets. In 1957, the state of Nebraska with the RC Laboratories presents a stretch of US Highway 77 and Interstate 2, wherein this passage was implanted sensors on the highways that feel electrical signals to the cars telling the position and velocity of metal objects that were transiting on the road [2, 14].

First regarded as a platform for the lunar rover, the Stanford Cart of the '60s and '70s was a simple buggy equipped with a camcorder and remote control with a very long cable, which over time this car was imbued with a larger capacity, intelligence, and image processing. The automotive concepts also had theirs at the 1962 General Motors exhibit Seattle World Fair, being an experimental Firebird III with a single control lever in place of a steering wheel and powered by a gasoline turbine engine, as shown in Fig. 5, which was just a part of General Motors' tests for cars of the future. In 1979, Hans Moravec achieved what was at the moment a signal achievement in robotic mobility, which the model successfully crossed a room full of chairs without human intervention in about five hours [2, 14].

What could be called the first truly autonomous car, a Volkswagen Touareg could process images of the road ahead in 1977 by S. Tsugawa and colleagues at the Tsukuba Mechanical Engineering Laboratory in Japan, which was able to accelerate up to 30 km/h, but was aided by a high rail [2, 14].

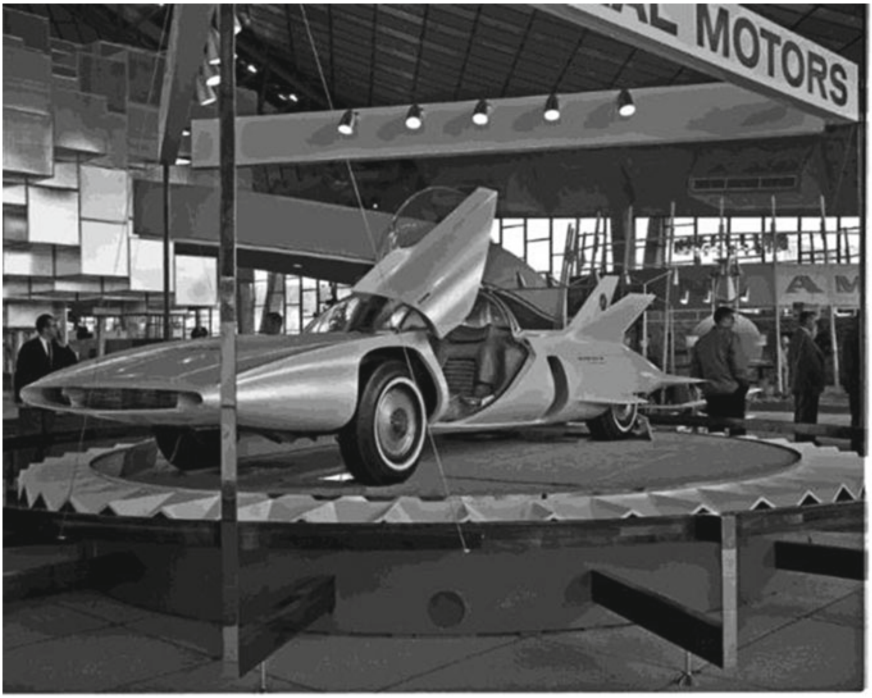


Fig. 5 Firebird III exhibited at the Seattle World's Fair, 1962 [17]

In 1987, the VaMoRs, equipped with two cameras, eight Intel 16-bit microprocessors, and a host of other sensors and software, running at 90 km/h for about 20 km. In 1994, a VaMP, with its two cameras processing 320 by 240 pixels per image at a distance of 100 meters, could recognize road markings, relative track positions, and the presence of other vehicles. In 1995, Carnegie Mellon University roboticists drove NavLab 5 from Pittsburgh to Los Angeles, the vehicle's add-ons included "a laptop, a windshield-mounted camera, a GPS receiver, as well as other equipment." where the car completed 98.2% autonomous driving percentage, needing a little help with preventing obstacles, among other things being almost 115 km, considered the longest period without human assistance [2, 14].

In the following decades, all advances made in cars were essentially sensors that somehow helped the user in driving. This situation was changed when the US military in order in war advances, with respect to Defense Advanced Research Projects Agency (DARPA), promoted an exclusive competition of autonomous cars. The Agency for Research and Military Development of states was pressed immediately after the attack of September 11, the twin towers, to reduce the use of American soldiers on battlefields. For this and develop this stimulus, the competition was created among the greatest American universities, where completely autonomous cars had to go through paths of different situations and come to certain points. It is worth highlighting the Grand Challenge, which was awarded $1000 lion for the car to travel 241 km in the Mojave Desert, which straddles the California and Nevada [15, 16].

As a result of technical problems, this first challenge in 2004, there was no winner. The car at Carnegie Mellon University was the best placed, but only managed to complete 5% of the race route. In this context, DARPA doubled the prize to $2 million and five cars of 195 participants were able to complete the entire route. The car's winner was Stanford University; students used a Volkswagen car to adapt a robot and used the same features as the previous year. But for change to address the problem, the software evolves over the previous year. The car can complete the route of 212 km in 6 h and 53 min [1], as shown in Figs. 6 and 7 [15, 16].

As of 2010, Audi's autonomous TTS was named after the first woman to win the Pikes Peak International Hill Climb award, Shelley, designed by Chris Gerdes and colleagues at Stanford University's Automotive Research Center, is packed with GPS, wheel sensors, and traction control algorithms that have allowed the car to limit boundaries, which has sprinted the 12th sprint, 42 miles to the top of Pike's Peak in 27 min [18].

Also as of 2010, technology giant Google's fleet of seven autonomous Toyota Prius hybrids has accumulated more than 140,000 miles (225,000 km) with only occasional human intervention since it hits the road in 2010, being part of the Google Driverless Car program led by Sebastian Thrun of the DARPA Grand Challenge. The cars used data from Google Street View, along with camera data, generating the ability to determine the car's position on a map, making the cars successfully transit San Francisco's Lombard Street, circling Lake Tahoe to Mountain View afterward to Santa Monica, California, USA [19].

Fig. 6 Red Jeep Carnegie Mellon University [17]

Fig. 7 Car of the University Sanford completed the course in the DARPA 2005 [17]

The technology behind the development of autonomous cars is one of the most complexes in the world and is advancing more and more. As sensor-related sensing technologies have become smaller, cheaper, and more powerful, they come a little closer to turning this horizon drawn throughout history into reality. In this regard, the US National Highway Traffic Safety Administration (NHTSA) currently defines autonomous cars in six categories. As shown in the following in Fig. 8 [20].

At level zero is basically cars that have no assistance to go. At level 1, the cars already have a basic form of some assistance, such as speed control, but in the end, it is the driver who can use or not care. In level 2, there is partial automation, where the car can maneuver and stop without the help of the driver; however, it is still necessary that the driver is present to take control on the course, thus at this level of automation in commercial level [20, 23].

At level 3, many manufacturers already have prototypes with this level of complexity, where the car can be in a stand-alone mode much of the route, but the driver must be present and attend the course so that if there is a need to take control. At level 4, the car can be autonomous in almost all conditions, leaving the driver the possibility of not having to take control. And finally, at level 5, the car is capable of driving in all possible conditions without the need for human help [20, 23, 24].

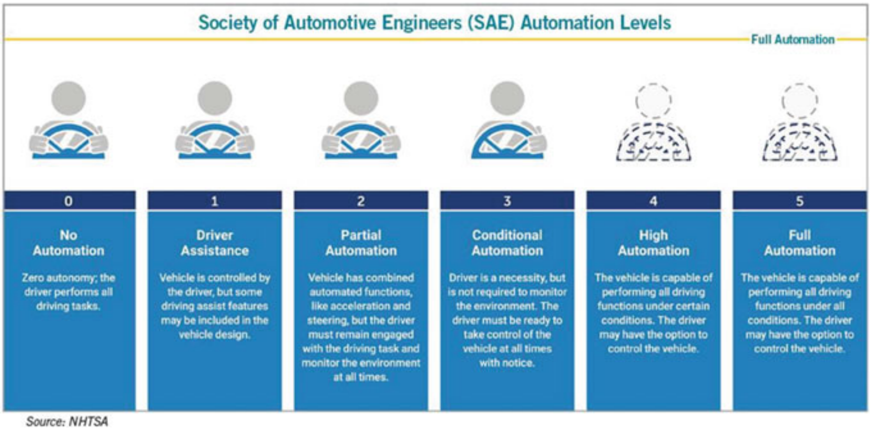


Fig. 8 Classification of autonomous cars [21, 22]

4 Results and Discussion

It is almost imperceptible but the present world is experiencing a second space race, yet in modern days the ships walk on the asphalt, and the astronauts do not need them. The race for the development of the first truly autonomous cars, which are those in which the driver gets in and out without touching the wheel, is happening, and basically, everyone who thinks about this future and conducts research in this direction is in the race.

Automotive companies such as Mercedes-Benz, Audi, BMW, and Tesla have launched over the years models that increasingly exempt drivers from their obligations. In the same vein, tech giants like Google and Uber are already testing vehicles with automated software on the street, while another tech giant like Apple is revealing more and more details of its design for self-driving cars. Where it is possible to state that a central technology, which consolidates all concepts and is considered as the mother of all projects, is artificial intelligence.

Many scientists estimate that it is possible to reach level 5 of automation of cars as long as there is the capacity to adapt roads and cities for that. The concept of smart roads [25, 26] is well known from the beginning of autonomous car studies. Many experts say that the set of autonomous cars, smart roads, and smart cities is the bridge to a future of great change in our society. In this context, it is possible to expect changes in the field of product logistics and urban mobility services to be discontinued for an obsession with new technologies as examples for insurance against car accidents [27].

Today, automakers and technology companies work with all five levels of automation. At level 1, the smallest, the car can control speed during a road trip or even turn the wheel for you to park. At level 5, the largest, the car does not even have to have a steering wheel. The development of current cars is currently at tier 2, where the vehicle can accelerate, brake, and steer by itself for certain periods of time, where a predictable next generation of cars will already be tier 3, completely hands-free.

Still taking into consideration that one of the technologies that are giving a great outlet for this future is the Internet of Things [28]. With increasingly connected and producing real-time data sets, it is no wonder the level of accuracy that cars and devices will exchange information. The Fourth Industrial Revolution is primarily based on data analysis [29] or the date that these same devices collect and transmit to the internet. All this information helped autonomous cars make more accurate decisions and short time [30].

However, one of the challenges that this technology must overcome is that driverless cars will only be truly safe when operated alongside other driverless cars in a controlled environment. Due to what is called the human factor, where the behaviors of pedestrians and non-robotic drivers (people) are unpredictable and difficult to replicate, it is noticeable that even the most modern models in traffic caused accidents when they interacted with each other real people in traffic.

The main difficulty is related to infrastructure. In order for a vehicle to run without human interference, there must be roads and telecommunications networks ready for it.

There is a long way to go yet still considering that discussions will be needed and understand more about permissions, laws on autonomous cars, but it is possible to agree beforehand as to the security that there will be when the modern world has made the transition completely [31].

5 Conclusions

Connecting cars with city infrastructure as they are today is a major challenge for integrating autonomous vehicles into society, since the presence of these cars full of embedded technologies can be very useful for safer, more efficient, and fluid traffic, Since this technology can collect data during tours, it can help to know which times are the hardest to travel around the city, but the reality is the opposite to this perspective, since the vast majority of cities is stagnant regarding the incorporation of these vehicles to your routine.

In this scenario, cities are not preparing for the arrival of autonomous vehicles at the same speed as technology companies and automakers develop their products and services, although in developed countries like the USA, for example, only some of the largest cities in the country are planning and adapting to the arrival of this new reality regarding autonomous vehicles.

While in developing countries, such as Brazil, it seems to be far from including autonomous vehicles in long-term traffic planning, as the country still faces major problems with common traffic control. As long the self drive car could bring some hope that all the tech that is build the cars will do all the adaptation work on badly signposted and bumpy streets, avoiding accidents and other problems.

Still considering some basic conditions regarding the proper functioning of these, among them, a primordial one regarding a stable connection to the internet and wide coverage of 3G and 4G networks, something that is still scarce in several regions of Brazil. However, in more developed areas 5G can also transfer more data than 4G, allowing cars to share sensor data such as video camera input and radar.

However, among the main trends for the future of vehicles, the development of autonomous cars is certainly one of the most anticipated and promising, while this innovation will usher in a new era in the automotive market, doubts about its operation are still very common.

References

1. Bimbraw, K.: Autonomous cars: past, present and future a review of the developments in the last century, the present scenario and the expected future of autonomous vehicle technology. In: 2015 12th International Conference on Informatics in Control, Automation and Robotics (ICINCO), vol. 1, pp. 191–198. IEEE, 2015
2. Khodayari, A., Ghaffari, A., Ameli, S., Flahatgar, J.: A historical review on lateral and longitudinal control of autonomous vehicle motions. In: 2010 International Conference on Mechanical and Electrical Technology, pp. 421–429. IEEE, 2010
3. "Phantom Auto" to Be Operated Here". The Free-Lance Star. 17 June 1932. Retrieved 14 Sept 2013
4. Delcker, J.: The man who invented the self-driving car (in 1986): long before Big Tech got behind the wheel, Ernst Dickmanns unleashed a driverless Mercedes onto the roads of Europe (July 19, 2018). Source: Politico: https://www.politico.eu/article/delf-driving-car-born-1986-ernst-dickmanns-mercedes/
5. West, D. M.: Moving Forward: Self-driving Vehicles in China, Europe, Japan, Korea, and the United States. Center for Technology Innovation at Brookings. Np, 12, 2016
6. França, R.P., Peluso, M., Monteiro, A.C.B., Iano, Y., Arthur, R., Estrela, V.V.: Development of a Kernel: a deeper look at the architecture of an operating system. In: Brazilian Technology Symposium, pp. 103–114. Springer, Cham, 2018
7. Padilha, R., Iano, Y., Monteiro, A.C.B., Arthur, R., Estrela, V.V.: Betterment proposal to multipath fading channels potential to MIMO systems. In: Brazilian Technology Symposium, pp. 115–130. Springer, Cham, 2018
8. Monteiro, A.C.B., Iano, Y., França, R.P., Arthur, R., Estrela, V.V.: A comparative study between methodologies based on the Hough transform and watershed transform on the blood cell count. In: Brazilian Technology Symposium, pp. 65–78. Springer, Cham, 2018
9. Monteiro, A.C.B., Iano, Y., França, R.P., Arthur, R.: Methodology of high accuracy, sensitivity and specificity in the counts of erythrocytes and leukocytes in blood smear images. In: Brazilian Technology Symposium, pp. 79–90. Springer, Cham, 2018
10. Kröger, F.: Autonomous Driving. Institut d'histoire moderne et contemporaine (IHMC), Paris (2016)
11. Available: http://obviousmag.org/archives/2008/06/futurama_uma_visao.html
12. Lutin, J., Kornhauser, AL.: The revolutionary development of self-driving vehicles and transportation implications for the engineering profession. ITE J. pp. 28–32, 2013
13. Available: https://www.engadget.com/gallery/self-driving-cars/#
14. Segrave, K.: The Electric Car in America, 1890–1922: A Social History. McFarland (2019)
15. Light, S.E.: The Military-environmental complex and the courts. J. Land Use Environ. Law **32**(2), 455–476 (2017)
16. Walker, M., Launchbury, J., Tompkins, S., Pallota, B., Hepburn, M., Sanchez, J.: Innovation at DARPA. Defense Advanced Research Projects Agency Arlington United States (2016)

17. Available: https://epocanegocios.globo.com/Tecnologia/noticia/2018/07/o-caminho-ate-o-carro-do-futuro.html
18. J. Funke, et al.: Up to the limits: autonomous Audi TTS. In: 2012 IEEE Intelligent Vehicles Symposium (IV), Alcal de Henares Madrid, Spain, pp. 541–547
19. Richtel, M., Dougherty, C.: Google's driverless cars run into problem: cars with drivers. New York Times, 1(September) (2015)
20. National Highway Traffic Safety Administration.: National Highway Traffic Safety Administration (NHTSA) notes. Trauma system agenda for the future. Ann. Emerg. Med. **41**(6), 798 (2003)
21. Available; https://www.howtogeek.com/401759/what-are-the-different-self-driving-car-levels-of-autonomy/
22. Available: https://atos.net/en/blog/autonomous-vehicles-myth-or-reality
23. Coben, J.H.: National Highway Traffic Safety Administration (NHTSA) notes. Contrasting rural and urban fatal crashes 1994–2003. Ann. Emerg. Med. **47**(6), 574–575, 2006
24. Anderson, J.M., Kalra, N., Stanley, K., Sorensen, P.: Autonomous Vehicle Technology A Guide for Policymakers. RAND Corporation, Santa Monica (2016)
25. Pindarwati, A.: SMART highway: highway near future systems in Jakarta. In: The Third Asia Future Conference, 2016
26. Smart transportation Alliance.: Smart Roads: Vision. transportation.org, 2015
27. Capgemini Research Institute.: The Autonomous Car: The Consumer Perspective (2018)
28. Understanding the Internet of Things (IoT): (2014). London: GSMA
29. Bloom, C., Tan, J., Ramjohn, J., Bauer, L.: Self-driving cars and collection date: privacy perceptions of networked autonomous vehicles. In: Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017), pp. 357–375, 2017
30. Cheruvu, R.: Big Data applications in self-driving cars. Source: Research Gate: https://www.researchgate.net/publication/318940413_Big_Data_Applications_in_Self-Driving_Cars
31. Gouvêa, B.V.: Strategies of Google and General Motors for Hybrid Car. Source: pantheon (Setmbro 2016). https://pantheon.ufrj.br/bitstream/11422/5029/1/Breno%20Vieira%20de%20Gouv%C3%AAa%20-%20As%20Estrat%C3%A9gias%20do%20Google%20e%20of%20General%20Motors20para%20%%%20Carro20Aut%C3%B4nomo.pdf

Escalators: A Technological Contribution in Reducing Failures and Maintaining

Ana Luiza de Souza Fracaro, Fernanda Antoneli Miyagi, Letícia Volpi, Luiz Vicente Figueira de Mello Filho, Raíssa Coutinho Tomaz Reganassi, and Roberta Luchiari Caixeta

Abstract Maintenance processes require systematic work to prolong the service life of the equipment. This article reports the most common failures and which types of maintenance are most used for escalators in order to make them safer and more efficient. The methodology used was the study of literature and interviews with companies that provide services in the area. The study reports the origin, functioning, and norms, so that it is possible to understand the main failures that occur and how the maintenance can be done in such a way that these errors do not hinder the operation of the ladder. Then, there is the fault and maintenance data provided by companies that provide this type of service. To conclude, improvements are proposed through current technologies for more satisfactory performance of escalators, both in the prevention of accidents and in the consumption of electricity.

