Oleg Sergiyenko *Editor*

Optoelectronic Devices in Robotic Systems

Springer

Optoelectronic Devices in Robotic Systems

Oleg Sergiyenko
Editor

Optoelectronic Devices in Robotic Systems

Springer

Editor
Oleg Sergiyenko
Universidad Autónoma de Baja California
Mexicali, Baja California, Mexico

ISBN 978-3-031-09790-4 ISBN 978-3-031-09791-1 (eBook)
https://doi.org/10.1007/978-3-031-09791-1

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

Preface

Optoelectronic devices are widely utilized to achieve a better performance in robotic systems with application in different practical problems solutions. A large number of different practical applications which are used in robotics can be listed. Among them are: optical machine vision systems for spatial sectors scanning, targets recognition, object surfaces coordinate measurement, robot navigation, medical scanning, automatic access control systems, military services, telecommunications, and structural health monitoring, to mention some.

Also can be mentioned the search, classification, industrial process automatic monitoring, visualization of various physical phenomena, rescue, vigilance, mapping, dangerous situations detection, and other areas where optoelectronic devices play an important role in machine control based on vision ability.

It is possible to mention many different kinds of optoelectronic devices such as photodiodes, phototransistors, photoresistors, cameras, lasers, solar cells, light-emitting diodes, laser diodes, and optical fibers. These devices usually are included in sophisticated schemes and circuits and are interacting between them to perform complex practical actions in order to reach the defined goals of automatic system functioning.

Very important parts, in this case, are methods and corresponding programs/algorithms to control their interaction, frequently aiming to optimize the results by careful and appropriated cooperative processing or parameter interpretation.

In such complex combined systems, many parameters and physical characters are magnitudes of influence. Research of these influences, their systematization, and functional description often is key to the general improvement of the desired industrial process. Sometimes the use of knowledge about these phenomena and their description application is able to change the quality of a whole automated process, or improve and add totally new features in practice. That is why the research in the field of optoelectronic devices application has greatly grown in a number of scientific laboratories worldwide. New technologies are constantly creating and changing drastically during very short periods of time.

The reasons why the use of optoelectronic devices and related technologies can be valuable for automated and robotic systems are obvious. Processing time is also a significant parameter on complex algorithms used to extract important patterns and trends to understand the data mining, exceeding human capabilities. The time of processing due to the complexity of algorithms to extract important patterns and trends to understand the data meaning it also significantly exceeds human capabilities.

Robotic systems design, using the variety of optoelectronic devices, is closely tied with many associated practical and theoretical tasks, and careful consideration of the corresponding theoretical aspects. We can mention among them several next important topics:

- Dynamic ability to capture video and extract the versatility of attributed information with its further analysis and interpretation
- Application of neural networks, deep learning, and statistical methods to filter and rectify the captured information; both for extended real-time and fine accuracy post-processing
- Matching and calibration in complex optical multi-component devices
- Field-of-view or frame margins restriction and establishment
- Multichannel simultaneous data collection with the possibility of posterior true/false data recognition and correction
- Objects shape and pose automatic estimation and digitization; lightweight, energy-saving, and robust robotic systems design
- Appropriate spatial information acquisition with optimized data mining aiming to facilitate further acquired data usage for application tasks
- Object-oriented cross-comparison of different physical approaches, such as two-phase flows, amplitude-shadow, Time-of-Flight (ToF), and Doppler methods, for quantitative/qualitative analysis of different aggregate states of matter
- Methods of reconstructions of 3D surfaces' maps from the data-sets remotely obtained by optoelectronic devices, as well as the software methods intended for these data analyses and investigations in digital form
- Methods of decision-making in robot functioning based on the analysis of obtained optical information, as well as different known theories' application for further control improvement, such as the fuzzy logic, robust controllers, and so on
- Juridical aspects of surroundings optical measurements and visualization, as well as the real cases of optoelectronic scanning and monitoring of higher-importance civil engineering structures, world's Cultural Heritage (CH) buildings and monuments, and so on
- Methods and models of safe and secure robot positioning and object detection, and their common cooperation within the defined spatial sector

This book is intended to consider all the above-mentioned theoretical problems and engineering tasks in the context of its topic for the appropriate practical application. Such a book perspective, in our opinion, can give to readers the maximal method/tool understanding for their own innovative and advanced application.

In this sense, *optoelectronic devices in robotic systems* are important to modern science and industrial practical implementation. Hence, it is necessary to create new algorithms and systems to improve their performance. The chapters in this book relate to contributions in optoelectronic devices for robotic systems and machine vision applications among them. It is intended that each book chapter shows the state of the art in explained methods and 3D and 2D technologies, as well as the novel strategies in optoelectronic devices' performance.

The book covers both theories and application of optoelectronic devices in robotic systems. In our opinion, this book should be attractive to potential consumers/citers because of a well-balanced source of novel technologies in the area of optoelectronics and robotics, with an explicit overview of recently existing systems. The book presents a comparative analysis of optoelectronic devices properties, advantages, and disadvantages. The topics are of interest to readers from a diverse group of audiences in different areas of specialty such as electrical, electronics, computer engineering, technologists, and even non-specialist readers. Thus, it is intended to be used as a text and reference work on advanced topics in machine vision and navigation. It is dedicated to academics, researchers, advanced-level students, and technology developers, who will find this text useful in furthering their research exposure to pertinent topics in *optoelectronic devices in robotic systems* and assisting in their future research efforts in this field.

An Overview of Optoelectronic Devices in Robotic Systems

The combination of optoelectronic devices and robotic systems is most promising nowadays, because in recent years, finally, the full-scale release of daily-use devices, such as robot-cleaners, robot-assistants for the elderly, and so on, has become a part of our everyday life. Based on the experience, this is a remarkable sign of the upcoming boom of demand for novel competitive technologies in the field. Coverage of all the topics of study in this book would be impossible; however, the most significant have been selected. The book contains 11 chapters, which have been classified and briefly described as follows.

Our first chapter proposes an innovative method for dynamic video-mapping a suit over a puppet with a projector for art shows. To follow the movements of the puppet, real-time tracking is done with a Red Green Blue plus Depth (RGBD) camera and pose computation algorithms that, at first, scan the silhouette of the target and define the pose and joint angles to precisely suit up the model. A comparison between Infra-Red (IR) reflectors and RGBD cameras is done, remarking the importance of the flexibility of cameras for different shapes and deformable surfaces over IR markers, which requires complex arrangements according to the object of interest's geometry.

Passing to a deep learning application, the second chapter is dedicated to the use of Fully Convolutional Neural Networks (FCNN) for weed detection on precision agriculture and compares different FCNN architectures with its optimal layers and

parameters. These architectures use information captured with an Unmanned Aerial Vehicle (UAV) that flies over the field, taking images and sending them to a base station. The advantages and disadvantages of each architecture with regard to the identification of weed are also reported in order to define which is optimal for real-time applications.

Following with image processing, the next chapter discusses the importance of the disparity map improvement generated by matching algorithms used in stereo vision and presents a focused method for this. This technique takes mismatching values and analyzes them with multiple zoom image pairs in order to restrict the search area. Also, the necessary process to obtain depth data with stereo vision is described step by step, with a special interest in comparison of disparity refinement methods.

Focused on obtaining redundant information, a combination of two vision systems is applied in a chapter where stereo vision depth estimation is taken as feedback for a Technical Vision System (TVS) laser control in order to point the laser beam to a desired point and complement or replace depth data, raising the accuracy and precision of 3D information. This synchronization between systems is reached using transformation matrices for same origin sharing and correcting the lectures according to the position and rotation of cameras regarding TVS.

With the goal of estimating hand pose using Convolutional Neural Networks (CNN), one of our authors worked in the development of lightweight architecture; in terms of processing, this CNN is formed with only 1.9 million parameters and extracts features from images and determines the position and rotation of each finger of a human. It is also compared with other popular architectures, proving its reliability in real-life applications.

Another chapter explains different methods to acquire spatial information for robot navigation using various techniques such as stereo vision, laser scanners, and Light Detection and Ranging (LiDAR). With appropriate data handling, it is possible to obtain enough odometry for decision-making in applications like self-driving cars, Surgical Navigation Robots (SNR), and object detection. The author also mentions the required sensors and processing equations to match these measuring systems with inertial systems for refinement of robot pose estimation.

Next, a chapter talks about the diagnosis of two-phase flows using non-invasive ultrasonic signals in a technique based on amplitude-shadow, Time-of-Flight (ToF), and Doppler methods for fluid characterization. The acquired information helps in monitoring the state of metal coolants that flow in heat transfer systems applied in nuclear reactors. Other ultrasonic diagnosis methods are presented and compared with X-ray and magnetic flow meter-based methods.

Opening the range of applications for laser scanners, another author presents reconstruction methods for human body parts, with the goal to identify anomalies in shapes. This is proposed from the idea that a point cloud obtained from depth estimation techniques could be enough to identify deformities such as pectus excavatum and pectus carinatum in feet; however, lectures should be reliable, so different post-processing methods are presented and analyzed to refine 3D models and raise the confidence in the diagnosis.

Keeping with the development in robot navigation, one chapter is dedicated to robot control based on fuzzy logic. The designed algorithm was tested in a programmed model that shows the behavior and decision taken by the robot under a variety of parameters and working conditions. This chapter also contains the basics of fuzzy logic applied to decision-making in robot navigation and the use of fuzzy sets for algorithm testing to obtain valid results, robust and usable under real-life conditions.

Also, in this book is presented a detailed review of different methods and techniques used to digitize the world's Cultural Heritage (CH) buildings and monuments, such as laser scanning, photogrammetry, and image processing, with the goal of documenting these important sites through 3D models and 2D visual digital technologies. A special emphasis is on the importance of the use of the mentioned techniques for interpretation, promotion, conservation, and monitoring of CH.

The ending chapter explains the wide variety of machine vision techniques actually applied to Mobile Robot Navigation (MRN) within controlled and harsh environments. The author also dives into the fundamentals of and common problems in MRN, as well as the practical and theoretical approaches to solve them, such as Neural Networks (NN), Fuzzy Logic (FL), and miscellaneous algorithms based on widely studied methods and models, with the goal of reaching a robust robot positioning and object detection method.

We believe that the present book will be helpful to a wide range of readers, with different levels of expertise: for students and professionals, PhD candidates, engineers, professors, and researchers, and finally for everybody who is interested in a deeper understanding of the tasks of modern optoelectronics and robotics. We would like to thank all chapter authors for the interesting material presented in collaborative team spirit. We would like to thank all qualified reviewers for their valuable comments and creative recommendations; they were very helpful in general updates to the book. Also, we would like to thank Editorial for the possibility to disseminate the latest knowledge in this field to a wide range of readers worldwide.

Mexicali, Baja California, México Oleg Sergiyenko

Acknowledgments

It is a pleasure to offer the acknowledgment to all the authors that have contributed with a chapter in this book *Machine Vision and Navigation*. Every author has put in the best effort in writing their most recent research findings. I feel satisfied with the academic product generated with an international vision. More than 30 researchers around the world have collaborated in this project, representing the participation of ten countries: France, Greece, Germany, India, Japan, Jordan, Mexico, Russia, Taiwan, and Ukraine. A whole list of the authors with the affiliations and biographies are provided in the "List of Authors" and "About Authors" sections of this book.

I am delighted and thankful to have received the support of researchers in the areas of machine vision, navigation, robotics, control, and artificial intelligence. I extend the acknowledgment to all the reviewers, who have done a great job reading and suggesting improvements for each chapter: Andrey Somov, Cesar Sepulveda-Valdez, Fabian N. Murrieta-Rico, Guillaume Caron, Julio Cesar Rodríguez-Quiñonez, Lars Lindner, Ruben Alaniz-Plata, and Wendy Flores-Fuentes.

Thanks also go to the editorial board and the officials at Springer International AG for their invaluable efforts, great support, and valuable advice for this project toward the successful publication of this book. I also want to thank the institution Universidad Autónoma de Baja California for providing me with a location and time to develop this project.

Mexicali, Baja California, México Oleg Sergiyenko

Contents

Editors and Contributors

About the Authors

Ruben Alaniz-Plata is a teacher at the Autonomous University of Baja California, where he obtained a bachelor's (2016) and a master's degree in Science (2019), both in the Engineering area. He is working to obtain a PhD from the Engineering Institute of UABC in the optoelectronics field, with special interest in laser scanner systems and stereo vision. He has presented two articles as a co-author in a national congress of Academia Journals in Michoacán, Morelia (2018), and is a reviewer of a book chapter for IGI Global (2020).

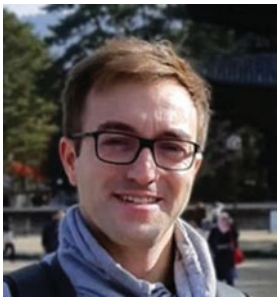
Humberto Andrade-Collazo received the Engineering degree in Mechatronics from the Autonomous University of Baja California (UABC), Mexico, in 2017. He is working to obtain a master's from the Engineering Institute of UABC in the optoelectronics field, with special interest in laser scanner systems and vision in mobile robots. He has been working in companies dedicated to the field of heat ventilation and air conditioning during his three years as a designer engineer.

Mounya Belghiti was a member of the Robotic Perception Group of the MIS Laboratory in 2015 at Universite de Picardie Jules Verne (UPJV), France. She received the MSc degree in Robotics from UPJV in 2015. She is now working on advanced driver-assistance systems (ADAS) as an engineer in a private group for the car industry.

Vikrant Bhateja is an associate professor in the Department of Electronics and Communication Engineering at Shri Ramswaroop Memorial Group of Professional Colleges (SRMGPC), Lucknow (U.P.), India, and is also the Dean of the same college. He has a doctorate in ECE (Bio-medical Imaging and Signal Processing), with a total academic teaching experience of 17 years and around 160 publications in reputed international conferences, journals, and online book chapter contributions, out of which 22 papers are published in SCIE indexed journals (summing up to a total JCR Impact Factor: 70). His areas of research include digital image and video processing, computer vision, image fidelity measurement, medical imaging, bio-medical signal processing and measurement, and machine learning. He has been chairing/co-chairing around 25 international conferences as publication or TPC Chair. He has edited 30 book volumes from Springer Nature as a corresponding/co-editor/author. He is a senior member of IEEE and life member of CSI, and has served in Exe-Com of IEEE U.P. Section. He has also chaired IEEE Young Professional Affinity Group in 2016–2017. He is Editor-in-Chief of IGI Global—*International Journal of Natural Computing and Research* (IJNCR): an ACM and DBLP indexed journal since 2017. He has guest-edited many Special Issues in reputed Scopus/SCIE indexed journals. He has received the "Certificate of Outstanding Contribution in Reviewing" for 2016 and 2017, respectively, in prestigious Elsevier journals such as *AEUE*, *Measurements*, and *Computer Methods and Programs in Biomedicine*. Two of his conference papers have received "Best Paper Award" in the international conferences ICICT-2015 and ICTIS-2015. In 2016, his article, published in *Review of Scientific Instruments Journal*

(AIP Press), was selected as "Editor-Choice" article of the issue. In January 2017, he received the "Paper Presenter Award at International Conferences: 2015–16" at 51st CSI Annual Convention, Coimbatore, India.

Artem Bronnikov received the bachelor's and master's degree in the field of Computerized and Robotics Systems at Kharkiv National University of Radioelectronics (NURE), Ukraine, in 2009 and 2010, respectively. In 2014 he joined the Computer-Integrated Technologies, Automation and Mechatronics Department of the Faculty of Automatics and Computerized Technologies at NURE, where he collaborated on researches in the robotic and automation fields. He received the PhD degree for his work "Methods and Models of Visual-Guided Control of Robots" in 2021 at NURE. Till the present time, he has written 42 publications for scientific journals and conferences, mainly in Ukraine. His research interests include mobile robot control, computer vision, and robotic application for bionics.



Guillaume Caron has been an associate professor since 2011 at Universite de Picardie Jules Verne (UPJV), France. He received the PhD degree in Robotics and the habilitation degree from UPJV in 2010 and 2019, respectively. He was the leader of the Robotic Perception Group of the MIS Laboratory (UPJV) from 2016 to 2020. Since 2019, he has been a CNRS (French National Scientific Research Center) delegate at the CNRS-AIST Joint Robotics Laboratory, Tsukuba, Japan. Before, he stayed at Inria, Rennes, France (2010–2011), and at the University of Osaka, Japan (April–May 2013). His research interests include artificial vision for robotics, real-time visual tracking and servoing, and digital heritage. He has been the author of 14 articles in international journals (T-RO, IJRR, RA-L, etc.) and 21 papers in proceedings of international conferences (ICRA, IROS, Digital Heritage, etc.). He received the "Best Runner-up Paper Award" at the IEEE/KROS URAI international conference in 2015. He also gave five invited talks in international workshops and co-organized three times the international workshop "E-

Heritage" (2017, 2019 with ICCV, and 2021 with IROS). He is actively involved in several national and international scientific societies, including the IAPR TC19 "Computer Vision for Cultural Heritage Applications," which he has been chairing since 2018.

Moisés J. Castro-Toscano was born in Baja California, Mexico, in March 1989. He received his BS degree from Instituto Tecnológico de Mexicali in 2011, master's degree in engineering from the Technological Institute of Mexicali in 2014, and PhD degree from Baja California Autonomous University, México, in 2019. He is a professor at the Technological Institute of Mexicali, teaching basic science to engineering students. He is also a professor and a researcher in the Faculty of Engineering, Universidad Autónoma de Baja California, and a member of the National Research System CONACYT. Moises is also involved with the development of navigation application through inertial navigation systems. He has worked on seven research articles and three book chapters on different topics such as stereo vision systems, laser vision systems, and inertial navigation systems. He has presented research results at different international conferences such as the IEEE 2017 in Edinburgh, UK; IECON 2018 in Washington DC, USA; and ISIE 2020 in Deft, the Netherlands. His current research interests include inertial navigation, stereo vision systems, control systems, robot navigation, and 3D laser scanners.

Yu-Ting Chen received the BS degree in Computer Science and Information Engineering from National Chin-Yi University of Technology and the MS degree in Electrical Engineering from National Chung Cheng University, Taiwan. His research interests include computer vision, pattern recognition, machine learning, and intelligent and embedded systems. He is with Silicon Motion Technology Corp., Hsinchu, Taiwan, as a senior engineer.

Anthony Dessaux was a member of the Robotic Perception Group of the MIS Laboratory in 2016 at Universite de Picardie Jules Verne, France. He received his Engineering degree from Universite de Technologie de Compiegne, France, in 2019. He is now working as a software engineer specialized in calculation tools for a mechanical engineering company.

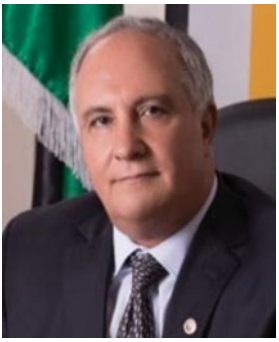
Sergey Dvoynishnikov was born on September 22, 1983 in the city of Berdsk, Novosibirsk Region. From 2000 to 2006 he studied at the Physics Department of Novosibirsk State University. His diploma work was awarded the medal of the Ministry of Education and Science of the Russian Federation. Since 2006, he has been holding scientific positions at Kutateladze Institute of Thermophysics SB RAS: 2006–2009 junior researcher, 2009–2011 researcher, 2011–2018 senior researcher, since 2019 leading researcher and head of the laboratory. From 2006 to 2009 he studied in the postgraduate course of the Institute of Thermophysics. S.S. Kutateladze SB RAS. S. Dvoynishnikov has been an associate professor at Novosibirsk State University since 2006. He teaches a course of lectures and seminars on data processing methods in scientific research. In 2009 he received his PhD in Technical Sciences. In 2016 he received the degree of Doctor of Technical Sciences. In 2017, he was awarded the Prize of the Government of the Russian Federation in the field of science and technology for young scientists as part of a youth team. Sergey S. Dvoynishnikov is the author of more than 200 publications and 20 patents.

Wendy Flores-Fuentes received the bachelor's degree in Electronic Engineering from the Autonomous University of Baja California in 2001, the master's degree in Engineering from Technological Institute of Mexicali in 2006, and the PhD degree in Science, Applied Physics, with emphasis on Optoelectronic Scanning Systems for SHM, from the Autonomous University of Baja California in June 2014. By now she is the author of 36 journal articles in Elsevier, IEEE, Emerald,

and Springer; 18 book chapters; and 8 books in Springer, Intech, IGI Global, and Lambert Academic; 46 proceedings articles in IEEE ISIE 2014–2022; IECON 2014, 2018, and 2019; the World Congress on Engineering and Computer Science (IAENG 2013); IEEE Section Mexico IEEE ROCC2011; and the VII International Conference on Industrial Engineering ARGOS 2014. Recently, she organized and participated as chair of the Special Session on "Machine Vision, Control and Navigation" at IEEE ISIE 2015–2021 and IECON 2018 and 2019. She has participated as a guest editor for the *Journal of Sensors* with Hindawi, the *International Journal of Advanced Robotic Systems* with SAGE, *IEEE Sensors*, and *Elsevier Measurement*. She holds one patent in Mexico and one in Ukraine. She has been a reviewer of several articles published by Taylor & Francis, IEEE, Elsevier, and EEMJ. She is a full-time professor at Universidad Autónoma de Baja California, Faculty of Engineering. She has been incorporated into CONACYT National Research System in 2015. She received the award of "Best Session Presentation" in WSECS2013, San Francisco, USA. She received as co-author the award of "Outstanding Paper in the 2017 Emerald Literati Network Awards for Excellence." Her interests include optoelectronics, robotics, artificial intelligence, measurement systems, and machine vision systems.

Antonios Gasteratos is Full Professor of "Robotics, Mechanotronics, Industrial Automation, Computer Vision, and Machine Learning" at the Department of Production and Management Engineering of the Democritus University of Thrace (DUTH). He has been with this Department since 2003 and has served as its Head (2015–2020). Prior to joining the Department he worked as a postdoctoral fellow in the Laboratory of Advanced Robotics of the Department of Informatics, Systems and Telematics, University of Genoa. In addition, he is a member of the Strategic Planning Committee of DUTH and has been a member of the Senate and the Research Committee of DUTH. He has also been a member of the Plenary of the National Telecommunications and Post Commission

(EETT), Chairman of the Service Council of EETT, and a member of standing committees. Today he serves as Vice-Head of the Department of Production and Management Engineering and Director of the Robotics and Automation Laboratory, as well as Director of the Postgraduate Program "Innovation, Technology and Business Administration." He holds a diploma and a doctorate from the Department of Electrical and Computer Engineering of DUTH (1994 and 1999, respectively). He has taught various courses and in recent years has been teaching the courses "Mechanotronics," "Electronics," "Robotics," "Automatic Control Systems," and "Machine Vision." He has supervised 13 doctoral dissertations and more than 70 diploma theses. During the last 15 years he has been scientifically responsible for many projects funded mainly by the European Commission, the European Space Agency (ESA), the Hellenic Secretariat for Research and Innovation, the industry, and other sources (total funding > €3.5 million). He has had over 250 articles published in scientific journals and international conferences. He is the Subject Editor-in-Chief of *Electronics Letters* and Associate Editor at the *International Journal of Optomechatronics* and the *International Journal of Advanced Robotics Systems*. He is a member of committees at international scientific conferences in the field of new technologies and has organized international conferences and workshops. He has been a reviewer in all international journals in the field of machine vision and robotics and in many others in the discipline of electrical and computer engineering. He is a project evaluator in various international and state organizations (European Commission, European Association of National Metrology Institutes (EURAMET) and (International Association for the Promotion of Cooperation with Scientists from the New Independent States of the Former Soviet Union) INTAS, Greece, Italy, Cyprus, Russia, Georgia, New Zealand, etc.). Finally, he participates in the following scientific associations as a fellow member of Industrial Engineering Technology (IET), Institute of Electrical and Electronics Engineers (IEEE), Association for Computing Machinery (ACM), Sociedad Española de Técnicas Neutrénicas (SETN).



Naif Haddad was born on May 5, 1959. He obtained his BSc and MSc (H. Diploma) in Architecture and Planning in 1985 and the PhD degree in History/Architectural Heritage and Conservation from the Aristotle University of Thessaloniki, Greece, in 1995. Naif Haddad is a full professor, recently was the Dean of the Faculty of Architecture and Design (2017–2021), and the Dean of Scientific Research and Graduate Studies (2019–2021) at the American University of Madaba (AUM). In addition, recently, he has served as Acting President (Interim President) of AUM. He is the founder and chairman of the Department of Conservation Science at Queen Rania's Faculty of Tourism and Heritage at Hashemite University (2001–2005) in Jordan, where he still serves as a faculty member. He is also a founding member of CulTech for Heritage and Conservation in Jordan (2012). In addition, he is a member of the National Committee for the Protection Architectural and Urban Heritage (2019 to present) by the Prime Ministry/Jordan. As a consultant and heritage expert, he was involved in different international and local conservation, restoration, and management projects. He worked as a historian architect and consultant in several excavations, research programs, and projects to document, interpret, restore, conserve, and reuse historic buildings and monuments in Jordan and Greece (as a permanent collaborator with Aristotle University excavations at Vergina and with I.Z. Trustees for Prehistoric and Classical Antiquities of the Greek Ministry of Culture). He has had more than 80 articles published in refereed and indexed international journals, books, and chapters in books. He is a reviewer of many international refereed journals. His publications and research are in the area of architectural heritage and classical studies, ancient science and technology, 3D digital documentation, edutainment and multimedia, conservation, and sustainable management of historical buildings, sustainable tourism planning, and development. Parallel to his scientific research and heritage studies and projects, Professor Haddad has also worked in multimedia as co-manager, consultant, creative director, art director, and scriptwriter in various TV productions, animated TV spots, documentaries, and outreach materials since 1997.

