IJOER
RESEARCH JOURNAL

International Journal of Engineering Research & Science

ISSN
2395-6992

www.ijoer.com
www.adpublications.org

Volume-7! Issue-6! June, 2021 www.ijoer.com ! info@ijoer.com

Preface

We would like to present, with great pleasure, the inaugural volume-7, Issue-6, June 2021, of a scholarly journal, *International Journal of Engineering Research & Science*. This journal is part of the AD Publications series *in the field of Engineering, Mathematics, Physics, Chemistry and science Research Development*, and is devoted to the gamut of Engineering and Science issues, from theoretical aspects to application-dependent studies and the validation of emerging technologies.

This journal was envisioned and founded to represent the growing needs of Engineering and Science as an emerging and increasingly vital field, now widely recognized as an integral part of scientific and technical investigations. Its mission is to become a voice of the Engineering and Science community, addressing researchers and practitioners in below areas

Chemical Engineering	
Biomolecular Engineering	Materials Engineering
Molecular Engineering	Process Engineering
Corrosion Engineering	
Civil Engineering	
Environmental Engineering	Geotechnical Engineering
Structural Engineering	Mining Engineering
Transport Engineering	Water resources Engineering
Electrical Engineering	
Power System Engineering	Optical Engineering
Mechanical Engineering	
Acoustical Engineering	Manufacturing Engineering
Optomechanical Engineering	Thermal Engineering
Power plant Engineering	Energy Engineering
Sports Engineering	Vehicle Engineering
Software Engineering	
Computer-aided Engineering	Cryptographic Engineering
Teletraffic Engineering	Web Engineering
System Engineering	
Mathematics	
Arithmetic	Algebra
Number theory	Field theory and polynomials
Analysis	Combinatorics
Geometry and topology	Topology
Probability and Statistics	Computational Science
Physical Science	Operational Research
Physics	
Nuclear and particle physics	Atomic, molecular, and optical physics
Condensed matter physics	Astrophysics
Applied Physics	Modern physics
Philosophy	Core theories

Chemistry	
Analytical chemistry	Biochemistry
Inorganic chemistry	Materials chemistry
Neurochemistry	Nuclear chemistry
Organic chemistry	Physical chemistry
Other Engineering Areas	
Aerospace Engineering	Agricultural Engineering
Applied Engineering	Biomedical Engineering
Biological Engineering	Building services Engineering
Energy Engineering	Railway Engineering
Industrial Engineering	Mechatronics Engineering
Management Engineering	Military Engineering
Petroleum Engineering	Nuclear Engineering
Textile Engineering	Nano Engineering
Algorithm and Computational Complexity	Artificial Intelligence
Electronics & Communication Engineering	Image Processing
Information Retrieval	Low Power VLSI Design
Neural Networks	Plastic Engineering

Each article in this issue provides an example of a concrete industrial application or a case study of the presented methodology to amplify the impact of the contribution. We are very thankful to everybody within that community who supported the idea of creating a new Research with IJOER. We are certain that this issue will be followed by many others, reporting new developments in the Engineering and Science field. This issue would not have been possible without the great support of the Reviewer, Editorial Board members and also with our Advisory Board Members, and we would like to express our sincere thanks to all of them. We would also like to express our gratitude to the editorial staff of AD Publications, who supported us at every stage of the project. It is our hope that this fine collection of articles will be a valuable resource for *IJOER* readers and will stimulate further research into the vibrant area of Engineering and Science Research.

Mukesh Arora

(Chief Editor)

Dr. M. Varatha Vijayan

Annauniversity Rank Holder, Commissioned Officer Indian Navy, Ncc Navy Officer (Ex-Serviceman Navy), Best Researcher Awardee, Best Publication Awardee, Tamilnadu Best Innovation & Social Service Awardee From Lions Club.

Dr. Mohamed Abdel Fatah Ashabrawy Moustafa

PhD. in Computer Science - Faculty of Science - Suez Canal University University, 2010, Egypt.

Assistant Professor Computer Science, Prince Sattam bin AbdulAziz University ALkharj, KSA.

Prof.S.Balamurugan

Dr S. Balamurugan is the Head of Research and Development, Quants IS & CS, India. He has authored/co-authored 35 books, 200+ publications in various international journals and conferences and 6 patents to his credit. He was awarded with Three Post-Doctoral Degrees - Doctor of Science (D.Sc.) degree and Two Doctor of Letters (D.Litt) degrees for his significant contribution to research and development in Engineering.

Dr. Mahdi Hosseini

Dr. Mahdi did his Pre-University (12^{th}) in Mathematical Science. Later he received his Bachelor of Engineering with Distinction in Civil Engineering and later he Received both M.Tech. and Ph.D. Degree in Structural Engineering with Grade "A" First Class with Distinction.

Dr. Anil Lamba

Practice Head – Cyber Security, EXL Services Inc., New Jersey USA.

Dr. Anil Lamba is a researcher, an innovator, and an influencer with proven success in spearheading Strategic Information Security Initiatives and Large-scale IT Infrastructure projects across industry verticals. He has helped bring about a profound shift in cybersecurity defense. Throughout his career, he has parlayed his extensive background in security and a deep knowledge to help organizations build and implement strategic cybersecurity solutions. His published researches and conference papers has led to many thought provoking examples for augmenting better security.

Dr. Ali İhsan KAYA

Currently working as Associate Professor in Mehmet Akif Ersoy University, Turkey.

Research Area: Civil Engineering - Building Material - Insulation Materials Applications, Chemistry - Physical Chemistry – Composites.

Dr. Parsa Heydarpour

Ph.D. in Structural Engineering from George Washington University (Jan 2018), GPA=4.00.

Dr. Heba Mahmoud Mohamed Afify

Ph.D degree of philosophy in Biomedical Engineering, Cairo University, Egypt worked as Assistant Professor at MTI University.

Dr. Aurora Angela Pisano

Ph.D. in Civil Engineering, Currently Serving as Associate Professor of Solid and Structural Mechanics (scientific discipline area nationally denoted as ICAR/08"–"Scienza delle Costruzioni"), University Mediterranea of Reggio Calabria, Italy.

Dr. Faizullah Mahar

Associate Professor in Department of Electrical Engineering, Balochistan University Engineering & Technology Khuzdar. He is PhD (Electronic Engineering) from IQRA University, Defense View, Karachi, Pakistan.

Prof. Viviane Barrozo da Silva

Graduated in Physics from the Federal University of Paraná (1997), graduated in Electrical Engineering from the Federal University of Rio Grande do Sul - UFRGS (2008), and master's degree in Physics from the Federal University of Rio Grande do Sul (2001).