Keywords Escalator · Maintenance · Sensors · Reducing · Failure

A. L. de Souza Fracaro · F. A. Miyagi (✉) · L. Volpi · L. V. F. de Mello Filho · R. C. T. Reganassi · R. L. Caixeta
Mackenzie Presbyterian University, Campinas, Brazil
e-mail: fernanda.antoneli@gmail.com

A. L. de Souza Fracaro
e-mail: anafracaro98@gmail.com

L. Volpi
e-mail: leticiavolpi1019@gmail.com

L. V. F. de Mello Filho
e-mail: luiz.mello@mackenzie.br

R. C. T. Reganassi
e-mail: reganassi@gmail.com

R. L. Caixeta
e-mail: ro.caixeta@gmail.com

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_57

1 Introduction

Currently, the most widely used transport method in public places, by supporting a large flow of people, is the escalator. This is the most widely used for transporting a larger number of people in the shortest possible time between floors of a building.

Its use may be to climb or descend different levels of floors, having its movement controlled according to the schedule basically by three forms: with a greater flow of people at peak times, sensoriality; with the application of presence sensors; or can be permanently connected, in places of the constant flow of people.

With the proper operation of the device, escalators reach a transport capacity from 4500 to 13,500 people per hour. Mass on average, between five and fifteen tons, provides greater comfort of displacement and safety to its users, and depending on its inclination, has greater efficiency, requiring less space and facilities for a lower price [1, 2].

Deepening the studies related to the operation of escalators, maintenance, their main failures and how to avoid them using technology, new methodologies are proposed to improve their operation and thus provide greater comfort for Users.

The aim of this research is to understand a little more about the failures and problems of maintaining escalators along with their operation, so as to understand the best form to prevent or decrease them.

2 Historic

In addition to the traditional staircases, the elevators and escalators are used to take us from one floor to the other, in various types of constructions. The escalators as we know today emerged from the joint work of some inventors like Jesse Reno, an American who around 1892 created and patented what would be the first version of an escalator, and Charles D. Seeberger, another American who years later developed the escalator design closest to what we have at present [3].

The initial creation of Reno consisted of a sloping mat, made of rubber-coated wood boards, in order to give more adherence to users' footwear, and driven by a motor with a speed of approximately 0.69 m per second. His intention at the time was to provide people with fun and not transportation, as we see these days. In the same year as Jesse's invention, George H. Wheeler designed and patented the first escalator with flat steps, but the project was not put into practice [4].

It was only in 1898 that Seeberger bought the patent of George and, together with the company Otis Elevator Company, developed and initiated the manufacture of escalators. The first time the general public had contact with the escalators was at the International Fair of Paris in 1901 [3, 4].

Even popular, the demand of escalators is not so great compared to elevators, causing a lack of research in the area and consequent scenario of technological stagnation. Although there are some improvements in the safety sensors, the basic design has not changed. The system used is basically the same for 50 years, being the same mechanisms employed worldwide [5].

3 Operation Principle

The operating mechanism that acts behind a treadmill is intelligible, i.e., an electric motor, usually 100 horsepower, depending on the conditioning value according to the number of people. Furthermore, calculations are made from the daily circulation, considering the transported mass, determined angle, length and width of the stairs. It is also responsible for activating all the equipment and the brake to turn off the device function. The engine's role is to enable a current that has coupled steps and maintain a continuous, circular motion to move the ladder 27–55 m per minute. Moreover, it has the handrail drive the impeller, so that it also moves. The steps run the length of the stairs and then bow, doing the opposite way underneath [1].

Overcoming elevators, escalators can transport persons up to 300 kg, and, depending on the width of the step and velocity reach a transport capacity of 4500–13,500 people per hour. On average, a treadmill has a mass between five and fifteen tons, depending on their size [1, 2].

Since they are a means of moving people, safety on escalators is critical, since any incorrect use of equipment, disregard for rules or failure might cause accidents to users. In order to avoid these accidental injuries, the stairs have multiple sensors and actuators throughout its length, which guarantee the device operation and the safety of users [6].

3.1 Standards

Thinking of the best security use of escalators, three ABNT (Brazilian Association of Technical Standards) standards were created for the installation, maintenance, and inspection. ABNT NBR 16083: 2012 concerning the requirements for maintenance of operating instructions of elevators, escalators, and moving walkways. ABNT NBR 10147: 2016 regarding periodic and routine acceptance tests and inspections to ensure the expected performances of escalators and moving walks. Finally, ABNT NBR 16734: 2019, which aims to establish safety requirements for construction and installation of escalators and moving walks, aligned to it, companies [7].

The latest Brazilian standard mentioned, came into effect on May 14, 2019, replacing the ABNT NBR NM 195: 1999, applied in the MERCOSUL countries. In addition to the regulation for the physical product, there is ABNT NBR 14364: 1999, establishing qualification requirements for activities of inspectors and supervisors who perform inspection and testing of lifts and escalators [8].

Table 1 Theoretical and effective transport capacity at 0.5 m/s

Step width (mm)	Theoretical transportation capacity v = 0.5 m/s (people/hour)	Theoretical transportation capacity at rated speed of v = 0.5 m/s		
		Uncrowded (people/hour)	Convenient (people/hour)	With crowding (people/hour)
600	4500	1800	2700	3600
800	6750	2400	3600	4800
1000	9000	3000	4500	6000

Source Planning Guide for escalators and moving walks Schindler

Within the mentioned standards, some prerequisites for escalator installations may be listed. In Table 1, the theoretical transport capacity of the escalators can be identified according to width and speed, as well as the effective capacity, which corresponds to the range between 40 and 80% of the theoretical capacity, which depends on the density of users and width of the steps [9].

3.2 Maintenance

The maintenance of equipment, machinery, and facilities is made in order to keep them as long as possible in order to decrease waste, satisfy customers, and entice users [10].

Thus, the concept of maintenance is given by a set of actions taken to develop the specific maintenance policies in the production organization. Thus, the design of the maintenance consists of a set of forms of maintenance work and the overall structure, in which these procedures are done [10].

There are four types of maintenance studied in the article: corrective, which can be planned and unplanned; preventive; predictive; and detective.

a. Corrective Maintenance:

   I. Planned corrective maintenance: "There is a failure or abnormal condition of operation of equipment and the correction depends on managerial decision, in the function of predictive monitoring or by the decision to operate until the break." According to PINTO and XAVIER, 2001 [11].

   II. Unplanned corrective maintenance: It consists of a failure that has already occurred, in which there is no time for planning to postpone the repair.

b. Preventive maintenance: It is given by interventions at certain time intervals with the purpose of reducing failures and avoiding a fall in the performance of a process, which is mandatory in the air sector.

c. Predictive maintenance: It is performed when there is a variation in the performance condition of equipment, which can be currently measured and evaluated by means of the technology.
d. Detective maintenance: aims to find hidden failures or that are not noticeable and are used for checking a system.
   There are also the maintenance approaches proposed in the literature and each of them has its particularity. The most popular are: (e) reliability-centered maintenance (MCC); (f) total productive maintenance and; (g) risk-centered maintenance (RBM) [10].
e. Reliability-centric maintenance (MCC): It is the combination of different techniques and tools for maintenance administration and can be decision trees and analysis of the mode of failure and effect to support decisions taken efficiently and effectively.
f. Total productive maintenance (TPM): The main objective is the maintenance of equipment with the participation of the production team, aiming at continuous improvement, because it is considered that the operator is the one who best knows the equipment and its operation.
g. Risk-based maintenance (RBM): It is based on the risk of accidents and aims to reduce it from the equipment. Where the risk is assessed as being high or medium, higher maintenance is concentrated, already in areas with a lower risk, the efforts are lower to minimize labor and maintenance costs.

3.3 Failures

Based on ABNT NBR 5462-1994, the concept of failure is an item's ability to perform the required function. It is the total or partial decrease in the capacity of a part, component or machine to perform its function during a period of time [11].

Several factors can cause an operating stop in the escalators during its operation. These factors may be failures of some mechanical, electrical, or electronic equipment present in its structure; correct warnings received by the safety sensors, such as situations in which the ladder should actually stop; incorrect warnings received by the sensors mentioned above, in which there is undue stop of the escalator or events outside the machine, such as vandalism [12].

Other possible failures that may cause accidents or discomfort of the user are the absence of operability of the stop device at the ends of the installation, precluding an immediate stop if necessary; the speed of the handrail is above the permissible in relation to the speed of the steps or plates; not immediate stopping with the existence of an indentation occasioned by objects or some end of the human body; extreme situation where there is no immediate immobilization if the installation of the handrail occurs; and untensioned tension current, causing malfunction and breaking limit [13].

Sudden stops coming from some failures, often arising from the safety sensors, are very abrupt and may cause an imbalance of users with consequent injuries such as cuts, bruises, entraining of fingers, falls, abrasions, fractures, and in certain cases crushes that can lead to suffocation. In addition, there may be material expenditures (maintenance team displacement) and intangible (user dissatisfaction) [6].

The handrail is also another point of attention, being able to launch children in the span of escalators and have risks of accidents with serious levels of injury and even fatal, being the physical protection of these fundamental points to stop any possibility in operation of the handrails.

3.4 Sensors and Energy Saving

Failures in escalators present in places with a high flow of people, like some lines of the São Paulo subway, are the so-called false failures. These are due to large variations in the vibrations perceived by the various safety sensors that exist throughout your structure. These can stop the escalator's operation when they receive information of something wrong in any part of the equipment, being present at certain points of its structure, such as in the footer that detects if there is a presence of foreign bodies trapped on a rung [6].

During the treadmill operation, some vibrations arise from your movement. To detect the causes of unexpected and unnecessary stops, one of the ways is in the installation of the vibration sensor on a rung of an escalator, who presented problems with great frequency. This way it is possible to analyze that the high flow of people on the same ladder and can cause greater vibrations in the sensors and be the reason to cause the stop. The vibrational increase leads to an understanding, on the part of the safety sensors, that something wrong could be happening in some part of the equipment, which may reflect on a stop of its often unnecessary operation [6].

The escalator, as stated above, is composed of several safety sensory devices. Taking this into consideration, coupled with the economy of energy and economic resources, new technological resources are used to minimize the energy consumption of escalators in moments of low flow of people [14].

The escalator control system (SCER) consists of a presence sensor, a microcontroller electronic circuit, and a frequency inverter. The SCER enters into a normal operating regime when the ladder starts operating. If bypassing an established time without the flow of people and without the actuation of the presence sensor, the speed of the ladder is reduced by means of the electronic circuit microcontroller and performed by a frequency inverter. If after entering a reduced regime, another period of time determined without passage of users, the ladder enters a stopped regime, i.e., with zero speed. If any user is present during any of these periods, the ladder initiates a smooth acceleration until it restores its normal operating regime. The three possible operating states of the ladder in the normal, reduced, or stopped regime are associated with the digital input of the frequency inverter [14].

The SCER can be a substitute for the main control panel, which together with the security panel, controls the mechanical devices by means of end-of-stroke sensors. With the test of its implementation on a ladder with a high flow of people, it is possible to obtain a considerable financial economy monthly, in addition to a good reduction in energy consumption, which impacts the energy economy. Therefore, the installation of this system can reduce expenses with the maintenance of the panels and facilitates the work of employees to shutdown the equipment. Adaptations like this can bring great benefits, which contemplate users [14].

4 Methodology

Ethical approval for the experiment in this study was deemed unnecessary by the Ethics Committee. As also the activities in the experiment do not pose risks greater than those ordinarily encountered in daily life and all the subjects remained anonymous. As well as all participants were given consent and were also informed about the academic objective of the research as well as the guarantee of anonymity giving the choice of participation or not. Still considering that as it is an opinion poll, according to Resolutions 466 and 510 of the National Health Council, there is no need to go through the University's Ethics Committee. This is following CNS Resolution 466/12 of the National Health Council, of December 12, 2012, (available at https://conselho.saude.gov.br/resolucoes/2012/Reso466.pdf). Idem with Article 1 of Resolution 510, of April 7, 2016, of Brazil's National Health Council (available at http://www.conselho.saude.gov.br/resolucoes/2016/Reso510.pdf) and also with item 8.0.5 of the Ethical Principles of Psychologists and Code of Conduct of the American Psychological Association (available at https://www.apa.org/ethics/code/principles.pdf).

To elaborate on the present work, we opted for the study of cases in the literature, along with the application of questionnaires to companies in the field of elevators and escalators.

According to Eisenhardt (1989), the case study is a research strategy that focuses on understanding the facts present in unique scenarios, combining methods of collection such as analysis of documents, interviews, questionnaires, and observations, and can quantitative, qualitative or both.

The questions of the questionnaire had the model of open questions, in which respondents had complete freedom to respond with their own words, without limitations of alternatives. They were:

- What is the frequency of failures?
- What are the most frequent failures?
- Do many accidents occur? What are the types and motives?
- What types of maintenance are performed?
- What is the frequency of maintenance and what is done in it?
- How is the cleaning of the escalators and how often it happens?
- Escalators that are close to the food court are subject to falling food and beverages. Does this affect your operation in any way?
- What are the prerequisites for installing an escalator? (structure, size, weight, angulation).

- Energy-saving escalator (with presence sensors) has more flaws?

The information from the questionnaires was of a qualitative nature and covered mainly the themes of failures and maintenance of the escalators.

5 Results and Analysis

As mentioned earlier, from interviews with specialized companies. As a result, it was identified that the faults are not constant, however, maintenance should be performed frequently, and preventive maintenance is recommended monthly.

Moving parts should be lubricated constantly, such as set currents, checked clearances, electrical contacts, and sensors tested, and finally, thorough cleaning should be done. Corrective maintenance is only performed when the stairs come to a standstill or if there is any symptom in the operation, which is opted for an intervention.

The most frequent failures are those caused by the user himself, related to clothes and personal effects that may get caught in the step combs. The most frequent accidents, in turn, are the falls of users who climb the handrails, an action prohibited by safety regulations, clothes, and shoes that get caught on the steps, and children who lock their fingers in the handrail entrances.

Cleanliness is something that should also be critical. Rubber parts such as handrails and steps should be cleaned with detergent and sponge and free of petroleum products or the like due to rubber wear. Internal cleaning should only be done with degreasing water-based and cloths lint-free. Another relevant topic is the fact that spillage of liquids in public places, for example, may affect stair components and may vary in the type and amount of liquid. If this is a solvent or corrosive chemical, it will clearly damage the model, but if it is just water, it does not affect anything, since the stairs have already been proven.

According to information obtained from the interviews, extreme cases of accidents in which the stairway opens, such as those in China, are due to possible safety contacts that may have been incorrectly joined, as well as brakes unregulated emergency, some essential parts not properly fixed. In other words, human failure and corporate irresponsibility in delaying maintenance are the main causes of failure.

6 Conclusion

After the data collected on the escalators discussed in this paper, it is possible to conclude that this equipment is the best option when it comes to effectiveness, safety, and agility. They are regulated by ABNT standards for better safety in their use, installation, maintenance, and inspection. Some prerequisites are required for your installation, thus avoiding the various factors that may cause a shutdown during operation.

Despite constant maintenance, control, compliance with required standards and necessary prerequisites, the device is still prone to failures and accidents may be related to mechanical, electrical or electronic equipment present in its structure. Also, as incorrect warnings received by the sensors causing improper stops, vandalism, and even improper use of the device. Intelligent cameras that identify sudden movements on escalators can be parameterized for momentary stopping and accident reduction.