Daniel Hernández-Balbuena was born on July 25, 1971. He received the BS degree from Puebla Autonomous University, Puebla, México, in 1996 and the MS degree from Ensenada Center for Scientific Research and Higher Education, Baja California, México, in 1999. He received the PhD degree from the Autonomous University of Baja California in 2010. His research interests are in the areas of time and frequency metrology, design and characterization of microwave devices and systems RF measurements, research applications of unmanned aerial vehicles, and image digital processing.

Ivan Kabardin was born on May 16, 1985. He received the BS and MS degrees from Novosibirsk State University, Novosibirsk, Russia, in 2006 and 2008, respectively. His diploma work was awarded the diploma by the Ministry of Education and Science of the Russian Federation. He received the PhD degree from Kutateladze Institute of Thermophysics SB RAS in the specialty "Liquid, Gas and Plasma Mechanics" in 2015. He has been the author of 1 book, 3 book chapters, and 49 papers indexed in SCOPUS (h-index 6), and holds 5 patents in Russia. Since 2004 time, he is represented by his research works in several congresses and conferences of IEEE, EUROMECH, and COARICE in Japan, Italy, Canada, and Russia. He is a senior researcher at the Institute of Thermophysics SB RAS, Novosibirsk, Russia, and director of several master's and doctorate theses. He was a member of the program committees of various international and local conferences. He is a member of EUROMECH. He received the "Russian Federation Government Award in Science and Technology" in 2017.

Huei-Yung Lin is a professor and the director of Robot Vision Lab at National Chung Cheng University. He received his PhD degree in Electrical and Computer Engineering from the State University of New York at Stony Brook, USA. In 2002 he joined the Department of Electrical Engineering, National Chung Cheng University, Taiwan, as an assistant professor and a full professor. He served as the director of Research Liaison Division from 2009 to 2013, and the director of Academic Devel-

opment Division from 2012 to 2014, both with the Office of Research and Development. He is the author of more than 170 journal and conference articles, and has written three book chapters. He also holds 11 US patents and 9 Taiwan patents. He serves as an organizing committee member and a program committee member of more than 50 international conferences. He is the recipient of the "Outstanding Robotics Engineer Award" from Robotics Society of Taiwan, "Outstanding Academic-Industry Cooperation Award" from Taiwan Association of System, and "Science and Engineering, and Excellent Research Award" from National Chung Cheng University. His research interests include robotics, computer vision, machine learning, and image processing. He is a senior member of IEEE/OSA and a fellow of IET.

Lars Lindner was born on July 20, 1981 in Dresden, Germany. He received his MS degree in Mechatronics Engineering from the Dresden Technical University in January 2009. He was working as a graduate assistant during his studies at Fraunhofer Institute for Integrated Circuits EAS in Dresden and also finished his master's thesis there, with the title "Support of Rapid Control Prototyping of Machine Controls by Transformation of Hierarchical State Machine Graphs into IEC 1131 Code of Programmable Logic Controllers." After finishing his studies, he moved to Mexico and started teaching engineering classes at different universities in Mexicali. He also teaches at the Engineering Faculty, Autonomous University of Baja California, campus Mexicali. Since August 2013 he began his PhD studies at the Engineering Institute of the Autonomous University of Baja California in Mexicali, with the project titled "Theoretical Method to Increase the Speed of Continuous Mapping of a Three-Dimensional Laser Scanner Using Servomotor Control," in which he worked in the development of optoelectronic prototypes for the measurement of 3D coordinates, using dynamic triangulation with laser. He concluded his doctoral studies with the thesis and the regulatory examination on January 30, 2017. As one of his PhD works, he completed a research stay at the Fraunhofer Institute IIS-EAS in Dresden,

taking part in the project "Development of a Software Driver for a 2D Laser Scanner" in June/July 2015. Its academic products include various original research articles, published in specialized international journals, articles in national and international congresses, and various book chapters. In the year 2017 his research article "Mobile Robot Vision System Using Continuous Laser Scanning for Industrial Application" in the international journal *Industrial Robot* of the publisher Emerald received the distinction "Outstanding Paper." In September 2017, he was appointed as a Level 1 National Researcher by the National System of Researchers CONACYT for the period 2018–2020. He is working full time at the Engineering Institute of the Autonomous University of Baja California, Department of Applied Physics and is part of the research group "Optoelectronics and Automatic Measurements."

Vladimir Meledin was born on November 24, 1959. He received the specialist degree from Novosibirsk State University, Novosibirsk, Russia, in 1972. His diploma work was awarded a medal by the Ministry of Education and Science of the Russian Federation. He received the Doctor of Technical Science degree from Kutateladze Institute of Thermophysics SB RAS, Novosibirsk, Russia, in 1996. He has been the author of 3 books, 7 book chapters, 99 papers indexed in SCOPUS (h-index 7), and holds 25 patents in Russia. He is the leading researcher at the Institute of Thermophysics SB RAS, director of several master's, and PhD theses. He was a member of the program committees of various international and local conferences. He received the State Prize of the Russian Federation in the field of science and technology in 2019. He received the "Russian Federation Government Award in Science and Technology" in 2014. He is the laureate of the prize of Academic Kosygin in the field of science, technology, and organization of production. He is the winner of the national award "Golden Mercury" by the RF CCI for innovation. He was awarded honorary diplomas of the RAS and A.M. Prokhorov orders 1 and 2 degrees. He is a Russian Skolkovo science foundation expert. He is a member of the Russian Supreme Union of Engineers.

He is a real state member (academic) of the Prokhorov Russian engineering academy. He is the director of the Institute of Electro-optic Information Technologies.

Alexander Menshchikov received his PhD in Data Science from Skolkovo Institute of Science and Technology (Skoltech), Russia, in 2020 and MS in Applied Physics and Mathematics at the Moscow Institute of Physics and Technology (MIPT), Russia, in 2014. He was a visiting student at the Aeroastro Department of Massachusetts Institute of Technology (MIT), USA, in a spring semester in 2015. Alexander's background is in Applied Mathematics and Physics. Before joining Skoltech he had worked as an engineer in Central Aerohydromechanis Institute (TsAGI), Russia (2011–2014), as a space systems engineer in Sputnix and Dauria Aerospace (2015–2016), as an engineer in the Mechatronics and Robotics Group in TOPCON (2017–2018). His research interests include robotics, artificial intelligence, and embedded systems.

Paolo Mercorelli received his PhD in Systems Engineering from the University of Bologna, Bologna, Italy, in 1998. In 1997, he was a visiting researcher for one year with the Department of Mechanical and Environmental Engineering, University of California, Santa Barbara, CA, USA. From 1998 to 2001, he was a postdoctoral researcher with Asea Brown Boveri Corporate Research, Heidelberg, Germany. From 2002 to 2005, he was a senior researcher with the Institute of Automation and Informatics, Wernigerode, Germany. From 2005 to 2011, he was Associate Professor of Process Informatics at Ostfalia University of Applied Sciences, Wolfsburg, Germany. Since 2012, he has been a distinguished full professor and the Chair of Control and Drive Systems at the Institute of Product and Process Innovation, Leuphana University of Lueneburg, Lueneburg, Germany. Since the winter semester of the academic year 2017/2018, he has been an international distinguished visiting professor at the Institute of Automatic Control, Lodz University of Technology, Lodz, Poland. In 1998 he was the recipient of a three-year scholarship from the Marie Curie Actions Research Fellowship Program,

which is one of the most competitive and prestigious European awards sponsored by the European Commission in 1998.

Jesús Elías Miranda-Vega was born in 1984. He received the BS degree in Electrical and Electronic Engineering from Tecnológico Nacional de México/IT de Los Mochis in 2007, the master's degree in Electronic Engineering from Tecnológico Nacional de México/IT de Mexicali in 2014, and the PhD degree in Science and Applied Physics from the Autonomous University of Baja California in December 2019, receiving honorable mention (Cum Laude). He has written five book chapters and several journal articles and conference proceedings. Since 2021 he has been with Tecnológico Nacional de Mexico at the Instituto Tecnologico de Mexicali (ITM) campus. He has been the thesis co-director of a master's student. He has been incorporated into CONACYT National Research System in 2021. His research interests include machine vision, machine learning, data signal processing, the theory and optoelectronics devices, and their applications.

Igor Nevliudov received the MS degree from Kharkiv National Polytechnical University in 1970. He received the PhD degree from the same institution in the specialty "Technology of Instrumental Design" in 1974. He received the postdoc degree (DsC, or habilitation thesis) from Kharkiv National University of Radioelectronics in 1985. He has been the author of 16 books of 9 textbooks, 29 workbooks, and 54 papers indexed in SCOPUS (h-index 4), and holds 16 patents in Ukraine. He is the head of the Computer-Integrated Technologies, Automation, and Mechatronics Department of Kharkiv National University of Radioelectronics and supervisor of several master's, doctorate, and habilitation thesis. He was a member of the program committees of various international and local conferences from 2014 to 2021. He is a member (academician) of the Academy of Applied Radio Electronics of Ukraine. He received the award of "Honored Scientist of Ukraine" in 2013, "State Prize of Ukraine in Field of Technology" in 2014, and

"State Prize of Ukraine in Field of Education" in 2019. His research interests include technology and manufacturing automation for radio-electronic devices.

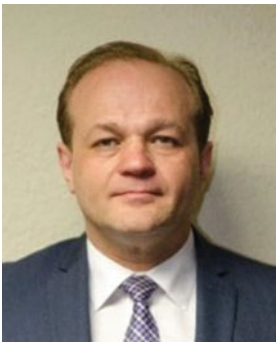
Luis Roberto Ramírez-Hernández was born on June 18, 1988. He received the bachelor's degree in Mechatronic Engineering from the Autonomous University of Baja California in 2010. He received the master's degree in Science in the area of artificial intelligence, with emphasis on the development of algorithms for GPU technology, from the Engineering Institute of the Autonomous University of Baja California in 2015. Also, he received the PhD degree in Science in the area of control and instrumentation, with emphasis on the development of a new method of camera calibration for a stereoscopic vision system, from the Autonomous University of Baja California in 2021. By now, he is the author of one journal article published by Sage in 2020, one journal article published by IEEE in 2021, one book chapter published by Springer and two proceedings articles published by IECON 2019 and ISIE 2020. He is an assistant professor at the Autonomous University of Baja California, teaching electronic science to engineering students. His research interests are camera calibration methods and servoing control applied on stereo vision systems.

Oscar Real-Moreno received the BS degree from the Autonomous University of Baja California in 2014. He received the Master's in Engineering degree from the same university in 2016. He is teaching at the Technological University of San Luis Rio Colorado and studying to obtain his doctorate degree from the Autonomous University of Baja California. In his master's studies he developed an optical scanning prototype as a project for his thesis; he also had articles published in two international conferences (ISIE 2017 and IECON 2018) and a book chapter for the book *Optoelectronics in Machine Vision-Based Theories and Applications*. His research work for his PhD thesis is based on "Spatial Object Detection with Stereo Vision Systems for Autonomous Navigation."

Julio Cesar Rodríguez-Quiñonez received the BS degree from CETYS, Mexico, in 2007. He received the PhD degree from the Autonomous University of Baja California, México, in 2013. He is a full-time researcher-professor in the Engineering Faculty of the Autonomous University of Baja California and member of the National Research System Level 1. Since 2016, he is a senior member of IEEE. He is involved in the development of an optical scanning prototype in the Applied Physics Department and is a research leader in the development of a new stereo vision system prototype. He has been thesis director of three doctoral students and four master's students. He has registered two patents on the dynamic triangulation method; has been the editor of four books, a guest editor of the *IEEE Sensors* journal, *International Journal of Advanced Robotic Systems*, and *Journal of Sensors*; has written over 80 papers, and eight book chapters; and has been a reviewer for the *IEEE Sensors* journal, *Optics and Lasers in Engineering*, *IEEE Transaction on Mechatronics and Neural Computing and Applications* (Springer); he participated as a reviewer and session chair of IEEE ISIE conferences in 2014 (Turkey), 2015 (Brazil), 2016 (USA), 2017 (UK), 2019 (Canada), IECON 2018 (USA), IECON 2019 (Portugal), ISIE 2020 (the Netherlands), and ISIE 2021 (Japan). His research interests include automated metrology, stereo vision systems, control systems, robot navigation, and 3D laser scanners.

Nicholas Santavas received the diploma degree from the Department of Production and Management Engineering, Democritus University of Thrace, Greece, in 2018. He is working toward the PhD degree in the Laboratory of Robotics and Automation (LRA), Department of Production and Management Engineering, Democritus University of Thrace (DUTH), Greece. His research interests include artificial neural networks and human-computer interaction applications.

Cesar Antonio Sepúlveda-Valdez received the Engineering and master's degree from the Autonomous University of Baja California (UABC), Mexico, in 2017 and 2020, respectively. In 2018 he joined the Physics Department of the Engineering Institute at UABC, where he collaborated on researches in the optoelectronics field. He is working to get a PhD degree from UABC. Until now, he has written two publications for international congresses of IEEE in Canada (ISIE 2019) and the Netherlands (ISIE2020) and a book chapter for IGI Global. He received the award of "Best Session Presentation" at ISIE 2019 in Vancouver, Canada. He also has been a reviewer for the mentioned congresses. His research interests include robot navigation and communication, signal processing, vision systems, and 3D laser scanners.



Oleg Sergiyenko was born on February 9, 1969. He received the BS (Honoris Causa) and MS degrees from Kharkiv National University of Automobiles and Highways, Kharkiv, Ukraine, in 1991 and 1993, respectively. He received the PhD degree from Kharkiv National Polytechnic University in the specialty "Tools and Methods of Non-destructive Control" in 1997. He received the postdoc degree (DsC, or habilitation thesis) from Kharkiv National University of Radioelectronics in 2018. He is the author of 1 book (and editor of 7 books), 23 book chapters, and 128 papers indexed in SCOPUS (h-index 15), and holds two patents in Ukraine and one patent in Mexico. Since 1994, he is represented by his research works in several international congresses of IEEE, ICROS, SICE, and IMEKO in the USA, England, Japan, Italy, Austria, Scotland, Portugal, Canada, Ukraine, Turkey, Brazil, and Mexico. He is the head of the Applied Physics Department, Engineering Institute, Autonomous University of Baja California, Mexico and director of several master's and doctorate theses. He was a member of the program committees of various international and local conferences, participating annually as session chair of IEEE ISIE and IECON conferences in 2014–2020. He is a recognized reviewer of *Elsevier*, constant reviewer of *IEEE Transaction on Industrial Electronics*, and so on. He is

a member (academician) of the Academy of Applied Radioelectronics of Ukraine, Russia, and Belorussia. He received the award of "Best Session Presentation" in IECON2014 in Dallas, USA, IECON2016 in Florence, Italy, and ISIE2019 in Vancouver, Canada. He received as co-author the award of "Outstanding Paper in the 2017 Emerald Literati Network Awards for Excellence."
He was the editor and author of several books by international publishers, including his last book with Springer Nature, *Machine Vision and Navigation*.

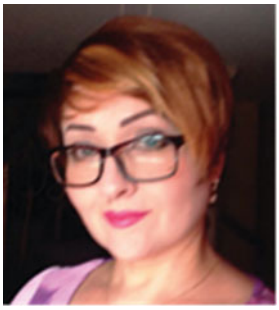
Andrey Somov is an assistant professor at Skolkovo Institute of Science and Technology (Skoltech), Russia. From 2018 to 2021 he served as Deputy Head of Skoltech's NTI Center of Excellence in Internet of Things (IoT) and Wireless Communication. He graduated from 'MATI'–Russian State Technological University, Russia (2004) and holds a diploma in Electronics Engineering from the same institution (2006). Andrey received his PhD (2009) from the University of Trento, Italy, for his work in the field of power management in wireless sensor networks (WSN). Before joining Skoltech (2017), he had worked as a senior researcher for FBK CREATE-NET Research Center, Italy (2010–2015) and as a research fellow for the University of Exeter, UK (2016–2017). In the fall 2008 he was a visiting scholar at the University of California, Berkeley, USA. Andrey has had more than 100 papers published in peer-reviewed international journals and conferences. His research interests include intelligent sensing, artificial intelligence, cognitive IoT and associated proof-of-concept implementation. Dr. Somov holds some awards in the fields of WSN and IoT, including the Google IoT Technology Research Award (2016) and the Best Paper Award at IEEE IoP conference (2019).

Alexey Strelnik was born on July 31, 1994 in Ust-Kamenogorsk, Kazakhstan. From 2012 to 2018 he studied at the Faculty of Physics of Novosibirsk State University. Since 2015, he has been working as a laboratory assistant and research engineer at Kutateladze Institute of Thermophysics SB RAS. He is a specialist in the field of digital signal processing and automation of physical and technical research. Strelnik participated in the study of gas phase detection in a liquid-metal coolant. He developed the approach for diagnosing the flow structure in optically opaque media.

Gabriel Trujillo-Hernández was born on October 16, 1992. He received the BS (Mechatronic Engineering) and MS degrees from Universidad Autónoma de Baja California, Mexicali, Baja California, in 2014 and 2020, respectively. Since 2018, he is represented by his research works in several international congresses of IEEE, Portugal, and the Netherlands. Since 2018, he is Professor of Mechanical Design, Calculus, and Metrology at the Universidad del Valle de Mexico. He has participated in regional technical posters sponsored by Sensata in Aguascalientes, México.

Oleksandr Tsymbal graduated from the Faculty of Radio Physics and Electronics, Kharkiv State University in 1992. He received the PhD degree from the Technological Institute of Instrument-Making in 1996 in the specialties "Computer-Aided Design" and "Automation of Technological Processes." In July 2008 he was invited as a guest researcher at Robinson-College, University of Cambridge, UK. He received the postdoc degree (DsC, or habilitation thesis) from Kharkiv National University of Radioelectronics (NURE) in 2015 for the thesis "Methods and Models of Intellectual Decision-making Support for Automatized Control of Flexible Integrated Manufacturing." He has been the author of three monographs, four workbooks, and eight papers indexed in SCOPUS. He is a professor in the Computer-Integrated Technologies, Automation, and Mechatronics Department of

Kharkiv National University of Radioelectronics and supervisor of several master's and doctorate theses. His research interests include logical programming, decision-making, and computer vision systems.



Vera V. (Vira) Tyrsa received the BS and MS degrees from Kharkov National University of Automobiles and Highways, Kharkov, Ukraine, in 1991 and 1993, respectively (Honoris Causa). She received the PhD degree from Kharkov National Polytechnic University in the specialty "Electric Machines, Systems and Networks, Elements and Devices of Computer Technics" in 1996. She has written 1 book, 7 book chapters, and more than 76 papers indexed in SCOPUS (h-index 20). She holds one patent in Ukraine and one patent in Mexico. From 1994 onward, she is represented by her research works in international congresses in the USA, England, Italy, Japan, Ukraine, and Mexico. In April 1996, she joined Kharkov National University of Automobiles and Highways, where she held the position of an associate professor in the Electrical Engineering Department (1998–2006). From 2006 to 2011, she was invited by the Polytechnic University of Baja California, Mexico, for a professor and researcher position. She is a professor teaching electronic topics at the Engineering Faculty, Autonomous University of Baja California. She was the co-author of the paper that received the "Emerald Literati Award of 2017 Outstanding Paper" in *Industrial Robot*, an international journal. Her research interests include automated metrology, machine vision systems, fast electrical measurements, control systems, robot navigation, and 3D laser scanners.

Jorge Alejandro Valdez-Rodriguez received the bachelor's degree in Mechatronics engineering from Tecnologico Nacional de Mexico/IT Mexicali in 2010 and the master's degree in Aerospace Engineering from CETYS in 2015. He is Professor of Engineering at CETYS since 2015 and started, in 2019, working toward the PhD degree in Engineering at the Autonomous University of Baja California, Faculty of Engineering, focused on inertial navigation systems. Since 2021 he is a student member of IEEE. He actively participates in SAE Aerospace Design since

2021 as faculty advisor of the Fox Force engineering student group. His research interests include navigation, signal processing, control systems, aerospace, and aerodynamics.

Contributors

Ruben Alaniz-Plata Engineering Institute of Autonomous University of Baja California, Mexicali, BC, Mexico

Humberto Andrade-Collazo Engineering Institute of Autonomous University of Baja California, Mexicali, BC, Mexico

Mounya Belghiti Université de Picardie Jules Verne, MIS Laboratory, Amiens, France

Vikrant Bhateja Shri Ramswaroop Memorial Group of Professional Colleges (SRMGPC), Lucknow, UP, India

Artem Bronnikov Department of Computer-Integrated Technologies, Automation and Mechatronics of Kharkiv National University of Radioelectronics, Kharkiv, Ukraine

Guillaume Caron Université de Picardie Jules Verne, MIS Laboratory, Amiens, FranceCNRS-AIST JRL (Joint Robotics Laboratory), National Institute of AIST, Ibaraki, Japan

Moisés J. Castro-Toscano Faculty of Engineering of Autonomous University of Baja California, Mexicali, BC, Mexico

Yu-Ting Chen Department of Electrical Engineering, National Chung Cheng University, Chiayi, Taiwan

Anthony Dessaux Université de Picardie Jules Verne, MIS laboratory, Amiens, France

Sergey Dvoynishnikov Kutateladze Institute of Thermophysics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russian Federation

Wendy Flores-Fuentes Engineering Faculty of Autonomous University of Baja California, Mexicali, BC, Mexico

Antonios Gasteratos Department of Production and Management Engineering, Democritus University of Thrace, Thrace, Greece

Naif Haddad Queen Rania Faculty of Tourism and Heritage, Department of Conservation Science, Hashemite University, Zarqa, Jordan

Daniel Hernández-Balbuena Faculty of Engineering of Autonomous University of Baja California, Mexicali, BC, Mexico

Ivan Kabardin Kutateladze Institute of Thermophysics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russian Federation

Huei-Yung Lin Department of Electrical Engineering, National Chung Cheng University, Chiayi, Taiwan

Lars Lindner Engineering Institute of Autonomous University of Baja California, Mexicali, BC, Mexico

Vladimir Meledin Kutateladze Institute of Thermophysics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russian Federation

Alexander Menshchikov Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia

Paolo Mercorelli Institute of Product and Process Innovation, Leuphana University of Lueneburg, Lueneburg, Germany

Jesús Elías Miranda-Vega Tecnológico Nacional de México/ IT de Mexicali, Mexicali, Baja California, México

Igor Nevliudov Department of Computer-Integrated Technologies, Automation and Mechatronics of Kharkiv National University of Radioelectronics, Kharkiv, Ukraine

Luis Roberto Ramírez-Hernández Faculty of Engineering of Autonomous University of Baja California, Mexicali, BC, Mexico

Oscar Real-Moreno Faculty of Engineering of Autonomous University of Baja California, Mexicali, BC, Mexico

Cesar Antonio Sepúlveda-Valdez Engineering Faculty of Autonomous University of Baja California, Mexicali, BC, Mexico

Nicholas Santavas Department of Production and Management Engineering, Democritus University of Thrace, Thrace, Greece

Cesar Sepulveda-Valdez Engineering Institute of Autonomous University of Baja California, Blvd. Benito Juarez y Calle de La Normal, Mexicali, BC, Mexico

Oleg Sergiyenko Engineering Institute of Autonomous University of Baja California, Mexicali, BC, Mexico

Andrey Somov Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia

Alexey Strelnik Kutateladze Institute of Thermophysics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russian Federation

Gabriel Trujillo-Hernández Faculty of Engineering of Autonomous University of Baja California, Mexicali, BC, Mexico

Oleksandr Tsymbal Department of Computer-Integrated Technologies, Automation and Mechatronics of Kharkiv National University of Radioelectronics, Kharkiv, Ukraine

Vera V. (Vira) Tyrsa Engineering Institute of Autonomous University of Baja California, Mexicali, BC, Mexico

Jorge Alejandro Valdez-Rodriguez Faculty of Engineering of Autonomous University of Baja California, Mexicali, BC, Mexico

Abbreviations

2D	2 Dimensions
3D	3 Dimensions
ACO	Ant Colony Optimization
ACS	Automated Control System
AD	Anteposterior Diameter
Adam	Adaptive Moment Estimation
ADC	Analog-Digital Converter
ADCS	Attitude Determination and Control System
AI	Artificial Intelligence
APF	Artificial Potential Field
API	Application Programming Interface
AUC	Area Under the Curve
BFO	Bacterial Foraging Optimization
BIM	Building Information Modeling
BM	Block Matching
BN	Batch Normalization
BP	Belief Propagation
BPR	Bad Pixel Rate
CAS	Active Stereo Coordinate Reference Frame
CCD	Charge-Coupled Device
CCW	Counter Clockwise
CD	Cell Decomposition
CH	Cultural Heritage
CMOS	Complementary Metal Oxide Semiconductor
CNN	Convolutional Neural Network
COCO	Common Objects in Context
CPM	Convolutional Pose Machines
CPU	Central Processing Unit
CS	Cuckoo Search algorithm
CS	Creative Suite
CW	Clockwise