Dr. S. Kannadhasan

Ph.D (Smart Antennas), M.E (Communication Systems), M.B.A (Human Resources).

Dr. Christo Ananth

Ph.D. Co-operative Networks, M.E. Applied Electronics, B.E Electronics & Communication Engineering Working as Associate Professor, Lecturer and Faculty Advisor/ Department of Electronics & Communication Engineering in Francis Xavier Engineering College, Tirunelveli.

Dr. S.R.Boselin Prabhu

Ph.D, Wireless Sensor Networks, M.E. Network Engineering, Excellent Professional Achievement Award Winner from Society of Professional Engineers Biography Included in Marquis Who's Who in the World (Academic Year 2015 and 2016). Currently Serving as Assistant Professor in the department of ECE in SVS College of Engineering, Coimbatore.

Dr. PAUL P MATHAI

Dr. Paul P Mathai received his Bachelor's degree in Computer Science and Engineering from University of Madras, India. Then he obtained his Master's degree in Computer and Information Technology from Manonmanium Sundaranar University, India. In 2018, he received his Doctor of Philosophy in Computer Science and Engineering from Noorul Islam Centre for Higher Education, Kanyakumari, India.

Dr. M. Ramesh Kumar

Ph.D (Computer Science and Engineering), M.E (Computer Science and Engineering).

Currently working as Associate Professor in VSB College of Engineering Technical Campus, Coimbatore.

Dr. Maheshwar Shrestha

Postdoctoral Research Fellow in DEPT. OF ELE ENGG & COMP SCI, SDSU, Brookings, SD Ph.D, M.Sc. in Electrical Engineering from SOUTH DAKOTA STATE UNIVERSITY, Brookings, SD.

Dr. D. Amaranatha Reddy

Ph.D. (Postdocteral Fellow, Pusan National University, South Korea), M.Sc., B.Sc. : Physics.

Dr. Dibya Prakash Rai

Post Doctoral Fellow (PDF), M.Sc., B.Sc., Working as Assistant Professor in Department of Physics in Pachhuncga University College, Mizoram, India.

Dr. Pankaj Kumar Pal

Ph.D R/S, ECE Deptt., IIT-Roorkee.

Dr. P. Thangam

PhD in Information & Communication Engineering, ME (CSE), BE (Computer Hardware & Software), currently serving as Associate Professor in the Department of Computer Science and Engineering of Coimbatore Institute of Engineering and Technology.

Dr. Pradeep K. Sharma

PhD., M.Phil, M.Sc, B.Sc, in Physics, MBA in System Management, Presently working as Provost and Associate Professor & Head of Department for Physics in University of Engineering & Management, Jaipur.

Dr. R. Devi Priya

Ph.D (CSE), Anna University Chennai in 2013, M.E, B.E (CSE) from Kongu Engineering College, currently working in the Department of Computer Science and Engineering in Kongu Engineering College, Tamil Nadu, India.

Dr. Sandeep

Post-doctoral fellow, Principal Investigator, Young Scientist Scheme Project (DST-SERB), Department of Physics, Mizoram University, Aizawl Mizoram, India- 796001.

Dr. Roberto Volpe

Faculty of Engineering and Architecture, Università degli Studi di Enna "Kore", Cittadella Universitaria, 94100 – Enna (IT).

Dr. S. Kannadhasan

Ph.D (Smart Antennas), M.E (Communication Systems), M.B.A (Human Resources).

Research Area: Engineering Physics, Electromagnetic Field Theory, Electronic Material and Processes, Wireless Communications.

Mr. Amit Kumar

Amit Kumar is associated as a Researcher with the Department of Computer Science, College of Information Science and Technology, Nanjing Forestry University, Nanjing, China since 2009. He is working as a State Representative (HP), Spoken Tutorial Project, IIT Bombay promoting and integrating ICT in Literacy through Free and Open Source Software under National Mission on Education through ICT (NMEICT) of MHRD, Govt. of India; in the state of Himachal Pradesh, India.

Mr. Tanvir Singh

Tanvir Singh is acting as Outreach Officer (Punjab and J&K) for MHRD Govt. of India Project: Spoken Tutorial - IIT Bombay fostering IT Literacy through Open Source Technology under National Mission on Education through ICT (NMEICT). He is also acting as Research Associate since 2010 with Nanjing Forestry University, Nanjing, Jiangsu, China in the field of Social and Environmental Sustainability.

Mr. Abilash

MTech in VLSI, BTech in Electronics & Telecommunication engineering through A.M.I.E.T.E from Central Electronics Engineering Research Institute (C.E.E.R.I) Pilani, Industrial Electronics from ATI-EPI Hyderabad, IEEE course in Mechatronics, CSHAM from Birla Institute Of Professional Studies.

Mr. Varun Shukla

M.Tech in ECE from RGPV (Awarded with silver Medal By President of India), Assistant Professor, Dept. of ECE, PSIT, Kanpur.

Mr. Shrikant Harle

Presently working as a Assistant Professor in Civil Engineering field of Prof. Ram Meghe College of Engineering and Management, Amravati. He was Senior Design Engineer (Larsen & Toubro Limited, India).

Zairi Ismael Rizman

Senior Lecturer, Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM) (Terengganu) Malaysia Master (Science) in Microelectronics (2005), Universiti Kebangsaan Malaysia (UKM), Malaysia. Bachelor (Hons.) and Diploma in Electrical Engineering (Communication) (2002), UiTM Shah Alam, Malaysia.

Mr. Ronak

Qualification: M.Tech. in Mechanical Engineering (CAD/CAM), B.E.

Presently working as a Assistant Professor in Mechanical Engineering in ITM Vocational University, Vadodara. Mr. Ronak also worked as Design Engineer at Finstern Engineering Private Limited, Makarpura, Vadodara.

Table of Contents

Volume-7, Issue-6, June 2021

Application of Data Mining Techniques using Internet of Things

Sania Talha[1], Devunuri Sharanya[2], Dr V Sravan Kumar[3]

[1,2]Student of computer science & engineering at Balaji Institute of Technology & Science, Warangal, Telangana, India

[3]Associate Professor, Dept. of. CSE, Balaji Institute of technology & science, Warangal district, Telangana, India.

Received: 7 June 2021/ Revised: 18 June 2021/ Accepted: 25 June 2021/ Published: 30-06-2021

Abstract— *The generation and growing power of computer science have boosted data collection, storage, and manipulation as data sets are broad in size and complexity level. Internet of Things (IOT) is the most popular term in describing this new interconnected world. The massive data generated by the Internet of Things (IoT) are considered of high business value, and data mining algorithms can be applied to IoT to extract hidden information from data. As more and more devices connected to IoT, the latest algorithms should be applied to IOT. This paper explores a systematic review of various data mining models as well as its applications in the Internet of things along with its advantages and disadvantages.*

Keywords— Internet of things (IOT), Data mining, Applications of Data mining.