Finally, in order to improve the quality of safety and well-being of users on escalators, it would be proposed a higher precision in the safety sensors for detecting foreign bodies before the stop trigger by parameterization according to the volume of users.

In addition, the installation of a kinetic energy converter from the pressure exerted by the footsteps of users through accelerometers, in conversion to electricity, can be used to contribute to the saving of electricity and even supply the escalators as a suggestion for new technological research.

References

1. Engetax elevators: How the escalator works. Available at. http://engetax.com.br/como-funciona-a-escada-rolante/. Accessed 13 May (2019)
2. Thyssenkrupp elevators: 125 years ago The World's first escalator went into operation. Available at. https://www.thyssenkruppelevadores.com.br/release/download/898/. Accessed 13 May (2019)
3. Villarta elevators: Do you know the story of the escalator? Where did she come from? No? Then go up slowly and enjoy the trip! Available at. https://villarta.com.br/blog/2017/07/31/voce-conhece-a-historia-da-escada-rolante-de-onde-ela-veio-nao-entao-suba-devagar-e-aproveite-a-viagem/. Accessed 13 May 2019
4. Otis elevators: Documentation about Escalators. Available at. http://www.otis.com/site/br/OT_DL_Documents/OT_DL_SiteDocuments/Bb4_AboutEscalators.pdf. Accessed 13 May (2019)
5. AECWEB: Well designed escalator avoids accidents. Available in: https://www.aecweb.com.br/cont/m/rev/escada-rolante-bem-projetada-evita-acidentes_7472_10_0Accessed 21 May (2019)
6. Alves, L.A: system for Identifying Failures in Escalators Using Artificial Neural Networks. 2015.84 F. Dissertation (Master in Automation and Process control.)-Federal Institute of Education, Science and Technology of São Paulo, IFSP, São Paulo (2015)
7. ABNT. Technical Dossier-Escalators and Treadmills. Available in: http://abnt.org.br/paginampe/biblioteca/files/upload/anexos/pdf/5a746d69f1d077097b3e8fb22a5bad4a.pdf. Accessed 17 May (2019)
8. ABNT Catalogue: Technical Standard. Available in: https://www.abntcatalogo.com.br/norma.aspx?id=3095. Accessed 17 May (2019)
9. Atlas Schindler: Planning Guide-Stairs and treadmills projects. Available in: https://www.schindler.com/content/dam/web/br/pdf/ni/guia-planejamento-escadaesteira.pdf. Accessed 16 May (2019)
10. Fuentes, F.F.E: Methodology for Industrial Maintenance Management Innovation. (2006). 208 F. Thesis (postgraduate degree of Ph.D. in Mechanical Engineering.)—Universidade Federal de Santa Catarina, Florianópolis (2006)
11. Pinto, A.K.; Xavier, J.N: Maintenance: Strategic Function Qualitymark Editora, Rio de Janeiro (1999)

12. Alves, L.A: System for Identifying Failures in Escalators Using Artificial Neural Networks: 2015.84 F. Dissertation (Master in Automation and Process control.)-Federal Institute of Education, Science and Technology of São Paulo, IFSP, São Paulo (2015)
13. Duarte, L.M.D.L: Risk Analysis and Safety Inspections in Elevators, Mechanical Staircases and Conveyors (2014).100 F. Dissertation (Master's degree in Mechanical engineering.)-Departmental Area of Mechanical Engineering, ISEL, Lisbon (2014)
14. Schröder, R.; Perussi, A.; Alcantara, C.L.M: Modernization of Control and Intermittent Operation for Escalators. (2011).103 F. Dissertation (undergraduate degree in technology in electronics)-academic Department of Electronics, UTFPR, Curitiba (2011)

Use of Data Acquisition Technologies to Optimize Construction Progress Monitoring Times in Residential Buildings

Marisol Aronés, Andrea Mendoza, Sandra Rodriguez, and Humberto Ramirez

Abstract The monitoring systems that are currently used as part of the management of construction projects consist of collecting the field information, documenting it manually and then digitizing it. Those responsible for carrying out the reports through this procedure dedicate between 28% and 41% of their time daily for their preparation. This traditional process is slow and inefficient because the information is dispersed in different documents, so data can be omitted, causing that no corrective actions are taken in a timely manner. Likewise, the construction sector has little use of technological resources, and therefore, there is no automated process, which makes it difficult to monitor construction projects in an efficient way. In this sense, this research establishes the optimization of the control procedure through the use of a data acquisition tool to reduce the man-hours used in advance control. With the application of this proposal, the hours spent by the personnel involved in project management were reduced by 30%.

Keywords Monitoring · Advance · Control · Construction · Projects · Optimization · Technology · Data · Residential buildings

M. Aronés · A. Mendoza · S. Rodriguez (✉) · H. Ramirez
Peruvian University of Applied Sciences, Av. Prolongación Primavera 2390, Santiago de Surco, Lima, Peru
e-mail: sandra.rodriguez@upc.edu.pe

M. Aronés
e-mail: u201513007@upc.edu.pe

A. Mendoza
e-mail: u201510232@upc.edu.pe

H. Ramirez
e-mail: pccihram@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_58

1 Introduction

Construction projects are segmented into stages such as initiation, planning, execution, closure and monitoring. At the control stage, those in charge of various areas of the project must coordinate the human resources, materials, equipment, time and cost determined to achieve the objectives set. The construction progress monitoring by the traditional method consists of collecting the field information, documenting it manually and then digitizing it using established formats. All this information is accumulated and allows those responsible to measure the advance of the project. However, this activity demands a lot of time which often causes delays and rework, thus causing low productivity of the work team dedicated to monitoring the progress of construction.

According to studies carried out, control tasks using the traditional method generate difficulties in handling information due to inaccurate manual introductions and deferred data collection, which leads to delays in decision-making [1]. It is common that the lack of timely information on the real state of the projects in the central office means that the problems remain unsolved and that senior engineering personnel cannot solve them in a timely manner [2].

One of the proposed solutions to optimize project control is the use of high definition video cameras. With these videos, the work hours of the crew and the teams are recorded, the productivity of the activities is analyzed and even the proposal helps identify possible safety problems in the workplace. The time of information collecting through this technique improves by 30–50%, thus facilitating the work of field engineers and/or project managers [3].

Also, for the monitoring of construction progress, [4] presents the use of low-precision 3D scans, which are small enough to be in workers' protective helmets and used machinery. In this way, workers capture all project workplaces in real time and record the information in a cloud. The information captured is taken to a model that is subsequently compared with a planned model to identify their differences and, therefore, the deviations from the schedule. This method allows more efficient project management because changes are constantly perceived and the model of how it is built is continuously updated throughout the construction process, thus reporting the differences in the schedule as soon as possible [4].

Another solution is to take advantage of advances in short-range photogrammetry to continuously monitor construction activities. With this technique, discrepancies between the planned and the executed schedule are detected, by means of a comparison between the Building Information Modeling (BIM) and the 3D point model. If the system detects deviations between the built and the planned, emails and SMS are sent to the staff. The investigation concludes that the whole process, from the initial storage of the photos to the sending of notifications, takes less than 60 min [5].

Similarly, another proposal is to import the modeling of an office into Autodesk BIM 360 Field[1] software to show the information that is required to monitor the

[1]Construction management software that combines mobile technologies from the construction site with reports in a cloud of information.

progress of the project. For the collection of real-time data, the percentage of advance is registered daily at the worksite using an iPad.[2] This data allows to view the progress of the project and makes the corresponding reports. This methodology offers significant benefits, such as being able to handle all the information in one place and also the efficiency in the communication with the personnel that is in the work and the office [6].

Deferred information and inefficient communication between project stakeholders limit efficiency in construction monitoring. Therefore, a collaborative work procedure is analyzed where the collection, analysis and dissemination of information are standardized. This system offers manual inspection assistance using a database and a registration template accessible on mobile devices. Similarly, it uses several smart sensors to achieve real-time monitoring of large equipment and facilities. This method demonstrates great potential to improve the evaluation of labor inspections and facilitate the development of collaborative management among those responsible [7].

Another proposal is a methodology that allows the user to monitor the current state of the construction by verifying the activities through a real-time comparison between the planned process represented by an nD object, based on telepresence, and a real image of the construction site, sent from a webcam on the site. With this method, the project manager can visualize the real state of work and also use it as a decision-making system to formulate the most appropriate project schedule [8].

Likewise, there is the integration between Building Information Modeling (BIM) and Reverse Engineering (RE) to improve the use of information in different phases in order to reduce errors and rework in projects. In addition, for a better understanding of design and construction, this methodology incorporates support technologies such as virtual reality, 3D printing and prefabrication. The implementation of this method eliminates design changes by 30%, project modifications by 25% and, finally, constant monitoring saves two months of work [9].

Currently, the use of technology is transforming project management into construction because it allows an improvement in data acquisition, and therefore, there is an improvement in decision-making in project management [10].

The previous research shows different techniques to optimize construction progress monitoring. Therefore, the present investigation proposes a work procedure using technological tools so that those responsible do not invest time and efforts greater than necessary in the collection of information, preparation of control reports and the exchange of this information for decision-making, but that can be dedicated to other activities that require more attention.

[2]Mobile electronic device, also known as a tablet, which performs similar functions to laptops.

Table 1 Characteristics of the analyzed sample

	Project A	Project B	Project C	Project D
Basements	0	3	5	5
Floors	7 (4 towers)	10	37	5
Total area (m^2)	13,304.10	12,498.63	36,244.93	7,249.32
Contractual term (months)	16	15	21.6	13
Stage project	Structures	Structures	Structures	Structures

Fig. 1 Research process

2 Methodology

For this research, the information was collected from a sample of residential buildings, an area of land between 7200 and 36,000 m^2, construction period between 13 and 22 months, and also, they were evaluated in the structure stage conformed by the activities of reinforcing steel, formwork and concrete. Projects A, B and C were used for data collection and project D is used for the implementation of the proposal. These buildings present the characteristics shown in Table 1.

To carry out the process, tools such as a flowchart were used to present the new process which includes their respective managers, and also, an evaluation matrix to determine the data acquisition tool that best suits the proposal developed. The digital tools evaluated were Procore,[3] IPSUM,[4] BulldozAIR[5] and FINALCAD.[6] In addition, comparative tables were included to show the optimization of the time of those responsible for monitoring the progress of the construction.

The authors propose the following methodology for the research shown in Fig. 1.

This proposed methodology is a logical sequence of procedures to optimize one of the construction project indicators.

As a first step, the information obtained from the sample is recorded and analyzed with respect to the monitoring of the construction progress by the traditional method, which includes the content of the reports, the time required to produce each report and the process of interaction between the areas of each project. Then, taking this

[3]Digital platform that allows project management, ensuring safety, quality and productivity in the field.

[4]Collaborative platform where the Last Planner tools are available in digital version for a better management of the progress of the project.

[5]Data acquisition tool that manages projects, collects data, generates a centralized collaboration and exports reports automatically.

[6]Digital tool aimed at controlling the quality of deliverables in construction projects.

information into account, the area that demands the most time in its monitoring was selected. After this, the current monitoring process of the selected area was determined. To optimize the time spent on monitoring the progress of the construction of the selected area, a redesign of the process was proposed, which uses a data acquisition tool as a means. To do this, the parameters that the work report must contain, the people involved and the most appropriate procedure are established. Once the parameters that are re-wanted for the new process were defined, the existing data acquisition tools were evaluated and it defined the one that best fits the proposed solution. Subsequently, the selected tool was adapted to the new process, which includes uploading the digital report to its platform and the access of those involved in this report. Finally, the implementation stage was carried out in the project.

3 Results

3.1 Registration and Analysis of Information

Taking into account the traditional method, the information for each project in the areas of quality, production and planning was recorded. Within the registered information, you have the time required to carry out your work control report. For the analysis of this information, the process was divided into two activities: the collection of information for the report and the transcription of this information in it.

Figure 2 shows the times invested daily in the collection of information and the transcription of this information in their respective reports using the traditional method. The quality area invests a total of 2.2 h, the planning area 2.8 h and the production area is a total of 3.3 h, thus being the area selected for this investigation. According to this analysis, it spends between 28 and 41% of their time daily for the preparation of their report.

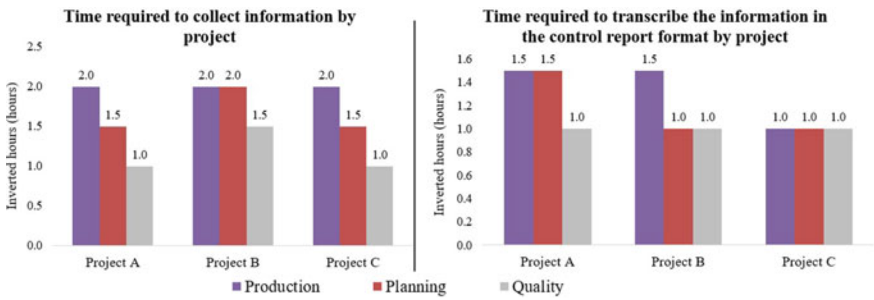


Fig. 2 Times for reporting by area and by project

3.2 Determination of the Current Process

Being the production area that demands the longest time, the traditional process for the realization of the advance control report was determined. It begins with the coordination and execution of activities in different areas of the project. Then, this information is collected from each work area to collect it in a single document. With this information the daily advance report is made and sent by mail to the work resident so that the report is signed and validated. The procedure is shown in Fig. 3.

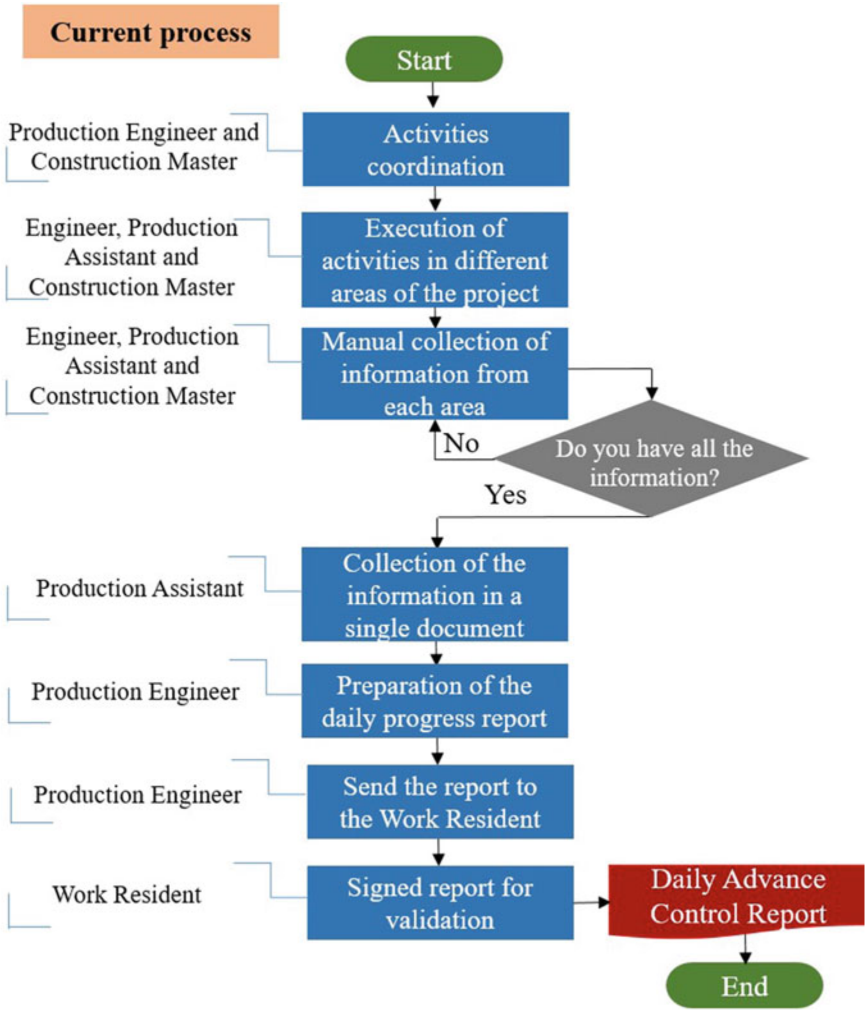


Fig. 3 Procedure for monitoring progress by the traditional method

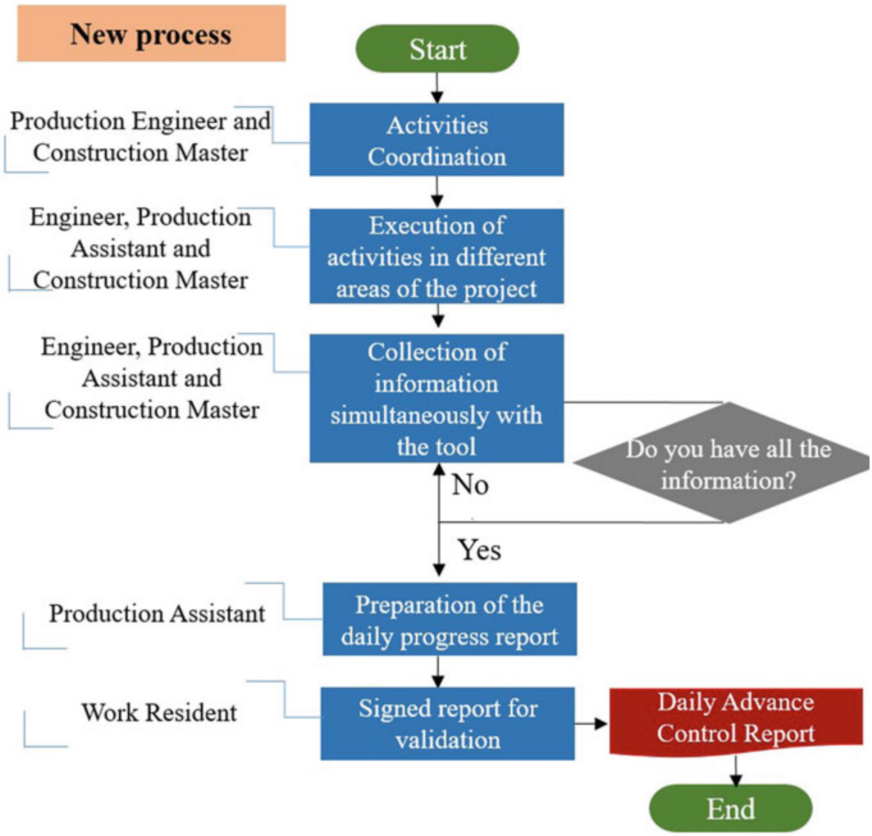


Fig. 4 Procedure for monitoring progress by the proposed method

3.3 Process redesign

As can be seen in Fig. 4, the change in the new process begins with the collection of simultaneous information from different work areas, due to the use of a data acquisition tool. Likewise, the use of this tool collects information about the different work areas in the same digital platform. Once the report has been made in this new process, it is not necessary to send this document by mail to the resident, because he has access to the report through the digital platform.