DC	Direct Current
DCD	Digital count distortion
DCM	Direction Cosine Matrix
DDJ	Data-Driven Jitter
DL	Deep Learning
DLT	Direct Linear Transformation
DoF	Degree-of-Freedom (plural: DoFs) IR Infra-Red
DSP	Digital Signal Processor
EKF	Extended Kalman Filter
EMF	Electro-Motive Force
FA	Firefly Algorithm
FCNN	Fully Convolutional Neural Network
FIS	Flexible Integrated Systems
FL	Fuzzy Logic
FLOPS	Floating Point Operation per Second
FOV	Field of View
FPGA	Field-Programmable Gate Array
FPS	Frames Per Second
FWFF	Fuzzy Well-Form Formula
GA	Genetic Algorithm
GC	Graph Cut
GCI	Getty Conservation Institute
GIS	Geographical Information System
GNSS	Global Navigation Satellite System
GPU	Graphics Processing Unit
GSD	Ground Sample Distance
HA	Hardware Acceleration
H-BIM	Heritage BIM
HCI	Human Computer Interaction
HMI	Hebal Malas Index
HPE	Human Pose Estimation
HSC	Hardware Software Complex
HSL	Hue, Saturation, Lightness
ICOMOS	International Council on Monuments and Sites
ICP	Iterative Closet Point
IMU	Inertial Measurement Unit
INS	Inertial Navigation System
IoT	Internet of Things
IR	Infra-Red
IT	Information Technology
LAA	Longitudinal Arch Angle
LMS	Least Mean Square
LSS	Laser Scanner System
MADS	Mesh Adaptive Direct Search
MCU	Micro Controller Unit

MHA	Multi-Head Attention
ML	Machine Learning
MM	Malleolus Medial
MR	Mobile Robot
MT1	First Metatarsal
N/A	Not Applicable
NAS	Neural Architecture Search
NCC	Normalized Cross Correlation
NCS	Neuro Compute Stick
NDVI	Normalized Difference Vegetation Index
NED	North, East, Down
NEU	North, East, Up
NLP	Natural Language Processing
NN	Neural Network
NNAPI	Android Neural Network API
ODM	Original Design Manufacturer
OLED	Organic Light-Emitting Diode
OMA	Other Miscellaneous Algorithm
OSV	Omnidirectional Stereo Video
OVINS	Omnidirectional Visual Inertial Navigation System
PA	Precision Agriculture
PC	Pectus Carinatum
PCK	Percentage of Correct Keypoints
PE	Pectus Excavatum
PJ	Periodic Jitter
PL	Positioning Laser
PM	Polygon Mesh
PSO	Particle Swarm Optimization
R.U.R.	Rossumovi Univerzální Roboti
RA	Roadmap Approach
RADAR	Radio Detection and Ranging
RAM	Random Access Memory
RBA	Routing by Agreement
ReLU	Rectified Linear Unit
RG	Robotic Group
RGB	Red Green Blue
RGBD	RGB plus Depth
RJ	Random Jitter
RM	Roadmap Method
RMIS	Robot-assisted Minimally Invasive Surgery
ROC	Receiver Operating Characteristic Curve
ROI	Region of Interest
ROS	Rotational Optical Scanner
RPi	RaspberryPi
SA	Scanning Aperture

SAD	Sum of Absolute Differences
SBC	Single Board Computer
SDA	Scale Distance Agreement
SFLA	Shuffled Frog Leaping Algorithm
SGM	Semi-Global Matching
SLAM	Simultaneous Localization and Mapping
SNR	Surgical Navigation Robot
SONAR	Sound Navigation and Ranging
SPA	Screened Poisson Algorithm
SSD	Sum of Squared Differences
SVS	Stereoscopic Vision System
TCA	Tourism and Culture Authority
TLS	Terrestrial Laser Scanner
TN	Tuberosity Navicular
TOF	Time of Flight
TPU	Tensor Processing Unit
TVS	Technical Vision System
UAS	Unmanned Aerial Systems
UAV	Unmanned Aerial Vehicle
UGV	Unmanned Ground Vehicle
VINS	Visual Inertial Navigation System
VIO	Visual Inertial Odometry
VP	Video-Projector (plural: VPs)
VPU	Visual Processing Unit
WFF	Well-Form Formula
WSN	Wireless Sensor Networks
WTA	Winner-Take-All
XAI	Explainable AI

3D Model-Based Tracking of Puppet in Depth Images for the Dynamic Video-Mapping of Its Suit

Guillaume Caron, Mounya Belghiti, and Anthony Dessaux

Acronyms

2D	Bidimensional
3D	Tridimensional
VP	Video-projector (plural: VPs)
DoF	Degree-of-Freedom (plural: DoFs)
IR	Infrared
RGB	Red Green Blue
RGBD	RGB plus Depth
GPU	Graphics Processing Unit
RAM	Random Access Memory

1 Introduction

1.1 Motivation

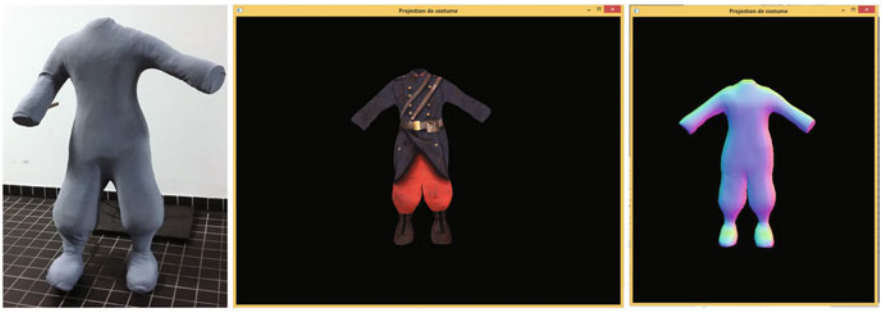
Video-mapping is well known for creative shows where images, animations, or movies are video-projected on buildings or objects [1], recently interactively [2]. Common video-mapping is done on static objects. The dynamic video-mapping, i.e., the coherent video-mapping of images or animations over moving objects, is

G. Caron (✉)
Université de Picardie Jules Verne, MIS laboratory, Amiens, France

CNRS-AIST Joint Robotics Laboratory, National Institute of AIST, Tsukuba, Japan
e-mail: guillaume.caron@u-picardie.fr

M. Belghiti · A. Dessaux
Université de Picardie Jules Verne, MIS laboratory, Amiens, France

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1_1



(a) (b) (c)

Fig. 1 (**a**) The puppet considered for the show, i.e., the "screen" for video-projection. (**b**) The textured 3D model of the puppet. (**c**) The 3D model (mesh with colors showing surface normal orientation) corresponding to the actual puppet

still an open research topic [3–5]. The challenge is to track the object and video-project augmentation data coherently on the actual object, both fast enough for the success of the video-mapping, i.e., perfect mapping on the object before it moves again.

This chapter fits the latter topic furthermore with articulated object manipulated by a human. More precisely, the aim is to detail a dynamic video-mapping framework for live puppet shows where a puppeteer moves and actuates a textureless puppet (Fig. 1a) on which a video-projector (VP) maps a suit (Fig. 1b).

The puppet considered in this chapter is designed by artists (acknowledged in Sect. 6) to correspond to a given creative puppet theater play. The puppet has neither head nor hands (Fig. 1a) to fit the play scenario where the puppeteer may put his own head on top of the puppet.

Each puppet arm or leg has three degrees of freedom (DoFs): two joints for each shoulder and each ankle and one joint for each elbow and each knee. Thus, each arm or each leg has three DoFs, leading to a total of twelve joints. Hence, in addition to the six Cartesian DoFs of the puppet tridimensional (3D) pose, this leads to estimate in real time 18 DoFs, used to animate a virtual model of the puppet (see Fig. 1c, shown untextured to better see the 3D shape) and then displayed by the VP. This chapter focuses on the estimation of the puppet DoFs by visual observation, i.e., relying on a vision sensor. So, to make the mapping coherent, the VP and the vision camera must be calibrated together.

1.2 Related Works

To handle the previously mentioned challenges, existing works have first put infrared (IR) reflectors on the object to be tracked by a motion capture system, e.g., a

Vicon system as for the Omote face tracking project.[1] Carefully selected IR markers are measured by the capture system but a little visible by spectators. Furthermore, marker-based techniques may deal with deforming object shapes even with a more basic RGB (red, green, and blue) color camera [6]. The main drawback of markers is the need to equip the object to track and to design smartly the arrangement of markers, depending on the object itself. Hence, the latter approaches need tedious prior setups to make dynamic video-mapping possible.

Inversely, marker-less techniques are faster to set up but require more processing to track the object to video-map. Geometric information is considered for this purpose such as contours in bidimensional (2D) IR imaging [3, 7], point clouds with RGBD (RGB Depth) imaging [8] or key points, in 2D RGB imaging, got from the video-projection over the object to track itself [9]. In all works mentioned above, once the object 3D pose (and deformation parameters for some) is estimated, the pose of the data to display is updated and video-projected on the object to augment (black color out of it). The object must, hence, be moved in the field of a static single or multiple VPs [10, 11].

Another group of approaches adapts the video-projection system field of view with multi-reflections on orientable mirrors in order to display only on the particularly considered object [12]. Combined with a high-speed camera-based tracking system, the latter work is very reactive and deals with highly dynamic moving objects, however constraining the shape of the object to the field of projection shape as circular [13] or spherical [5] targets. The latter is illustrated in Fig. 2 where the video-projector alone, at coordinate frame F_p, displays a lava planet in a direction agnostic to the scene content: the display does not consider the actual pose of the

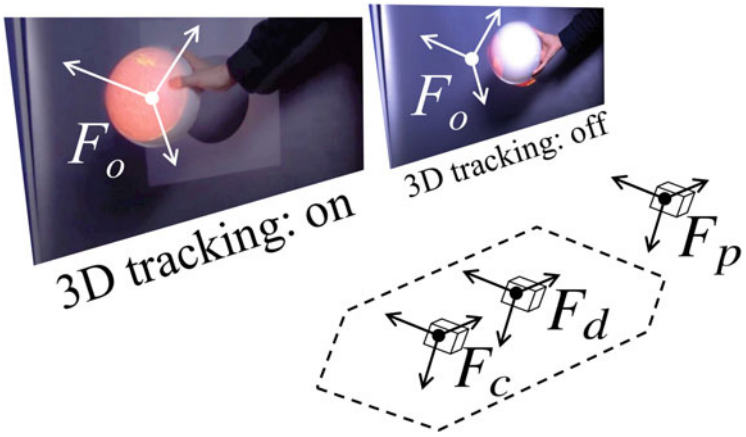


Fig. 2 Dynamic video-mapping system principle with examples of video-mapping on a sphere [12]

[1] http://projection-mapping.org/omote.

white sphere when the 3D tracking is off, and thus the lava planet is misaligned with the white sphere (Fig. 2 background, right). Instead, when the 3D tracking system (bordered with dashed lines in Fig. 2, here with a depth camera, of frame F_d, rigidly attached to a color camera, of frame F_c) is on, the pose of the white sphere, of frame F_o, is computed and used to update the virtual lava planet video-projected on the scene, thus on the white sphere (Fig. 2 foreground, left).

The video-projection depth of field can also be adapted dynamically to ensure the video-mapping is always sharp, however only shown on planar objects fronto-parallel to the VP [14].

Siegl et al. [10] and Lange et al. [2, 11] manage to track the object of interest of a shape that is not constrained by the projector field (the statue of a head) thanks to projective iterative closest point, applied to 3D point clouds got from successive RGBD acquisitions, at the price of a Graphics Processing Unit (GPU) implementation to overcome the algorithm complexity bottleneck. Direct acquisition of the depth information in front of the camera is generally preferred to stereo vision [15] in this field, although both require calibration to obtain precise depth [16, 17]. Indeed, video-mapping changes the appearance of the object to track, obviously making stereo-matching algorithms, which also spend time to run, to fail. Laser scanning also enables depth acquisition independently of the object illumination [18]. However, both the capture of every depth information at once in the RGBD camera field of view and its multimodal (RGB and depth) capture in a single setup ease the whole video-mapping system calibration and the object tracking process compared to point-by-point depth capture [19].

The auto-calibration of the RGBD camera/VP pair can further be handled extending structured light-based techniques [20] but neither for deformed nor for articulated objects as some elements must remain constant to make the 3D tracking converge in a short period of time. The tracking algorithms complexity is thus generally avoided by dealing with planar surfaces [21, 22] or deformed ones but able to be fitted on parametric surfaces, e.g., Bezier surfaces [3]. With such constraints, articulated objects cannot be considered. One can imagine other 3D tracking of articulated object to be integrated in a video-mapping framework but despite recent progresses allowing to track an articulated object at 350 Hz [23], such high frequency is reached only considering 4 joint DoFs, whereas an entire puppet, as the one considered in this chapter, features 18 DoFs.

1.3 Real-Time Estimation of Every Puppet DoF

To estimate every puppet DoF, depth images of RGBD camera are considered because such camera brings 3D depth measurements not disturbed by the VP light, as previous works highlight [2, 8, 10, 11]. Indeed, since a lowly textured suit is displayed on the puppet, one cannot consider RGB images. Other advantages of considering RGBD vision are regarding compactness, i.e., the tracking system must

be invisible to spectators, and regarding cost as puppetry is a craft art with low budgets (see the consumer-grade sensor Kinect v2, in experiments of Sect. 4).

However, even if many RGBD-based human tracking algorithms are available [24], the lack of head and hands of the puppet makes existing software to fail (no skeleton detected at all on the puppet). Thus, an alternative tracking approach in the depth images of the puppet moving randomly on the theater stage exploits the silhouette of the 3D model of the puppet (Fig. 1c). Considering the silhouette only to recover the 3D pose of the puppet is one of the key elements enabling the real-time tracking since the processing is fast.

The closest related work exploiting the silhouette of some 3D model to track and estimate the 3D pose of an object is [25]. It considers every object contour in a virtual depth image and minimizes the distances between tangent lines to these contours and contour points detected in the RGB image plane. Thus, a combination of geometry and appearance is made for the nonlinear optimization of the object 3D pose. The puppet 3D tracking algorithm uses close image processing (Sect. 3.2), but, compared to [25], the main differences are:

- Using a captured depth image, contrary to an RGB image (no appearance)
- Considering the silhouette only (no internal contour)
- Depth image processing leads to matched 3D points, instead of 2D–3D matches
- Robust and time efficient registration between 3D–3D correspondences

Another work [26] tackled the estimation of robotic arm joint angles from the edges of its observation by an RGB camera. In the latter work, salient edges characterize the robot shape, and these edges are considered to form the wireframe 3D model of the robotic arm. Point-to-edge 3D model-based tracking is implemented with two articular DoFs. If the problematic is very similar to the puppet tracking context, however focusing on a unique arm of static base, the main differences of the puppet 3D tracking algorithm (Sect. 3) are in the following items, compared to [26]:

- Considering two arms and two legs of three joints each
- Considering a floating base
- Considering limb silhouette only (no salient edge on the puppet, see Fig. 1a)
- Registration is made in 3D space, from points only.

Directly considering correspondences of 3D points, leading, thus, to the direct computation of the 3D pose and joint angles in the tracking process, costs less computation power than in 2D [25, 26], since projections in the image plane at each iteration of the pose optimization are avoided.

Finally, none of the other 3D model-based tracking approaches using contour [25, 26] considers the dynamic video-mapping challenging context.

1.4 Chapter Outline

The rest of the chapter is organized as follows. First, the geometrical modeling of the dynamic video-mapping setup is introduced as well as its calibration (Sect. 2). Then, Sect. 3 presents the 3D model-based tracking from the object silhouette. Finally, results are shown (Sect. 4) before discussion (Sect. 5) and the conclusion (Sect. 6).

2 Video-Mapping Setup Modeling and Calibration

2.1 Geometrical Modeling

The dynamic video-mapping setup is geometrically described by four frames, of which three are fixed, i.e., F_p, F_d, and F_c, respectively, the VP frame, the depth camera frame, and the color camera frame (Fig. 2). Both latter frames are in the RGBD camera. The last frame F_o is associated with the object (i.e., the puppet), thus mobile. Then, geometrically, the set of *extrinsic parameters*[2] of the latter setup is made of the following frame changes:

- ${}^{c}\mathbf{M}_{d}$: the frame change from F_d to F_c
- ${}^{p}\mathbf{M}_{c}$: the frame change from F_c to F_p
- ${}^{d}\mathbf{M}_{o}$: the frame change from F_o to F_d

Both former frame changes are constant since the RGBD camera is static as well as the VP. Thus, the puppet tracking (partly) consists in computing ${}^{d}\mathbf{M}_{o}$ in real time.

Finally, intrinsic parameters of depth and color cameras are also considered as well as for VP. Assuming the perspective projection fits these three devices, three sets ${}^{d}\boldsymbol{\gamma}$, ${}^{c}\boldsymbol{\gamma}$, and ${}^{p}\boldsymbol{\gamma}$, respectively, of common intrinsic parameters are considered. With superscript $^{\bullet}$ and subscript $_{\bullet}$ indicating frames d, c, or p, each ${}^{\bullet}\boldsymbol{\gamma}$ includes four parameters. ${}^{\bullet}\alpha_u$ and ${}^{\bullet}\alpha_v$, for combinations of the focal length and the pixel ratio, and $({}^{\bullet}u_0, {}^{\bullet}v_0)$ for the principal point, leading to the perspective projection function expression ($pr_{{}^{\bullet}\boldsymbol{\gamma}}$ ()):

$$ {}^{\bullet}\mathbf{u} = pr_{{}^{\bullet}\boldsymbol{\gamma}}\left({}^{\bullet}\mathbf{X}\right) = \begin{pmatrix} {}^{\bullet}\alpha_u \frac{{}^{\bullet}X}{{}^{\bullet}Z} + {}^{\bullet}u_0 \\ {}^{\bullet}\alpha_v \frac{{}^{\bullet}Y}{{}^{\bullet}Z} + {}^{\bullet}v_0 \end{pmatrix}, \tag{1} $$

with ${}^{\bullet}\mathbf{X} = ({}^{\bullet}X, {}^{\bullet}Y, {}^{\bullet}Z, 1)^{\mathsf{T}}$, a 3D point expressed in frame $F_{\bullet}$, and ${}^{\bullet}\mathbf{u} = ({}^{\bullet}u, {}^{\bullet}v)^{\mathsf{T}}$ its projection in the digital image plane.

[2] Extrinsic parameters are meant for the parameters of the system made of the RGBD camera and the VP, which are external to each element (RGB camera, depth camera, VP) of that system [27]. They are the Euclidean rigid transformations between frames of pairs of elements of the system. Inversely, *intrinsic parameters* refer to parameters that are internal to each element of the RGBD camera and VP system. They count parameters about lenses, sensor, etc.

2.2 Calibration

Two steps are considered for offline calibration: the RGBD camera calibration itself and the VP intrinsic and extrinsic calibration with respect to the RGBD camera frame.

2.2.1 RGBD Camera Calibration

First, the RGBD camera is calibrated exploiting the calibration toolbox[3] of [16], considering 24 pairs of color and depth images of a sphere at various positions. It computes intrinsic parameter sets ${}^{d}\boldsymbol{\gamma}$ and ${}^{c}\boldsymbol{\gamma}$, and the extrinsic frame change ${}^{c}\mathbf{M}_{d}$, that, farther in the chapter, is represented as a vector, ${}^{c}\mathbf{r}_{d} = [t_X, t_Y, t_Z, \theta w_X, \theta w_Y, \theta w_Z]^{\mathsf{T}}$. t_X, t_Y, and t_Z are the translations along the three axes and θ the rotation angle around the unit axis of rotation of coordinates $[w_X, w_Y, w_Z]^{\mathsf{T}}$. Formulas of Rodrigues [28] allow to compute ${}^{c}\mathbf{r}_{d}$ from ${}^{c}\mathbf{M}_{d}$ and ${}^{c}\mathbf{M}_{d}$ from ${}^{c}\mathbf{r}_{d}$.

2.2.2 Video-Projector Calibration

To compute the VP intrinsic parameters ${}^{p}\boldsymbol{\gamma}$ and extrinsic ones ${}^{p}\mathbf{M}_{c}$, a color-coded calibration pattern is considered. It is made of dots [29] in order to make possible the automatic matching between the 2D coordinates ${}^{p}\mathbf{u}_{i}^{*} = \left({}^{p}u_{i}^{*}, {}^{p}v_{i}^{*}\right)^{\mathsf{T}}$ of their centers (Fig. 3a) and the 3D coordinates ${}^{c}\mathbf{X}_{i} = ({}^{c}X_{i}, {}^{c}Y_{i}, {}^{c}Z_{i}, 1)^{\mathsf{T}}$, expressed in F_c, of their observation by the RGBD sensor. Indeed, each 3×3 block of color dots is unique in the pattern.

Concretely, a white board is set at various orientations in both VP and RGBD camera fields and several RGBD images are acquired. Then, each color image is considered for automatic dots matching (Fig. 3), and the associated depth image provides the ${}^{c}Z_{i}$ value corresponding to each ${}^{c}\mathbf{u}_{i} = ({}^{c}u_{i}, {}^{c}v_{i})^{\mathsf{T}}$ dot center in the color image. Then, the full-scale 3D coordinates are computed thanks to ${}^{c}\boldsymbol{\gamma} = \{{}^{c}\alpha_{u}, {}^{c}\alpha_{v}, {}^{c}u_{0}, {}^{c}v_{0}\}$ and the well-known inverse perspective projection equations:

$$\begin{cases} {}^{c}X_{i} = {}^{c}Z_{i}\frac{{}^{c}u_{i} - {}^{c}u_{0}}{{}^{c}\alpha_{u}} \\ {}^{c}Y_{i} = {}^{c}Z_{i}\frac{{}^{c}v_{i} - {}^{c}v_{0}}{{}^{c}\alpha_{v}}. \end{cases} \tag{2}$$

[3] Available on request to the authors.

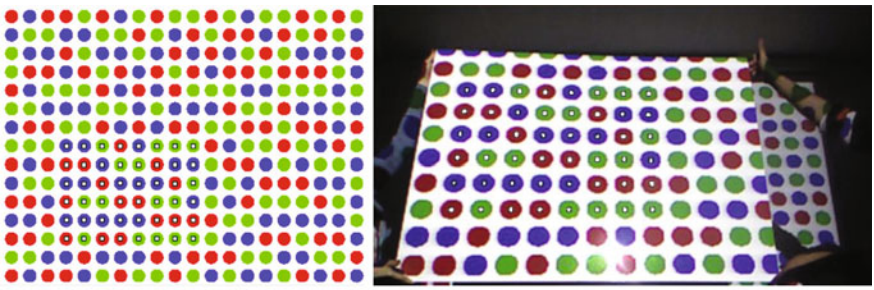


Fig. 3 (**a**) Color-coded pattern for automatic matching for video-projector calibration. (**b**) Automatic matching result for visible dots (white filled squares at the centers of matched dots, corresponding to the ones of (**a**))

Finally, the entire set of 2D–3D correspondences, i.e., pairs of ${}^{p}\mathbf{u}_i^*$ and ${}^{c}\mathbf{X}_i$, is considered together in the following optimization problem:

$$\begin{bmatrix} {}^{p}\hat{\mathbf{r}}_c \\ {}^{p}\hat{\boldsymbol{\gamma}} \end{bmatrix} = \arg \min_{{}^{p}\mathbf{r}_c, {}^{p}\boldsymbol{\gamma}} ||pr_{{}^{p}\boldsymbol{\gamma}} \left({}^{p}\mathbf{M}_c \; {}^{c}\mathbf{X}_i \right) - {}^{p}\mathbf{u}_i^*||. \tag{3}$$

The latter is solved by the Gauss–Newton optimization as for classical camera calibration [27], only considering that the 2D digital image points are virtually "observed" by the VP and the corresponding calibration target 3D points are the ones got from the RGBD camera as explained above.

3 Silhouette-Based Visual Tracking

The goal is to compute ${}^{d}\mathbf{M}_o$, i.e., the puppet pose in the depth camera frame, as well as the four arms and legs configurations $\mathbf{q}_m = [q_1, q_2, q_3]^{\mathsf{T}}$ ($m = 1, 2, 3$ or 4), from depth image measurements. The algorithm considers the captured depth image (Fig. 4a) only, to track the puppet and to compute its pose and joint angles. Indeed, recall that the dark environment of the theater stage and the video-projection of the puppet suit has no impact on the captured depth image, contrary to the RGB image, as theoretically awaited and experimentally observed.

This section focuses on the tracking and pose estimation but not on the detection problem that is solved by imposing to the puppeteer a first puppet pose, approximately fronto-parallel to the RGBD camera and with the puppet chest approximately aligned with the RGBD camera main axis.

Thus, we consider the tracking and pose estimation problems under the family of 3D model-based tracking (see [30, 31] for detailed surveys). The main idea is to consider an initial pose of the tracked object at which its 3D model can be projected

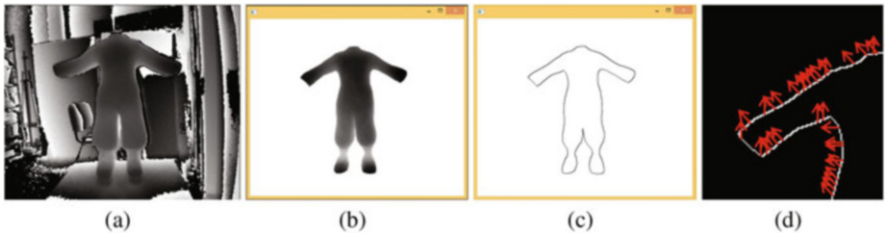


Fig. 4 (**a**) A depth image of the puppet from the Kinect v2 RGBD camera. (**b**) A Z-buffer image of the puppet 3D model with the same intrinsic parameters as the actual depth camera. (**c**) The resulting silhouette pixels computation from (**b**) (black pixel is a silhouette pixel, white not). (**d**) Zoom on the local orientations of the silhouette

in the image near the observed object. The error between the projected 3D model and the real object is then used for computing updates of 3D pose and joints allowing the projection of the 3D model to fit perfectly the object.