I. INTRODUCTION

The term Internet of Things is 16 years old. But the actual idea of connected devices had been around longer, at least since the 70s Back then, the idea was often called “embedded internet “or “pervasive computing”. But the actual term “Internet of Things” was coined by “Kevin Ashton” in 1999 during his work at Procter & Gamble. The internet was the hottest new trend in 1999 and because it somehow made sense, he called his presentation “Internet of things (IOT)”. The Internet of things describes the network of physical objects “things” that are embedded with sensors, software, and other technologies for the purpose of connecting and exchanging data with other devices and systems over the internet.

One of the most valuable technologies is data mining. Data mining helps in discovering novel, interesting and potentially useful patterns from large data sets and applying algorithms to the extraction of hidden information. Data Mining began in the 1990s and is the process of discovering novel, interesting, and potentially useful patterns from large data sets and applying algorithms to the extraction of hidden information.

In order to make IoT smarter, lots of analysis technologies are introduced into IoT; one of the most valuable technologies is data mining. Data mining overlaps with other fields like statistics, machine learning, artificial intelligence, databases but mainly it focuses on automation of handling large heterogeneous data, algorithm and scalability of number of features and instances research ions for in process of big data mining from IoT comes with its own set of challenges such as disparate datasets, large volumes of data, and the integrity of data sources. With the increasing popularity of IoT, new solutions and data mining algorithms are being developed to tackle such problems. [1,2]

On the basis of the definition of data mining and the definition of data mining functions, a data mining process includes the following steps:

- **Data preparation:** prepare the data for mining. It includes 3 steps: integrate data in various data sources and clean noise from the data; extract some parts of data into the data mining systems; pre-process the data to facilitate the data mining
- **Data Mining:** apply algorithms to the data to find the patterns and evaluate patterns of discovered knowledge.

Data presentation: visualize the data and represent mined knowledge to the user.

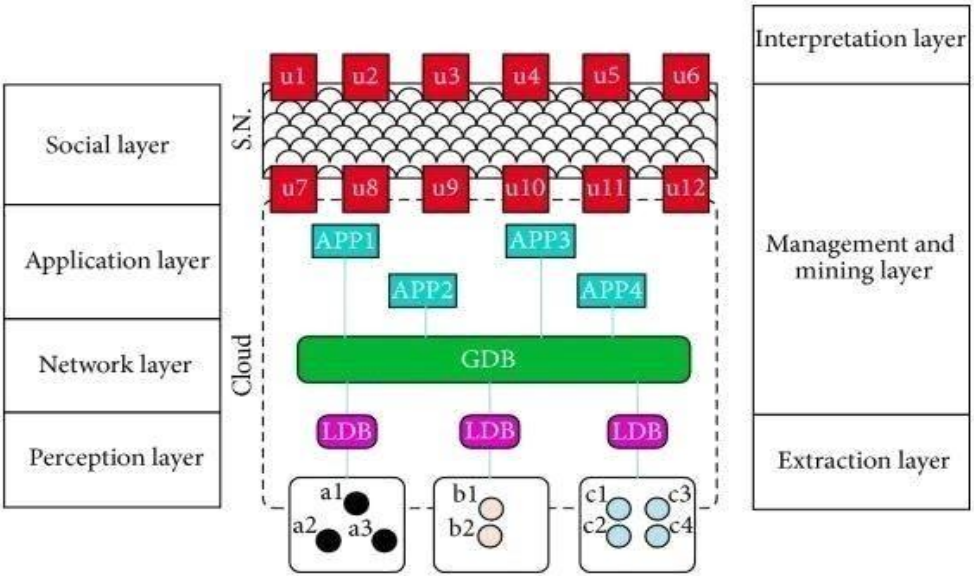

FIGURE 1: Architecture for data mining process

II. DATA MINING FUNCTIONALITIES

There are number of Data mining functionalities, they include:

- Characterization and discrimination
- Mining of frequent patterns, associations, and correlations.
- Classification and regression
- Clustering analysis
- Outlier analysis.

The key contribution of this paper includes:

- Introduction to IOT and Data mining.
- Process of Data mining.
- Functionalities of Data mining.
- Applications of data mining techniques in IOT.
- Advantages and Disadvantages

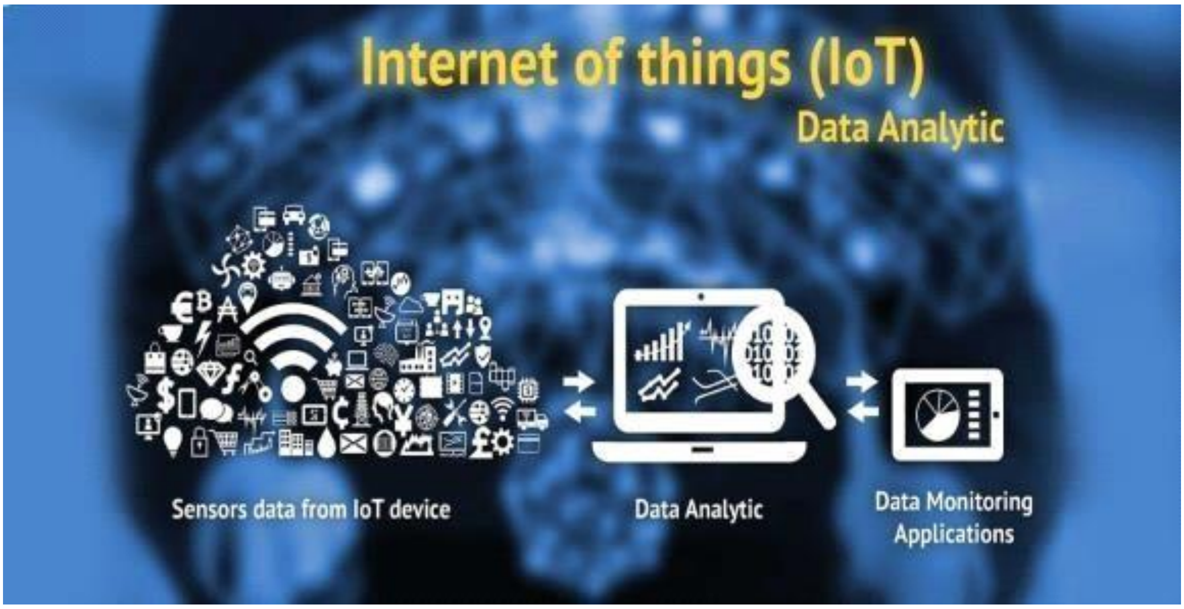

FIGURE 2: Data transfer through Internet of things (IOT)

Data Mining through IOT is primarily used today by companies with a strong consumer for retail, financial, communication, and marketing organizations, to drill down into their transactional data and determine pricing, customer preferences and product positioning, impact on sales, customer satisfaction and corporate profits.