3.4 Evaluation of the existing data acquisition tools

Among the most important requirements that the data acquisition tool must have is to be able to count on the information of the technical file on the digital platform, disseminate the information of the work among all those involved, generate reports and statistical tables automatically, notification of atypical processes, segmentation of work teams and real-time monitoring of activities. Likewise, the tool must adapt to the previously defined parameters for the production area report. Therefore, an evaluation matrix was made considering the following digital tools existing in the local market such as Procore, IPSUM, BulldozAIR and FINALCAD, which is shown in Table 2.

Table 2 Matrix of evaluation of data acquisition tools

Attributes	Procore	IPSUM	BulldozAIR	FINALCAD
Information on the technical file on the digital platform	1	1	1	1
Information disseminated among those involved in various areas	1	1	1	1
Automatic generation of reports, work reports and statistical tables	1	1	1	1
Notification of atypical processes	0	1	1	1
Segmentation of work teams	0	0	1	1
Real-time activities tracking	1	1	1	1
Resource management for the project	1	1	1	1
Changes without connection to Internet networks	1	1	1	1
User editing	1	1	0	1
Observations linked to plans, photographs and freehand drawings	1	1	1	1
Total	8	9	9	10

The assignment of the value "1" corresponds to the fulfillment of the requirement and "0" otherwise. The tool selected for adapting the new process is FINALCAD because it meets all the selected requirements, with a total score of 10.

3.5 Adaptation of the tool to the new process

The digital report was uploaded to the FINALCAD platform and was provided access to those involved in this report. The mentioned report is presented in Fig. 5.

The report contains general information, as well as the theoretical metered of the activity and then its field metered. Likewise, there are comments, photographs and, finally, the signature of the resident engineer to conclude the report.

Once the data is taken, the information shown in Fig. 6 is generated.

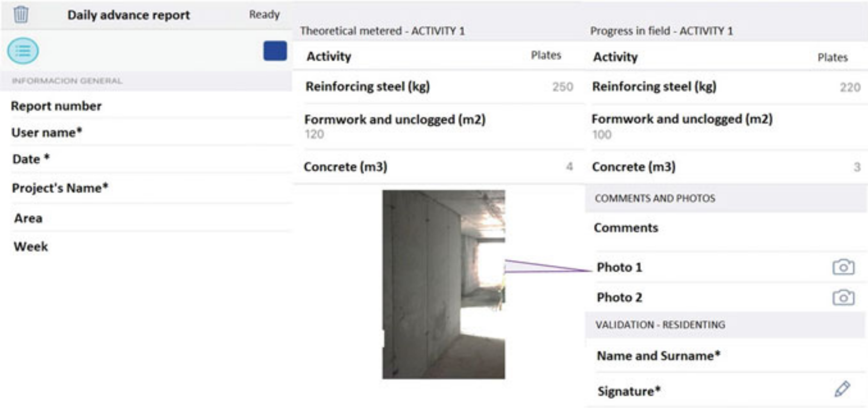
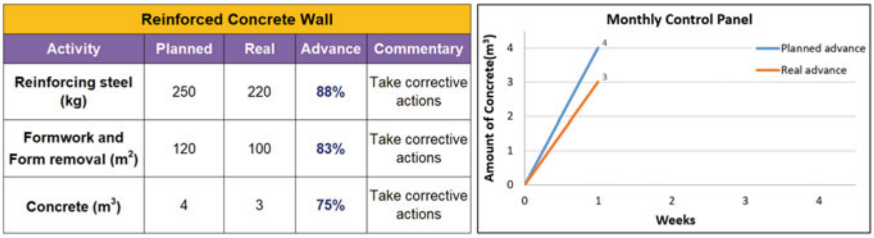


Fig. 5 Report format daily advance control on the digital platform



Reinforced Concrete Wall				
Activity	Planned	Real	Advance	Commentary
Reinforcing steel (kg)	250	220	88%	Take corrective actions
Formwork and Form removal (m^2)	120	100	83%	Take corrective actions
Concrete (m^3)	4	3	75%	Take corrective actions

Fig. 6 Information generated with the data of the advance control format

The percentage of the advance of the items of steel, formwork and concrete for the element and their respective comments are observed. Likewise, a variation curve between the planned and the executed in the field is shown.

3.6 *Implementation*

The proposal applied to project D shows the results of the time required to make the daily advance report by the production area.

Figure 7 shows the comparison of times between project D and the projects analyzed previously. Regarding the collection of information, it is appreciated that the time spent is the same in the four projects. However, the time taken to prepare the advance control report decreases significantly, from an approximate time of 1.3–0.3 h.

Finally, Table 3 shows the times that the production area invests in making its report by the traditional and the proposed method.

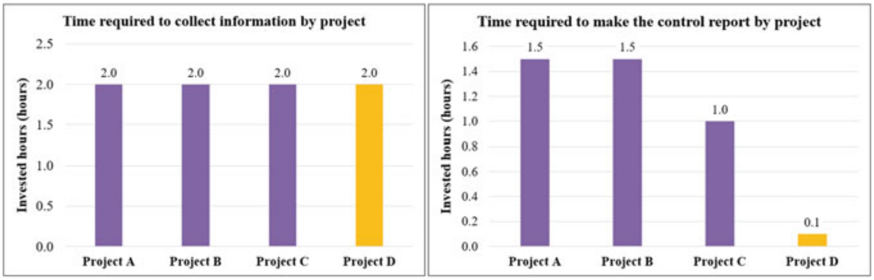


Fig. 7 Times for the realization of the reports of the production area by project

Table 3 Comparative chart of the times invested in the realization of the daily advance report by the traditional and the proposed method

Activities	Traditional method	Proposed method
Information collection	2.0	2.0
Make the control report	1.3	0.1
Total (h)	3.3	2.3

Table 3 shows that for an 8-hour workday, it has a reduction of 3.3–2.3 h, which is equivalent to a 30% time saving in a workday. This happens because, in the traditional method, the collection of information and the realization of the report are isolated and sequential activities. However, in the proposed method, these two activities are carried out in parallel.

4 Conclusions

The proposal mainly focuses on optimizing the time spent monitoring the progress of construction in residential buildings. Currently, there is no data acquisition tool for production personnel, so an existing tool was adapted to allow daily progress reports to measure the progress of construction projects. The time saving of the production team for a working day is 30%, which is one month equals 30 working hours. As the project structure stage lasts 10 months, a total of 300 h would be saved for a project. For future research and taking into account that the proposed methodology can be implemented in other project areas such as planning and quality, the optimization would save 750 h throughout the project.

References

1. You, Z., Wu, C.: Advanced engineering informatics a framework for data-driven informatization of the construction company. Adv. Eng. Inform. **39**(February), 269–277 (2019)
2. Behnam, A., Wickramasinghe, D.C., Ghaffar, M.A.A., Vu, T.T., Tang, Y.H., Isa, H.B.M.: Automated progress monitoring system for linear infrastructure projects using satellite remote sensing. Autom. Constr. **68**, 114–127 (2016)
3. Zhu, Z., Ren, X., Chen, Z.: Integrated detection and tracking of workforce and equipment from construction jobsite videos. Autom. Constr. **81**(June 2015), 161–171 (2017)
4. Pučko, Z., Šuman, N., Rebolj, D.: Automated continuous construction progress monitoring using multiple workplace real time 3D scans. Adv. Eng. Inform. **38**(October 2017), 27–40 (2018)
5. Omar, H., Mahdjoubi, L., Kheder, G.: Computers in industry towards an automated photogrammetry-based approach for monitoring and controlling construction site activities. Comput. Ind. **98**, 172–182 (2018)
6. Matthews, J., Love, P.E.D., Heinemann, S., Chandler, R., Rumsey, C., Olatunj, O.: Real time progress management: Re-engineering processes for cloud-based BIM in construction. Autom. Constr. **58**, 38–47 (2015)
7. Zhong, B., et al.: Automation in construction mapping computer vision research in construction : developments, knowledge gaps and implications for research. Autom. Constr., 107(July), 102919 (2019)
8. Seok, L., Seung, H., Seok, H., Kim, S.: Automation in construction managing construction schedule by telepresence: integration of site video feed with an active nD CAD simulation. Autom. Constr. **68**, 32–43 (2016)
9. Ding, Z., Liu, S., Liao, L., Zhang, L.: Automation in construction a digital construction framework integrating building information modeling and reverse engineering technologies for renovation projects. Autom. Constr. 102(December 2018), 45–58 (2019)
10. Xu, Q., Chong, H., Liao, P.: Automation in construction collaborative information integration for construction safety monitoring. Autom. Constr. **102**(January), 120–134 (2019)

Daily Control System for the Construction of Multi-family Building Projects

Diego Fernandez, Yeffry Zamora, Sandra Rodriguez, and Jorge Vilcapoma

Abstract Construction projects for buildings, in the execution stage, show a variation of the real with respect to what is planned in the budget. The consequence of this lag presented usually causes large economic losses. Given this situation, in the construction industry, various control systems have been developed that have as a final deliverable a frequent management report, which serves to make decisions regarding the analyzed, whose results are applied and evaluated after a specific period of time. These management reports have good results at the time of their application; however, data processing, analysis execution, and decision making take approximately 30 days. In this sense, the present investigation proposes to develop a control system focused in real time, generated by the daily control in the execution stage. The results show a reduction of waste in the main resources of multi-family building projects, and this is in labor and materials, identifying the causes that originate them and suggesting generic solutions, at the end of the workday.

Keywords Building projects · Management · Cost reduction variability

D. Fernandez · Y. Zamora · S. Rodriguez (✉) · J. Vilcapoma
Peruvian University of Applied Sciences, Av. Prolongación Primavera 2390, Santiago de Surco, Lima, Peru
e-mail: sandra.rodriguez@upc.edu.pe

D. Fernandez
e-mail: u201412203@upc.edu.pe

Y. Zamora
e-mail: u201313013@upc.edu.pe

J. Vilcapoma
e-mail: pccijvil@upc.edu.pe

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_59

1 Introduction

Nowadays, one of the most frequent problems in the construction sector is the unpredictable of the real construction plan against the scheduled plan, and this causes an increase in the budget, which has negative effects on construction projects for buildings that result in economic losses assumed by the contractor. Therefore, it is essential that these types of projects have a control system during the execution stage; that is, to achieve an adjusted project execution, based on identifying and formulating recommendations to eliminate or reduce waste from the root [1], in order to control the waste of resources that have a greater cost influence on the budgets identified as materials and labor as well as find the main causes of these losses. There are several reasons that cause loss of productivity; an agent that always intervenes in construction projects are interruptions at the time of execution, which vary the continuous workflow, causing an over-cost [2], and these losses of efficiencies occur due to deviations from normal work, which develop in different phases of a project [3]. Subsequently, having generic solutions facilitate the reduction of construction waste in the shortest time.

Liu et al. [4] conducted a study that aimed to identify the operating systems used by workers in the construction sector in the works, identify barriers for which technology is not applied in building projects, and by lastly, determine the popularity of the existing applications that are currently applied in constructions. The purpose of the mentioned study is to identify the operational processes in different types of construction, in order to be able to define in which stages control operating systems are applied, as well as the reason why these systems are not applied in other processes of the construction sector. This flow of information is very important in the present study in order to define the aspects to be controlled by the proposed control system.

The result is that the main system that workers use is iPphone's iOS, with respect to the barriers to the use of technology, and six factors were identified, the main one being the economic one and PlanGrid is one of the most applied technological applications in the cloud of construction projects. It is known that there are control systems and Building Information Modeling applications that are executed in construction projects with the purpose of modeling and representing in a virtual way everything planned to have a better vision in the execution stage to increase productivity and reduce waste, which are responsible for improving estimates of yields, man-hours, and cost [5]; however, the final deliverable is wide because they are controlled by specialty that house a large number of items. It is taken into account in the construction industry that there are specific objectives in the execution stage, mainly in the aspects of construction at the time of execution such as taking care of the correct administration of resources and materials, as well as the area of production. It is precisely in these stages and activities mentioned that the use of technology must be established as a tool to improve its execution. Construction information modeling technology (BIM) is currently being actively developed in order to significantly reduce construction waste [6].

Also, in the current market of the construction industry, there is software that is in charge of helping the project manager to find the yield of the project that determines the quantitative and qualitative productivity, using the tool of the Earned Value Management indices (EVM), and this whole process is adopted in the Super Decision software; the result of which indicates that it is an effective software because it has an immediate and easy-to-use database and development [7].

On the other hand, the authors Chang and Yu [8] conducted studies on the triple restriction: scope, time, and cost. The studies were about the variances of some aspects such as the yield of workers in construction works, and the main purpose was to analyze the yields and schedules established in the projects, which resulted in different variances; it is concluded that the yields established resources which may vary before, during, and after a stage of project execution. These data were analyzed because they specify that the real yield obtained in the execution stage is variable and depends on the construction stages, so it is inferred that it is important to know data in a smaller amount of time.

One of the important aspects of the control system is also the estimated duration to complete the construction in the planning stage. Chen [9] indicates in his study that the control systems that are linked to the earned value method (EVA) estimate the duration to complete a project at the stage of execution, but not at the stage of planning, therefore in their investigations determines that to define the final time of a project it is important to take into account the following aspects: scope, equipment, communication, risk, and performance to have a reasonable estimate accuracy of the duration of a project in the planning stage. The present author indicates an approximate accuracy of 85.49–88.13%.

Finally, Sujatha Gospel Ramani and Sruthi [7] indicate that one of the important factors when carrying out the control of a project is its performance. This aspect reveals that the most effective way to control the cost of a project is through the value earned, whose performance and execution rates are necessary when it comes to being responsible for execution.

Given that the results of the previous research are still inconclusive in the development of a daily control system in multi-family building projects, the objective of this study is to reduce waste of labor and material resources through the implementation of a technological tool.

2 Methodology

The authors propose the following methodology for the research shown in Fig. 1. This proposed methodology is a logical sequence of procedures to control the waste of materials and workforce resources.

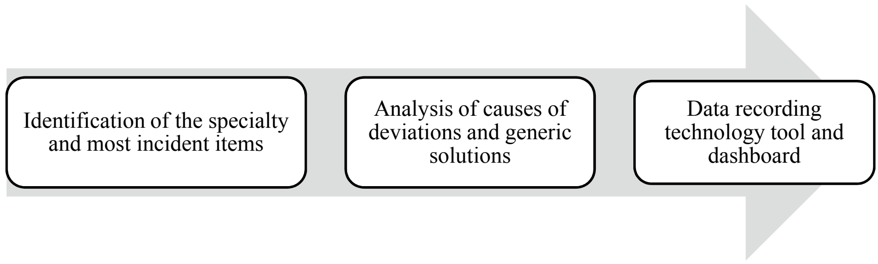


Fig. 1 Research methodology

Table 1 Characteristics of the selected projects

	Project A	Project B	Project C
Total budget (S/)	26,681,291	62,028,856	22,465,759
Time limit (days)	240	300	200
Number of floors	20	37	17
Roofed area (m^2)	16,576	33,733	13,522
R. Architecture (S/./m^2)	450	482	474

First, the project budget is analyzed in order to determine the most incident specialty in it; with this specialty, a Pareto diagram[1] is developed to identify in turn the items with the greatest influence on that specialty. In order to carry out the analysis of causes of the deviations in the selected item, a qualitative evaluation is carried out on-site, where the main causes of waste are determined both in materials and in labor. With the causes identified, two or three generic solutions are proposed based on the previous experiences validated by expert judgment.[2] With this information, causes, and generic solutions, a database is generated that is used to create a technological tool that allows the data to be recorded in the project on a daily basis and provides us with reports of real daily yields, accumulated and planned. Simultaneously, a visual organizer known as "dashboard" is generated, which takes the information from the technological tool and develops the value gained and more complete reports including projections at the end of the work.