For such a problem, the key is to identify the feature adapted to the data that enables matching the reference 3D model to the actual object in the image. In the context of this work, the object (puppet) silhouette is considered as the tracking feature.

3.1 3D Model Silhouette Computation

The initial silhouette is computed from the 3D rendering of the puppet 3D model. In order to have similar image geometry with respect to the depth images, the virtual camera parameters are set to the same values as the ones got from the RGBD camera calibration, ${}^{d}\boldsymbol{\gamma}$ (Sect. 2.2.1). Then, a virtual depth image of the puppet 3D model, at initial pose ${}^{d^{(0)}}\mathbf{M}_o$, is obtained from the *Z-Buffer*[4] (Fig. 4b) of the graphics engine, e.g., Ogre 3D (https://www.ogre3d.org).

Elementary image processing is applied to detect every Z-buffer image point ${}^{d}_{V}\mathbf{u}_j$ number j (left subscript V means the 2D coordinates are corresponding to a point in the virtual depth image, i.e., the Z-buffer) that is at the border between the puppet body and the background, which is empty (Fig. 4c). Concretely, each bone of the virtual puppet is regularly sampled to produce starting points for a search of a high slope of the Z-buffer derivative along the normal axis to the considered bone. The latter method allows to keep the link between each sample point and the puppet bone to which it corresponds. Keeping such link is necessary for the estimation of

[4] The Z-Buffer (or depth buffer) is a 2D grid used by the hidden surface elimination technique of common graphics rendering pipelines as OpenGL [32]. When the rendering is done, each element of the Z-Buffer 2D grid contains the depth of the object rendered at the same pixel coordinates in the virtual image.

the articular joint angles (Sect. 3.4). The local orientation $\phi_S\left({}^d_V\mathbf{u}_j\right)$ of the border at every silhouette pixel ${}^d_V\mathbf{u}_j$ is also easily got from the latter step (Fig. 4d).

3.2 Silhouette Samples Tracking

The next step of the algorithm is to look for image points ${}^d\mathbf{u}_j$ (in the captured depth image), corresponding to silhouette sample points ${}^d_V\mathbf{u}_j$. Since the virtual camera of the graphics engine considers the same intrinsic parameters as the actual depth camera, the ${}^d_V\mathbf{u}_j$ 2D coordinates can directly be considered in the actual depth image as starting points in the search for their correspondence. The mobile contour elements algorithm [33] applies the latter search along an axis of orientation $\phi_S\left({}^d_V\mathbf{u}_j\right)$. The corresponding contour point ${}^d\mathbf{u}_j$ is the one at which the response of the convolution operation between *masks of oriented contours*[5] and the depth image is maximal. To make more robust the latter search, the depth range of the depth image is adapted to surround the puppet 3D model initial pose, including a safety margin, to avoid clipping the puppet data itself. Figure 5 shows a silhouette sample points tracking result.

Fig. 5 Silhouette (green line) sample points tracking: the resulting points are displayed in red. One must note that the range of the depth image is automatically reduced around the depth of the 3D model for more robustness in the image processing

[5] Masks of oriented contours are 2D grid samples of precomputed derivative kernels rotated in every direction [34], for a given angular step.

3.3 Pose Computation

Corresponding silhouette 2D points ${}^{d}_{V}\mathbf{u}_j$ and ${}^{d}\mathbf{u}_j$ are defined in depth images, the virtual one for the former and the actual one for the latter. Hence, using Eq. (2), for intrinsic parameters ${}^{d}\boldsymbol{\gamma}$, we get two sets of corresponding points $\{{}^{d^{(0)}}\mathbf{X}_j\}$ and $\{{}^{d^{(*)}}\mathbf{X}_j\}$, describing, respectively, silhouette 3D points at the initial pose ${}^{d^{(0)}}\mathbf{M}_o$, and their corresponding silhouette 3D points at the desired pose ${}^{d^{(*)}}\mathbf{M}_o$, which is unknown. Hence, the next step of the silhouette-based tracking algorithm is to compute the relative transformation ${}^{d^{(*)}}\mathbf{M}_{d^{(0)}}$.

Every pair ${}^{d^{(0)}}\mathbf{X}_j$ and ${}^{d^{(*)}}\mathbf{X}_j$ is linearly related by ${}^{d^{(*)}}\mathbf{M}_{d^{(0)}}$:

$$ {}^{d^{(*)}}\mathbf{X}_j = {}^{d^{(*)}}\mathbf{M}_{d^{(0)}}\, {}^{d^{(0)}}\mathbf{X}_j, \tag{4} $$

and considering the entire set of 3D correspondences, ${}^{d^{(*)}}\mathbf{M}_{d^{(0)}}$ can directly be computed using a standard rigid transformation estimation algorithm [35].

The latter one produces a precise frame change, in the sense of mean squares, if the data is not noisy and free of outliers. Practically, it is inherent to the silhouette samples tracking approach described in Sect. 3.2 that the precision of these samples is limited, due to limited depth image resolution and imprecision of measured depths. Furthermore, some outlier samples usually exist if the puppet is partially occluded, for instance. To deal with those issues, the linearly computed ${}^{d^{(*)}}\mathbf{M}_{d^{(0)}}$ is considered as a good initial guess of the puppet pose for a nonlinear optimization scheme, which includes a weighting function to deal with outliers. Thus, considering ${}^{d^{(*)}}\mathbf{r}_{d^{(0)}}$, the vector representation of pose ${}^{d^{(*)}}\mathbf{M}_{d^{(0)}}$, and $\rho(.)$ a robust weighting function, the robust pose optimization problem is written as

$$ {}^{d^{(*)}}\hat{\mathbf{r}}_{d^{(0)}} = \arg \min_{{}^{d^{(*)}}\mathbf{r}_{d^{(0)}}} \sum_k \rho\left(e_k\right), \tag{5} $$

with e_k the k-th element of the error vector $\mathbf{e}$, made by stacking every 3D point-to-point error, i.e.,

$$ \mathbf{e}_{k:k+2} = {}^{d^{(*)}}\mathbf{M}_{d^{(0)}}\, {}^{d^{(0)}}\mathbf{X}_j - {}^{d^{(*)}}\mathbf{X}_j, \;\; k = 3j. \tag{6} $$

To solve the optimization problem of Eq. (5), we consider the robust iterative closest point nonlinear optimization method of [36], without the point matching stage, since it is already done as described in Sect. 3.2. Then, the $\rho(.)$ weighting function is a Huber one [37], weighting less uncertain measures based on the mean and standard deviation of the error vector $\mathbf{e}$.

Finally, ${}^{d}\mathbf{M}_o = {}^{d^{(*)}}\mathbf{M}_o$ is obtained by the usual transformation composition:

$$ {}^{d}\mathbf{M}_o = {}^{d^{(*)}}\mathbf{M}_{d^{(0)}}\, {}^{d^{(0)}}\mathbf{M}_o. \tag{7} $$

3.4 Joint Angles Computation

This section focuses on the estimation of joint angles $\mathbf{q}_m$ of an arm of the puppet ($m = 1$), with, recall, two orthogonal joints (q_1 and q_2) at the shoulder and one joint (q_3) at the elbow. This is straightforward to apply the same method for the second arm and the legs.

First of all, the arm geometry is got from the standard representation of Denavit–Hartenberg [28], based on relative orientation of joint axes and arm bones length (standard phrasing does not use *bones*, but this word fits better the puppet description). Then, thanks to ${}^{d}\mathbf{M}_o$ (Eq. (7)) and the rigid transformation between F_o and the arm basis frame F_m, i.e., ${}^{o}\mathbf{M}_m$, each pair of silhouette 3D points is transformed to be expressed in the arm basis frame as ${}^{m}\mathbf{X}_j$, for the initial (got from the virtual depth image) and ${}^{m}\mathbf{X}^*_j$, for the desired (got from the captured depth image). Only 3D points associated with the bones B_1 and B_2 of the arm m (Sect. 3.1) are considered to estimate the arm configuration $\mathbf{q}^*_m$, which is unknown (corresponding to silhouette points ${}^{m}\mathbf{X}^*_j$), starting from the known initial configuration $\mathbf{q}_m$ (corresponding to points ${}^{m}\mathbf{X}_j$). Then, it is straightforward to express points ${}^{m}\mathbf{X}_j$ in bone frames F_{B_1} and F_{B_2}. Indeed, points ${}^{m}\mathbf{X}_j$ become ${}^{B_b}\mathbf{X}_j$ ($b = 1$ or 2) by composition with transformation matrices ${}^{m}\mathbf{M}_{B_1}$ and ${}^{m}\mathbf{M}_{B_2}$ computed by forward kinematics of the arm [28], using its joint angles $\mathbf{q}_m$.

Thus, the $\mathbf{q}^*_m$ estimation is expressed as the following robust optimization problem:

$$\hat{\mathbf{q}}^*_m = \arg\min_{\mathbf{q}} \sum_j \rho\left({}^{m}\mathbf{X}_j - {}^{m}\mathbf{X}^*_j\right), \tag{8}$$

where ${}^{m}\mathbf{X}_j$ depends on frame changes ${}^{m}\mathbf{M}_{B_1}$ and ${}^{m}\mathbf{M}_{B_2}$, themselves depending on joint variables $\mathbf{q}_m$. Thus, more precisely,

$${}^{m}\mathbf{X}_j = {}^{m}\mathbf{M}_{B_b}(\mathbf{q}_m)\ {}^{B_b}\mathbf{X}_j, \tag{9}$$

and since no bone exists between joints 1 and 2, ${}^{m}\mathbf{M}_{B_1}$ depends on q_1 and q_2 and ${}^{m}\mathbf{M}_{B_2}$ depends on q_1, q_2, and q_3.

Equation (8) is solved by nonlinear optimization using the Gauss–Newton method, which core issue is to express the Jacobian matrix of ${}^{m}\mathbf{r}_{B_b}$, the vector representation of ${}^{m}\mathbf{M}_{B_b}$, over $\mathbf{q}_m$, which is very common in robotics [38]. Each optimization iteration it leads to an arm configuration update $\dot{\mathbf{q}}_m$ so that

$$\mathbf{q}^{it}_m = \mathbf{q}^{it-1}_m + \dot{\mathbf{q}}_m. \tag{10}$$

Then, all ${}^{m}\mathbf{X}_j$ are updated thanks to new ${}^{m}\mathbf{M}_{B_b}$ got from $\mathbf{q}^{it}_m$ and the algorithm loops until convergence.

3.5 Solving the 3D Pose and Joint Angles Together

If the initial pose is not close enough to the optimal one, several iterations of the algorithm described from Sects. 3.1–3.4 might be applied on the same depth image, to reach a higher pose precision, due to the assumption that 2D points on the actual puppet contour exactly correspond to the silhouette sample points got from the virtual depth image. In practice, the number of iterations is limited to the depth image acquisition rate to avoid time delay in the video-mapping. In our experiments, we found two iterations to be sufficient for slow motion of the puppet.

4 Puppet Dynamic Video-Mapping Demonstrations

This section applies the entire set of methods presented in Sects. 2 and 3 to allow the dynamic video-mapping of the suit of a puppet. Section 4.1 describes the hardware devices and the software setup together with the calibration results of a system made of an RGBD camera and a VP. Then, combined with the tracking of articulated puppet, the precision of the video-mapping is quantitatively evaluated on a static puppet (Sect. 4.2). Finally, the dynamic video-mapping is demonstrated and qualitatively evaluated in Sect. 4.3, for both a rigid and an articulated puppet. An external link to a demonstration video is also provided (Sect. 4.3.3).

4.1 Hardware Setup and Software

The hardware setup (Fig. 6) is made of:

- The puppet itself, which can be moved by the puppeteer in a volume of 2 m × 1.5 m × 1.5 m
- A Vivitek D940VX video-projector, set to enlighten the puppet on stage
- The Kinect v2 RGBD camera, facing, at most, the stage volume, while being out of the VP field

To reach the required precision, the Kinect v2 is calibrated with the method of [16], as recalled in Sect. 2.2.1. The values of optimal intrinsic and extrinsic parameters of the Kinect v2 (Table 1) show great coherency with respect to the device dimensions themselves. One can also easily consider these values to know expected orders of magnitude of intrinsic and extrinsic parameters when calibrating another Kinect v2, which is a very common sensor.

Then, the calibration of the VP relatively to the Kinect v2 needs image processing and calibration optimization (Sect. 2.2.2) which are done with the ViSP [39] C++ library, even if the automatic matching of dots (Fig. 3) is a re-implementation of a part of [29]. RGB and depth images are captured thanks to the Microsoft's Software

Fig. 6 Panoramic view of the hardware setup: the Kinect v2 RGBD camera (central green rectangle) is pointed to the puppet (right blue rectangle) on which the video-projector (left red rectangle) must project its suit

Table 1 Kinect v2 intrinsic (top) and extrinsic (bottom) parameters. The pixel reprojection mean error for these parameters is 1.50 (standard deviation: 0.88)

Intrinsic parameters (pixel unit)						
Camera	frame	$^{\bullet}\alpha_u$	$^{\bullet}\alpha_v$	$^{\bullet}u_0$	$^{\bullet}v_0$	
Depth	F_d	359.90	359.21	239.80	208.67	
Color	F_c	1065.92	1063.69	944.65	549.32	
Extrinsic parameters ($t.$ in mm, $\theta w.$ in rad)						
Frame change	t_X	t_Y	t_Z	θw_X	θw_Y	θw_Z
$^{c}\mathbf{M}_d$	−55.64	0.95	7.04	−0.02	−0.01	−0.00

Development Kit[6] for the Kinect v2. VP calibration results are shown in Table 2, only valid for the given setup. Indeed, $^{p}v_0$, for instance, is negative because the video-projection field is not equiangular around the optical axis of the VP lens. This is needed by the configuration depicted in Fig. 6, where the VP is set on one side of the room ceiling and the usual video-projection surface is on the opposite wall.

Finally, the rest of the tracking and mapping process is done online thanks to the Kinect v2 that acquires 30 color and depth images per second and thanks to the video-mapping process that takes, for a unique iteration of the tracking in one depth image acquisition, 13 ms to 17 ms. Thus, when needed, it makes possible to run two iterations of the tracking in one depth image to improve precision. Moving the actual puppet and having the real-time dynamic video-mapping of its suit are, hence, possible. The considered computation hardware is a HP laptop with Intel i7 4800MQ 2.7–3.7 GHz microprocessor, 16 GB of RAM (approximately 150MB

[6] https://www.microsoft.com/en-us/download/details.aspx?id=44561.

Table 2 Video-projector intrinsic (top) and extrinsic (bottom) parameters obtained for the setup (Fig. 6) considered in the experimental results (Sect. 4)

Intrinsic parameters (pixel unit)						
${}^{p}\alpha_u$		${}^{p}\alpha_v$		${}^{p}u_0$		${}^{p}v_0$
2145.99		2138.92		478.96		−47.94
Extrinsic parameters ($t.$ in mm and $\theta w.$ in rad)						
Frame change	t_X	t_Y	t_Z	θw_X	θw_Y	θw_Z
${}^{p}\mathbf{M}_c$	80.19	1456.58	2612.01	0.00	0.00	−0.00

used for the dynamic video-mapping) and an NVIDIA Quadro K610M graphics card (only used once or twice per acquired depth image, as stated in Sect. 3.5). The entire program is written in C++, without particular code optimization, using ViSP and Ogre3D libraries.

4.2 Video-Mapping

First of all, in order to evaluate the video-mapping precision, a first experiment is made with a rigid version of the puppet, to ease the evaluation process. Then, once the current puppet pose ${}^{d}\mathbf{M}_o$ is computed from the depth image processing, a color virtual image of the textured 3D model of the puppet is rendered (Fig. 1b) from the VP point of view F_p, thanks to ${}^{c}\mathbf{M}_d$ and ${}^{p}\mathbf{M}_c$ frame changes and ${}^{p}\boldsymbol{\gamma}$ intrinsic parameters. This virtual image is simply displayed by the VP and the puppet virtual suit is mapped on the actual puppet (Fig. 7).

In such a setup, qualitative evaluation is rather easy since the suit is displayed on the puppet at the correct position and orientation as well as the correct scale, globally (see Fig. 7 where the top, bottom, left, and right parts of the suit fit well the puppet). On the other hand, quantitative evaluation is more difficult to obtain.

A basic mapping precision evaluation is led in photographs observing video-mapping results, from viewpoints obviously different from the VP. In the photo of Fig. 7, manual selection of pixels corresponding to the puppet, to the suit projected on the puppet and to the suit projected outside the puppet (on the background wall), is done. Then, counting the number of image pixels in these selections leads to the quality evaluation of video-mapping. A perfect mapping could be defined as "every pixel of the puppet in the photo is one of the suit *and* not any pixel of the suit is projected out of the puppet." Pixel counts reported in Table 3 for the photograph of Fig. 7 indicate a correct mapping at 90%, the main source of incorrect mapping being the difference between the 3D model and the actual puppet (80% of the remaining 10% since suit parts out of the puppet represent only 2%).

Fig. 7 Puppet suit video-mapping result (brightness of the video-projector is set to the maximum value to see the projected suit and the actual puppet of median on the photograph). The face of the person has been pixelized afterward to prevent copyright issues

Table 3 Quantitative evaluation of video-mapping correctness (total number of pixels in the considered photo of Fig. 7: 2662 × 3712 pixels)

Selection	Number of pixels in the photo	Ratio w.r.t. puppet number of pixels
Puppet	3 026 107	100.00%
Suit on the puppet	2 742 388	90.62%
Suit outside the puppet	67 380	2.23%

4.3 Dynamic Video-Mapping

4.3.1 Rigid Puppet

Still with the rigid puppet, the tracking and pose estimation approach proposed in this chapter is applied. Thus, the puppet is moved in the field shared by both the Kinect v2 and the VP. Figure 8 shows that the suit follows well the puppet for pure translations (Fig. 8a–g) as well as for three axes rotations (Fig. 8h–l). Obviously, since held by a human, the Cartesian 6 DoFs of the puppet pose are mandatory to be considered to ensure a correct tracking and mapping, and all of them are used in this experiment.

Fig. 8 Puppet suit dynamic video-mapping images (**a**)–(**l**) extracted from a single sequence of 1 min in which various possible poses with correct automatic mapping are demonstrated, filmed with a third-party camera. The face of the person has been pixelized afterward to prevent copyright issues

4.3.2 Articulated Puppet

The third result considers the articulated puppet for which the joint angles estimation is necessary to display correctly the suit arms and legs. Figure 9 shows both legs are manually moved by the puppeteer with a coherent update of the joint angles. Figure 9a–c shows the dynamic video-mapping result when moving the puppet left leg. Then, the puppeteer moves it back to its initial state (almost vertical) and moves the other leg (Fig. 9d). Finally, both legs are actuated simultaneously (Fig. 9e–f).

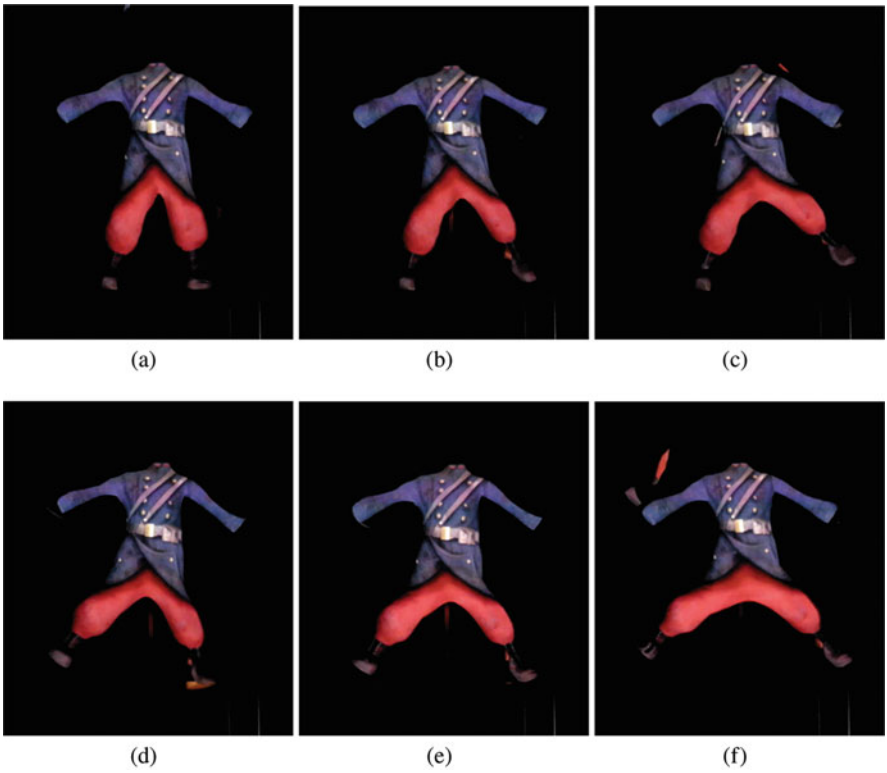


Fig. 9 Articulated puppet suit dynamic video-mapping images (**a**)–(**f**) extracted from a single sequence of 40 s in which various possible legs motions are encountered with correct automatic mapping are demonstrated, filmed with a third-party camera. The legs are moved manually by the puppeteer, thanks to wooden sticks at the back of each foot

4.3.3 Video

A video showing live captures of the dynamic video-mapping of both the rigid and the articulated puppets is available at http://mis.u-picardie.fr/~g-caron/videos/puppet_dynamic_vm.mp4.

5 Discussion

Beyond the success of the dynamic video-mapping of the suit of two puppets, some limits are identified. Indeed, shifts appear in some images of the presented sequences between the suit and the puppet. It is mainly due to two facts:

- Latency of devices and processing: indeed, the difference between the time at which the puppet reaches a given pose and the time at which the suit is displayed at that pose is not null (Kinect v2 minimum latency is 20 ms, dynamic video-mapping algorithm minimum processing time is 13 ms, and video-projection latency is between 33 ms and 160 ms).
- Puppet motion along an axis approximately tangent to a large proportion of the silhouette: indeed, in that case, several mismatched silhouette points are considered in the pose update computation, concretely leading to a low error at those points and an underestimated pose update, needing several iterations of the tracking algorithm to well match the puppet in the depth image.
- Considering a black background may not always be relevant since when slight registration error appears, puppet suit parts can be displayed on the background stage (Fig. 9f: a part of the red trousers and of the puppet right shoe is displayed on the background stage, leading to undesired visual effects).

Both former facts are particularly visible when the puppeteer moves a bit faster (Fig. 8h). However, when he slows down the puppet motion, the suit maps again the puppet (Fig. 8i).

6 Conclusion

This chapter has reviewed the dynamic video-mapping state-of-the-art and introduced the new video-mapping of an articulated puppet. The silhouette feature only is considered in the 3D model-based tracking of the puppet, exploiting depth images of an RGBD camera. The proposed calibration procedure as well as the efficiency of the puppet tracking, in precision and computational load, leads to precise dynamic video-mapping in real time, without any dedicated hardware or GPU implementation (the GPU is only used to render images of the textured 3D model of the puppet).

Future works will handle the identified limits of the current algorithm as the latency issue due to devices and the algorithm itself for a fast set of motions.

Acknowledgments The authors wish to highlight the puppeteers company named “Le Tas de Sable - Chés Panses Vertes” at the origin of this work: Sylvie Baillon, for directing the puppet show, and Eric Goulouzelle, for making the puppets used in this chapter. The puppet 3D model was designed by Alexis Leleu and textured by Margot Briquet, from the 3D animation school “Waide Somme.”

References

1. A. Grundhöfer and D. Iwai. Recent advances in projection mapping algorithms, hardware and applications. In *Computer Graphics Forum*, volume 37, pages 653–675, 2018.