With data mining, a retailer can use point-of-sale records of customer purchases to develop products and promotions to appeal to specific customer segments. [3,4]

III. DATA MINING TECHNIQUES IN FRAUD DETECTION IN CREDIT-DEBIT CARD TRANSACTIONS

Billions of dollars have been lost to the action of frauds. Traditional methods of fraud Detections are time consuming and complex. Data mining aids in providing meaningful patterns and turning data into information. Any information that is valid and useful is knowledge. A perfect fraud detection system should protect information of all the users. A supervised method includes collection of sample records. These records are classified fraudulent or non-fraudulent. A model is built using this data and the algorithm is made to identify whether the record is fraudulent or not. A fuzzy logic system incorporated the actual fraud evaluation policy using optimum threshold values. The result showed the chances of fraud and the reasons why an insurance claim is fraudulent. Another logic system used two approaches to imitate the reasoning of fraud experts, i) the discovery model, uses an unsupervised neural network to find the relationships in data and to find clusters, then patterns within the clusters are identified, and ii) the fuzzy anomaly detection model, which used Wang- Mendel algorithm to find how health care providers committed fraud against insurance companies.

Classification techniques have proved to be very effective in fraud detection and therefore, can be applied to categorize crime data. The distributed data mining model (Chen et al. 1999) uses a realistic cost model to evaluate CART, and naïve Bayesian classification models. The method was applied to credit card transactions. The neural data mining approach uses rule-based association rules to mine symbolic data and Radial Basis Function neural network to mine analog data. The approach discusses the importance of use of non-numeric data in fraud detection. It was found that the results of association rules increased the predictive accuracy. The Bayesian Belief Network (BBN) and Artificial Neural Network (ANN) study used the STAGE algorithm for BBN in fraud detection and back propagation for ANN. The different types of fraud detection are: internal, insurance, credit card, and telecommunications fraud detection. [6]

IV. BAYESIAN BELIEF NETWORK

Bayesian Belief Networks provide a graphic model of causal relationships on which class membership probabilities are predicted, so that a given instance is legal or fraud. Naïve Bayesian classification assumes that the attributes of an instance are independent, given the target attribute. The aim is to assign a new instance to the class that has the highest posterior probability. The algorithm is very effective and can give better predictive accuracy when compared to decision trees and back propagation However, when the attributes are redundant, the predictive accuracy is reduced. For the purpose of fraud detection and we construct two Bayesian networks to describe the behaviour of auto insurance. First, a Bayesian network is constructed to model behaviour under the assumption that the driver is fraudulent (F) and another model under the assumption the driver is a legitimate user (NF). The 'fraud net' is set up by using expert knowledge. The 'user net' is set up by using data from non-fraudulent drivers.[7]

During operation user net is adapted to a specific user based on emerging data. By inserting evidence in these networks (the observed user behaviour x derived from his toll tickets) and propagating it through the network, we can get the probability of the measurement x under two above mentioned hypotheses. This means, we obtain judgments to what degree observed user behaviour meets typical fraudulent or non-fraudulent behaviour. These quantities we call $p(x|NF)$ and $p(x|F)$. By postulating the probability of fraud $P(F)$ and $P(NF) = 1- P(F)$ in general and by applying Bayes' rule, we get the probability of fraud, given Journal of Digital Forensics, Security and Law, the measurement x,

$$P(F|x) = P(F)p(x|F)/ p(x)$$

Where, the denominator $p(x)$ can be calculated as

$$P(x) = P(F)p(x|F) + P(NF)p(x|NF)$$

The chain rule of probabilities is: Suppose there are two classes C1, C2 for fraud and legal respectively. Given an instance $X = (X1, X2, \ldots, Xn)$ and each row is represented by an attribute vector $A = (A1, A2, \ldots, An)$ The classification is to derive the maximum $P(Ci|X)$ which can be derived from Bayes' theorem as given in the following steps:

- P (fraud |X) = [P (fraud | X) P(fraud)] / P(X) P (legal |X) = [P (legal | X) P(legal)] / P(X)

  As P(X) is constant for all classes, only [P(fraud | X) P(fraud)] and [P(legal | X) P(legal)] need to be maximized.

- The class prior probabilities may be estimated by:

  P(fraud) = si / s

  Here, s is the total number of training examples and si is the number of training examples of class fraud.

- A simplified assumption of no dependence relation between attributes is made. Thus,

  P (X |fraud) = $\prod$= n k 1 P (x k |fraud) and

  P (X |legal) = $\prod$= n 1 P (x k |legal)

The probabilities P (x1 |fraud), P (x2 |fraud) can be estimated from the training samples:

P (x k |fraud) = si k / si

Here, si is the number of training examples for class fraud and si k is the number of training examples of class with value x k for Ak.

V. OUTPUT

We present Bayesian learning algorithm to predict occurrence of fraud. Using the "Output" classification results for Table 1. there are 17 tuples classified as legal, and 3 as fraud. To facilitate classification, we divide the age of driver attribute into ranges.

TABLE 1
TRAINING SET

	Name	Gender	Age_driver	fault	Driver_ rating	Vehicle_ age	Output
1	David Okyere	M	25	1	0	2	legal
2	Beau Jackson	M	32	1	1	5	fraud
3	Jeremy Dejuan	M	40	0	0	7	legal
4	Robert Howard	M	35	1	0.33	1	legal
5	Crystal Smith	F	22	1	0.66	8	legal
6	Chibuike Penson	M	36	0	0.66	6	legal
7	Collin Pyle	M	42	1	0.33	3	legal
8	Eric Penson	M	39	1	1	2	fraud
9	Kristina Green	F	29	1	0	4	legal
10	Jerry Smith	M	33	1	1	5	legal
11	Maggie Frazier	F	42	1	0.66	3	legal
12	Justin Howard	M	21	1	0	2	fraud
13	Michael Vasconic	M	37	0	0.33	4	legal
14	Bryan Thompson	M	32	1	0.33	4	legal
15	Chris Wilson	M	28	1	1	6	legal
16	Michael Pullen	M	42	1	0	5	legal
17	Aaron Dusek	M	48	1	0.33	8	legal
18	Bryan Sanders	M	49	1	0	3	legal
19	Derek Garrett	M	32	0	0	3	legal
20	Jasmine Jackson	F	27	0	1	2	legal
X	Crystal Smith	F	31	1	0	2	?