3 Results

For the following research, building projects have been evaluated, which are for residential use. The sample evaluated for this study includes the projects in Table 1.

[1] Bar graph where the graphic values are ordered from highest to lowest.

[2] An informed opinion of people with experience, who can provide information, evidence, judgments, and assessments in a subject.

The results shown below correspond to those obtained in project B as a sample to validate the previously proposed method.

3.1 Specialty Identification and Most Incident Items

As it is known the planned budget, we proceed to make a bar chart in descending order to find the specialty that has the greatest influence on the budget. The corresponding information is detailed in Fig. 2. The following nomenclature is presented:

A	Architecture
S	Structure
E	Electrical Inst.
PW	Provisional works
M	Mechanical Inst.
SAN	Sanitation
WAF	Water against fire
C	Communication Inst.
SEC	Security

Subsequently, using the Pareto diagram tool, the identification of the most representative activities within the most influential specialty in the budget is carried out, since these will be grouped to be able to perform the control using the focused system proposed (Fig. 3).

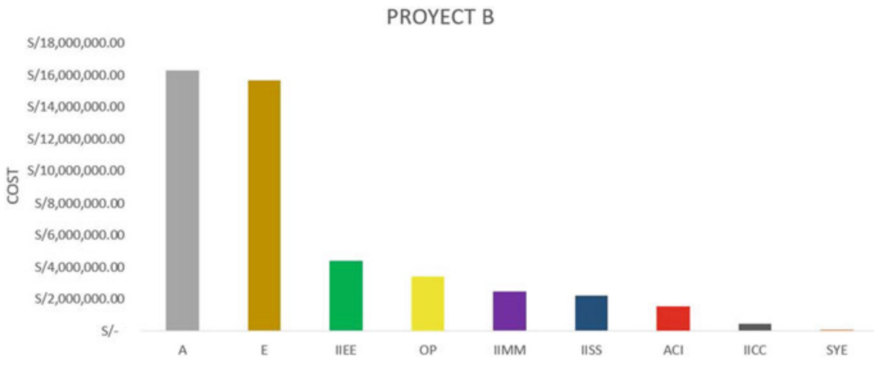


Fig. 2 Descending bar diagram of specialties in project B

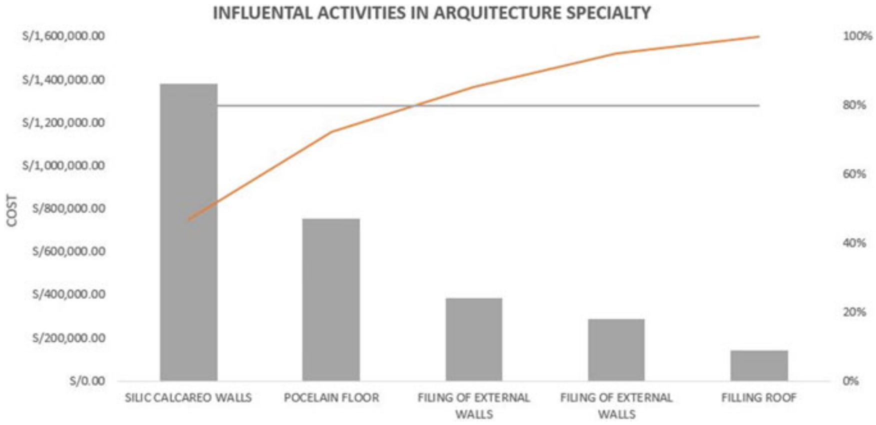


Fig. 3 Pareto diagram of influential activities in architecture specialty

3.2 Analysis of Causes of Deviations and Generic Solutions

After carrying out the qualitative evaluation on-site, and based on the previous experiences according to the experts, the generic causes and solutions for the item "Porcelain floor" identified in the previous step are presented, in Table 2 for materials and in Table 3 for labor.

In the identification of causes of waste in the resource of materials, it was determined that the main factor is the bad construction process that belongs to the execution stage, as well as the least influential factor, that of inventory, which belongs to the logistics. Likewise, for the process of identifying causes of waste of the labor resource, it is determined that one of the main factors that cause it is that of idle time.

3.3 Data Recording in the Technological Tool and Dashboard

It is important to mention that before the implementation of the methodology, and according to the measurements made in the field, the information is shown in Table 4.

The information obtained in the previous stage, causes, and solutions is included in the technological tool in order to be able to carry out the daily records, obtain statistics on the causes, propose solutions to the possible waste, and have the yield curves.

Figure 4 shows the daily record of man-hours for the "Porcelain floor" activity and the yield curves that are generated from this information, with the best workforce yield of 0.76 mh/m^2 and the lowest yield of 0.83 mh/m^2, thus having a total of 37.04 man-hours lost in the project for the week the tool is implemented. Figure 4 shows how the application organizes the information to be registered. The app is configured in Spanish, but the language can be changed.

Table 2 Causes and generic solutions in materials resource

Material		
Analysis of incidence	Cause of waste	Generic solutions
39%	Bad constructive process	Random hours inspection of the construction process
		Hire trained personnel
		Training before starting the daily wage
26%	Rework	Correct supervision of the works
		Increase the quality of finishes
		Hire trained personnel
17%	Transport	Plan a correct distribution in plant
		Decrease production batches
		Transportation schedule at the beginning of each wage
10%	Defective product	Mark all defective products and inform the supplier
		Perform quality process when receiving materials
		Identify and classify the reason for defects
8%	Inventory	Plan necessary materials
		Good warehouse distribution
		Administration and daily verification

Table 3 Causes and generic solutions in labor resource

Workforce		
Analysis of incidence	Cause of waste	Generic solutions
35%	Idle time	Identify restrictions at the beginning of the wage
		Detail daily goals at the beginning of the wage
25%	Complexity	Prior training in the initial stage of execution
		Perform an activity execution as a test
22%	Break	Organize the required breaks by the time
		Plan weekly goals detailing break times
18%	Collaboration with other areas	Request presence of more staff
		Collaboration change for break time

Table 4 Initial project information

Daily average advance (m^2/day)	71.53
Cumulative advance (m^2)	429.00
Workforce (h)	Materials (bags)
Average yield (h/m^2) 0.85	Average yield (bags/m^2) 0.31
Cumulative consumption of 366.08	Cumulative consumption of 132.99
Expected consumption of 304.59	Expected consumption of 107.25
Waste to date 61.49	Waste to date 25.74
Waste at the end of the project 1351.88	Waste at the end of the project 565.90

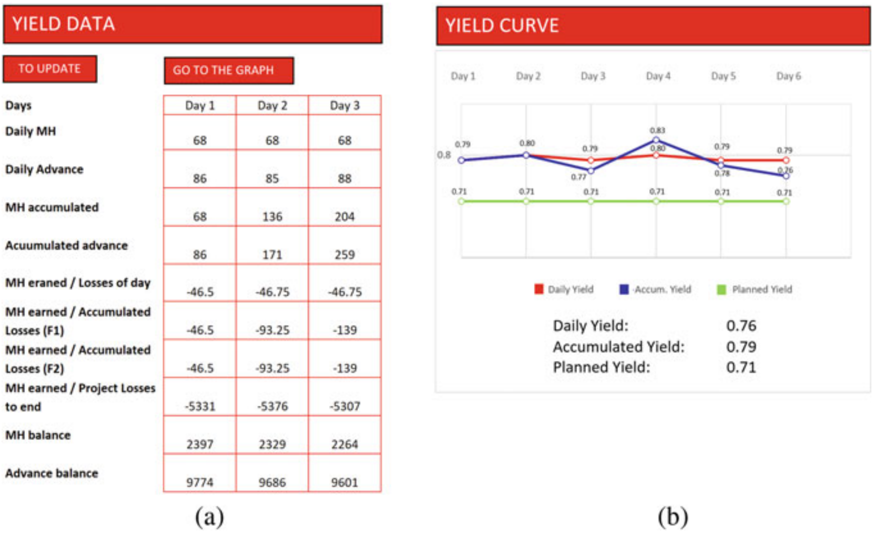


Fig. 4 **a** Data recording and **b** graph of daily, accumulated and planned yields of man-hours applying the daily control system

The information on daily work is registered after measuring the real work at the construction site. For the materials of the evaluated item, there is an average yield of 0.27 bags/m^2, finally having a glue consumption of 130 bags accumulated in the project, which makes a total of 8 bags of waste.

The technological tool also allows recording the causes that generate waste for both labor and materials as shown in Fig. 5, where the main cause of waste for the material corresponds to the "Bad construction process" for which are generated as possible solutions: random inspections and training.

To complement the results obtained from the technological tool, it is necessary to use a dashboard, where the details of the daily control through the yield curves and speedometers for the workforce and for the materials evaluated are known by means of graphics, as shown in Figs. 6 and 7.

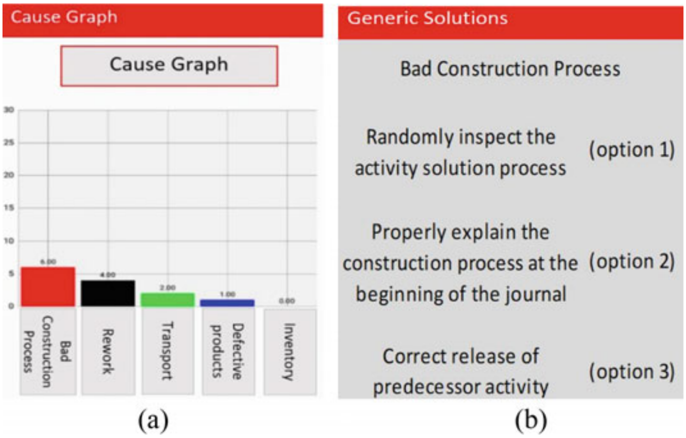


Fig. 5 **a** Causes identification and **b** evaluation of generic solutions—materials

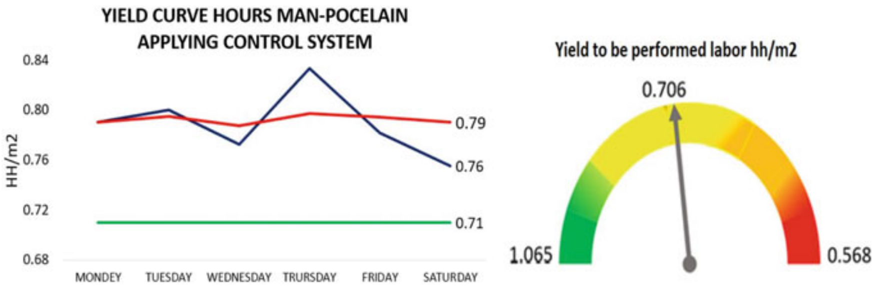


Fig. 6 Daily, accumulated and planned yield curve of man-hours using the dashboard

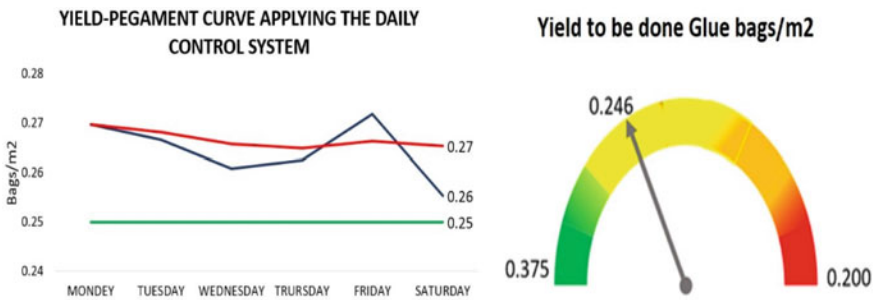


Fig. 7 Daily yield curve, accumulated and planned glue using the dashboard

Table 5 Final project information

Daily average advance (m^2/day)	77.17
Cumulative advance (m^2)	490.00
Workforce (h)	Materials (bags)
Average yield (h/m^2) 0.79	Average yield (bags/m^2) 0.27
Cumulative consumption of 386.28	Cumulative consumption of 128.85
Expected consumption of 347.90	Expected consumption of 122.50
Waste to date: 38.38	Waste to date: 7.35
Waste at the end of the project 734.04	Waste at the end of the project 140.56

When put into practice the solutions proposed from the proposed daily control system, Table 5 shows the results after having made the measurements.

4 Conclusions

The proposal raised through the use of the technological tool obtains an improvement in production yields which affects the consumption of man-hours and the consumption of the main materials. For the sample evaluated in the item "porcelain floor," there is a reduction of 0.06 h/m^2, which makes a total of 617 man-hours less used for a total of 9860.72 m^2 of the porcelain floor installed in the project. Likewise, from the material evaluated, a reduction in the consumption of glue of 0.04 bags/m^2 is obtained, which is equivalent to 425 bags of glueless waste, all this after implementing the daily control system. Finally, if we quantify the optimizations, a saving of US$3620 for man-hours and US$4225 for glue which makes a total of US$7845 only for the control carried out on the porcelain floor activity which represents the 3.4% for this activity. Taking into account the results shown previously, it is important to adapt the proposed control system to other types of projects such as roads, bridges, drainage works, and railways.

References

1. Karningsih, P.D., Anggrahini, D., Karim, M.B., Adhitama, R.: Improving project efficiency using lean construction. In: International Conference on Advanced Mechatronics, Intelligent Manufacture, and industrial Automation (ICAMIMIA), 2017
2. J.-S. Lee. Estimating cumulative damages due to discruptions in repetitive construction. J. Constr. Eng. Manag. (2016)
3. Randolph Thomas, H.: Schedule acceleration, work flow, and labor productivity. J. Constr. Eng. Manag. (2016)
4. Liu, T., Mathrani, A., Mbachu, J.: Hunting the popular Construction Apps. Computer Science and Engineering (2016)

5. Koo, B., Shin, B., Lee, G.: A cost plus estimating framework for BIM related design and engineering services. KSCE J. Civ. Eng. (2017)
6. Aleksanin, A.: Potential for the use of information systems in the management of construction waste. In: MATEC Web of Conferences, 2018
7. Ramani S.E., Sruthi, N.: Application of Earned Value Management to Compute the Project Performance Using Analytical Network Process. Eng. Manag. J. (2016)
8. Chang, C.J., Yu, S.W.: Three variance approach for updating earned value management. J. Constr. Eng. Manag. (2018)
9. Chen, H.L.: Early prediction of project duration. Eng. Manag. J. (2018)

The Role of Cyberspace and the New Information and Communication Technologies in the Improvement of the Sustainability of the Urban Habitat

Roberta Betania Ferreira Squaiella, Roberto Righi, and Maria Victoria Marchelli

Abstract This chapter discusses the emergence of cyberspace and the impact of new information and communication technologies in the configuration of cities in order to improve the sustainability of urban habitat. The population growth and the emerging concentration of people in urban centers, combined with the degradation of natural resources, have increased the concern about the environment and sustainability in the cities. From the global interconnection of computers occur the new relations of communication. The advances in new technologies are seen as promising aid for the organization of spaces/activities, aiming the rationalizing of energy consumption and other natural resources, and mitigation of negative effects of the accentuated urbanization that impact on the planet. In the urban thematic, the issues of density and mobility stand out as transformations of the landscape and the environment. Currently, the amount and speed of information exchanges, with new and ubiquitous ways of interaction, enhance and expand the space and contemporary experiences, leading to the reconfiguration of urban spaces and social practices, governed by the real and immediate time in the global world.

Keywords Cities · Cyberspace · Sustainability · Landscape · Environment

R. B. F. Squaiella (✉) · R. Righi
Mackenzie Presbyterian University–Faculty of Architecture and Urbanism, São Paulo, Brazil
e-mail: robssquaiella@gmail.com

R. Righi
e-mail: roberto.righi@mackenzie.br

M. V. Marchelli
University of Florida–Citylab Orlando, Orlando, Florida, USA
e-mail: mv.marchelli@ufl.edu

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*,
Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_60

1 Introduction

Population increase and concentration in ever-growing urban centers have increased the concerns about the environment and the sustainability of the planet and, in particular, the cities. This growth brings about many social and economic losses such as traffic jams, air, water, and soil pollution, insecurity, and many others. To minimize these issues, more sustainable solutions must be sought for urban development, possible through technology development [1].