2. V. Lange, P. Kurth, B. Keinert, M. Boss, M. Stamminger, and F. Bauer. Proxy painting: Digital colorization of real-world objects. *ACM J. Comput. Cult. Herit.*, 13(3), July 2020.
3. M. T. Ibrahim, G. Meenakshisundaram, and A. Majumder. Dynamic projection mapping of deformable stretchable materials. In *ACM Symposium on Virtual Reality Software and Technology*, Virtual Event, Canada, 2020.
4. K. Maeda and H. Koike. MirAIProjection: Real-Time Projection onto High-Speed Objects by Predicting Their 3D Position and Pose Using DNNs. In *ACM Int. Conf. on Advanced Visual Interfaces*, Salerno, Italy, 2020.
5. Y. Mikawa, T. Sueishi, Y. Watanabe, and M. Ishikawa. Projection mapping system to a widely dynamic sphere with circumferential markers. In *IEEE Int. Conf. on Multimedia and Expo (ICME)*, pages 1–6, 2020.
6. Y. Fujimoto, R. T. Smith, T. Taketomi, G. Yamamoto, J. Miyazaki, H. Kato, and B. H. Thomas. Geometrically-correct projection-based texture mapping onto a deformable object. *IEEE Trans. on Visualization and Computer Graphics*, 20(4):540–549, April 2014.
7. N. Hashimoto, R. Koisumi, and D. Kobayashi. Dynamic Projection Mapping with a Single IR Camera. *Int. Journal of Computer Games Technology*, 2017, Feb 2017.
8. Y. Zhou, S. Xiao, N. Tang, Z. Wei, and X. Chen. Pmomo: Projection Mapping on Movable 3D Object. In *Conf. on Human Factors in Computing Systems (CHI)*, pages 781–790, Santa Clara, CA, USA, May 2016. ACM.
9. C. Resch, P. Keitler, and G. Klinker. Sticky Projections-A Model-Based Approach to Interactive Shader Lamps Tracking. *Trans. on Visualization and Computer Graphics (TVCG)*, 22(3):1291–1301, 2016.
10. C. Siegl, M. Colaianni, L. Thies, J. Thies, M. Zollhöfer, S. Izadi, M. Stamminger, and F. Bauer. Real-time pixel luminance optimization for dynamic multi-projection mapping. *ACM Trans. Graph.*, 34(6):237:1–237:11, October 2015.
11. V. Lange, C. Siegl, M. Colaianni, M. Stamminger, and F. Bauer. Robust Blending and Occlusion Compensation in Dynamic Multi-Projection Mapping. In *Eurographics - Short Papers*, 2017.
12. T. Sueishi, H. Oku, and M. Ishikawa. Robust high-speed tracking against illumination changes for dynamic projection mapping. In *IEEE Virtual Reality (VR)*, pages 97–104, March 2015.
13. Y. Mikawa, T. Sueishi, Y. Watanabe, and M. Ishikawa. VarioLight: Hybrid dynamic projection mapping using high-speed projector and optical axis controller. In *ACM SIGGRAPH Asia Emerging Technologies*, Tokyo, Japan, 2018.
14. L. Wang, H. Xu, Y. Hu, S. Tabata, and M. Ishikawa. Dynamic depth-of-field projection for 3D projection mapping. In *ACM Conf. on Human Factors in Computing Systems (CHI) - Extended Abstracts*, page 1–4, Glasgow, Scotland UK, 2019.
15. J.C. Rodríguez-Qui nonez, O. Sergiyenko, W. Flores-Fuentes, M. Rivas-lopez, D. Hernandez-Balbuena, R. Rascón, and P. Mercorelli. Improve a 3d distance measurement accuracy in stereo vision systems using optimization methods' approach. *Opto-Electronics Review*, 25(1):24–32, 2017.
16. A. Staranowicz, G. Brown, F. Morbidi, and G.-L. Mariottini. Practical and accurate calibration of RGB-D cameras using spheres. *Computer Vision and Image Understanding (CVIU)*, 137:102–114, August 2015.
17. L. R. Ramírez-Hernández, J. C. Rodríguez-Quiñonez, M. J. Castro-Toscano, D. Hernández-Balbuena, W. Flores-Fuentes, R. Rascón-Carmona, L. Lindner, and O. Sergiyenko. Improve three-dimensional point localization accuracy in stereo vision systems using a novel camera calibration method. *International Journal of Advanced Robotic Systems*, 17(1), 2020.
18. C. Sepulveda-Valdez, O. Sergiyenko, V. Tyrsa, W. Flores-Fuentes, J. C. Rodríguez-Quiñonez, F. N. Murrienta-Rico, J. E. Miranda-Vega, D. Hernandez-Balbuena, P. Mercorelli, and M. Kolendovska. Geometric analysis of a laser scanner functioning based on dynamic triangulation. In *IEEE 29th International Symposium on Industrial Electronics (ISIE)*, pages 1398–1403, 2020.

19. O.Yu. Sergiyenko, M.V. Ivanov, V.V. Tyrsa, V.M. Kartashov, M. Rivas-López, D. Hernández-Balbuena, W. Flores-Fuentes, J.C. Rodríguez-Quiñonez, J.I. Nieto-Hipólito, W. Hernandez, and A. Tchernykh. Data transferring model determination in robotic group. *Robotics and Autonomous Systems*, 83:251–260, 2016.
20. P. Kurth, V. Lange, C. Siegl, M. Stamminger, and F. Bauer. Auto-calibration for dynamic multi-projection mapping on arbitrary surfaces. *IEEE Trans. on Visualization and Computer Graphics*, 24(11):2886–2894, 2018.
21. S. Kagami and K. Hashimoto. Animated stickies: Fast video projection mapping onto a markerless plane through a direct closed-loop alignment. *IEEE Trans. on Visualization and Computer Graphics*, 25(11):3094–3104, 2019.
22. S. Kagami and K. Hashimoto. Interactive stickies: Low-latency projection mapping for dynamic interaction with projected images on a movable surface. In *ACM SIGGRAPH Emerging Technologies*, Virtual Event, USA, 2020.
23. Y. Liu and A. Namiki. Articulated object tracking by high-speed monocular RGB camera. *IEEE Sensors Journal*, pages 1–1, October 2020.
24. J. Zhang, W. Li, P. O. Ogunbona, P. Wang, and C. Tang. Rgb-d-based action recognition datasets: A survey. *Pattern Recognition*, 60:86–105, 2016.
25. A. Petit, E. Marchand, and K. Kanani. Tracking complex targets for space rendezvous and debris removal applications. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 4483–4488, Vilamoura, Portugal, October 2012.
26. A. Comport, E. Marchand, and F. Chaumette. Complex articulated object tracking. *Electronic Letters on Computer Vision and Image Analysis*, 5(3):20–30, 2005.
27. G. Caron and D. Eynard. Multiple camera types simultaneous stereo calibration. In *IEEE International Conference on Robotics and Automation*, pages 2933–2938, 2011.
28. P. I. Corke. *Robotics, Vision & Control: Fundamental Algorithms in Matlab*. Springer, 2011.
29. J. Pages, C. Collewet, F. Chaumette, and J. Salvi. An approach to visual servoing based on coded light. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 4118–4123, May 2006.
30. V. Lepetit and P. Fua. Monocular model-based 3d tracking of rigid objects. *Found. Trends. Comput. Graph. Vis.*, 1(1):1–89, January 2005.
31. E. Marchand, H. Uchiyama, and F. Spindler. Pose Estimation for Augmented Reality: A Hands-On Survey. *IEEE Transactions on Visualization and Computer Graphics*, 22(12):2633–2651, December 2016.
32. M. Segal and K. Akeley. *The OpenGL Graphics System: A Specification (v. 1.2.1)*. Silicon Graphics Inc., 1999.
33. P. Bouthemy. A maximum likelihood framework for determining moving edges. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11(5):499–511, May 1989.
34. M. Jacob and M. Unser. Design of steerable filters for feature detection using canny-like criteria. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1007–1019, 2004.
35. D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall Professional Technical Reference, 2002.
36. S. Fantoni, U. Castellani, and A. Fusiello. Accurate and automatic alignment of range surfaces. In *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, pages 73–80, Oct 2012.
37. P. J. Huber. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35(1):73–101, March 1964.
38. R. P. Paul and B. Shimano. Kinematic control equations for simple manipulators. In *IEEE Conference on Decision and Control*, pages 1398–1406, Jan 1978.
39. E. Marchand, F. Spindler, and F. Chaumette. ViSP for visual servoing: a generic software platform with a wide class of robot control skills. *IEEE Robotics and Automation Mag.*, 12(4):40–52, December 2005.

Aerial Robotics for Precision Agriculture: Weeds Detection Through UAV and Machine Vision

Alexander Menshchikov and Andrey Somov

Acronyms

Adam	Adaptive Moment Estimation
ADCS	Attitude Determination and Control System
AI	Artificial Intelligence
API	Application Programming Interface
AUC	Area Under the ROC Curve
BN	Batch Normalization
CMOS	Complementary Metal Oxide Semiconductor
CNN	Convolutional Neural Network
COCO	Common Objects in Context
CPU	Central Processing Unit
DL	Deep Learning
DSP	Digital Signal Processor
FCNN	Fully Convolutional Neural Network
FOV	Field of View
FPS	Frames Per Second
GPU	Graphics Processing Unit
GSD	Ground Sample Distance
HA	Hardware Acceleration
HSL	Hue, Saturation, Lightness
IoT	Internet of Things
MCU	Micro Controller Unit
ML	Machine Learning
N/A	Not Applicable
NAS	Neural Architecture Search

A. Menshchikov · A. Somov (✉)
Skolkovo Institute of Science and Technology (Skoltech), Moscow, Russia
e-mail: alexander.menshchikov@skolkovotech.ru; a.somov@skoltech.ru

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1_2

NCS	Neuro Compute Stick
NDVI	Normalized Difference Vegetation Index
NN	Neural Network
NNAPI	Android Neural Network API
PA	Precision Agriculture
RAM	Random Access Memory
RBA	Routing By Agreement
ReLU	Rectified Linear Unit
RGB	Red, Green, and Blue
ROC	Receiver Operating Characteristic Curve
RPi	RaspberryPi
SBC	Single Board Computer
SDA	Scale Distance Agreement
TPU	Tensor Processing Unit
UAS	Unmanned Aerial Systems
UAV	Unmanned Aerial Vehicles
UGV	Unmanned Ground Vehicles
VPU	Visual Processing Unit
WSN	Wireless Sensor Networks
XAI	Explainable AI

1 Introduction

The term *aerial robotics* was coined by Robert Michelson [88] to define intelligent flying machines. This definition covers several concepts, e.g., robotic flying machines and robotics that uses flying machines. Nowadays, the robotic flying machines are typically called unmanned aerial vehicles (UAVs). At the same time, the entire infrastructure together with systems and human elements required to run and operate these machines to accomplish the operational goal are usually referred to as unmanned aerial systems (UASs). In this chapter, we define the aerial robot as a UAV with a camera on board enriched with artificial intelligence (AI) capabilities.

Myriads of potential applications exist for robots characterized by mobility offered by flight. The lion share of applications is aerial monitoring including environmental surveys, 3D scanning, target identification, and crop monitoring. Other applications are payload delivery, e.g., to remote areas, military operations, and a number of civilian and private applications, e.g., film making, air pollution detection. However, real implementations are often limited due to stringent air and safety regulations.

Another promising field for aerial robotics is precision agriculture (PA). PA is a management strategy, according to the International Society of PA [37] that collects, processes, and analyzes temporal, spatial, and individual data and combines it with other information to support management decisions based on estimated variability to improve the production, productivity, quality, profitability, and sustainability of agri-

cultural resources. This research field includes new technologies that have recently been successfully implemented, e.g., remote sensing [116], laser scanners [38], sensor networks [23], robotics [14] and swarm robotics [95], AI [50], and Internet of Things (IoT) [3]. These promising innovations allow food safety to be safeguarded, reduce the effect on the environment, and ensure the economy's profitability.

Using various and multimodal sensor data [42, 74], research in the area usually focuses on improving the growing conditions, agricultural performance, and field monitoring. At the same time, the state-of-the-art research focuses primarily on "how to maximize" yield. However, there is a lack of research addressing natural challenges, such as the dissemination of *weeds* [80], which continues to hinder the growth of PA. In the scope of PA, precise and targeted weed control has a significant impact on the environment as well as economy.

Although the advanced UAVs provide high-quality multispectral imaging, processing these images still poses a daunting challenge for the machine vision algorithms [94] because of the complex nature of the plant topology and the large range of background types. Partly by using fully convolutional neural networks (FCNNs) for semantic segmentation, this issue can be overcome. Due to the complexity of the problem proposed, however, current algorithms cannot effectively carry out semantic segmentation of different plants in field conditions. In addition, there are still problems in optimizing high-performance computational algorithms for mobile platforms/embedded systems, and it is highly important to investigate and incorporate this new fundamental approach.

In this chapter, we touch upon the problem of weeds detection. In particular, we consider a real-life problem of Hogweed of Sosnowski (*lat. Heracleum*) detection using intelligent aerial robotics. Therefore, the detection task imposes several challenges: (i) detection, without human presence, should be carried out in order to prevent the seeds dissemination, (ii) monitoring is required to cover a wide range, and (iii) the location of the data should be accessible and "real time" for fast action of the findings on the data analysis. To address this problem, we discuss the following points: relevant UAV architecture, the corresponding image capture cameras, computer vision methods, and computer vision hardware for UAV. Until now, this challenge has not been fully addressed—it requires further research activities for designing a novel monitoring platform with intelligent capabilities on board allowing for the automation of weed detection routine which is currently carried out manually. For many countries, in situ detection of the hazardous plants is an immense challenge in the context of PA.

2 Monitoring Scenario

Weeds distribution, and particularly Hogweed of Sosnowsky, is a growing agriculture problem in many countries. This fast-growing weed spreads rapidly across Eurasia: from Germany to the Siberian part of Russia, and from year to year, its distribution area grows. Hogweed has a 3–5 m high that usually reaches 12 cm in

Fig. 1 Photos of the blooming hogweed: (**a**) captured from the ground and (**b**) captured by the UAV

diameter, with a straight, firm stem. Its root is extremely strong with a diameter of up to 30 cm; at the end of each stalk, an inflorescence is a large umbel (see Fig. 1). It produces thousands of seeds that are dispersed easily by wind and water during its blooming. Hogweed is toxic to humans and detrimental to agricultural crops and local habitats. That is why this hazardous weed must be properly identified and removed. The weed detection includes the use of aerial robotics to prevent the spread of seeds, the coverage of vast areas during aerial observation, and real-time data analysis to ensure rapid inference and intervention. Until now, these problems have not been fully addressed—they need additional research activities to design a novel monitoring platform with onboard intelligent capabilities to automate the routine for the identification of hogweed that is currently performed manually.

In agriculture, the observation is carried out through a number of various methods. The field and the crops are usually manually inspected by the farmer, who may be fitted with different handheld instruments. The use of agricultural facilities, such as tractors, may also be investigated. The farmer can, for example, mount nitrogen sensors on the machinery to measure fertilization nutrition demands while driving over the field. While these methods may still be advantageous for farmers engaged in the field of PA, there is a move toward autonomous systems allowing for monitoring of larger areas and preventing dissemination of seeds by feet and wheels. Wireless sensor networks (WSNs), satellites, unmanned ground vehicles (UGVs), and UAV are the most popular. WSN consists of sensor nodes that are wireless battery-powered devices [39] that are deployed over a wide (agricultural) area. The nodes perform the monitoring tasks and transmit the data over the wireless network to a user/server. Sensor nodes are fitted with low-power sensors and inherently avoid applications requiring the processing and transmission of images or video because it means stringent criteria for power consumption. Although some work already exists on the application of AI to image processing in WSN [96], this approach does not permit comprehensive area coverage, long-term, and autonomous sensing system operation. Usually, the implementations of WSNs with AI on board are limited to smart greenhouse monitoring [102]. At the same time, UGV cannot cover truly wide areas as well.

Two promising solutions are satellites and UAVs: they may guarantee successful detection of weeds and monitor wide areas. The choice of altitude above the ground for satellites is often a trade-off between field of view (FOV) and ground sample distance (GSD). GSD is the ground distance (L) reached by a camera's single pixel (l). This value is also equivalent to the altitude ratio (H) and the camera's focal length (f). It is expressed in $cm/pixel$ and varies for various sensors and altitudes of the camera.

That is why a satellite, with a low GSD usually equal to tens or hundreds of meters per pixel, can capture thousands of km^2 per orbit pass lasting for about 90 min. 30 cm/px [98] is the best available resolution for a satellite image. Typical fixed-wing UAVs equipped with a multispectral camera with several centimeters of GSD can capture tens of km^2 per flight. With subcentimeter GSD, a typical UAV with a similar payload can cover up to 1 km^2 per flight. For a number of agricultural tasks including the classification of weeds [59], vegetation indices [106], and crop disease identification [115], it is critical. The hogweed thickets can be effectively recognized by the satellite imagery. For an individual plant recognition, however, this GSD is not enough. In view of the natural capacity of hogweed to grow rapidly (it reaches up to 1.7 m within 40 days in spring) and disseminate thousands of seeds rapidly, it may be crucial to miss even a single plant in terms of weed spread. That is why high GSD is also needed for the proper hogweed detection issue.

In contrast to the satellite option, another crucial benefit of UAVs is high temporal resolution. Satellites normally maintain the Sun-synchronous or polar orbits for Earth observation, which means that the same position on the ground is revisited regularly at a particular local time. Only the satellite constellation, spread over various orbits, can solve the problem of capturing the same location at an arbitrary local time by satellites. The satellite monitoring could, however, be wiped out by cloudy weather. By demand, the UAVs may conduct a mission at any local time along an arbitrary trajectory over the same area, even under the cloudy conditions. Since hogweed is a fast-growing plant, it is an important factor for its detection, and space imagery could miss the flowering stage. We discuss the requirements for drones in detail in the next section.

3 Aerial Robotic System

3.1 UAV

During the mission, the state-of-the-art UAVs acquire images relying on the absolute coordination system. After the observation task, the postprocessing of these images, e.g., segmentation and classification, happens on a personal computer. These operations have to be done daily or even hourly, depending on the mission. It is an inefficient task that is labor-intensive. Recent research has shown the potential for autonomous systems to be developed and implemented in agriculture [6].

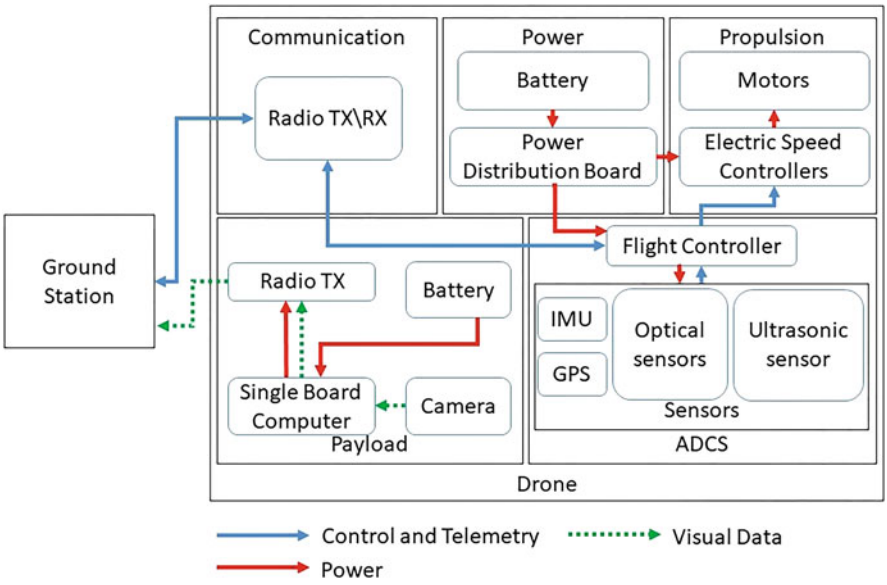


Fig. 2 The diagram of the proposed platform

Although none of the systems completely perform semantic segmentation on board, a number of demonstrated applications require the image segmentation and the object classification during the flight.

As it is shown in Fig. 2, the platform required for the weeds detection is a typical UAV with the following units: communication, power, propulsion, payload, and Attitude Determination and Control System (ADCS). With the aid of experienced agriculturalists, the platform criteria for real-time hogweed monitoring are summarized and are as follows:

- Minimum Area Covered: 15 ha
- Minimum ROC AUC (AUC: Area Under the ROC Curve and ROC: Receiver Operating Characteristic Curve): 0.90
- Maximum Power Consumption: 12 W

Four metrics are considered for the evaluation of the convolutional neural networks (CNNs) performing the image processing (see Sect. 4): *ROC AUC* is the quality metric, and *FPS* is the measure of the speed of computations and finally *power consumption* and *area coverage*. In fact, from practical feasibility point of view to estimate the performance of proposed solution's, it is important to figure out how many km^2 can be observed by the UAV. This is the reason why the *Area Covered* parameter is important to measure. For example, the Nvidia Jetson Nano, hardware for embedded systems to run neural networks (see Sect. 5), can be used as a payload for the DJI Matrice 200 drone. The images can be gathered with a 12 MP camera, the recommended flight altitude is 10 m, and the mission duration is limited

to 40 min. There is no resemblance between the images of neighbors. The drone takes a shot and then flies to the next spot to take the next one. We calculate the total area as follows:

$$S = GSD^2 \times x \times y \times FPS \times t \tag{1}$$

where

$$GSD = \frac{H}{f} = \frac{L}{l} \tag{2}$$

For instance, if we use a 12 MP camera as a payload and the flight altitude is 10 m, the GSD is 0.43 cm/px, x and y are the quantity of pixels along the x and y axes of the image, FPS is the frames per second, and t is the 40 min time period. The high performance of neural networks (NNs) enables the higher frame rate. It allows for flying at higher speed and, therefore, covering wider area during a single flight.

3.2 RGB, Multispectral, and Hyperspectral Cameras and Laser Scanners

Colors in a photographic image are typically based on the three primary colors: red, green, and blue (RGB). Combining these three colors in different proportions, you can get new color in visible spectrum. An RGB camera is usually equipped with a standard complementary metal-oxide semiconductor (CMOS) sensor through which the colored images are acquired. The acquisition of static photos is usually expressed in megapixels defining the amount of pixels that compose the image. With the help of a process called interpolation, the RGB camera computes the actual color of each pixel [111].

As a cost-effective solution that meets the job requirements, an RGB camera can be used for hogweed detection. The system is able to recognize hogweed at the vegetation stage which is one of the most complicated detection problems. The point is that at this stage we can only recognize the hogweed shape as it is green and blends in with the background. It is much simpler for the AI algorithms to detect the hogweed in the later phases, e.g., flowering, because its white flowers can be easily identified in the RGB range even without the application of complex FCNN.

In contrast to RGB cameras that capture the images in visible spectrum, multispectral cameras are able to capture the image data at specific frequencies in electromagnetic spectrum. For the precise recognition of plants and phenotyping [22, 35, 101], multispectral imaging is highly promising. The key difference between multispectral plant detection and common RGB is as follows: certain plants have special wavebands of reflection that may be beyond the visible range (see Fig. 3). The multispectral approach allows simple classification algorithms to be implemented instead of complex plant shape detection FCNNs. The pixels are identified

Fig. 3 Example of aerial imagery for (**a**) RGB camera and (**b**) multispectral camera

by these algorithms and those with the intensity representing the appropriate waveband and belonging to a certain plant are found. Several multispectral plant phenotype research studies were devoted to the remote detection of plant disease [66].

The use of multispectral and hyperspectral digital cameras for the identification of plant diseases opens wide vista for researchers, as a large number of diseases have their particular spectral characteristics at the initial and next stages. This method helps farmers to recognize the diseases at an early stage and prevent from spreading infections [28, 79]. Multispectral imaging is also used to detect physical stress [34]. The cost of high-quality commercial multispectral cameras compared to the cost of traditional RGB cameras is one of the drawbacks. The low-cost multispectral solutions were mainly designed to measure the normalized difference vegetation index (NDVI) [46, 47] and could not measure the necessary wavebands for the detection of specific plants.

Hyperspectral cameras measure spectrum in a broad range and are characterized by a spectral resolution, but as a payload, they are too expensive and heavy to place them on the UAV [99]. At the same time, it has been already demonstrated that the NDVI index from the RGB images is reconstructed by a CNN [45]. In contrast to the application of multispectral camera, it is commercially efficient to use the RGB camera enriched with the CNN algorithm.

Laser scanners can be used as an alternative solution for the cameras for surface monitoring tasks [4]. A laser scanner is based on the principle of triangulation. The scanner includes two units: a positioning laser unit emits the laser ray onto the surface, while a scanning aperture unit is rotating with the required velocity and receives the reflected beam of the laser energy. The triangle is set for a small period of time, so that the system could calculate the 3D coordinates of the laser point relying on the law of sines [58]. Laser scanners, in contrast to camera vision, are characterized by the ability to function in complete darkness and obtain the 3D coordinates of any point on the surface. However, they need a certain time to accomplish 3D scanning of a required sector, and they cannot process all the scanned surfaces at the same time. At the same time, *dead zones* between two adjacent points

of scanning may appear due to the technical limitation of a stepper motor driving the positioning laser. It prevents the detection of obstacles between these points. One of the solutions is to engage the robotics swarm to scan the required surface properly [38].

4 Computer Vision

Different techniques focused on the handcrafted features [61, 62] or the end-to-end learning theory [21, 89] has been applied for the localization and phenotyping of different plants for several years. A fairly simple approach is based on statistics extraction from the color space of the HSL (hue, saturation, lightness) and using them as an input for the NN [10]. Advanced methods are based on classical approaches of machine learning (ML) and include the Random Forest classification [31] and variety of the Random Markov Field optimization [61]. For the aerial imagery classification, which only involves the RGB channels [62], the same approach can be applied. The algorithms behind remote sensing research in PA are shifting to end-to-end approaches due to enormous potential. They depend to a large extent on different methods of deep learning. These methods include CNN and FCNN. CNNs have demonstrated their efficiency when applied to detection and classification tasks. The cascades of CNNs proved high efficiency in the generalization and processing of previously unseen data [25] in the PA domain. FCNNs are typically applied for instance and semantic segmentation [7, 87]. Images captured by either RGB [60] or multispectral cameras [70, 89] can be perfectly processed by FCNNs.

While CNNs provide the state-of-the-art performance, *capsule networks* [90] have recently gained popularity as they resolve some CNN limitations. For example, capsule networks retain the position of the object and are capable of creating a hierarchy of object parts that make up the object while enhancing the state-of-the-art results [82]. This hierarchical ability makes capsule networks more transparent than CNNs, i.e., they behave less like a black box comparing to CNNs. Capsules in capsule networks are groups of neurons in a parse tree which represent an object or a part of an object. A capsule output is a so-called vector of activation or instantiation describing the properties of this object. However, there is major downside of capsule networks: they are computationally intensive NNs. It significantly limits their application in mobile devices and embedded systems, therefore, requiring further research prior to adopt them in aerial robotics.