Table 1 shows the counts and subsequent probabilities associated with the attributes. With these simulated training data, we estimate the prior probabilities:

The classifier has to predict the class of instance to be fraud or legal.

P(fraud) = si / s = 3/20 = 0.15

P(legal) = si / s =17/20 = 0.85

TABLE 2
PROBABILITIES ASSOCIATED WITH ATTRIBUTES

Attribute	Value	Count		Probabilities	
		legal	fraud	legal	Fraud
Gender	M	13	3	13/17	3/3
	F	4	0	4/17	0/3
age driver	(20, 25)	3	0	3/18	0
	(25, 30)	4	0	4/18	0
	(30, 35)	3	1	3/18	1/2
	(35, 40)	3	1	3/18	1/2
	(40, 45)	3	0	3/18	0
	(45, 50)	2	0	2/18	0
fault	0	5	0	5/17	0
	1	12	3	12/17	3/17
driver rating	0	6	1	6/17	1/3
	0.33	5	0	5/17	0
	0.66	3	0	3/17	0
	1	3	2	3/17	2/3

By using these values and the associated probabilities of gender and driver age, we obtain the following estimates:

P (X |legal) = 4/17 * 3/18 = 0.039

P (X |fraud) = 3/3 * 1/2 = 0.500

Thus, Likelihood of being legal = 0.039*0.9=0.0351

Likelihood of being fraud = 0.500 *0.1= 0.050

We estimate P(X) by summing up these individuals' likelihood values since X will be either legal of fraud:

P(X) = 0.0351 + 0.050 = 0.0851

Finally, we obtain the actual probabilities of each event:

P (legal | X) = (0.039 *0.9)/0.0851= 0.412

P (fraud |X) = (0.500 *0.1) / 0.0851= 0.588

Therefore, based on these probabilities, we classify the new tuple as fraud because it has the highest probability. Since attributes are treated as independent, the addition of redundant ones reduces its predictive power. To relax this conditional independence is to add derived attributes which are created from combinations of existing attributes. Missing data cause problems during classification process. [8,9]

Naïve Bayesian classifier can handle missing values in training datasets. To demonstrate this, seven missing values appear in dataset. The naïve Bayes approach is easy to use and only one scan of the training data is required. The approach can handle missing values by simply omitting that probability when calculating the likelihoods of membership in each class. Although the approach is straightforward, it does not always yield satisfactory results. The attributes usually are not independent. We could use subset of the attributes by ignoring any that are dependent on others. The technique does not handle continuous data. Dividing the continuous values into ranges could be used to solve this problem, but the division of the continuous values is a tedious task, and this is done can impact the results.

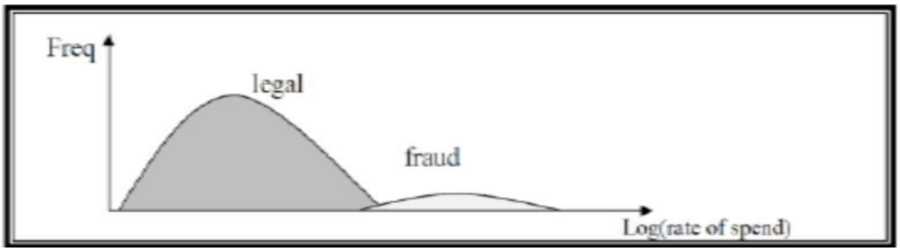

FIGURE 3: frequency distribution of legal and fraud transaction

VI. RESEARCH ANALYSIS

History shows that we have witnessed revolutionary changes in research. Data mining is helpful in data cleaning, data pre-processing and integration of databases. The research can find any similar data from the database that might bring any change in the research. Identification of any co-occurring sequence and correlation between any activities can be known. Data visualization and visual data mining provide us with a clear view of data. Any technology available today has not reached its 100 % capability. It always has a gap to go. So, we can say that data mining through the Internet of Things has a significant technology in a world that can help other technologies to reach its accurate and complete 100 % capability as well.

6.1 Advantages of Mining through IOT:

Data mining has a lot of advantages when using it in a specific industry. Besides those advantages, e.g., Privacy, security, and misuse of information.

FIG 4: Industrial internet of things (IOT)

Data mining through the internet of things facilitates the several advantages in day-to-day life in the business sector. Some of its benefits are given below:

- **Efficient resource utilization:** If we know the functionality and the way that each device works, we definitely increase the efficient resource utilization as well as monitor natural. Minimize human effort: As the devices of data mining through IoT interact and communicate with each other and do a lot of tasks for us, then they minimize resources. The human effort.
- **Save time:** As it reduces the human effort then it definitely saves out time. Time is the primary factor which can save data mining through IoT platforms.
- **Enhance Data Collection:** Improve security: Now, if we have a system where all these things are interconnected then we can make the system more secure.

6.2 Disadvantages of Data Mining through IOT:

As the data mining through the Internet of things facilitates a set of benefits, it also creates a significant set of challenges. Some of the IoT challenges are given below:

- **Security:** As the data mining through IoT systems are interconnected and communicated over networks. The system offers little control despite any security measures, and it can lead the various kinds of network attacks.
- **Privacy:** Even without the active participation of the user, the system provides substantial personal data in maximum detail.
- **Complexity:** The designing, developing, and maintaining and enabling the large technology to system is quite complicated.

FIG 5: IOT offers Security to the system

VII. CONCLUSION

The output counts subsequent probabilities associated with the attributes. With these simulated training data, we estimate the prior probabilities. With simulated training data, we estimate the prior probabilities in fraud detection. Finally, we obtain the actual probabilities of each event. Due to seamless integration of classical networks with IOT. It enables a great vision that all things can be easily monitored and controlled which results in voluminous data. As a vital improvement of the next age of internet, the internet of things pulls in numerous considerations by industry world and scholarly circles. This makes the issue of information mining in IOT turn into a test process.