The Internet of Things resources tend to minimize the bad effects of pronounced urbanization and reduce further impacts on the planet and the cities. Today, an important item for urban planning is a forecast on the effects caused by the mobility of people, goods, services, and information. Currently, urban mobility in metropolitan cities is done in non-sustainable ways, causing environmental problems, while changing the climate, and jeopardizing their inhabitants' quality of life. In such scenario, the new technologies applied to urban mobility show themselves as promising for helping reduce such problems [2].

On the other hand, with the invisible but growing incorporation of Information and Communication Technologies (ICT's) into various systems and structures of the urban and local life [3], new technologies represent a counterpoint and allow for decentralization, creating centrifugal forces. Current cities aim to adapt to technological innovations, as potentials to transform urban territories and cause an impact on the employment market, on the structures of the transportation networks, and as a result to the environmental quality. With technical and social changes at such a pace, new lifestyles and functions emerge, associated more with telematic networks.

The urban landscape is also transformed, since it is a result of an interaction between the urban space and human actions [4], which are governed by rising new technologies. According to Ref. [5], society is every day more combined with the city, and the latter is where quick transformations that occurred in our civilization are gathered. New ways of life arise as well as new forms of urban agglomeration due to phenomena such as financial globalization, cultural heterogeneity, debility of the industrial economy, transformation of family arrangements, and the advancing information technologies. Regardless of what the future holds for cities, it will always be important to care for the environment so that it can be ensured a better quality of life for the population. Thus, sustainability should be at the center of the discussion, which must entail, among other factors, orderly population growth, restructuring of the consumption system, preservation of natural resources, and reduction of environmental pollution.

2 Urban Habitat Sustainability

Ever since men started exploiting the natural resources and speed up their occupancy on urban environments, many profound changes have taken place in the environment.

Such changes were mainly intensified following the industrial revolution, when a large mass of people migrated into the cities looking to satisfy their needs and improve their quality of life. The accelerated, continuous migration process happened at a much higher speed than planning did, impacting on the organization of the urban space and the public utilities of cities. The result is the formation of cities without any infrastructure, failing to cope with the high population demand. In that scenario, urban sustainability arises from a need for a better quality of life for the population, through legal and planning instruments, aiming to reduce the socio-environmental impacts on cities [6].

There is no sustainable city where all of such factors are not interconnected. Indeed, approaching these is key so that the city will not collapse, as there is a certain tendency to run out of resources, throughout improper planning. Human societies often end up making disastrous decisions due to their inability to predict and realize the problem, and consequently to solve it [7]. Thinking about a city's sustainability requires approaching factors such as mobility, accessibility, technology, and densification, aiming to improve its inhabitants' quality of life. The fast-paced growth of cities and urban centers, in parallel to the economic-spatial decentralization of the businesses, has led to a greater distance between the living and working, resulting in longer displacements. The city, transportation, mobility, technology, and densification equation shapes the urban dynamic and complexity.

3 Habitat Development and Appropriation of Technologies

The term habitat is understood as an adaptation of the space inhabited, not only it's shaped as a domestic space, but also all the surrounding space, alluding to urban life and the living conditions of a population [8]. Habitat is held as one of the ways of appropriating space and shaping territories, while helping shape mankind's memory and giving meaning to urban intervention strategies.

Cities ever since the most ancient civilizations, have been the most effective way of conveying knowledge and obtaining information. Proximity within the urban environment and virtual space allow for a connection between cultures, enabling cooperation, joint production of knowledge, and increasing community intelligence. With technology integrating the various objects used on a daily basis, a major network is created. Objects are connected in a sensorial and smart manner and start communicating with each other, controlling and being controlled remotely, by producing and issuing information through a system known as the Internet of Things (IoT). This concept, which became more important as from the mid-2000s, relates to smart cars, houses, and cities, radiofrequency labels, and geolocation [9]. This technological evolution stands for the future of computing and communication, the development of which depends on dynamic technical innovation in so important fields such as, wireless sensors and nanotechnology. According to Ref. [10]; "In the early days of 2015, the world was reporting 25 billion devices connected to the Internet."

Successful cities have always invested in their inhabitants, increasing the wealth of human energy. There is no successful city without human capital. American theorist and urban planner Richard Florida created the concept of the creative economy and defined it as the agglomeration and concentration of talented, productive people who drive the true economic force of the cities [11]. The proximity between inhabitants boosts not only new ideas, but also productivity. ICTs are believed to have the power to catalyze the spatial and experimental transformations of cities, where the space enlarged by ICTs has gained new qualitative and collective dimensions [3].

Through network communications, cities are thought of as ways of reestablishing the public space and reinforcing community bonds, strengthening contemporary democracy, according to various collective intelligences. Throughout them, public and private spaces are redefined, with exchanges among their citizens and the occupancy of concrete spaces is promoted.

4 Structuring of Cyberspace and Sharing in the Cities

Cyberspace is defined as the new means of communication, arising from the global interconnection of computers, which is likely to become the most important infrastructure for economic production, transaction, and management [12]. The expanding communication capacity has been associated with the possibility to unlimitedly extend the human body and the forms of perceiving and experimenting space. In urban studies, researchers try to redefine the city as the preferred support for material and nonmaterial exchanges that characterize our contemporary way of life [3].

The main factor-involving cyberspace is not the consumption of information or interactive services, but rather the participation in a social collective intelligence process [12]. Exercising citizenship in the current urban space is associated with a general connection through the large communication networks, from "nomad technologies"—notebooks, tablets, and cell phones to Wi-Fi networks. It is necessary to acknowledge the establishment of a new dynamic of reconfiguring the urban space, enabling effective forms of communication and re-appropriation of the physical space, while valuing the public space, and strengthening contemporary democracy. Through geolocation resources, a set of wireless digital devices, sensors, networks and their corresponding databases, enable a dynamic exchange in the urban space. New meanings of places are created, and the social practices in the urban space are re-dimensioned [13].

The new technologies are favoring collection speed and processing of information, enabling interactivity with any part of the world and, consequently, a higher flow of knowledge. The concept of sharing, which highlights the sense of community and retrieves cooperative social practices, allows for goods, spaces, and services to be collectively used [14]. "Cyberspace organization is the result of a particular form of urban planning or architecture, a nonphysical one, the importance of which will only increase" [12]. Articulating the physical space and cyberspace aims to set off, as much as possible, the inertia, slowness, and rigidity of territories, solving the city's problems by sharing competencies, resources, and ideas.

5 Smart Cities

The concept of smart city is related to the use of new information and communication technologies to provide higher interaction between the town and its citizens. This means correlation between the technical aspects (smart technologies), human resources (smart people management), and governance (smart cooperation), to promote further accessibility and efficiency of the general infrastructure and public services, with higher commitment to their historical and cultural heritage, as well as concerns for the environment.

The two important factors driving smart cities are the increasing world population and the growing urbanization. Such factors point out to the concern about the scarcity of natural resources, which compromises the global supply to the world population, in addition to problems concerning the environment and the ongoing climate changes these days. Such scenario, places the challenge of minimizing the consumption of natural energy resources, promoting renewable energies, and reducing CO_2 emissions. The concept of a smart city reveals itself as a potential tool to efficiently manage the infrastructure and services of the city [15]. A smart city must be thought of as an interconnected organic system, with each of its subsystems interrelating: transportation, power, education, healthcare buildings, physical infrastructure, food, water resources, and public security. The consistent interconnection across those fields exceeds the simple deployment of technology applications. The better the connections are, the higher the information intelligence will be for decision making, either for governments or society [16].

In order to think about the contemporary challenge of what the urban space is, need to reflect on the action of a hybrid terrain, between the real and the virtual universe, and understand that one exists in the other and both are inseparable. To understand the contemporary reality, it is necessary to take into consideration the mobility and interdependence between technology resources, where architecture cannot be limited by the static conditions of time and place [3]. Organization of territory, defined by its borders and center, is determined by the physical or geographic proximities. Cybernaut practices tend to prioritize transverse modes of relationship and fluidity of the structures, whereas territory institutions are hierarchic and rigid. The technical possibilities of cyberspace make new forms of direct democracy, practical on a large scale [12].

Information and communication technologies change the contemporary urban territories, and in this regard, technology is believed to favor politics [4]. There is currently an infiltration of information and communication technologies in the city, that bring epistemological and sensorial changes to urban living and point to a radical change in what is understood as a city. The technical aspects of smart cities include, especially, the issue of urban mobility, understood as an important factor for the development of the cities aiming at a better quality of life.

6 Urban Mobility in the Age of Connection

The so-called digital cities are regarded as innovative because they aim to improve the quality of life and work standards, according to the current technology application paradigms [8]. In a twenty-first century city, the culture of mobility is characterized by the physical and virtual displacement of people, objects, culture, technologies, and information. Indeed, it is a great challenge for contemporary metropolitan areas to ensure quality in urban mobility. In large urban centers, the chaotic traffic leads to many negative impacts such as environmental pollution, huge financial and energetic expenditure due to the countless congestions, and worse quality of life for people. The world's car fleet is expected to grow by 3% a year on average by 2030, and 75% of the world population will be by 2050 living in urban areas [17].

Currently, urban mobility in large cities is revealed to be a non-sustainable solution, which causes problems in the environment and climate changes due to the emission of pollutants from vehicles, as well as noise, congestions, and accidents, which jeopardize the quality of life of their inhabitants. The study of new transportation technologies and traffic management systems points to some trends that smart cities will solve these problems, such as: optimizing car and people location services; installing smart access points for cars in urban centers; providing support to drivers and detecting in a cooperative manner collisions and other events; building autonomous cars waiving drivers [15].

A noteworthy challenge for urban mobility is overpopulation and the effects of climate change. The new technologies are indicated as potential tools to solve the inefficiency of the means of transport, particularly in cargo logistics [18]. With the development of cars with a wireless sensorial and communication capacity, every day new applications are developed and used by using ad hoc networks (or vehicular ad hoc networks), capable of obtaining information on the environment as well as exchanges messages with each other and/or between access points distributed all over the city. The data collected by these networks is able to provide information on the road conditions, traffic, climate, and vehicle and driver behavior [15].

7 Construction of the Technology Landscape and Urban Densification

The use of technology changes the space–time relationship in the contemporary world, and when thinking about ICT's and the urban phenomena, need to consider the perception of space and landscape and the new relations between technology and the actions over the space [3]. Today, time is not associated with time zones only, and space does not mean just a geographic distance. The human–machine interface takes place in the immaterial telematic networks, where the webs are not those of the built space only. On the other hand, with real-time electronic systems, the urban space loses its geopolitical reality and time instantaneity exceeds a physical population of space. The new "office" ceases to be the architectural apparatus and becomes a simple screen, a nodal center for the technological society [19]. Thus, with the use of new information and communication technologies, it is no longer necessary to physically go somewhere to perform many activities. The main existing limitation for this enormous potential for paradigmatic transformation to fully materialize is in fact, a cultural one.

In this new technological context, it is needed to reflect on the landscape. Currently, the observer's active participation through new technologies allows for new rhythms and dynamics in urban territories. The landscape cannot be described as a final product [19], due to the fact that it is in an ongoing process of transformation and depends on creativity, participation, and mobility of the people using the place, in either a physical and/or a virtual manner. The city's transformation may take place through a dense information and communication network, and through diversification of lifestyles supported on a spatial base of internal homogeneity [8].

With network cooperative work instruments it is possible to take part in the local and even international economics, from nearby centers or our home itself. The noteworthy benefits of this include attenuated pollution through the reduced circulation of cars and urban traffic jams, better distribution of populations over territories, and improved quality of life. The global social cost of conference calls is lower than an actual trip. A teleworking position will not need any office in the city. In spite of the growing cyberspace these days, there is still no intensive decentralization of the major urban centers, since "a study of the statistics reveals that the highest densities of cyberspace access and digital technology use coincide with the world's main scientific research, economic activity, and financial transaction centers" [12].

Cyberspace growth can increase control over economic, technological, and human networks, each day farther and bigger, as well as minimize the physical separation between interlocutors, through a new way of collectively organizing communication, sharing and valuing intelligence across simultaneously connected communities. It is evident that with the information technologies, the dimensions between physical and virtual spaces have been changing. The growth of virtual space connections and the consequential space autonomy are at the foundations of the compact city model. The latter was introduced by American mathematicians, who conducted studies in an attempt to efficiently look for better use of resources [20]. This model is relatively characterized by high density, that is, the concentration of a large number of people within a reduced space, a combined use type, efficient public transportation, resulting in an urban outline that can encourage pedestrian circulation and bicycle use.

A compact city is shaped based on the interconnection between house, work and leisure, while promoting activity concentration and making people's access to services easier by reducing distances and, as a result, decreasing displacements. The compact model, in lieu of sprawl over the urban territory, allows for smaller infrastructures, which accounts for reduced costs, less interventions and impacts on the physical space, by enabling higher proximities and contacts [21].

8 Final Considerations

Currently, sustainability and environmental concerns are emergent issues in the global world, where population growth tends to concentrate on large urban centers. New ways of life are experienced in which the expansive use of new information and communication technologies, in their potential, can be used towards reducing the impacts caused on the environment.

The evolution of the Internet of Things is expanding from the dimension of a physical means of communication and starts connecting objects and people to a virtual system. Through sensors, assembly lines, transport networks, home electric power networks, offices, and cars are connected. There is a tendency that, through technologies, products and services will be more and more used in a shared manner. In this regard, building a technology-compatible urban habit becomes a challenge, because technology must converge with the city rather than appropriate it, since the virtual environment is believed to not fully replace the real environment, but instead to interact with it.

The compact city model directly relates to introducing technologies into urban spaces, since it is based on less physical space with high densities. As a result, it is ensured better infrastructure and environmental impacts are minimized. In a virtual world, where physical barriers are removed, space takes on new meanings, and it is realized that urban sprawl needs not to be promoted, because it is the technologies that expand centrifugally and ensure connections.

Computerized resources can bring improvements into the urban space, particularly for urban mobility, because, with information, sharing is possible. A society that is more connected, active, and participative will decrease their demands and be able to mitigate their problems. Therefore, the use of technology in the urban infrastructure contributes towards the sustainable development of cities, ensuring a better future for its inhabitants.

References

1. Pavez, C.M., Detroz, D., Viana. A.P.: Cidades sustentáveis, inteligentes e inclusivas: reinvenção das cidades. Revista de Extensão e Iniciação Científica SOCIESC-REIS (2014)
2. Squaiella, R.B.F., Marchelli, M.V., Righi, R.: Tecnologias da informação e comunicação na mudança da mobilidade urbana na região Metropolitana de São Paulo/SP. Revista Nacional de Gerenciamento de Cidades **03**(15), 39–59 (2015)
3. Firmino, R., Duarte. F.: Cidade infiltrada, espaço ampliado: as tecnologias de informação e comunicação e as representações das especialidades contemporâneas. Arquitextos (2008)

4. Simmel. G.: Filosofia da paisagem. In: G. Simmel. La tragédie de la culture eu autres essais. Faculdade de Arquitetura e Urbanismo da Universidade de São Paulo, São Paulo (2013)
5. Moreno, J.: O futuro das cidades. Senac, São Paulo (2002)
6. Andrade, F.V.: (IN) Sustentabilidade urbana e impactos socioambientais: uma abordagem acerca da ocupação humana desordenada no espaço urbano. Contribuciones a las Ciencias Sociales (2013)
7. Diamond, J.: Colapso: Como as sociedades escolhem o fracasso ou sucesso, 9th edn. Editora Record Ltda, Rio de Janeiro (2013)
8. Ferreira, M.J.: A inovação nos habitats humanos e na organização das cidades. Revista CeoInova, n.10 (2004)
9. Singer. T.: Casas, Carros e Cidades Inteligentes: um estudo do enquadramento midiático da internet das coisas. Revista Geminis **1**(2), Year 4, 57–74 (2013)
10. Loes. J.: O que é a internet das coisas e como ela mudará a sua vida. Portal Terra (2015)
11. Florida, R.: The Rise of the Creative Class. And How It's Transforming Work, Leisure and Everyday Life. Basic Books, New York (2005)
12. Lévy, P.: Cibercultura, pp. 185–196. Editora 34, São Paulo (1999)
13. Lemos, A.: (org). Cibercidade: As cidades na cibercultura, pp. 19–26. Editora epapers, Rio de Janeiro (2004)
14. Santos, C.M.N.: Coworking: contribuições de um modelo de consumo colaborativo e da arquitetura corporativa para o gerenciamento das cidades. Revista Nacional de Gerenciamento de Cidades **2**(12) (2014)
15. Aquino, A.L.L., Ramos, H.S., Pereira, L.V., Frery. A.C.: Cidades Inteligentes, um Novo Paradigma da Sociedade do Conhecimento, pp. 165–178. Blucher, São Paulo (2015)
16. Strapazzon, C.L.: Convergência tecnológica nas políticas urbanas: pequenas e médias – cidades inteligentes. Revista Jurídica **22**(6), 89–108 (2011)
17. Oliveira. A.J.: Infográfico mostra futuro intermodal e inteligente da mobilidade urbana. Revista Galileu (2015)
18. Virilio, P.: O espaço crítico, 2. ed. Editora 34, São Paulo (1999)
19. Pronsato, S.A.D.: Arquitetura e paisagem projeto participativo e criação coletiva. Annablume, Fapeso, Fupam, São Paulo (2005)
20. Dantzig, G.B., Saaty, T.L.: Compact City: A Plan for a Liveable Urban Environment. W.H Freeman, San Francisco (1967)
21. Rogers, R., Gumuchdjian, P.: Cidades para um pequeno planeta, pp. 33–39. G. Gili, Barcelona (2011)

An Overview of Internet of Things Security from a Modern Perspective

Reinaldo Padilha França, Ana Carolina Borges Monteiro, Rangel Arthur, and Yuzo Iano

Abstract The number of devices connected to the Internet and among themselves is increasing. Internet of Things (IoT) is a large Web that creates interconnection of devices and objects to the Internet, making an immense sea of information and data. This concept is fundamental in business strategy; however, as security risks increase with so many "things" connected, the security challenges become immense, making these devices vulnerable to cyber-attacks. It is possible to suggest some solutions for protecting networks that are often used also to transmit information from these devices. Since IoT enables connections, information security must be treated with a focus on ensuring greater efficiency in operational processes and security of user data. Thus, this paper aims to provide an updated review of the security in the Internet of Things showing and approaching its particularities, through a concise background, categorizing and synthesizing the potential of technologies and best security practices and strategies focused on this topic.