4.1 Fully Convolutional Neural Networks

FCNN, especially when running on low-power embedded systems, is slow for real-time processing tasks. In certain cases, it is impossible to make an inference using it.

In fact, the optimization techniques exist for making inference relying on the FCNN architectures running on the low-power embedded systems. The following methods are widely recognized for optimizing ML algorithms for low-power embedded systems: quantization [54], information distillation, pruning, quick model evaluation methods, neural architecture search (NAS), freezing, and fusion. Reducing the weights' size by the computer number format is the main idea behind the quantization. For example, the transformation from float (32 bits) to integer (8 bits) of all weights in the NN would result in 4 times less weight memory occupation. This simple approach results in a major decrease in the time of inference and the reduction of network size. The key concept behind the pruning technique is to wipe out those weights within the network, which are grouped as filters, blocks, and layers [56]. This technique helps to overcome the issue of over-parameterization. The knowledge distillation is a process of moving knowledge from one NN to another one. It enables us to increase the generalization and accuracy while keeping the architecture. This approach results in improved precision of small networks. In the absence of initial training data, it allows retraining of the NN. NAS methods [117] work well on training multiple candidate models concurrently. All the candidates are developed by the so-called supermodel and evaluated during the training. There are thousands of graphics processing unit (GPU) hours needed to accomplish this process.

Using a convolution network followed by the deconvolution layers is the general concept behind the NN architecture for the semantic segmentation. These two sections are typically connected by a 1×1 convolution layer. A classification network, e.g., VGG [30] or ResNet [33], is the convolution part. The task of deconvolution is to project onto the high-resolution image the features created by the convolution part. In this section, we discuss the NN architectures that are used for the weeds detection in our research: SegNet [7], a variety of U-Net [87], and RefineNet [55] with the ResNet backbone. Figure 4 shows modified SegNet architecture which is designed for the semantic segmentation task. Figure 5 demonstrates the output by all the present FCNN architectures.

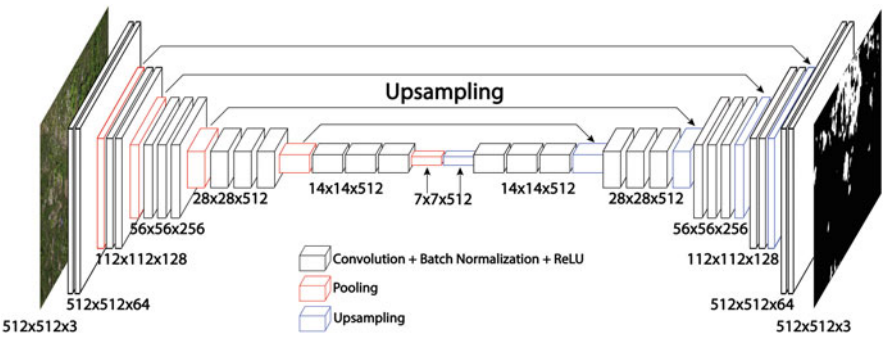


Fig. 4 Modified SegNet architecture used for weed detection

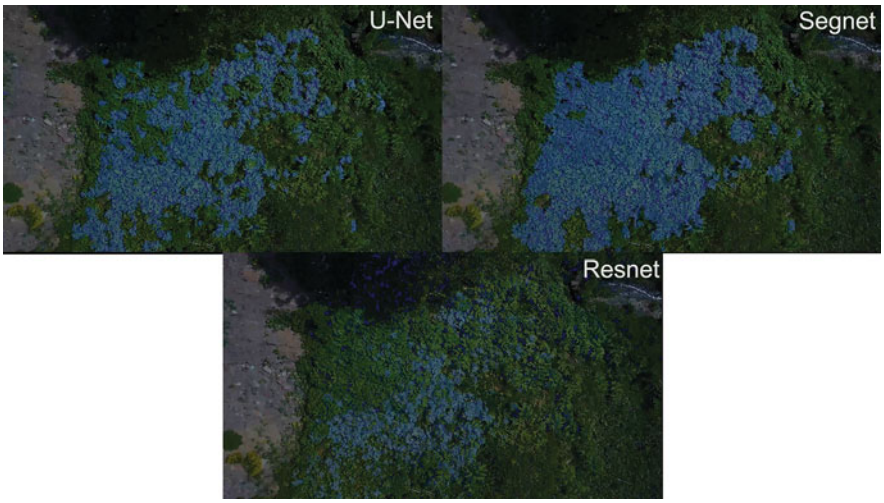


Fig. 5 The output examples of all the described FCNNs: U-Net, SegNet, and ResNet

4.1.1 SegNet

SegNet [7] is an NN architecture for multi-class pixel-wise segmentation. It contains a series of encoding and corresponding decoding blocks followed by the classifier as it is shown in Fig. 4. A convolutional layer, batch normalization (BN), and rectified linear unit (ReLU) nonlinearity are included in each encoder block. In the encoder, the first 13 convolution layers are the first 13 convolution layers from the VGG-16 network. For training SegNet, the following augmentation techniques are applied: 512×512 crops with 50% chance of horizontal and vertical flips and random rotation up to 30°. Next, *LogSoftmax* layer is a good option for predicting the hogweed class after the last convolution. The reason is that *LogSoftmax* heavily penalizes the model when it predicts the class incorrectly [63]. Afterward, the negative log-likelihood loss is considered for calculation of the derivatives and make step by Adaptive Moment Estimation (Adam) optimizer [114] with default parameters ($\alpha = 0.001$, $\beta_1 = 0.9$, and $\beta_2 = 0.999$). Where α stands for the stepsize (or so-called "learning rate"), β_1 and β_2 are hyper-parameters, which control the exponential decay rates of exponential moving averages of the gradient and the squared gradient. We used 4 Nvidia 1080Ti to train this network. It took approximately 9 h to reduce loss enough after 1500 epochs. Time-consuming training procedure happens due to a high number of network trainable parameters of up to 30M.

4.1.2 U-Net

U-Net [87] is a CNN architecture applied for semantic segmentation. U-Net and SegNet are the encoder–decoder architectures. For obtaining good results, several

modifications were applied to the original U-Net architecture. First, the BN layer after each convolutional layer was added. For example, U-Net consists of 9 double convolution blocks. Each block is composed of two convolutions, BNs, ReLU activations, and the output convolution layer. For providing an opportunity ensuring the width scaling technique [104], several convolutional channels were left adjustable. It is controlled by the parameter W denoting the number of output channels in the first convolutional layer. Furthermore, for enabling the depth scaling, one encoder block and one decoder block were made removable. This option is specified by the parameter D which is responsible for pointing out how many blocks exist in the encoder. During the training procedure, the following augmentation techniques were used: random crop with the size of 512×512, random horizontal flip, and random rotation on the angle from $[-\pi/2; \pi/2]$. Also, a weighted binary cross-entropy as the loss function was applied. Classes of hogweed and background were taken into account with weights 1 and 2 for guarantying the compensation of imbalance. Adam optimizer was used with the default learning parameters.

4.1.3 RefineNet with ResNet Backbone

If one considers the large scale images (2250×4000 pixels), fast and accurate image processing requirements must be met. To achieve this goal, several unique architectures for solving these problems were considered. For instance, RefineNet is a well-known general framework which performs the building of high-resolution segmentation networks on top of some other networks (backbone), e.g., ResNet's family. RefineNet is able to fuse various levels of details on the different layers of convolutions. On contrary, it allows us for maintaining the fine structures of the original pictures. At the same time, it does not rely on the large intermediate features' maps, which do not occupy much memory for operation. ResNet is another popular architecture featured by the residual connections between functional blocks [33]. The advantage of residual connections is the construction of extremely deep networks helping achieve good quality. RefineNet with a pretrained ResNet-50 network as backbone was used [75]. According to its original design, the RefineNet output is $1/4$ of the original input size. For restoring the original mask size, the interpolation was applied. In terms of data augmentation, the following approaches were used: random rotations, random crops, and horizontal and vertical flips with 0.5 probability.

4.2 Capsule Neural Networks

In general, CNNs include pooling layers for minimizing the computational complexity and guarantying the translational invariance. Their disadvantage is the loosing of information on the location of features. It is necessary for neurons to make good use of the precise locations of different lower level sections of this object

in order to detect objects in the upper level layers. It would also be beneficial to retain the position information of detected features, but in the case of CNNs, it is difficult to achieve. Position is among several characteristics of objects in a scene. Other features unlike CNNs, these characteristics are encoded in a direct and elegant way by Capsules Networks. Scalar-output neurons are exchanged with the vector-output capsules, which represent the sets of neurons. A capsule is an object or part of an object, and the capsule's activity vector encodes the instantiation parameters of that part. Capsule networks' key weakness is what is called the routing algorithm. It performs routing of output vectors to upper level capsules from the lower level capsules. It consists of an iterative algorithm which selects the most suitable parent capsule in the way that nodes in a parse tree are represented by the active capsules in the network. The point is that the routing algorithms are processing intensive tasks. In order to minimize this effect, quantization or replacing by a more efficient type of algorithm while performing a similar task could be applied. Classical routing algorithms such as routing by agreement (RBA) [90] or scale distance agreement (SDA) [78] are not translational invariants; it is highly important to investigate the implementation of convolutional capsule layers in mobile devices and embedded systems. If implemented and used in real-world scenarios, an NN or a capsule network may achieve high training, validation, and test-set accuracy, but still underperform. For instance, Lapuschkin et al. [51] showed that the winning method in the PASCAL VOC competition was not the one that concentrated on detecting the object, using context or background correlations to correctly classify images. Explainable AI (XAI) may assist in detecting these problems, understand a trained neural or capsule network's limitations, and ensure good real-world scenario performance.

5 Edge Computing

Edge computing is an approach in the design of computing systems. In contrast to cloud computing, it implies bringing the data collection, computation, and data storage on board the end device. Therefore, edge computing reduces the network load and the data processing time. It is efficient in applications with lagging or completely inaccessible networks. Edge computing is of use for data protection from hacking.

Many embedded systems and mobile devices fall into the vast domain of edge computing. However, there is another subdomain - "Edge AI". It implies using AI algorithms on board the edge computing devices. Modern AI is closely related to ML and deep learning (DL). The application of such high-load algorithms requires the multicore architecture of the computer processor. Therefore, the edge AI needs both advanced ML/DL algorithms and advanced hardware for algorithm processing. Real-time aerial image processing is critical in the scope of the modern PA. Edge AI is capable to resolve such a crucial task.

This section highlights all the core aspects of edge computing in PA, including the origin of the edge computing paradigm, the application of modern computer vision in PA, and modern commercially available edge AI hardware.

5.1 Origin of Edge Computing

Edge computing paradigm appeared at the beginning of the 2000s due to the development of two core technologies—mobile electronics and artificial intelligence. There is a rapid advancement of mobile electronics in modern science and technology today. The number of transistors mounted on an integrated circuit increases twice every 2 years, following Moore's law. It leads to a decrease in electronics size and power usage, which implies the advent of increasingly high-performance miniature devices such as wearable electronics, tablets, smartphones, mobile images, video cameras, and IoT devices. There is also a related development in the market, leading to the emergence of embedded systems such as UAV miniature autopilot [67, 68, 107], nanosatellite microelectronics and their subsystems [77, 92], smart cities [72], smart houses [48], unmanned vehicles [29], and electronics for automated production in the context of Industry 4.0 [16].

At the same time, the rapid growth of artificial intelligence technology has occurred in the last decade. The pervasive creation of the Internet (which is important for the collection of large quantities of data for the training of neural networks) and multiprocessing and multi-threaded processors have made it possible. For example, graphics chip is an essential component for the effective implementation of neural networks. In turn, this enabled the development of the CNNs, as well as methods of object segmentation using computer vision. These include K-means methods [20], semantic segmentation [97], segmentation using the watershed method [8], methods of growing regions [108], and many other methods [32]. CNNs have also made significant progress over the past decades, starting with handwritten number classification [53] and ending with facial expression classification of emotions [24], slow motion frame interpolation [40], UAV control [12, 100], and other applications [91]. In addition, the use of semantic segmentation technology in image detection and classification tasks enables real-time determination of the limits of an object with pixel precision [49]. Later on, a deep learning area has arisen, which greatly extends and deepens neural network capabilities [93].

Further progress of embedded electronics and AI technologies has led to the creation and production of cost-effective multiprocessing and multi-threaded video processors for systems characterized by low computational power [65, 109]. Moreover, the advent of special architectures of NNs for efficient and rapid data classification on mobile platforms: YOLO [84], YOLT [105], Darknet [85], Fast RCNN [27], MobileNet SSD [110], R-FCN [19], etc.

Apart from that, edge computing is getting popular due to the continuous progress of mobile platforms and devices, e.g., IoT, robotics, embedded systems,

and wearables. Edge computing has a number of advantages over cloud computing and fog computing [65] because of the following technological limitations [36]:

- Delays associated with network latency
- Dependency on an internet connection
- Bottleneck problems: the number of possible clients depends on the server's computational capabilities
- Privacy issues

5.2 Edge Computing in Precision Agriculture

The development of the state-of-the-art tools for acquiring and processing aerial images has been greatly influenced by these trends. The development of embedded electronics has led to the widespread development of various UAVs for aerial observation, monitoring of agriculture, and inspection of infrastructure. The technique may be based on stereo pairs or lidars to obtain depth maps, as well as on different thermal and multispectral cameras. It has led to a significant amount of remote sensing data and the following datasets: DOTA [112], INRIA Aerial Image Labeling Dataset [64], SpaceNet [113], VIRAT Video Dataset [76], CLIF 2007 Dataset [17], Stanford Drone Dataset [86], and EPFL MMSPG Video Drone Dataset [9]. It should be noted that a collection of extensive dataset has a great positive impact on the usage of NNs to address the image processing problems in different industries more effectively. As an example, it is relevant for the task of germinating seeds [83] and monitoring agricultural areas using drone UAVs [11].

Typically, the UAVs collect images relative to the absolute coordination system throughout the flight. Post-factum processing, as well as classification of objects on them, occurs after the flight on a personal computer or server. These operations have to be done daily or even hourly, costly, and labor-intensive depending on the mission. Therefore, no commercially available UAVs are currently available that identify images on board during the flight. It should be noted that there are already a number of tasks that include semantic image segmentation and object classification during the mission [13, 26, 41].

Hogweed detection is among real-time semantic segmentation tasks [69]. The key reason for that is the extremely high growth rate of the weed. One plant can produce up to 100,000 seeds per year, which are easily dispersed by the wind. This means that a single plant accidentally left behind makes the clearing operation utterly pointless. That is why the detection and elimination of every single plant are of high importance. In case the operation in traditional paradigm "data collection - stitching orthophoto map - data processing" (see Fig. 6a) may take up to 24 h, which is not practical for the elimination of such a fast-spreading weed. Application of edge computing paradigm to the hogweed detection allows receiving instant information about the geolocation of the poisonous plant. The operation goes in the following way: the edge device on board the UAV collects the data by the

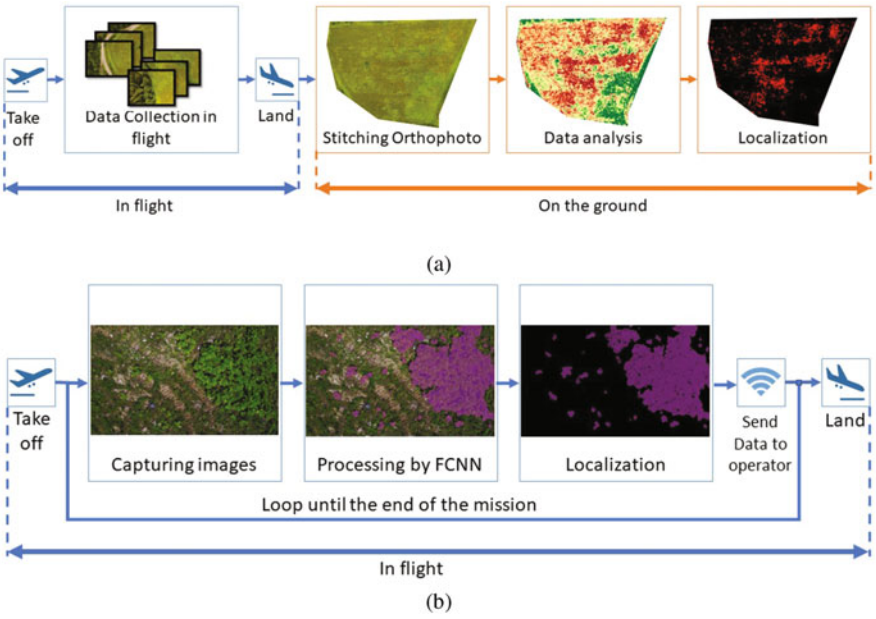


Fig. 6 Object detection on aerial imagery according to (**a**) traditional paradigm and (**b**) edge computing paradigm

camera, and then the image is processed by the FCNN. The output of FCNN is the semantic mask—it contains clusters of pixels, which cover the hogweed and the geolocation of the centroids of these clusters. This mask data is converted to text format and occupies several kilobytes of memory instead of hundreds of megabytes (for 4K images) or even gigabytes (in the case of 4K video stream). Then, the post-processed data is sent to the operator, who can manage the hogweed elimination operation (see Fig. 6b). Thus, the elimination could be started even before the end of the UAV flight. The "In-flight" phase takes the same time for both approaches (see Fig. 6); therefore, the edge computing approach saves a significant amount of time, spent for post-processing on the ground phase. Hence, the application of the edge computing paradigm to the hogweed detection task significantly improves the efficiency of the elimination of this plant.

5.3 Edge Computing Hardware for Computer Vision Tasks

Design and implementation of an intelligent computer vision system on board of a low-power embedded system with the constrained size is a challenging task. The key problem is the limited energy stored in the battery-powered systems and their low computational capabilities. A high number of research works reported on the

approaches addressing the long-term operation problem of low-power embedded sensing devices from a different perspective [43, 44, 71], but the problem has not been solved as yet.

Although most of embedded sensing devices are based on a low-power Micro Controller Unit (MCU), the research on adopting AI for energy-constrained devices exists [50], but still these works lack the experimental-based outcome [15]. Among the solutions that have been already tested is the application of a single-board computer with a powerful CPU, embedded or external GPU, or visual processing unit (VPU) for running AI methods optimized for embedded systems.

It brings us to the point that there is a strong need of fitting approaches for ML algorithms to embedded systems to address the problems mentioned earlier. These approaches are as follows: Hardware Acceleration (HA) by Digital Signal Processor (DSP) [18] or GPU [52]. The DSP HA is typically applied to mobile platforms due to its high performance and low power consumption even comparing to CPUs and GPUs. Next libraries could be used for realizing it into practice: Android Neural Network API (NNAPI) [1], TensorFlow Mobile [2], RenderScrpit-based CNNdroid [52], and RSTensorFlow. The last example is the GPU-based accelerator for matrix operations. It increases the matrix multiplication up to 3 times [5]. Moreover, it was reported that RenderScript could be used together with CPUs imprecise computing modes for reducing the execution time of computationally intensive models [73].

At present, a single-board computer (SBC) is a good trade-off for realizing computer vision methods in aerial robotics. The SBC is a tiny, lightweight, and cheap system. Obviously, their random access memory (RAM) size and processor performance are far worse than desktop computers. Nevertheless, SBCs are widely used in applications, e.g., greenhouses, flight tests, where reasonable AI performance, power consumption, and compact form factor are critical. In this section, we discuss the following commercial SBCs which have already been exploited in research projects and proof-of-concept implementation: RaspberryPi (RPi), Movidius Neuro Compute Stick (NCS) (version 1 and 2) [81], Nvidia Jetson Nano, Google Coral, and Odroid XU4. We compared these SBCs against a commercial laptop used as a reference. The key idea is to find the best SBC candidate. It should be:

- The lightest and the smallest, to be used as a payload for the drone (less than 200 grams and size less than $150 \times 150 \times 150$ mm).
- It should have the lowest possible power consumption, to minimize energy loss by the battery (less than 12 W).
- It should have the highest computational performance in terms of floating-point operations (at least 62 MFLOPS).
- It should have a sufficient amount of RAM to perform high-load FCNN computations (at least 1 Gb).

This technical comparison allows filtering some SBCs, which are capable of performing FCNN processing on board. After that, the platforms will be tested in terms of framerate (FPS: frames per second) in different computer vision tasks

to find which fits most to the hogweed detection application. The results of the comparison are reported in Table 1.

All the SBCs summarized in Table 1 have multiple cores and adequate RAM size for making inference of the state-of-the-art CNNs and FCNNs. However, for processing video in real-time conditions, it is not the only requirement: for realizing reasonably fast inference, an AI acceleration coprocessor is required as well. From this perspective, RPi equipped with different combinations of external VPU, Jetson Nano with powerful multiple cores GPU, and Google Coral with mobile Tensor Processing Unit (TPU) appear to be the most suitable platforms for the computer vision tasks in aerial robotics. Also, to fully satisfy the requirements of aerial robotics for weeds detection, the onboard computer is expected to comply with the data-intensive computing capabilities, low power consumption, small size, and weight.

For assessing the performance of SBCs, we carried out experiments aimed at solving various computer vision tasks including classification, object detection, and segmentation. The results are summarized in Table 2. In some experiments, we could not get the final numbers and specified the result with "N/A" (not applicable) which could occur due to some reasons, e.g., not sufficient memory size, unsupported layers or frameworks, hardware and/or software limitations. We also note that Odroid UX4 was discarded from this experiment since it does not significantly outweigh other platforms.

Experimental methodology is the following. The pretrained Mobile Net v2 [110], and Inception v1 [103] were used in classification test; Tiny YOLO v3 [85] and MobileNet SSD v2 [110] were used in the object detection test. These NNs were trained on the Common Objects in Context (COCO) dataset [57]. We were sending the test image with ten recognizable objects to the NNs input 10,000 times and computed the average framerate or FPS. The same test was performed using UNet [87]. We note here that it was pretrained on the original dataset for cells' membranes segmentation.

Table 2 demonstrates that the performance of Google Coral is the best. However, Jetson Nano and Raspberry Pi equipped with various NCSs ensure a better capability to run different types of CNNs. In contrast, Google Coral mentioned earlier is not able to run some types of object detection and semantic segmentation NNs. The problem is that the current version of application programming interface (API) does not support the upsampling operation, which is necessary for FCNN functionality. Moreover, for most of the relevant PA tasks, the framerate higher than 30 FPS is an excellent result even for aerial observation. Next, Google Coral does not support upsampling operation, which is the subject to discard it from the list of platforms candidates for semantic segmentation of plants from aerial imagery. Furthermore, according to the comparative study results summarized in Table 1, Google Coral is characterized by higher power consumption comparing to other SBCs. It can be a truly limiting factor for all the PA applications mentioned earlier.

In terms of metrics for the recognition quality evaluation, ROC AUC can be considered as a primary one. This metric shows a detailed view of some of the trained models (see Fig. 7). Also, it demonstrates that the best models in ROC AUC

Table 1 Comparative study on SBCs

Criterion	Laptop ASUS UX305CA	RPi 3B	RPi 3B + NCS	RPi 3B + NCS2	Jetson Nano	Google Coral	Odroid XU4
Architecture	x86_64	ARM	ARM	ARM	ARM	ARM	ARM
Type	Intel Core m3-6Y30	Cortex A53	Cortex A53	Cortex A53	Cortex A57	Cortex A53	Cortex A15
Cores	4	4	4	4	4	4	8
RAM, Gb	**4**	1	1	1	**4**	1	2
AI acceleration coprocessor	N/A	N/A	ncs with 12 128-bit shave cores @ 0.6 GHz	ncs2 with 16 128-bit shave cores @ 0.7 GHz	128-core NVIDIA Maxwell @ 0.921 GHz	Google Edge TPU ML accelerator coprocessor	N/A
GFLOPS	41.9	3.62	103.62	1003.62	472	**4000**	8.3
Average Power, W	15.9	**2.4**	3.4	3.4	10	15	13.9
GFLOPS/W	2.63	1.51	30.48	**295.18**	47.2	266.67	0.6
Size, WxDxH, mm	325 × 226 × 12.7	85.6 × 56.5 × 17	145.6×56.5×17	145.6×56.5×17	100 × 80 × 29	88.1 × 59.9 × 22.4	83 × 59 × 18
Weight, grams	907	**45**	80	80	140	138	60

Bold values illustrate the better results

Table 2 Comparative study of SBC performance in computer vision tasks

Model	Jetson Nano	RPi 3B	RPi 3B + NCS	RPi 3B + NCS2	Google Coral
Classification					
MobileNet v1 (300 × 300)	64 FPS	2.5 FPS	7.5 FPS	30 FPS	**130 FPS**
Inception v1 (224 × 224)	55 FPS	N\A	5.8 FPS	12 FPS	**243.9 FPS**
Object detection					
Tiny YOLO v2 (416 × 416)	**25 FPS**	0.5 FPS	2.57 FPS	5.1 FPS	N\ A
MobileNet SSD v2 (300 × 300)	39 FPS	1 FPS	4.5 FPS	11 FPS	**48 FPS**
Segmentation					
UNet (1 × 512 × 512)	**18 FPS**	N \A	N\A	5 FPS	N\A

Bold values illustrate the better results

are characterized by the highest true-positive rate and the lowest false-negative rate comparing to other models with similar threshold value. True-positive rate and false-positive rates are defined as follows:

$$TPR = \frac{TP}{TP + FN} \tag{3}$$

$$FPR = \frac{FP}{FP + TN} \tag{4}$$

where TP, TN, FP, and FN are the numbers defining true-positive, true-negative, false-positive, and false-negative classifications for target class, respectively.