REFERENCES

[1] Mining with Big data: Jampalachaitanya, Fazi Ahmed parvez. International journal for technological research in engineering. Volume 4 issue to oct 2016.
[2] Data Mining for the Internet of Thin: Literature Review and Challenge S. Feng chen, Pandeng, JiafuWan, Athanasios V.Vasilakos, Xiaohui.
[3] Saral Nigam, Shikha Asthana, and Punit Gupta. Iot based intelligent billboard using data mining. In Innovation and Challenges in Cyber Security (ICICCS-INBUSH), 2016 International Conference on pages 107–110. IEEE, 2016.
[4] Alexander Muriuki Njeru, Mwana Said Omar, Sun Yi, Samiullah Paracha, and Muhammad Wannous. Using iot technology to improve online education through data mining. In Applied System Innovation (ICASI), 2017 International Conference on, pages 515–518. IEEE, 2017.
[5] Sebastian Scholze Claudio CenedeseOliviuMatei, Carmen Anton. Multi- layered data mining architecture in the context of the internet of things. In IEEE. IEEE, 2017.
[6] Bhargava, B., Zhong, Y., & Lu, Y. (2003). Fraud Formalization and Detection. Proc. of DaWaK2003, 330-339.
[7] Bentley, P., Kim, J., Jung., G. & Choi, J. (2000). Fuzzy Darwinian Detection of Credit Card Fraud. Proc. of 14th Annual Fall Symposium of the Korean Information Processing Society.
[8] Bolton, R. & Hand, D. (2001). Unsupervised Profiling Methods for Fraud Detection. Credit Scoring and Credit Control VII.
[9] Brockett, P., Derrig, R., Golden, L., Levine, A. & Alpert, M. (2002). Fraud Classification using Principal Component Analysis of RIDITs. Journal of Risk and Insurance 69(3): 341-371.
[10] Burge, P. &Shawe-Taylor, J. (2001). An Unsupervised Neural Network Approach to Profiling the Behavior of Mobile Phone Users for Use in Fraud Detection. Journal of Parallel and Distributed Computing 61: 915-925.
[11] Bentley, P. (2000). Evolutionary, my dear Watson: Investigating Committee based Evolution of Fuzzy Rules for the Detection of Suspicious Insurance Claims. Proc. of GECCO2000.
[12] Ezawa, K. & Norton, S. (1996). Constructing Bayesian Networks to Predict Uncollectible Telecommunications Accounts. IEEE Expert October: 45-51.

Fusing Optimal Odometry Calibration and Partial Visual Odometry via A Particle Filter for Autonomous Vehicles Navigation

S.A. Ávila-Martínez[1], J.C. Martínez Romo[2], F.J. Luna Rosas[3], M.A. Rodríguez-Díaz, R[4], Mendoza-González[5], L.C. Rodríguez-Martínez[6]

[1]Tecnológico Nacional de México, campus Instituto Tecnológico de Aguascalientes. Graduate Studies and Research Division. Av. Adolfo López Mateos No. 1801, Aguascalientes, Ags. 20256. México.

[2]Tecnológico Nacional de México, campus Instituto Tecnológico de Aguascalientes.

[3,4,5,6]Department of Electrical and Electronics Engineering, Av. Adolfo López Mateos No. 1801, Aguascalientes, Ags. 20256. México.

Received: 10 June 2021/ Revised: 19 June 2021/ Accepted: 24 June 2021/ Published: 30-06-2021

***Abstract*—** *Autonomous vehicles are increasingly becoming ubiquitous in the 21st century; they find application in agriculture, industry, airplanes, cars, service robotics, and others; in order to display autonomous guidance, a vehicle needs to estimate its position and orientation relative to an arbitrary coordinate system; to do so, several sources of information can be used, including images, global positioning systems, inertial measurements or odometry, each according to the application; methods, such as Kalman Filter can be used to combine the several sources of information; however, the more accurate each source of information is, the better the estimation of vehicle position and orientation will be; therefore, the calibration of the parameters of the odometrical systems in autonomous terrestrial vehicles is a must; visual guidance is also an important technology used for vehicle guidance. In this paper, it is presented an off-line method for odometry calibration using a genetic algorithm and the fusion of odometry data with heading information from camera data; a particle filter is used to fuse the data from the optical encoder and the camera. This method was tested in an Automated Guided Vehicle (AGV) with tricycle topology, demonstrating high accuracy in position estimation and guidance through arbitrary paths.*

***Keywords*— Autonomous Guided Vehicle (AGV).**

I. INTRODUCTION

Autonomous vehicles are a relatively new technology with hundreds of potential applications in many aspects of the human life [1]; they have the potential of becoming the everyday driver of the people, and also the automatic guided vehicles AGV of the new generation industrial plant, not requiring magnetic tracks under the floor to follow a predetermined path; they exhibit the capability of interpreting data from sensors to determine their current position with respect to a predefined coordinate frame, responding at any time to the question where am I? [2]; the pose of a vehicle is comprised of the x,y coordinates of its position plus the heading or yaw angle [x, y, ϴ]. When two different sensors provide information from the same variable, it is necessary to decide at which extent one is more reliable than the other in order to provide a weighted estimate of the variable; such is the working of the Kalman Filter [3]. Sources of the same variable are, for instance, an inertial measurement unit (IMU) [4] providing yaw rate information -which integrated over time provides yaw angle- and optical encoder ticks which can be counted and the count converted into the yaw angle of the vehicle whose wheels they are attached to.

A navigation strategy, or algorithm, is usually required to move the vehicle from one point to another; the navigation algorithm is fed with the current position and a next desired position, and its outputs convey the commands to motors in charge moving and steering the vehicle. Again, the accuracy of the movement of the vehicle depends on the accuracy of the pose estimation; this leads us to the conclusion that among the first work to be done is to accurately estimate the pose of vehicle in order to provide reliable autonomous guidance.

Odometry, refers to the measurement of the distance traveled by a wheel as it turns over the terrain; when it comes to vehicles, three and four wheeled vehicles use odometry in at least, two opposite wheels simultaneously. Usually, an optical encoder is used to measure, in ticks, the advance of the wheel; the encoder ticks can be easily converted to the distance traveled by the wheel using a simple formula that assumes the radius of the wheel to be constant; however, systematic and non-systematic factors contribute to provide errors in the measurement of the pose of the vehicle, including unequal wheel diameter, misalignment of wheels, limited encoder resolution, travel on uneven floors, wheel slippage, and others; therefore, it is necessary to calibrate the odometry minimizing the impact of the unwanted errors already mentioned.

Visual odometry (VO) is the process of estimating the movement of a vehicle using images from a camera(s) attached to it; it is based on the incremental estimate of the position of the vehicle through examination of the changes in the images from onboard camera(s); like terrain odometry, VO needs to be calibrated; however, the advantage of VO with respect to terrain odometry is that it is not affected by wheel slippage or uneven terrain. In VO, the vehicle motion between the current and previous images is computed from monocular or stereo images [5]. The main components of a VO system are: image feature detection, feature matching, and motion estimation.