Keywords Internet of Things security · Network security · Data security · Risk management · IoT

R. Padilha França (✉) · A. C. Borges Monteiro (✉) · Y. Iano
School of Electrical and Computer Engineering (FEEC), University of Campinas—UNICAMP, Av. Albert Einstein 400, Barão Geraldo, Campinas, SP, Brazil
e-mail: padilha@decom.fee.unicamp.br

A. C. Borges Monteiro
e-mail: monteiro@decom.fee.unicamp.br

Y. Iano
e-mail: yuzo@decom.fee.unicamp.br

R. Arthur
School of Technology (FT), University of Campinas—UNICAMP, R. Paschoal Marmo - 1888, Jardim Nova Italia, Limeira, SP, Brazil
e-mail: rangel@ft.unicamp.br

© Springer Nature Switzerland AG 2021

Y. Iano et al. (eds.), *Proceedings of the 5th Brazilian Technology Symposium*, Smart Innovation, Systems and Technologies 201,
https://doi.org/10.1007/978-3-030-57548-9_61

1 Introduction

The IoT brings many changes at a global level transforming the way people relate to the world directly impacting our lives, the environment, business, and security, since the interconnection of systems allows a more responsive, intelligent, and efficient world [1, 2].

The IoT can serve several as allies, as a means of improving quality of life and better use of resources, in addition to a more responsive world, as well as targets in attacks, related to the degree of risk in which systems are exposed increases as more devices are connected to the network [2–4].

IoT encompasses several types of accessories, electronic locks, door sensors, sensors for capturing sound and heat, that can be operated remotely, among countless other types of devices that collect and send information (Fig. 1). What is possible from this to obtain the interpretation of these data, captured and stored previously, in the same way, that it is also possible to make these objects able to dialogue with users, and so, promoting greater productivity, improvements in quality of life, and agility in processes [5, 6].

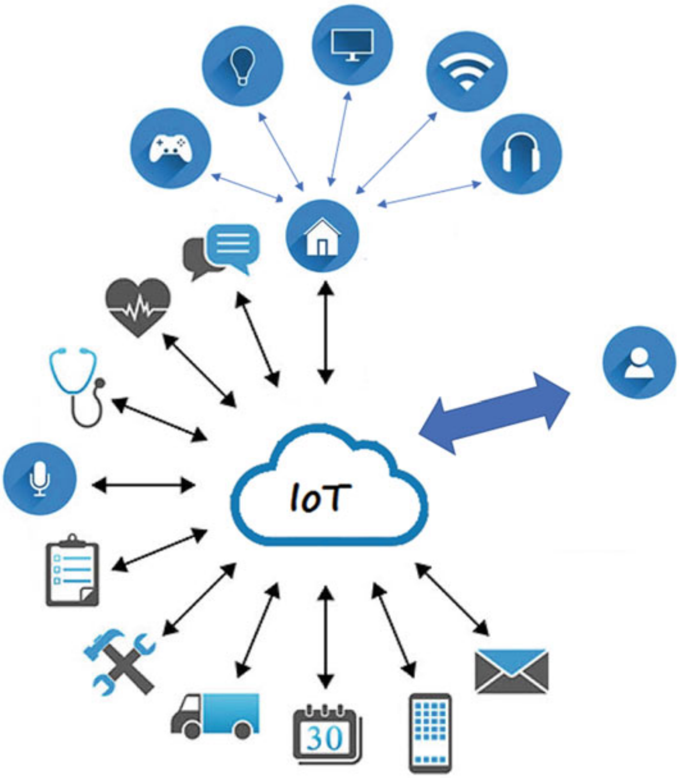


Fig. 1 Internet of Things

At the same time, the interconnected devices could present an opportunity for cybercriminals to be able to break into diverse systems in various segments, from industry to homes, if a highly efficient protection barrier is not placed in this ecosystem [7–9].

Therefore, IoT manufacturers are expected to create a mechanism to interpret the data generated and use it to their advantage when it comes to understanding the needs of this new market, offering not only more suitable products and services, but also a better experience and innovation for your consumers [10, 11].

However, to reduce production costs, in general parts are removed from the devices, which in most cases compromise their safety, since they have been able to optimize processes by reducing the number of security personnel characteristics with the function that promoted greater security, which is exposed to unnecessary risks. Some manufacturers can't even perform security updates on their products [10–12].

Besides the need to offer users new, innovative, and competitive products, the IoT device industry also needs to pay attention to the most basic security principles. What with proper application, it is necessary that manufacturers voluntarily adhere to the required safety standards [12, 13].

Another aspect of security is concerning the increasing amount of data generated by IoT devices, considering their integration with the data management systems. What both the IoT industry and consumers/users need to keep in mind that cybercriminals are always aware of the degree of exposure of these systems [10, 11, 13].

It is evident, therefore, that the risks to information security are real and can generate a series of damages to end-users. Thus, it is clear that the IoT provides a sea of opportunities that must be observed closely by all industries and markets; however, for connected things to bring more benefits, security issues must be present in all stages of implementing this technology [10–13].

Thus, this paper aims to provide an updated review of the security in the Internet of Things. Showing and approaching its particularities, through a concise background, as well as categorizing and synthesizing the potential of technologies and best security practices and strategies focused on this topic.

2 Methodology

This survey carries out a bibliographic review of the main research of scientific articles related to the theme of security in the Internet of Things, published in the last 5 years on renowned bases.

Fig. 2 IoT connectivity

3 Internet of Things (IoT)

The purpose of the Internet of Things (IoT) is for the physical and digital world to become one when using devices that communicate with each other, data centers, and their clouds. IoT is a term used to concern the set of devices that uses wireless communication interfaces to collect, send, and receive digital data. Considering all the devices that can "see, hear, and even capture movements" with high precision as well as send these records to software through the Internet, as shown in Fig. 2 [5, 6, 14, 15].

IoT is a technology allowing physical hardware devices, installed in any type of equipment to have the ability to communicate and even interact with each other. The dissemination of this technology comes from applications that depend mainly on the facility generated for users and confidence in your security, as many of these communications and actions are generated automatically [1, 3, 15, 16].

In an indoor environment, even TVs can execute voice commands due to sensors and processors that analyze spoken messages in an instant, still considering the presence of sensors acting in conjunction with thermostats that automatically set the temperature of business environments and homes, reducing cooling costs [13–16].

While the benefits of IoT technology are great, they also reveal serious information security challenges, since, over the past few years [1–3, 17, 18].

Thus, the IoT is like a great Web that interconnects devices and objects to the Internet, creating an immense sea of information and data, as well as new resources and functionalities that make it possible to improve processes and even create the business. Significantly bringing facilities to companies, allowing greater control over corporate routine and transforming business models as well as automating operational processes, reducing errors, and facilitating decision making, which is represented through this process connection, IoT makes the productivity more interactive and dynamic [13, 15, 16, 19].

4 Internet of Things (IoT) Security

Recent advances in computing have shown that inadequate antivirus compatibility and the frequent use of Universal Serial Bus (USB) have increased the attack surface across the multi-vertical IoT security landscape. What affects cyber-physical systems for industrial protection, retail, device control systems IoT among others (Fig. 3) [20, 21].

Security is critical in this transformation, since if companies and suppliers do not adopt security measures quickly, the consequences can be tragic, which include risks and cybercrime. In addition to remote attacks that even cause physical destruction, which can contaminate connected devices. Still considering to DDoS attacks and other malware that breaches sensitive data, they are also able to have access to communications on an Internet of Things network, capturing information that is used for spying, causing the captured data to be erased, exposed, or even modified, causing damage [22, 23].

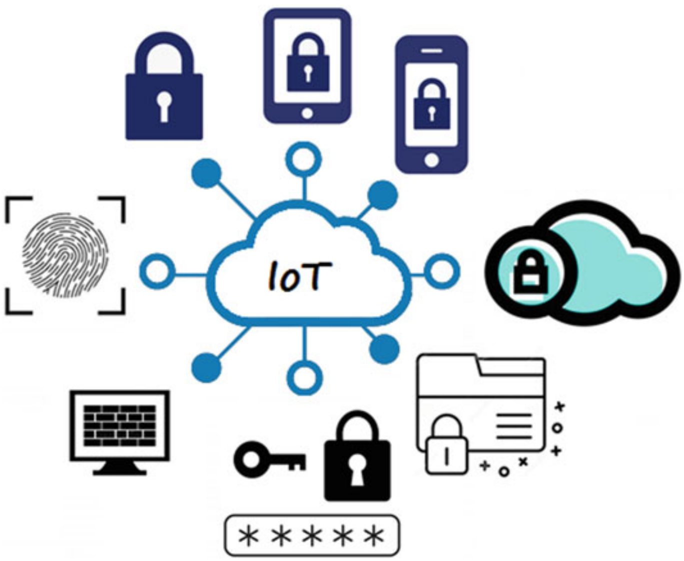


Fig. 3 Internet of Things security

The attacks in general target outdated cyber-physical systems that use the same protocols defined by the manufacturers against them, listing an IoT attack including illegally collected information, invalid data displayed in operations, operations downtime, as well as invalid and non-programmed programming sent to the controller [22–24].

The security policy must be adapted to constant changes, which IoT security solutions must guarantee the freedom and privacy of data, as well as regulating the treatment of personal information, requiring companies to adapt to the new security requirements present in the market [25, 26].

Keeping the software up to date is a recommended practice to control devices connected to IoT, reducing the risk of possible data exposure, as well as identifying possible targets of cyber-attacks by using strong authentication solutions (multiple authentication factors). Maintaining strong passwords and using multifactor authentication are means of inhibiting invalid access to the information that is available on systems connected to IoT [26–28].

So, companies that supply IoT devices need to develop a strategy based on security risks, and know how to identify, analyze, and classify them by priority, taking into account the probability and the impact they can cause to the business, creating conditions that facilitate the definition of the strategy and action plans to reduce exposure to risks. Many devices connected to IoT are accessed through the cloud, and thus, it is necessary to invest in data and media encryption [28, 29].

Through IoT devices, it is possible to traffic a considerable amount of information and valuable data; for this reason, they are generally targets of invasion. It is essential to look at security in a preventive and early manner, establishing a culture of security. So that it is possible to take advantage of the full potential of the IoT without becoming vulnerable to cybercriminals. Ensuring greater efficiency in operational processes and security of user data [26, 27, 29].

5 Results and Discussion

IoT impacts the way people interact with the world, since they are considered billions of "things" talking to each other, from refrigerators and cars, TVs, to smart meters, health monitors, and wearable devices. These devices are not only used by ordinary people, but also by companies, revolutionizing the corporate environment by changing the way businesses deal with their digital data [30–33].

IoT promises incomparable convenience; however, the huge amount of connected devices that collect personal information has become a recurring concern of companies with IoT security. It is essential to gain and maintain consumer confidence on issues such as privacy and security, reflecting on the large-scale information analysis infrastructure [34–36].

The great challenge will be to protect this information. Since the data transmitted by the IoT will form a graphical representation of each one of us in the digital communication Web, with the need to have digital security solutions and a solid policy in this regard. What with some practical measures, as previously briefly mentioned, it is possible to make the infrastructure more reliable, dealing with the different digital threats [37–39].

IoT's mission is to facilitate people's daily lives by increasing the efficiency and productivity of companies and employees, where the collected data helps to make smarter decisions. The main objective of using IoT in the corporate environment is to increase the level of business automation and expand the amount of information available for business analysis. So, IoT gadgets need to capture data continuously, which may involve collecting private records. This also has an impact on privacy expectations, which correlates in cases of compromised data collected by these devices. What results in damage users' confidence in the companies that use this data.

One of the biggest enemies of cybersecurity is overconfidence, regardless of IoT devices use, complexity, or degree. They collect private information about user behavior in areas such as education, finance, and health, among others.

IoT is also synonymous with investment in cloud computing, being one of the environments that have the greatest investment in security, which from the moment that smart devices are connected, it is necessary to have a safe way where this information will flow.

Being linked to the fact that there are many ways for an attacker to access resources or data on a connected device, considering three general targets. The main one of illegal access is the *device*. Since billions of connected devices will increase the use of software and data on business assets and consumer gadgets, meaning new points of attack. The *cloud infrastructure*, concerning the most insistent forms of threat, which these smart devices are connected, should be considered solutions such as data encryption and the security of the cloud structure offering comprehensive portfolio service providers and organizations to protect business assets. And the *network*, reflecting that across the spectrum of the cloud, lifecycle management of security components on devices is a critical element in a robust and long-term digital security strategy. Thus, the insertion of new devices, the integrating devices with a new cloud ecosystem or vice versa, disabling devices at the end of their lives such as controlling secure downloads of firmware and software are activities that need comprehensive management of identities, keys, and tokens [11, 40, 41].

Security is not an isolated activity, but a dynamic part of the IoT ecosystem. One of the biggest problems of IoT concerns insecurity about access interfaces. Since it is already known that IP cameras, smart TVs, and other devices can be easily hacked, whether for information theft or DDoS attacks, the possibility of accessing and managing all the physical components of an IoT device, or confidential information in the cloud, such as physical location or movements, represents a great risk to data privacy and even to the user's security [42–44].

The main challenge of ensuring the security of IoT systems is that some devices have limited resources and, therefore, cannot perform traditional security functions, with many of the IoT devices not being designed with security, as they are designed with the primary purpose to provide functionality at a low cost. In the same way, that there is no standard protocol, both for the security of IoT devices and their communication. What ends up opening loopholes for the infection by malware, for example, which infects the network.

Still considering that if the security structure is inadequately configured, IoT devices can represent a major problem for companies that need to keep their data safe. So, the use of complex passwords or the creation of rules for accessing the system ends up becoming safer entry doors to the corporate network, creating more security, and privacy.

In this sense, it is possible to highlight measures that must be taken to maintain a secure infrastructure concerning IoT devices. Monitoring the entire network infrastructure is to search for failures and intruders. The creation of a hidden Wi-Fi network is only for IoT equipment to isolate them from the rest of the infrastructure. The adoption of safer standards is to protect wireless networks and the agile update of the firmware of each gadget. The definition of device access and control rules is to delimit the number of sensitive areas of the infrastructure that they can access. The implementation of a firewall system is with the adoption of data encryption to store information collected by devices and exchange data between devices securely.

These measures can mitigate a large part of today's virtual attacks. Taking these precautions, the network starts to have a high level of reliability, since all connections are dynamically monitored. Taking into account, the equipment will be isolated against vulnerabilities of the devices of the IoT. Even if attacks are carried out, a determined company will not have its data exposed.

Still, making it impossible that external agents may have the ability to detect IoT devices connected to a network. In the same sense, they do not have the visibility to know which of these devices are communicating. This prevents certain IoT "blind spots" from being able to be explored.

An example of breach exploitation is related in the context of smart home and its smart appliances. In particular, a smart refrigerator with the ability to order food at the local supermarket containing bank details of the homeowner stored on the refrigerator network, which in turn, an attack focusing on this device may be carried out to access these bank details.

In another context, in contrast to the above, with human-controlled devices, gateways go through an unique authentication process.

Thus, the correct security strategy helps to protect both personal and company data and maintaining the user's trust in IoT devices, as well as in the security of the environment in which they are installed. Finally, the right security solutions for IoT can help a company to activate new business models and save costs.

6 Cybersecurity Education

Cybersecurity education is also a key factor within the context of IoT security that helps to provide safer technological environments.

The use of technology exposes people to a large number of digital threats and education is the main tool to be prevented. Since exposure to cyber risk is part of the perception of security and reliability that affects the relative advantage of IoT products, which requires education for safety and sensitivity to the problem [45–47].

Formal education is fundamental in the process of insertion of new generations in society. Currently, devices become increasingly part of human daily life, toward a scenario of smart cities. What starts with the requirement for the existence of a safety standard that comes from the factory, from the refrigerator with Wi-Fi access to the computer that works as a personal assistant. Ditto for IoT devices, operating by the software of physical devices through network, appliances, vehicles, among many others. Even the owner of a smart TV, a smart refrigerator or a smart car, who puts his data on these devices with the necessary safety precautions [46–48].