For achieving reasonable performance on an embedded system, it is vital to investigate the networks. If we run the original U-Net at NVIDIA Jetson Nano, it results in 0.025 FPS which does not make sense in the real application. In fact, the networks include many parameters and may be the subject to the model scaling without a significant loss of accuracy. Width scaling, W, is about reducing the number of output channels in the convolutional layers. W indicates the degree of scaling. Original U-Net's W is 64. Depth scaling, D, is about reducing the number of convolutional blocks. Original U-Net'd D is 5. Compound scaling is about applying both the width and depth scaling.

A comparative study on the models in terms of recognition quality, inference performance, the number of parameters, and power consumption is shown in Table 3. This study shows that the more parameters are in the model under consideration, more time for making inference it takes. As noticed earlier, the recognition quality is featured using the area under the ROC curves.

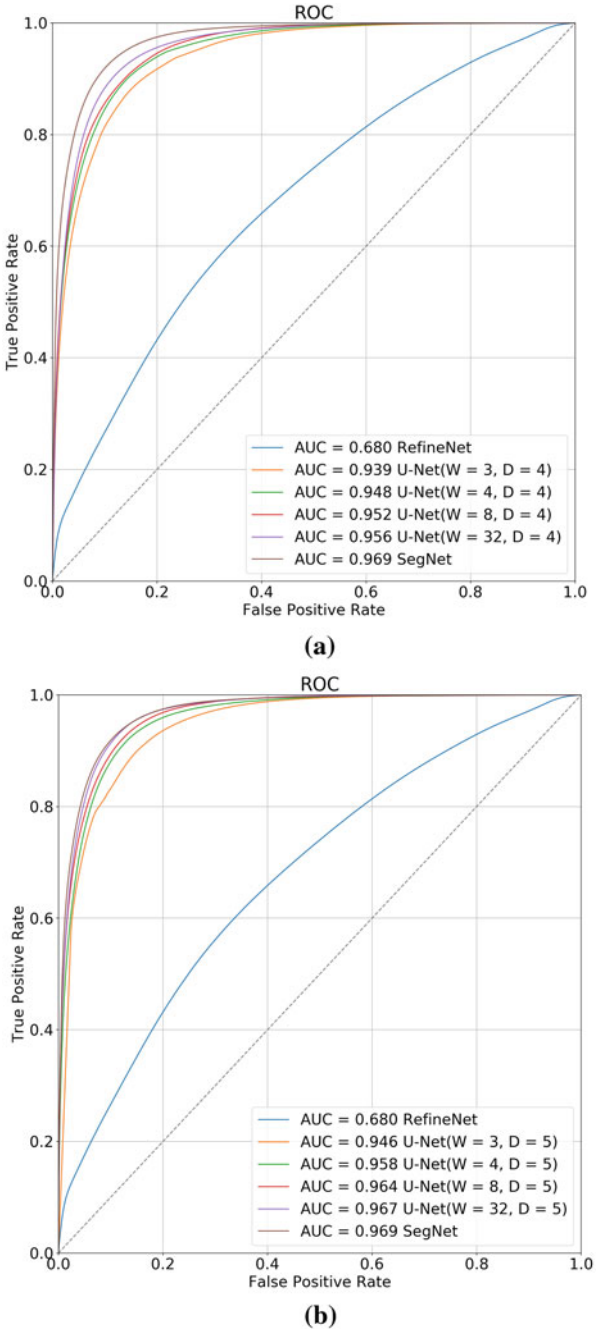


Fig. 7 ROC curves for different NN architectures: (**a**) Depth scaling, D, is 4 and (**b**) D is 5

Table 3 Assessment of models on Nvidia Jetson Nano

Model	FPS	ROC AUC	Parameters ($\times 10^6$)	Area covered, ha	Power, W
U-Net (W=64,D=5)	0.024	0.968	13.7	0.97	7.5
U-Net (W=32,D=5)	0.074	0.967	3.4	3.0	7.0
U-Net (W=16,D=5)	0.15	0.965	0.84	6.0	7.0
U-Net (W=8,D=5)	0.27	0.963	0.21	10.8	6.5
U-Net (W=4,D=5)	0.46	0.958	0.053	18.4	5.5
U-Net (W=3,D=5)	0.53	0.946	0.030	23.5	5.5
U-Net (W=2,D=5)	0.68	0.938	0.013	27.6	5.5
U-Net (W=32,D=4)	0.085	0.956	0.84	3.4	6.5
U-Net (W=16,D=4)	0.16	0.954	0.21	6.4	6.5
U-Net (W=8,D=4)	0.28	0.952	0.053	11.2	6.0
U-Net (W=4,D=4)	0.47	0.948	0.014	18.8	5.5
U-Net (W=3,D=4)	0.55	0.939	0.0077	24.0	5.5
U-Net (W=2,D=4)	0.70	0.938	0.0035	28.3	5.0
SegNet	0.020	0.969	29.4	0.8	7.0
RefineNet	0.20	0.968	27.3	8.2	7.5

6 Conclusions

Weed detection is getting the problem of high priority in precision agriculture. In this chapter, we have reported on and compared some monitoring approaches (UGV, WSN, satellites) based on their published characteristics and came to the conclusion that aerial robotics is the best option. It offers the potential of avoiding the distribution of weeds, ensures the coverage of huge areas, and performs real-time data analysis.

Next, we detailed the advantages of UAV and compared video cameras technologies which can be used as a payload for image acquisition. Although multispectral cameras possess much more important features in terms of data collection, their application is still expensive, i.e., RGB cameras equipped with FCNN can solve the segmentation task well. Regarding the FCNNs, we reported on several FCNN

architectures, as well as providing experimental results, and compared them with the capsule neural networks based on their published characteristics. Similarly to the case with cameras, capsule networks could guarantee better performance, but their application to embedded systems is still limited.

The presented outlook is promising for effectuating the detection and removal of weeds and enables disruptive innovation in precision agriculture.

References

1. NNAPI. https://developer.android.com/ndk/guides/neuralnetworks (2019). [Online; accessed 08-May-2021]
2. TensorFlow Mobile. https://www.tensorflow.org/lite (2019). [Online; accessed 09-May-2021]
3. Alavi, A.H., Jiao, P., Buttlar, W.G., Lajnef, N.: Internet of Things-enabled smart cities: State-of-the-art and future trends. Measurement **129**, 589–606 (2018)
4. Alexander, C., Moeslund, J.E., Bøcher, P.K., Arge, L., Svenning, J.C.: Airborne laser scanner (lidar) proxies for understory light conditions. Remote Sensing of Environment **134**, 152–161 (2013)
5. Alzantot, M., Wang, Y., Ren, Z., Srivastava, M.B.: RSTensorFlow: GPU enabled TensorFlow for deep learning on commodity android devices. In: Proceedings of the 1st International Workshop on Deep Learning for Mobile Systems and Applications, pp. 7–12. ACM (2017)
6. Andrew, W., Greatwood, C., Burghardt, T.: Aerial animal biometrics: Individual Friesian cattle recovery and visual identification via an autonomous UAV with onboard deep inference pp. 237–243 (2019)
7. Badrinarayanan, V., Kendall, A., Cipolla, R.: SegNet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence **39**(12), 2481–2495 (2017)
8. Bai, M., Urtasun, R.: Deep watershed transform for instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5221–5229 (2017)
9. Bonetto, M., Korshunov, P., Ramponi, G., Ebrahimi, T.: Privacy in mini-drone based video surveillance. In: 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), vol. 4, pp. 1–6. IEEE (2015)
10. Burks, T., Shearer, S., Gates, R.S., Donohue, K.: Backpropagation neural network design and evaluation for classifying weed species using color image texture. Transactions of the ASAE **43**(4), 1029 (2000)
11. Candiago, S., Remondino, F., De Giglio, M., Dubbini, M., Gattelli, M.: Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. Remote sensing **7**(4), 4026–4047 (2015)
12. Carrio, A., Sampedro, C., Rodriguez-Ramos, A., Campoy, P.: A review of deep learning methods and applications for unmanned aerial vehicles. Journal of Sensors **2017** (2017)
13. Chamoso, P., Raveane, W., Parra, V., González, A.: UAVs applied to the counting and monitoring of animals. In: Ambient Intelligence-Software and Applications, pp. 71–80. Springer (2014)
14. Chaudhury, A., Ward, C., Talasaz, A., Ivanov, A.G., Huner, N.P., Grodzinski, B., Patel, R.V., Barron, J.L.: Computer vision based autonomous robotic system for 3D plant growth measurement. In: 2015 12th Conference on Computer and Robot Vision (CRV), pp. 290–296. IEEE (2015)
15. Chauhan, J., Seneviratne, S., Hu, Y., Misra, A., Seneviratne, A., Lee, Y.: Breathing-based authentication on resource-constrained IoT devices using recurrent neural networks. Computer **51**(5), 60–67 (2018). DOI 10.1109/MC.2018.2381119

16. Cheng, G.J., Liu, L.T., Qiang, X.J., Liu, Y.: Industry 4.0 development and application of intelligent manufacturing. In: 2016 international conference on information system and artificial intelligence (ISAI), pp. 407–410. IEEE (2016)
17. Clif, A.: Dataset Over Ohio State University (2007)
18. Codrescu, L., Anderson, W., Venkumanhanti, S., Zeng, M., Plondke, E., Koob, C., Ingle, A., Tabony, C., Maule, R.: Hexagon DSP: An architecture optimized for mobile multimedia and communications. IEEE Micro **34**(2), 34–43 (2014)
19. Dai, J., Li, Y., He, K., Sun, J.: R-FCN: Object detection via region-based fully convolutional networks. In: Advances in neural information processing systems, pp. 379–387 (2016)
20. Dhanachandra, N., Manglem, K., Chanu, Y.J.: Image segmentation using k-means clustering algorithm and subtractive clustering algorithm. Procedia Computer Science **54**, 764–771 (2015)
21. Di Cicco, M., Potena, C., Grisetti, G., Pretto, A.: Automatic model based dataset generation for fast and accurate crop and weeds detection. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5188–5195. IEEE (2017)
22. Dutta, S., Cruz, J.A., Jiao, Y., Chen, J., Kramer, D.M., Osteryoung, K.W.: Non-invasive, whole-plant imaging of chloroplast movement and chlorophyll fluorescence reveals photosynthetic phenotypes independent of chloroplast photorelocation defects in chloroplast division mutants. The Plant Journal **84**(2), 428–442 (2015)
23. Eugster, P., Sundaram, V., Zhang, X.: Debugging the internet of things: The case of wireless sensor networks. IEEE Software (1), 1–1 (2015)
24. Fan, Y., Lu, X., Li, D., Liu, Y.: Video-based emotion recognition using CNN-RNN and C3D hybrid networks. In: Proceedings of the 18th ACM International Conference on Multimodal Interaction, pp. 445–450. ACM (2016)
25. Fawakherji, M., Youssef, A., Bloisi, D., Pretto, A., Nardi, D.: Crop and weeds classification for precision agriculture using context-independent pixel-wise segmentation. In: 2019 Third IEEE International Conference on Robotic Computing (IRC), pp. 146–152. IEEE (2019)
26. Flammini, F., Naddei, R., Pragliola, C., Smarra, G.: Towards automated drone surveillance in railways: State-of-the-art and future directions. In: International Conference on Advanced Concepts for Intelligent Vision Systems, pp. 336–348. Springer (2016)
27. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580–587 (2014)
28. Große-Stoltenberg, A., Hellmann, C., Thiele, J., Werner, C., Oldeland, J.: Early detection of GPP-related regime shifts after plant invasion by integrating imaging spectroscopy with airborne lidar. Remote sensing of environment **209**, 780–792 (2018)
29. Gurghian, A., Koduri, T., Bailur, S.V., Carey, K.J., Murali, V.N.: DeepLanes: End-to-end lane position estimation using deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 38–45 (2016)
30. Hammad, I., El-Sankary, K.: Impact of approximate multipliers on VGG deep learning network. IEEE Access **6**, 60438–60444 (2018)
31. Haug, S., Michaels, A., Biber, P., Ostermann, J.: Plant classification system for crop/weed discrimination without segmentation. In: IEEE winter conference on applications of computer vision, pp. 1142–1149. IEEE (2014)
32. He, H.J., Zheng, C., Sun, D.W.: Image segmentation techniques. In: Computer Vision Technology for Food Quality Evaluation, pp. 45–63. Elsevier (2016)
33. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778 (2016)
34. Hong, K., Liu, X., Liu, G., Chen, W.: Detection of physical stress using multispectral imaging. Neurocomputing **329**, 116–128 (2019)
35. Humplík, J.F., Lazár, D., Husičková, A., Spíchal, L.: Automated phenotyping of plant shoots using imaging methods for analysis of plant stress responses–a review. Plant methods **11**(1), 29 (2015)

36. Ignatov, A., Timofte, R., Chou, W., Wang, K., Wu, M., Hartley, T., Van Gool, L.: AI benchmark: Running deep neural networks on android smartphones. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
37. ISPAG: International society of precision agriculture (2020). URL https://www.ispag.org/
38. Ivanov, M., Sergyienko, O., Tyrsa, V., Lindner, L., Flores-Fuentes, W., Rodríguez-Quiñonez, J.C., Hernandez, W., Mercorelli, P.: Influence of data clouds fusion from 3D real-time vision system on robotic group dead reckoning in unknown terrain. IEEE/CAA Journal of Automatica Sinica **7**(2), 368–385 (2020)
39. Ivanov, S., Bhargava, K., Donnelly, W.: Precision farming: Sensor analytics. IEEE Intelligent Systems **30**(4), 76–80 (2015)
40. Jiang, H., Sun, D., Jampani, V., Yang, M.H., Learned-Miller, E., Kautz, J.: Super slomo: High quality estimation of multiple intermediate frames for video interpolation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9000–9008 (2018)
41. Jordan, B.R., et al.: A bird's-eye view of geology: The use of micro drones/UAVs in geologic fieldwork and education. GSA Today **25**(7), 50–52 (2015)
42. Kamilaris, A., Prenafeta-Boldú, F.X.: Deep learning in agriculture: A survey. Computers and Electronics in Agriculture **147**, 70–90 (2018)
43. Kaup, F., Gottschling, P., Hausheer, D.: PowerPi: Measuring and modeling the power consumption of the Raspberry Pi. In: 39th Annual IEEE Conference on Local Computer Networks, pp. 236–243. IEEE (2014)
44. Khaled, F., Ondel, O., Allard, B.: Optimal energy harvesting from serially connected microbial fuel cells. IEEE Transactions on Industrial Electronics **62**(6), 3508–3515 (2015)
45. Khan, Z., Rahimi-Eichi, V., Haefele, S., Garnett, T., Miklavcic, S.J.: Estimation of vegetation indices for high-throughput phenotyping of wheat using aerial imaging. Plant methods **14**(1), 20 (2018)
46. Kim, J., Ryu, Y., Jiang, C., Hwang, Y.: Continuous observation of vegetation canopy dynamics using an integrated low-cost, near-surface remote sensing system. Agricultural and forest meteorology **264**, 164–177 (2019)
47. Kitić, G., Tagarakis, A., Cselyuszka, N., Panić, M., Birgermajer, S., Sakulski, D., Matović, J.: A new low-cost portable multispectral optical device for precise plant status assessment. Computers and Electronics in Agriculture **162**, 300–308 (2019)
48. Kodali, R.K., Jain, V., Bose, S., Boppana, L.: IoT based smart security and home automation system. In: International conference on computing, communication and automation (ICCCA), pp. 1286–1289. IEEE (2016)
49. Kulikov, V., Yurchenko, V., Lempitsky, V.: Instance segmentation by deep coloring. arXiv preprint arXiv:1807.10007 (2018)
50. Lane, N.D., Bhattacharya, S., Mathur, A., Georgiev, P., Forlivesi, C., Kawsar, F.: Squeezing deep learning into mobile and embedded devices. IEEE Pervasive Computing **16**(3), 82–88 (2017)
51. Lapuschkin, S., Binder, A., Montavon, G., Müller, K., Samek, W.: Analyzing classifiers: Fisher vectors and deep neural networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2912–2920 (2016)
52. Latifi Oskouei, S.S., Golestani, H., Hashemi, M., Ghiasi, S.: CNNdroid: GPU-accelerated execution of trained deep convolutional neural networks on android. In: Proceedings of the 24th ACM international conference on Multimedia, pp. 1201–1205. ACM (2016)
53. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. Neural computation **1**(4), 541–551 (1989)
54. Li, Y., Dong, X., Wang, W.: Additive powers-of-two quantization: An efficient non-uniform discretization for neural networks. In: International Conference on Learning Representations (2019)
55. Lin, G., Milan, A., Shen, C., Reid, I.: RefineNet: Multi-path refinement networks for high-resolution semantic segmentation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5168–5177 (2017)

56. Lin, M., Ji, R., Wang, Y., Zhang, Y., Zhang, B., Tian, Y., Shao, L.: HRank: Filter pruning using high-rank feature map. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2020)
57. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision, pp. 740–755. Springer (2014)
58. Lindner, L., Sergiyenko, O., Tyrsa, V., Mercorelli, P.: An approach for dynamic triangulation using servomotors. In: IEEE 23rd International Symposium on Industrial Electronics (ISIE), pp. 1926–1931 (2014)
59. Liu, M.W., Ozdogan, M., Zhu, X.: Crop type classification by simultaneous use of satellite images of different resolutions. IEEE Transactions on geoscience and remote sensing **52**(6), 3637–3649 (2013)
60. Lottes, P., Behley, J., Milioto, A., Stachniss, C.: Fully convolutional networks with sequential information for robust crop and weed detection in precision farming. IEEE Robotics and Automation Letters **3**(4), 2870–2877 (2018)
61. Lottes, P., Hoeferlin, M., Sander, S., Müter, M., Schulze, P., Stachniss, L.C.: An effective classification system for separating sugar beets and weeds for precision farming applications. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 5157–5163. IEEE (2016)
62. Lottes, P., Khanna, R., Pfeifer, J., Siegwart, R., Stachniss, C.: UAV-based crop and weed classification for smart farming. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 3024–3031. IEEE (2017)
63. Luo, Y., Wong, Y., Kankanhalli, M., Zhao, Q.: $\mathcal{G}$ -SoftMax: Improving intraclass compactness and interclass separability of features. IEEE Transactions on Neural Networks and Learning Systems **31**(2), 685–699 (2020)
64. Maggiori, E., Tarabalka, Y., Charpiat, G., Alliez, P.: Can semantic labeling methods generalize to any city? The INRIA aerial image labeling benchmark. In: IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3226–3229. IEEE (2017)
65. Marantos, C., Karavalakis, N., Leon, V., Tsoutsouras, V., Pekmestzi, K., Soudris, D.: Efficient support vector machines implementation on Intel/Movidius Myriad 2. In: 2018 7th International Conference on Modern Circuits and Systems Technologies (MOCAST), pp. 1–4. IEEE (2018)
66. Martinelli, F., Scalenghe, R., Davino, S., Panno, S., Scuderi, G., Ruisi, P., Villa, P., Stroppiana, D., Boschetti, M., Goulart, L.R., et al.: Advanced methods of plant disease detection. a review. Agronomy for Sustainable Development **35**(1), 1–25 (2015)
67. Menshchikov, A., Ermilov, D., Dranitsky, I., Kupchenko, L., Panov, M., Fedorov, M., Somov, A.: Data-driven body-machine interface for drone intuitive control through voice and gestures. In: 45th Annual Conference of the IEEE Industrial Electronics Society, vol. 1, pp. 5602–5609 (2019)
68. Menshchikov, A., Lopatkin, D., Tsykunov, E., Tsetserukou, D., Somov, A.: Realizing body-machine interface for quadrotor control through Kalman filters and recurrent neural network. In: 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 595–602 (2020)
69. Menshchikov, A., Shadrin, D., Prutyanov, V., Lopatkin, D., Sosnin, S., Tsykunov, E., Iakovlev, E., Somov, A.: Real-time detection of hogweed: UAV platform empowered by deep learning. IEEE Transactions on Computers pp. 1–1 (2021)
70. Milioto, A., Lottes, P., Stachniss, C.: Real-time semantic segmentation of crop and weed for precision agriculture robots leveraging background knowledge in CNNS. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 2229–2235. IEEE (2018)
71. Minakov, I., Passerone, R.: PASES: An energy-aware design space exploration framework for wireless sensor networks. Journal of Systems Architecture **59**(8), 626 – 642 (2013)
72. Misbahuddin, S., Zubairi, J.A., Saggaf, A., Basuni, J., Sulaiman, A., Al-Sofi, A., et al.: IoT based dynamic road traffic management for smart cities. In: 12th International Conference on High-capacity Optical Networks and Enabling/Emerging Technologies (HONET), pp. 1–5. IEEE (2015)

73. Motamedi, M., Fong, D., Ghiasi, S.: Cappuccino: efficient CNN inference software synthesis for mobile system-on-chips. IEEE Embedded Systems Letters **11**(1), 9–12 (2019)
74. Muangprathub, J., Boonnam, N., Kajornkasirat, S., Lekbangpong, N., Wanichsombat, A., Nillaor, P.: IoT and agriculture data analysis for smart farm. Computers and Electronics in Agriculture **156**, 467–474 (2019)
75. Nekrasov, V., Shen, C., Reid, I.: Light-Weight RefineNet for Real-Time Semantic Segmentation. In: 29th British Machine Vision Conference (BMVC), pp. 1–15 (2018)
76. Oh, S., Hoogs, A., Perera, A., Cuntoor, N., Chen, C.C., Lee, J.T., Mukherjee, S., Aggarwal, J., Lee, H., Davis, L., et al.: A large-scale benchmark dataset for event recognition in surveillance video. In: Computer Vision and Pattern Recognition (CVPR), pp. 3153–3160. IEEE (2011)
77. Osman, D.A., Mohamed, S.W.: Hardware and software design of onboard computer of israsat1 CubeSat. In: 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE), pp. 1–4. IEEE (2017)
78. Peer, D., Stabinger, S., Rodriguez-Sanchez, A.: Limitation of capsule networks (2021)
79. Picon, A., Alvarez-Gila, A., Seitz, M., Ortiz-Barredo, A., Echazarra, J., Johannes, A.: Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. Computers and Electronics in Agriculture **161**, 280–290 (2019)
80. Pignatti, S., Casa, R., Harfouche, A., Huang, W., Palombo, A., Pascucci, S.: Maize crop and weeds species detection by using UAV VNIR hyperspectral data. In: IEEE International Geoscience and Remote Sensing Symposium, pp. 7235–7238 (2019)
81. Prutyanov, V., Melentev, N., Lopatkin, D., Menshchikov, A., Somov, A.: Developing IoT devices empowered by artificial intelligence: Experimental study. In: Global IoT Summit (GIoTS), pp. 1–6 (2019)
82. Rajasegaran, J., Jayasundara, V., Jayasekara, S., Jayasekara, H., Seneviratne, S., Rodrigo, R.: DeepCaps: Going deeper with capsule networks. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10717–10725 (2019)
83. Rasti, P., Demilly, D., Benoit, L., Belin, E., Ducournau, S., Chapeau-Blondeau, F., Rousseau, D.: Low-cost vision machine for high-throughput automated monitoring of heterotrophic seedling growth on wet paper support (2018)
84. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788 (2016)
85. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7263–7271 (2017)
86. Robicquet, A., Sadeghian, A., Alahi, A., Savarese, S.: Learning social etiquette: Human trajectory understanding in crowded scenes. In: European conference on computer vision, pp. 549–565. Springer (2016)
87. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241. Springer (2015)
88. Russel, J., Cohn, R.: International Aerial Robotics Competition. Book on Demand (2013)
89. Sa, I., Popović, M., Khanna, R., Chen, Z., Lottes, P., Liebisch, F., Nieto, J., Stachniss, C., Walter, A., Siegwart, R.: WeedMap: a large-scale semantic weed mapping framework using aerial multispectral imaging and deep neural network for precision farming. Remote Sensing **10**(9), 1423 (2018)
90. Sabour, S., Frosst, N., Hinton, G.E.: Dynamic routing between capsules. In: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (eds.) Advances in Neural Information Processing Systems, vol. 30, pp. 3856–3866. Curran Associates, Inc. (2017)
91. Sanchez-Castro, J.J., Rodríguez-Quiñonez, J.C., Ramírez-Hernández, L.R., Galaviz, G., Hernández-Balbuena, D., Trujillo-Hernández, G., Flores-Fuentes, W., Mercorelli, P., Hernández-Perdomo, W., Sergiyenko, O., González-Navarro, F.F.: A lean convolutional neural network for vehicle classification. In: IEEE 29th International Symposium on Industrial Electronics (ISIE), pp. 1365–1369 (2020)