Fusing data is the process of using noisy data from two or more sources related to one variable (one-dimensional) or more (n-dimensional). Typical fusing data techniques are the Kalman Filter for linear and gaussian systems and its variants, Extended Kalman Filter and Unscented Kalman Filter, for non-linear system. Kalman Filters are a kind of stochastic observers [6]. Particle filters (PF) is another valuable technique for sensor fusion; PF are useful when dealing with non-linear systems.

Visual odometry, Kalman filter and Particle filters (PF) are described elsewhere in the literature of the topic, therefore they will not be described here in more detail.

II. SYSTEM DESCRIPTION

In this research a tricycle topology vehicle was used. Here is described the architecture and operation principles of the approach presented in this paper for autonomous vehicle guidance; the proposed method is comprised of two stages: calibration (off-line) and operation. The calibration stage is depicted in figure 1.

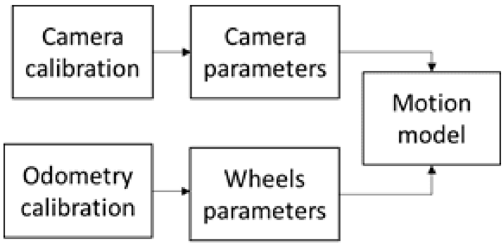

FIGURE 1: Off-line calibration

Calibration stage. In this stage, two procedures are developed; the first is camera calibration; here, the extrinsic and intrinsic parameters are found using typical methods; the second procedure is odometry calibration; as mentioned earlier, the odometry is calibrated in an evolutionary manner using a genetic algorithm; details can be found in [7]; in brief, the steps for odometry calibration are:

a) **Data gathering:** Consists of obtaining encoder pulses as the vehicle moves on predefined linear and curved paths; paths ranging from 1.5 to 3 meters are used and are very attractive when compared to more complex calibration paths presented in the literature [8]. Initial and final positions as well as encoder pulses are recorded.

b) **Evolutionary calibration:** The data recorded is used in a genetic algorithm, whose parameters can be found in Table 1. The aim of using a genetic algorithm is to find the calibration constants [k1, k2] and separation between wheels b, which are part of the kinematic model that will be described here later.

The parameters of the camera as well as the effective radius of the wheels are put together in a motion model, basically the kinematic model of the tricycle topology vehicle [9] and the camera model. It should be noticed that the modeling approach can be applied to other vehicle topologies.

In the case of the camera model, it is required to compute only one parameter: the yaw angle; therefore, it is a partial visual odometry approach. This reduces the computational complexity of the image processing operation of the system. The algorithms related to the calibration state where developed using Matlab®.

TABLE 1
GENETIC ALGORITHM PARAMETERS FOR ODOMETRY CALIBRATION.

Parameter	Value
Population	50 individuals
Mating	Roulete
Selection	Elitist, 2 individuals
Cross-point	Scatter
Mutation	Gaussian
Restrictions	3.0e-4 < [k1, k2] < 4.5e-4 0.35 > b > 0.39 (m)
Goal	To minimize pose error
Stop criterion	150 generations

The kinematic model of the tricycle topology is as follows:

$$\Delta s_l = \frac{2\pi r_l . pl_l}{R} = pl_l . k_1 \quad (1)$$

$$\Delta s_r = \frac{2\pi r_r . pl_r}{R} = pl_r . k_2 \quad (2)$$

$$\Delta\theta = (\Delta s_r + \Delta s_l) . b \quad (3)$$

$$\Delta u_{x,y} = \frac{\Delta s_r + \Delta s_l}{2} \quad (4)$$

$$\theta_1 = \theta_0 + \frac{\Delta\theta}{2} \quad (5)$$

$$x_k = x_{k-1} - \Delta u_{x,y} . sin\theta_1 \quad (6)$$

$$y_k = y_{k-1} - \Delta u_{x,y} . cos\theta_1 \quad (7)$$

Where:

pl_l, Count of encoder pulses from the left, rear wheel.

pl_r, Count of encoder pulses from the right, rear wheel.

k_1, Odometric compensation constant of the left wheel.

k_2, Odometric compensation constant of the right wheel.

R, Optical encoder resolution (pulses per revolution).

2.1 Operation stage

Figure 2 shows the operation stage; when tracking a specific path, the vehicle moves forward and produce encoder pulses from both rear wheels; at the same time, the camera captures images. From one current image and the previous, the digital image processing step computes θ (the yaw angle). As can be seen, two versions of θ are at the disposal of the particle filter; each computation of the particle filter produce the estimated version [x', y', θ'], which is the best estimate (in an optimization sense) of the pose of the vehicle.

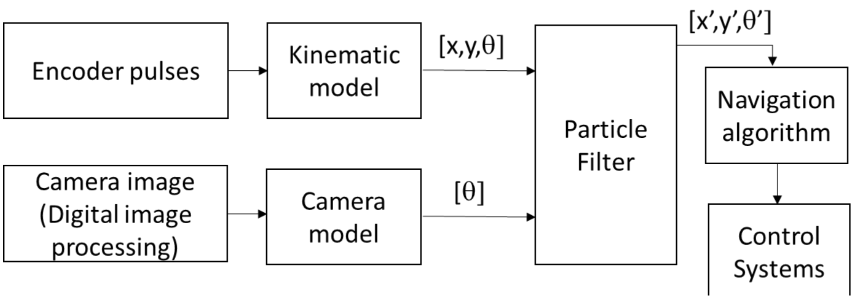

FIGURE 2: Operation stage of the vehicle.

2.2 Implementation

The system described in the previous section was implemented according to the following:

Physical system: A three-wheeled vehicle, shown in Figure 3. Front wheel steering.

Vehicle main-board: Jetson-nano, NVIDIA. With camera, artificial vision enabled.

Software: Embedded linux. OpenCV.

Control Systems: Microcontrollers and DSP based. Digital PID control.

Motor: DC type, both motion and steering.

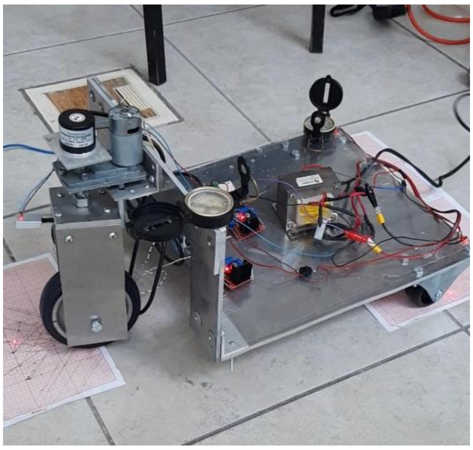

FIGURE 3: Vehicle used showing the tricycle topology.