In this sense, the educational system of each country should seek to provide citizens not only with basic academic knowledge, but also to help them develop the skills to face digital life. Taking into account, the challenges of each age on issues were related to security on the Internet since young people are often unaware of the risks involved in that excessive sharing of photographs and the publication of confidential information on social media. Still considering bad digital habits can be the gateway to an internal domestic network that may contaminate devices smart provoking phishing, sexting, and even cyberbullying [46–48].

The proliferation of bad cyber-habits creates the need for educational institutions with basic computer classes to include aspects of good cybersecurity practices. Since children are often taught about earthquake prevention, however, they must also be educated about responsible use of information technologies. Considering online behavior, seeking to instruct the importance of security education is concerning privacy and security, regarding various aspects of data security, protection, encryption, and prevention of identity and information theft and Web-based cyber-attacks [46–48].

Just as many people use technology without thinking too much about the risks of having their personal information stolen, until they or their relatives become victims of cybercrime. Related to the lack of insight into the importance of cybersecurity education for the citizens of the world current, what extends from password security, avoiding email phishing and malicious downloading including fraudsters posing as companies, sending fraudulent emails containing viruses and deploying malware, suspicious emails. Or mobile devices concerning trojans, viruses, spyware, adware. Even making use of the smartphone or IoT device on open Wi-Fi and may become victims of espionage attacks, data tampering, intermediary attack, and other intrusion methods, resulting from the device's discredit as well as financial losses [47–49].

Although threats in cybersecurity education can be divided into two large groups, the logic of the *computer* recognizes threats well and can make decisions about blocking them autonomously. Reflecting on antivirus protection, traffic filtering, protection against attacks DDOS, which are tasks that specialized solutions deal with automatically, and the logic *human*, for which there is no fully automated solution. In general, these threats are treated in an integrated manner by configuring access to the components of the IT infrastructure, ranging from the level of technical training of the employees of a company. So, that information confidential is sometimes found in smartphone notes, even in personal emails, which inadvertently end up in the hands of malicious third parties. It is important to monitor the movement of confidential data that can be transmitted outside the “perimeter” of a company, preventing most information leaks as well as resulting in loss of assets and intellectual property, reaching companies of any size, caused by the human factor [47–49].

Continuing cybersecurity training is beneficial to the business as a whole, even with the best professionals and technologies. The weakest link in an organization, when it comes to security, is usually its employees. What together with the IoT, it can be leaving a network vulnerable because people are using devices and connections that are not under a security posture [47–49].

Thus, proactive investment in ongoing cybersecurity training will protect the business in the future. Since attackers employ increasingly sophisticated techniques, intelligently leveraging innocent employees, such as through the use of social engineering, concerning security training cybernetics is not only an advantage, but a necessary investment [47–49].

As the threats increase, the number of people and professionals educated in cybersecurity must also increase, helping to reduce this deficiency. It is essential to have professionals with cybersecurity skills, to help organizations reduce vulnerabilities, damage, and recovery time in the event of an attack [47, 48, 50].

For consumers, the smart devices follow the concept of easily and highly adaptable, which are convenient and useful. However, without the minimum cybersecurity setup, personal or corporate information can be exposed or an attacker can take advantage of it. Still considering that many people remain skeptical or unconcerned about this level of cybersecurity. This makes them not worry about security until they are victims of an attack [49, 50].

Considering the relative infancy of many IoT markets and IoT companies, and the growing demand for cheap and affordable IoT devices, education and awareness of industrial cybersecurity are critical; it is up to manufacturers to make decisions about how best to deal with cybersecurity [48–51].

For some IoT developers, this may mean the choice between product usability and product security, where cybersecurity is often seen as a cost and not an investment. Even though these products and services are still placed on the market, there is a whole defensive analysis, where companies pay to discover flaws and exploit them. This factor does not rule out the need for safety education, digital education is needed from the first steps, as well as must be provided by the industry through products with clearer instructions and safety manuals for better use of the device [48–51].

However, it is necessary to understand the concept between education and training. Because education offers learning and understanding of the subject, while training makes employees effective by performing essential functions [48–51].

7 Trends

IoT applications require enhanced security, since the sensors and integrated devices transmit information between themselves and over the internet. In this scenario, it is important to take into account blockchain technology to ensure the protection of these communications. Preventing IoT devices from being compromised cyber-attacks and user behavior patterns are revealed [52, 53].

The blockchain is a technology of storage and transmission of information, safe and transparent, working without central control. Since it is not necessary a "master" computer that contains the entire data chain. It is capable of allowing the exchange of information, without a central server validating requests, even considering that all participants have a copy of it. So, this decentralization adds more security, since all network participants need to reach a consensus to effect the message transactions. That is, the users themselves participate in the validation of all transactions on the network [53, 54].

Using blockchain and IoT together, it becomes possible to track billions of connected devices, process transactions, and coordinate communication securely. Blockchain offers a standardized method to accelerate data exchange, allowing processes to run between IoT devices without intermediaries and safely [54, 55].

The blockchain principles that make the IoT more secure are the possibility of anonymity. The security features are reinforced, and the decentralization of transactions is carried out. Provided that in a distributed blockchain IoT network, the devices are in a point-to-mesh network point. Executing and authenticating transactions based on predetermined rules, without the need for a central server, maintaining an immutable record. This allows the autonomous operation of smart devices without the need for centralized authority. This results in the blockchain solution which opens the door to a series of IoT scenarios that were notably difficult or even impossible to implement [54–56].

By combining blockchain and IoT, companies can reduce challenges through innovative processes. Considering the cryptography and decentralized control offered by the blockchain as highly effective alternatives to replace traditional security mechanisms. This distributed architecture helps in the resolution of many of the problems caused by the increased use of IoT devices. Since the data are protected, where the participants (network "nodes") are responsible for the validation and registration of the transactions. What does not allow them to be falsified or altered in the performance of a malicious agent, successfully controlling all nodes [55, 57, 58].

For even though blockchain technology may be counterintuitive for understanding at a basic level. It is probably best understood as a kind of distributed book that monitors various transactions, where each "block" in that chain contains transactional records or other data to be protected against tampering which are linked to the previous one by a cryptographic hash. This means that any tampering with the block will invalidate that connection, allowing for distributed inventory and automated audits [57, 58].

8 Conclusions

The IoT is gradually revolutionizing the way companies deal with their processes and their market planning routines. Since businesses not only start to have services targeted according to market demands, but their capacity to optimize processes is maximized.

Insecure communication is not the only way in which attackers can collect personal information about users. Considering that all data is transferred via the cloud and the hosted services can also suffer external attacks. This is related to the possibility of data leaks and both about the communication between each gateway of own devices and in the cloud storage location of the data.

Thus, an awareness of issues related to device security and the potential implications of any type of vulnerabilities being exploited should be created, with a focus on the significant results that this will affect concerning consumer purchase decisions who will see that the online environment they are in is not safe.

Just as it is necessary to equip smart devices with the right security solutions, companies, as well as factories in Industry 4.0. Considering the gateways that connect IoT devices to networks of companies and manufacturers need to be protected, from the beginning of each communication process.

Just as manufacturers need to have a process to correct errors efficiently and safely, they need also take necessary measures to properly address known vulnerabilities which if not, consumers will lose confidence in the IoT devices themselves. It is the duty of managers and the information technology team to identify where their company may be vulnerable. Once they are identified, they will understand how security should be incorporated into the entire product design process. From coding, testing, to its final assessment, is having a clearer view of cyber threats.

References

1. Wortmann, F., Flüchter, K.: Internet of things. Bus. Inf. Syst. Eng. **57**(3), 221–224 (2015)
2. Madakam, S., et al.: Internet of Things (IoT): a literature review. J. Comput. Commun. **3**(05), 164 (2015)
3. Nord, J.H., Koohang, A., Paliszkiewicz, J.: The Internet of Things: review and theoretical framework. Expert Syst. Appl. (2019)
4. Lee, I., Lee, K.: The Internet of Things (IoT): applications, investments, and challenges for enterprises. Bus. Horiz. **58**(4), 431–440 (2015)
5. Khan, W.Z., et al.: Industrial internet of things: recent advances, enabling technologies and open challenges. Comput. Electr. Eng. **81**, 106522 (2020)
6. Farooq, M.U., et al.: A review on internet of things (IoT). Int. J. Comput. Appl. **113**(1), 1–7 (2015)
7. Radanliev, P., et al.: Design Principles for Cyber Risk Impact Assessment from Internet of Things (IoT). University of Oxford (2019)
8. Krishnan, D., Adesh, M.: A survey on security attacks in Internet of things and challenges in existing countermeasures. In: Proceedings of International Conference on Wireless Communication. Springer, Singapore, 2020

9. Dorri, A., et al.: Blockchain for IoT security and privacy: the case study of a smart home. In: 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops). IEEE, 2017
10. Iureva, R.A., et al.: Measures to Design Secure Cyber-Physical Things. Intelligent Decision Technologies 2019. Springer, Singapore, 2020. 315–322
11. Mahmoud, R., et al.: Internet of things (IoT) security: current status, challenges and prospective measures. In: 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST). IEEE, 2015
12. Liyanage, M., et al.: IoT Security: Advances in Authentication. Wiley (2020)
13. Hwang, Y.H.: IoT security & privacy: threats and challenges. In: Proceedings of the 1st ACM Workshop on IoT Privacy, Trust, and Security. 2015
14. Russell, B.: IoT Cyber Security. Intelligent Internet of Things, pp 473–512. Springer, Cham (2020)
15. de Matos, E., et al.: Context information sharing for the Internet of Things: a survey. Comput. Netw. **166**, 106988 (2020)
16. Dehkordi, S.A., et al.: A survey on data aggregation techniques in IoT sensor networks. Wirel. Netw. **26**(2), 1243–1263 (2020)
17. Li, S., Li, D.X., Shanshan, Z.: The internet of things: a survey. Inf. Syst. Front. **17**(2), 243–259 (2015)
18. Al-Emran, M., Sohail, I.M., Al-Kab M.N.: A survey of Internet of Things (IoT) in education: opportunities and challenges. In: Toward Social Internet of Things (SIoT): Enabling Technologies, Architectures and Applications, pp. 197–209. Springer, Cham, 2020
19. Neshenko, N., et al.: Demystifying IoT security: an exhaustive survey on IoT vulnerabilities and a first empirical look on internet-scale IoT exploitations. IEEE Commun. Surv. Tutor. **21**(3), 2702–2733 (2019)
20. Miettinen, M., et al.: IoT sentinel: automated device-type identification for security enforcement in IoT. In: 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2017
21. Wangyal, S., et al.: A preliminary study of multi-viewpoint risk assessment of IoT. Bull. Netw., Comput., Syst., Softw. **9**(1), 40–42 (2020)
22. Md Mahmud, H., Maziar, F., Hasan, R.: Towards an analysis of security issues, challenges, and open problems in the internet of things. In: 2015 IEEE World Congress on Services. IEEE, 2015
23. Shadab, A., et al.: Internet of Things (IoT) enabling technologies, requirements, and security challenges. In: Advances in Data and Information Sciences, pp. 119–126. Springer, Singapore, 2020
24. Wurm, J., et al.: Security analysis on consumer and industrial IoT devices. In: 2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC). IEEE, 2016
25. Dhanda, S.S., Brahmjit, S., Poonam, J.: IoT security: a comprehensive view. In: Principles of Internet of Things (IoT) Ecosystem: Insight Paradigm, pp. 467–494 Springer, Cham (2020)
26. Sivaraman, V., et al.: Network-level security and privacy control for smart-home IoT devices. In: 2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). IEEE, 2015
27. Wang, P., et al.: The Internet of Things: a security point of view. Internet Res. (2016)
28. Nawir, M., et al.: Internet of Things (IoT): taxonomy of security attacks. In: 2016 3rd International Conference on Electronic Design (ICED). IEEE, 2016
29. Viswanath, S.K., et al.: System design of the internet of things for residential smart grid. IEEE Wirel. Commun. **23**(5), 90–98 (2016)
30. Zhang, L., et al.: Research of the key technologies of the smart home based on IoT. In: Applied Mechanics and Materials, vol. 713. Trans Tech Publications Ltd. (2015)
31. Kim, J.: HEMS (home energy management system) base on the IoT smart home. Contemp. Eng. Sci. **9**(1), 21–28 (2016)
32. Park, E., et al.: Comprehensive approaches to user acceptance of Internet of Things in a smart home environment. IEEE Internet of Things J. **4**(6), 2342–2350 (2017)

33. Singh, S., Nirmala, S.: Internet of Things (IoT): security challenges, business opportunities & reference architecture for E-commerce. In: 2015 International Conference on Green Computing and Internet of Things (ICGCIoT). IEEE, 2015
34. Asplund, M., Nadjm-Tehrani, S.: Attitudes and perceptions of IoT security in critical societal services. IEEE Access **4**, 2130–2138 (2016)
35. Atlam, H.F., Gary, B.W.: IoT Security, privacy, safety and ethics. In: Digital Twin Technologies and Smart Cities, pp 123–149. Springer, Cham (2020)
36. Ioannis, A., Chrysostomou, C., Hadjichristofi, G.: Internet of Things: security vulnerabilities and challenges. In: 2015 IEEE Symposium on Computers and Communication (ISCC). IEEE, 2015
37. Pacheco, J., Hariri, S.: IoT security framework for smart cyber infrastructures. In: 2016 IEEE 1st International Workshops on Foundations and Applications of Self* Systems (FAS* W). IEEE, 2016
38. Alfaqih, T.M., Al-Muhtadi, J.: Internet of Things Security based on devices architecture. Int. J. Comput. Appl. **975**, 8887 (2016)
39. Sheng-Lung, P., Pal, S., Huang, L.: Principles of Internet of Things (IoT) Ecosystem: Insight Paradigm. Springer (2020)
40. Das, R., Gündüz, M.Z.: Analysis of cyber-attacks in IoT-based critical infrastructures. Int. J. Inf. Secur. Sci. **8**(4), 122–133 (2020)
41. Kolias, C., et al.: DDoS in the IoT: Mirai and other botnets. Computer **50**(7), 80–84 (2017)
42. Bhardwaj, K., Miranda, J.C., Gavrilovska, A.: Towards IoT-DDoS prevention using edge computing. In: {USENIX} Workshop on Hot Topics in Edge Computing (HotEdge 18). 2018
43. Zhang, C., Green, R.: Communication security in internet of thing: preventive measure and avoid DDoS attack over IoT network. In: Proceedings of the 18th Symposium on Communications & Networking. 2015
44. Radanliev, P., et al.: Future developments in standardisation of cyber risk in the Internet of Things (IoT). SN Appl. Sci. **2**(2), 169 (2020)
45. Lenhart, P., et al.: Challenges in integrating cybersecurity into existing development processes. No. 2020–01-0144. SAE Technical Paper, 2020
46. Siegel, C., Sweeney, M.: Cyber Strategy: Risk-Driven Security and Resiliency (2020)
47. Venkatesan, R., Nagaveni, B.R.: Cyber security in it modern era. Our Herit. **68**(54), 1331–1336 (2020)
48. Langer, A.M.: Cyber Security in Analysis and Design. Analysis and Design of Next-Generation Software Architectures, pp 181–199. Springer, Cham (2020)
49. Humayun, M., et al.: Cyber security threats and vulnerabilities: a systematic mapping study. Arab. J. Sci. Eng. 1–19 (2020)
50. Neave, N., et al.: Digital hoarding behaviours: implications for cybersecurity. In: Cyber Influence and Cognitive Threats. 77–95, Academic Press (2020)
51. Khan, M.A., Khaled, S.: IoT security: review, blockchain solutions, and open challenges. Future Gener. Comput. Syst. **82**, 395–411 (2018)
52. Kshetri, N.: Can blockchain strengthen the internet of things? IT Prof. **19**(4), 68–72 (2017)
53. Banerjee, M., Lee, M., Kim-Kwang, R.C.: A blockchain future for internet of things security: a position paper. Digit. Commun. Netw. **4**(3), 149–160 (2018)
54. Minoli, D., Occhiogrosso, B.: Blockchain mechanisms for IoT security. Internet of Things **1**, 1–13 (2018)
55. Dorri, A., Salil, S. K., Raja, J.: Towards an optimized blockchain for IoT. In: 2017 IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI). IEEE, 2017
56. Huh, S., Cho, S., Soohyung, K.: Managing IoT devices using blockchain platform. In: 2017 19th international conference on advanced communication technology (ICACT). IEEE, 2017
57. Yu, Y., et al.: Blockchain-based solutions to security and privacy issues in the Internet of Things. IEEE Wirel. Commun. **25**(6), 12–18 (2018)
58. Singh, M., Abhiraj, S., Shiho, K.: Blockchain: A game changer for securing IoT data. In: 2018 IEEE 4th World Forum on Internet of Things (WF-IoT). IEEE, 2018