III. RESULTS

To test this approach to autonomous vehicle navigation, the odometry calibration was carried out using calibration 15 runs of trajectories similar to the one shown in Figure 4. The genetic algorithm produced k_1=3.9447008587e-004, k_2= 3.9698427203e-004 and b= 0.3688034 as the calibrated parameters for the set of equations 8 through 7. The camera parameters are shown in the calibration matrix Q:

$$Q = \begin{bmatrix} 436.5401 & 0 & 206.2603 \\ 0 & 436.0246 & 151.2045 \\ 0 & 0 & 2 \end{bmatrix} \tag{8}$$

To test the system a set of irregular shaped trajectories was designed and programmed to be followed by the vehicle; for comparison purposes, three scenarios were tested:

a. Using uncalibrated odometry parameters,

b. Using only calibrated odometry parameters,

c. Using calibrated odometry and visual odometry with particle filter.

The results of autonomous path tracking are shown in Table 2. Asterisks denote that the task could not be completed by the vehicle. The trials were executed in a controlled environment, with uneven floor but not abrupt changes. The error reported is the mean error value of 10 trials at each scenario.

IV. DISCUSSION

It is evident that the worst-case response was the one of no calibration at all. Here, the only trajectory completed by the vehicle was of 22 meters long, but with a large error, up to about 12%. This is due to the fact that uncalibrated data produce a rapidly growing accumulation of error. It was observed a variable deviation of the vehicle from the predefined trajectory, which implies that it is most likely the dominant error, was due to random or non-systematic errors.

When using only terrain odometry calibrated parameters, a remarkable improvement was obtained, reducing the error dramatically with respect to the uncalibrated scenario; the computed calibration parameters using the genetic algorithm were capable of significantly reduce the error, although it is not clear at what extent the reduced error was of the type systematic or non-systematic.

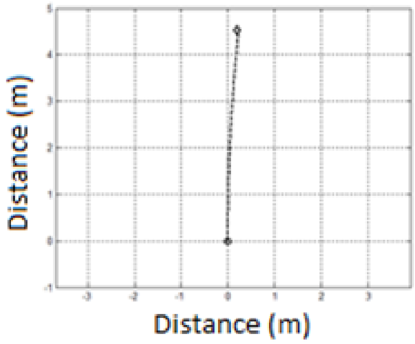

FIGURE 4: Example of calibration trajectory

TABLE 2
RESULTS OBTAINED IN 4 TRAJECTORIES.

Scenario	Length (m)	Absolute Error (%)
Uncalibrated odometry	22	12.242
	47	*
	100	*
Odometry calibration only	22	0.234
	47	0.333
	100	*
Odometry calibration + VO + PF	22	0.1025
	47	0.1236
	100	0.592

Finally, the full approach introduced here clearly separates from the other two scenarios since the vehicle was capable of successfully complete a 100 meters long trajectory. When compared to the two previous scenarios it showed to be remarkable better. To ensure success of the PF, the calibration data was used to simulate it and varying levels of noise were introduced to both, the calibration data and a simulated version of the θ yaw angle from the camera. Other approaches in literature reported similar results but using a more complex framework or tested on shorter trajectories.

V. CONCLUSION

This paper presents an approach to the problem of autonomous vehicle guidance; the use of optimization in a first stage of odometry calibration is one of major steps towards a real-world applicable system because it helped to reduce systematic and non-systematic errors, as seen in Table 2.

Combining data from the odometry system and the camera with the particle filter, as shown in Figure 2, significantly supported the dramatic reduction in the percentage of error shown as compared to the simple use of odometry calibration. Although many sophisticated algorithms have been published in the past and recent years, including auto-calibration and pose, they are complex and computational expensive, a key difference with the approach presented here. It is not discarded as future work to include some continuous optimal calibration method during the operational stage of the vehicle.

Finally, the goal of this project is to produce industrial grade AGV autonomous systems, so the efforts in such direction will continue from these research groups.

Combining data from the odometry system and the camera with the particle filter, as shown in Figure 2, significantly supported the dramatic reduction in the percentage of error shown as compared to the simple use of odometry calibration. Although many sophisticated algorithms have been published in the past and recent years, including auto-calibration and pose, they are complex and computational expensive, a key difference with the approach presented here. It is not discarded as future work to include some continuous optimal calibration method during the operational stage of the vehicle.

Finally, the goal of this project is to produce industrial grade AGV autonomous systems, so the efforts in such direction will continue from these research groups.

ACKNOWLEDGEMENTS

This work was supported by the Tecnológico Nacional de México/Campus Instituto Tecnológico de Aguascalientes, in the Doctorate in Engineering Sciences program and by the CONACYT México.

REFERENCES

[1] R. Kümmerle, G. Grisetti, and W. Burgard, "Simultaneous calibration, localization, and mapping," in IEEE International Conference on Intelligent Robots and Systems, 2011, pp. 3716–3721.

[2] L. Feng, J. Borenstein, and H. Everett, "'Where am I?' sensors and methods for autonomous mobile robot positioning. TR UM-MEAM-94-12," 1994.

[3] G. Fu, J. Zhang, W. Chen, F. Peng, P. Yang, and C. Che, "Precise Localization of Mobile Robots via Odometry and Wireless Sensor Network," Int. J. Adv. Robot. Syst., vol. C, p. 1, 2013.

[4] J. Borenstein, H. R. Everett, L. Feng, and D. Wehe, "Mobile robot positioning: Sensors and techniques," J. Robot. Syst., vol. 14, no. 4, pp. 231–249, 1997.

[5] D. Scaramuzza and F. Fraundorfer, "Visual Odometry : Part I: The First 30 Years and Fundamentals," IEEE Robot. Autom. Mag., vol. 18, no. 4, pp. 80–92, 2011.

[6] P. Zarchan and H. Musoff, Fundamentals of Kalman Filtering: A Practical Approach, Third. Reston, VA: American Institute of Aeronautics and Astronautics, Inc., 2013.

[7] D. E. Goldberg, Genetic Algorithm in Search, Optimization, and Machine Learning. Addison Wesley Publishing Company, 1989.

[8] J. Borenstein, "UMBmark: a benchmark test for measuring odometry errors in mobile robots," Proc. SPIE, vol. 2591, pp. 113–124, 1995.

[9] J. C. Martínez-romo et al., "Genetic Optimization for Odometry Calibration of Tricycle Topology mobile robots," pp. 1–20, 2016.

AD Publications

Sector-3, MP Nagar, Bikaner,

Rajasthan, India

www.adpublications.org, info@adpublications.org