92. Sánchez-Macián, A., Reviriego, P., Tabero, J., Regadío, A., Maestro, J.A.: SEFI protection for nanosat 16-bit chip onboard computer memories. IEEE Transactions on Device and Materials Reliability **17**(4), 698–707 (2017)
93. Schmidhuber, J.: Deep learning in neural networks: An overview. Neural networks **61**, 85–117 (2015)
94. Sergiyenko, O., Flores-Fuentes, W., Mercorelli, P.: Machine Vision and Navigation. Springer (2019)
95. Sergiyenko, O.Y., Tyrsa, V.V.: 3D optical machine vision sensors with intelligent data management for robotic swarm navigation improvement. IEEE Sensors Journal **21**(10), 11262–11274 (2021)
96. Shadrin, D., Menshchikov, A., Ermilov, D., Somov, A.: Designing future precision agriculture: Detection of seeds germination using artificial intelligence on a low-power embedded system. IEEE Sensors Journal **19**(23), 11573–11582 (2019)
97. Shadrin, D.G., Kulikov, V., Fedorov, M.: Instance segmentation for assessment of plant growth dynamics in artificial soilless conditions. In: 29th British Machine Vision Conference (BMVC), pp. 1–11 (2018)
98. Shean, D.E., Alexandrov, O., Moratto, Z.M., Smith, B.E., Joughin, I.R., Porter, C., Morin, P.: An automated, open-source pipeline for mass production of digital elevation models (DEMs) from very-high-resolution commercial stereo satellite imagery. ISPRS Journal of Photogrammetry and Remote Sensing **116**, 101–117 (2016)
99. Shuaibu, M., Lee, W.S., Schueller, J., Gader, P., Hong, Y.K., Kim, S.: Unsupervised hyperspectral band selection for apple Marssonina blotch detection. Computers and Electronics in Agriculture **148**, 45–53 (2018)
100. Smolyanskiy, N., Kamenev, A., Smith, J., Birchfield, S.: Toward low-flying autonomous MAV trail navigation using deep neural networks for environmental awareness. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 4241–4247. IEEE (2017)
101. Sodhi, P., Vijayarangan, S., Wettergreen, D.: In-field segmentation and identification of plant structures using 3D imaging. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5180–5187. IEEE (2017)
102. Somov, A., Shadrin, D., Fastovets, I., Nikitin, A., Matveev, S., Oseledets, I., Hrinchuk, O.: Pervasive agriculture: IoT-enabled greenhouse for plant growth control. IEEE Pervasive Computing **17**(4), 65–75 (2018)
103. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1–9 (2015)
104. Tan, M., Le, Q.V.: EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946 (2019)
105. Van Etten, A.: You only look twice: Rapid multi-scale object detection in satellite imagery. arXiv preprint arXiv:1805.09512 (2018)
106. Vega, F.A., Ramirez, F.C., Saiz, M.P., Rosua, F.O.: Multi-temporal imaging using an unmanned aerial vehicle for monitoring a sunflower crop. Biosystems Engineering **132**, 19–27 (2015)
107. Venkatesh, G.A., Sumanth, P., Jansi, K.: Fully autonomous UAV. In: 2017 International Conference on Technical Advancements in Computers and Communications (ICTACC), pp. 41–44. IEEE (2017)
108. Vo, A.V., Truong-Hong, L., Laefer, D.F., Bertolotto, M.: Octree-based region growing for point cloud segmentation. ISPRS Journal of Photogrammetry and Remote Sensing **104**, 88–100 (2015)
109. Wang, Y., Li, H., Li, X.: Re-architecting the on-chip memory sub-system of machine-learning accelerator for embedded devices. In: Proceedings of the 35th International Conference on Computer-Aided Design, p. 13. ACM (2016)
110. Wang, Y., Yan, J., Sun, Q., Li, J., Yang, Z.: A MobileNets convolutional neural network for GIS partial discharge pattern recognition in the ubiquitous power internet of things context: Optimization, comparison, and application. IEEE Access **7**, 150226–150236 (2019)

111. Weerasinghe, C., Kharitonenko, I., Ogunbona, P.: Method of color interpolation in a single sensor color camera using green channel separation. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 4, pp. 3233–3236 (2002)
112. Xia, G.S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L.: Dota: A large-scale dataset for object detection in aerial images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3974–3983 (2018)
113. Yuan, J.: Learning building extraction in aerial scenes with convolutional networks. IEEE transactions on pattern analysis and machine intelligence **40**(11), 2793–2798 (2017)
114. Zhang, Z.: Improved Adam optimizer for deep neural networks. In: 2018 IEEE/ACM 26th International Symposium on Quality of Service (IWQoS), pp. 1–2 (2018)
115. Zhao, Y., He, Y., Xu, X.: A novel algorithm for damage recognition on pest-infested oilseed rape leaves. Computers and electronics in agriculture **89**, 41–50 (2012)
116. Zhou, L., Chen, N., Chen, Z., Xing, C.: ROSCC: An efficient remote sensing observation-sharing method based on cloud computing for soil moisture mapping in precision agriculture. IEEE Journal of selected topics in applied earth observations and remote sensing **9**(12), 5588–5598 (2016)
117. Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V.: Learning transferable architectures for scalable image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 8697–8710 (2018)

Zooming Assisted Stereo Matching

Huei-Yung Lin and Yu-Ting Chen

Acronyms

GC	Graph Cut
BP	Belief Propagation
SGM	Semi-Global Matching
FOV	Field-of-View
SAD	Sum-of-Absolute Difference
NCC	Normalized Cross Correlation
WTA	Winner-Take-All
BM	Block Matching
BPR	Bad Pixel Rate

1 Introduction

One essential problem of sensing technologies is to measure the distance of an object. Its application areas range from industrial inspection [23] and reverse engineering to autonomous robot navigation [13, 24] and computer graphics [15]. There exist active sensing techniques such as laser range scanning [22, 25]. Typically, the visual cues observed in the recorded images are used for depth perception of the scene. An important depth measurement method using computer vision-based techniques is binocular vision. The images captured by two cameras at different positions will introduce a visual parallax [16]. We can then use this image disparity to derive the depth of a 3D point by triangulation. The accurate and dense correspondence matching is essential for practical applications such as

H.-Y. Lin (✉) · Y.-T. Chen
Department of Electrical Engineering, National Chung Cheng University, Chiayi, Taiwan
e-mail: lin@ee.ccu.edu.tw

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1_3

3D reconstruction, augmented reality, scene recognition, and 3D pose estimation. In the past few decades, many stereo matching algorithms have been proposed by academic researchers and industrial practitioners. Many computational stereo algorithms have been proposed in the late 1970s and till now. The objective is usually to obtain an accurate disparity map from a stereo image pair [21]. Since the availability of Middlebury stereo datasets for public benchmarking in the last decade [18], the development of stereo matching techniques has been accelerated and tested extensively. Based on the well-investigated camera and imaging geometry [7], the 3D structure of the scene can then be derived using the image formation parameters.

To ease stereo matching, the pair of images acquired by the cameras is usually rectified for further processing [6]. The geometric relation between two cameras is used to generate a new image pair with the epipolar lines parallel to the image scanlines [17]. It can be considered as taking the image pair using a new stereo rig with fixed centers of projection and parallel optical axes. The image rectification process greatly simplifies the correspondence search to a one-dimensional space along the image scanlines. More importantly, it constructs the form of standard input images for the development of state-of-the-art stereo techniques. The matching algorithms can focus more on the image content itself, instead of the camera system calibration.

Given a rectified image pair, a stereo matching algorithm usually consists of four steps: cost initialization, cost aggregation, disparity selection, and refinement. In the cost initialization step, the stereo similarity measures are generally performed at the pixel level. The use of different similarity measures will result in different performances and should be adopted under suitable conditions [11]. The computed initial matching costs represented in the disparity image are then aggregated over a local region. This cost aggregation step is used to improve the accuracy of the disparity measurements. In the disparity selection step, the winner-take-all algorithms are commonly used to assign the disparity of a pixel. Finally, the disparity refinement is performed with various filtering methods to remove or mitigate the incorrect matches.

The disparity of binocular vision can be derived by stereo matching techniques [21]. In general, the methods are classified into local or global approaches. The disparity of a pixel might be determined by color or intensity values within a fixed or adaptive support region in the local stereo matching algorithms [12]. A window or support region has to be sufficiently large to provide enough intensity variation, but sufficiently small to mitigate the depth discontinuities for reliable disparity estimation. This property might cause some problems such as noisy disparities obtained in textureless regions and on object region boundaries. As for the global stereo matching algorithms, a global optimization is applied for the disparity computation on all pixels in the image pair. In the past few decades, many global optimization methods such as graph cut (GC) [2, 4], belief propagation (BP) [5, 28], and semi-global matching (SGM) [9, 10] have been proposed and achieved some good results.

One major aspect of the stereo matching improvement is to reduce the correspondence mismatching rate. The classic work by Okutomi and Kanade presented

an approach to increase the correctness of stereo matching by using more cameras for image acquisition from different viewpoints [20]. More recently, Zbontar and LeCun trained and adopted some specific descriptors for stereo matching [29], and Buades and Facciolo proposed to use multiple masks and scales to derive more information for disparity computation [3]. On the other hand, Kim and Kim restricted the disparity search range by the use of texture information [14]. Similar to other approaches, the above methods are able to reduce the correspondence matching errors, but at the cost of increasing the hardware requirement. In this work, the idea is to improve the stereo matching results using multiple zoom image pairs. The stereo image pairs captured with different focus lengths are adopted for disparity computation. In addition to the correspondence matching on each stereo image pair, a zoom rectification technique is proposed to restrict the search range on the multiple zoom images. The matching costs from stereo and zoom image pairs, respectively, are then aggregated to remove the unreliable correspondence results. Our approach incorporates the additional zooming constraint to provide a robust disparity reliability check over stereo matching. Furthermore, due to the nature of its technical construction flow, the existing or newly developed stereo matching algorithms can be adopted with the proposed zoom-stereo framework. One important application of the proposed stereo with zooming technique is for the depth recovery. It is a passive method to measure the distance with multiple image captures.

The system flowchart of the proposed stereo with zooming technique is shown in Fig. 1. Given multiple image pairs acquired by a stereo camera system with different focal length settings, an image preprocessing stage is first adopted to detect the reliable points. Stereo matching is carried out on the image pair taken with the same focal length to derive an initial disparity map. It is then followed by *zoom matching* on the series of images captured by the same camera to identify the point correspondences among the zooming image sequence. Finally, the cost aggregation which combines the matching from stereo and zoom is performed to refine the disparity map.

2 Zoom-Stereo Image Formation

The field of view (FOV) of a camera determines the visible area of a scene and is controlled by the focal length or zoom factor. When the camera location is fixed with respective to the static scene, the scale of the image content is proportional to the zoom setting. The way of image change induced by different focal length settings is not the same as the imaging process due the camera motion such as structure from motion. Since the 3D structure of the scene remains intact under camera zooming, there is no further perspective distortion to be considered when the focal length is changed. Consequently, the optical flow induced by zooming is scene independent and homogeneous. Moreover, this is true not only for the optical zoom but also for the digital zoom.

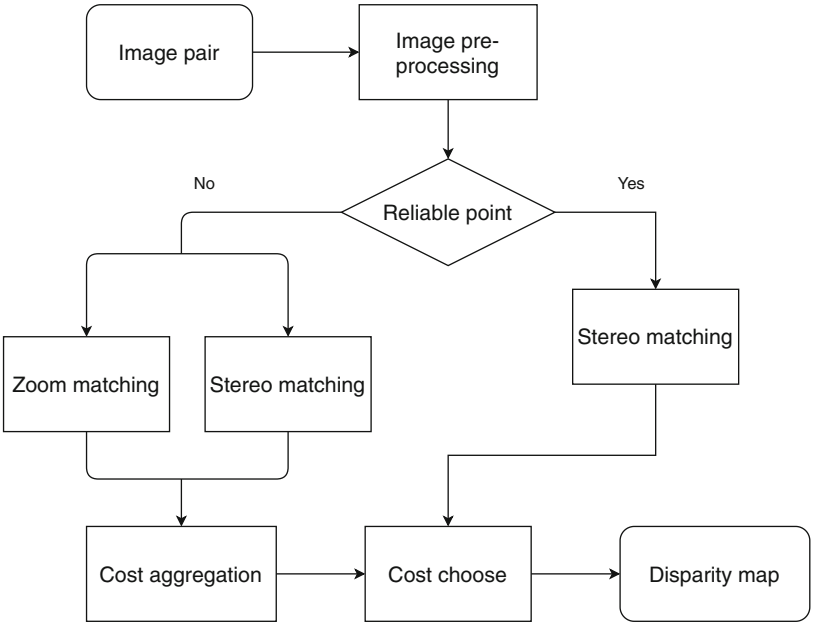


Fig. 1 The system flowchart of the proposed stereo with zooming technique. Given multiple image pairs acquired by a stereo camera system with different focal length settings as inputs, the disparity map is derived with the fusion of stereo and zoom information

It is well known that an image magnified by the digital zoom is created by spatial resampling with interpolation. Therefore, digital zoom images are de facto synthetic and generally come with lower image quality and less details. To ensure that the image contents fully reflect the real scene information, we adopt optical zoom in our framework for image acquisition. This involves the physical zoom lens motion and is able to provide independent image recording. Nevertheless the image formation is identical for both optical and digital zooms, the easily generated digital zoom images will also be adopted for our algorithm verification.

We first consider the relation among the series of images acquired with different zoom settings. For an ideal case when the optical axis of the camera is perpendicular to the image plane, the zoom matching constraint is imposed by the radial directions centered at the principal point on the image. These can be considered as *zoom vectors* defined as the motion vectors induced by the zooming process. Figure 2 shows the examples of zoom vectors observed as the optical flows on zoom-in and zoom-out images [30]. Let f_i and f_j be the focal length of the ith and jth zoom images, respectively. Then, the feature correspondences between the zoom image pair are constrained by the equation

$$\mathbf{v}_j = \lambda \mathbf{v}_i, \tag{1}$$

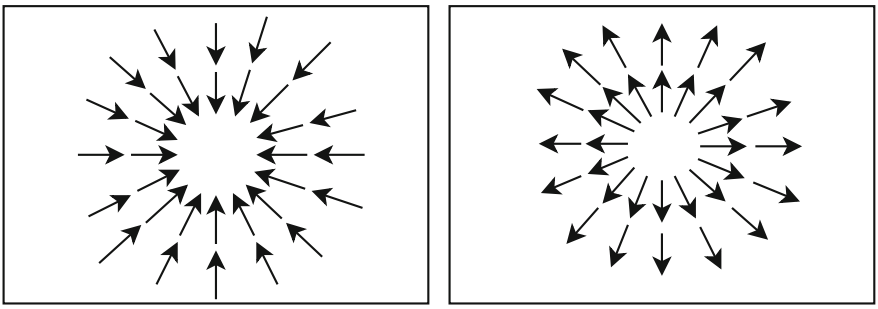


Fig. 2 The motion vectors generated from zooming. The centers are focus of contraction and focus of expansion, respectively

where $\mathbf{v}_i$ and $\mathbf{v}_j$ are the vectors originated from the principal point, and λ is the focal length ratio, i.e.,

$$\lambda = \frac{f_j}{f_i}. \tag{2}$$

2.1 Image Rectification

In conventional stereo matching, image rectification is an important procedure which transforms the image pair to a new one with the epipolar lines parallel to the image scanlines [6]. It reduces the correspondence search between the left and right images from the two-dimensional space to one-dimensional. Since the relationship between the viewpoints is essential, both the intrinsic and external camera parameters need to be obtained for image rectification [1]. In the proposed stereo with zooming, it requires multiple zoom image pairs for both stereo and zoom matching. However, due to the non-ideal lens movement (not coaxial with the optical axis) from the focal length change, the image rectification results are not identical for different zoom pairs even a suitable scale factor is applied. Figure 3 shows two pairs of zoom images after rectification. Although the camera position (i.e., the viewpoint) is fixed (but the FOV is changed), a slight change on the zoom factor results in different rectified geometry. In this work, we present a zoom rectification approach to make the pairs of rectified images consistent under different zoom factors. Based on the homography transformation between the zoom images, it ensures the search of matching points constrained correctly.

In the ideal zoom lens camera model, the optical axis is orthogonal to the sensor plane during the zooming process, so the principal point location is fixed. Moreover, the extrinsic camera parameters remain the same since the camera coordinate system does not change. In this case, only the image rectification is

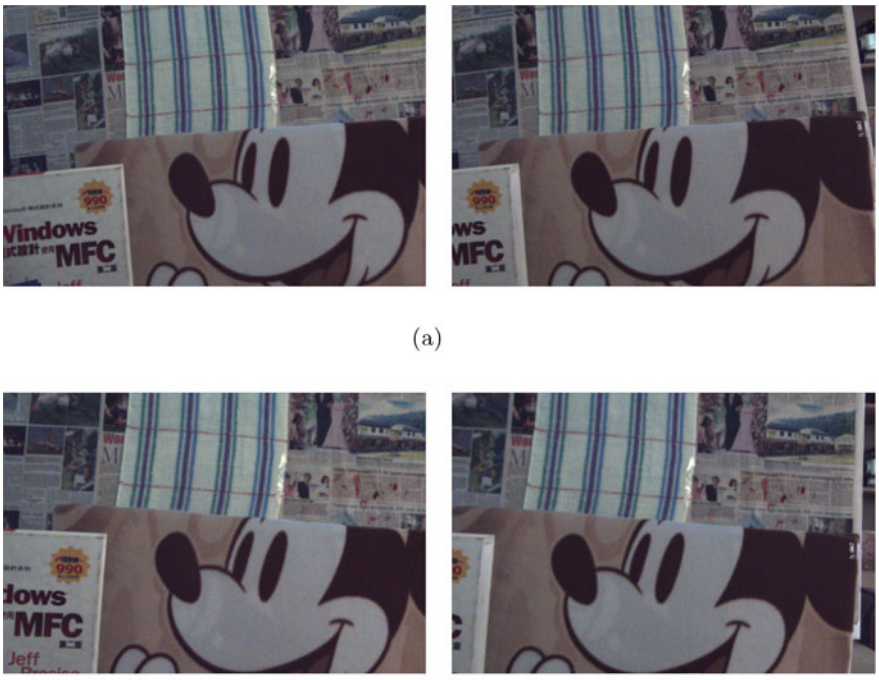


Fig. 3 Two pairs of zoom images after rectification. Although the camera position is fixed, a slight change on the zoom factor results in different rectified geometry. (**a**) Zoom1 image after image rectification. (**b**) Zoom2 image after image rectification

Fig. 4 The stereo pair with only the image rectification. Left: the left input image. Right: the right input image

required for the stereo from zooming framework. However, if the consistency of image rectification for different zoom settings is considered, the stereo pairs are transformed by homography for further *zoom rectification* [8]. Figures 4 and 5 show the stereo pair with only the image rectification and with an additional

Fig. 5 The stereo pair with additional zoom rectification. Left: the left input image. Right: the right input image

zoom rectification, respectively. To compute the homography matrix, the feature correspondences between the zoom images are extracted by SIFT [19] and K-NN algorithms. The correspondence matching for stereo and zoom is then carried out on the homography transformed zoom images. In the implementation, the zoom image with larger FOV is used to perform homography since it is able to provide more visible area.

3 Cost Aggregation

The main objective of this work is to improve the correctness of disparities obtained from stereo matching algorithms. Thus, it is necessary to detect the incorrect correspondences (or called unreliable points) for further updates. One important disparity reliability check technique is cross-checking or left–right consistency constraint. First, the left and right disparity maps are derived using the left and right images as references, respectively. For each pixel in the images, these two disparity maps are used to verify the left and right matching consistency of the correspondences. The reliability of the point is then represented by

$$R(x, y) = \begin{cases} 1, & D_L(x, y) = D_R(x - D_L(x, y), y) \\ 0, & \text{otherwise,} \end{cases} \tag{3}$$

where $D_L(x, y)$ and $D_R(x, y)$ are the disparities computed using the left and right images as reference, respectively. Suppose the correspondences between the left and right images are (x_l, y_l) and (x_r, y_r), then $D_L(x, y)$ and $D_R(x, y)$ are defined by

$$D_L(x, y) = (x_r - x_l, y_r - y_l) \quad \text{and} \quad D_R(x, y) = (x_l - x_r, y_l - y_r),$$

respectively. Figure 6 illustrates the disparity reliability check. Figure 6a,b show the disparity maps obtained using the left and right images as reference, respectively.

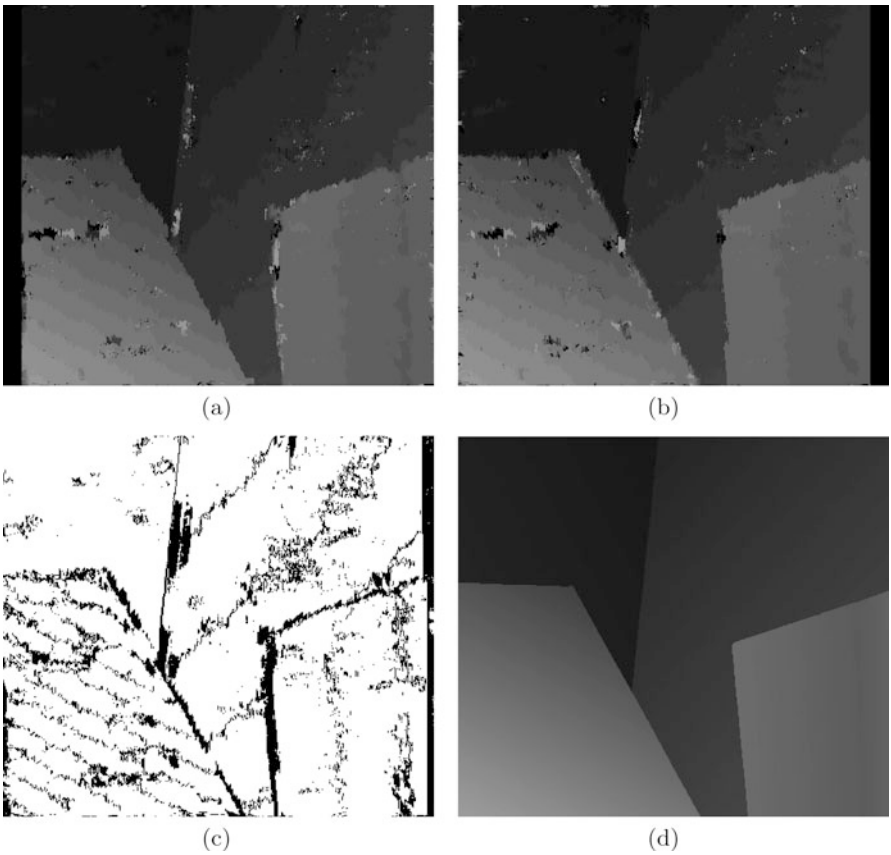


Fig. 6 The disparity reliability check. (**a**), (**b**) The disparity maps obtained using the left and right images as reference, respectively. (**c**) The unreliable points detected by cross-checking. (**d**) The ground-truth

The unreliable points detected by cross-checking are shown in Fig. 6c, compared to the ground-truth as shown in Fig. 6d.

To find the corresponding points between the zoom images, an adaptive support window with local matching method is used in this work. It is based on Veksler's approach stereo correspondence search [26, 27]. The adaptive support window is designed to capture sufficient intensity variation in order to handle the regions with poor texture information and reject the pixels with different disparities. In our method, the adaptive support window for each pixel is given by a line segment with its length determined by the average intensity value. Let L be a line segment centered at p, and then we search for the closest endpoints $p \pm r$ such that

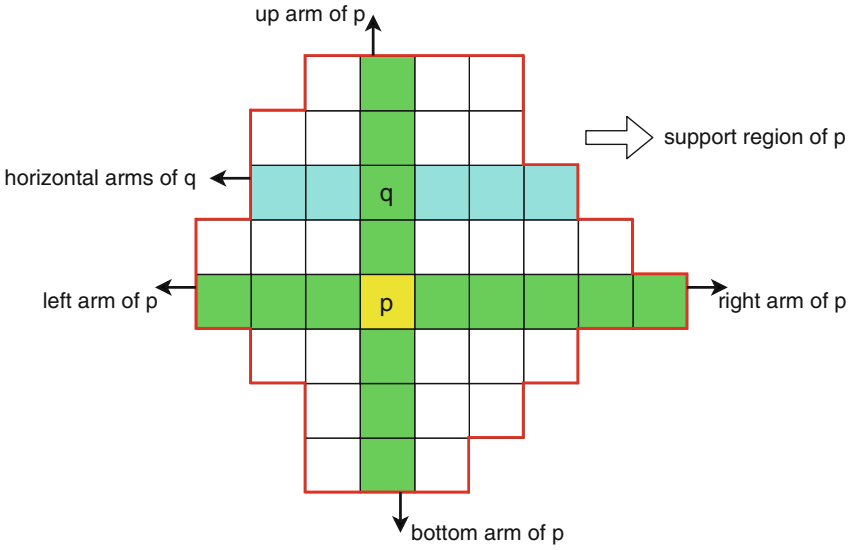


Fig. 7 The adaptive support window determined by the collection of all image scanline segments containing the pixels within the vertical line segment of the pixel

$$\left| \frac{1}{2r+1} \sum_{-r \le m \le r} I(m) - I(r) \right| < \tau, \tag{4}$$

where $I(\cdot)$ is the intensity of the pixel and τ is a threshold. Next, the cross-checking-based aggregation is employed to reduce the matching ambiguities. For a given pixel p, a line segment in the vertical direction is obtained by Eq. (4), similar to the image scanline direction. The adaptive support window is then determined by the collection of all image scanline segments containing the pixels within the vertical line segment of the pixel, as illustrated in Fig. 7.

The aggregation matching cost for the pixels p and p' in the adaptive support window is defined by

$$Local_{ASW}(p, p') = \sum_{q,q' \in AdaptiveWindow} Local(q, q'), \tag{5}$$

where $Local(\cdot, \cdot)$ is the general local cost such as SAD (Sum-of-Absolute Difference) and NCC (Normalized Cross Correlation). In general, the error matching happens frequently for local stereo methods. Thus, the correspondence search range in zoom images is further constrained by homography to increase the matching correctness. It should be noted that, because the scene is magnified in the zoom-in image, there will be additional matching issues if the same mask size is adopted. To

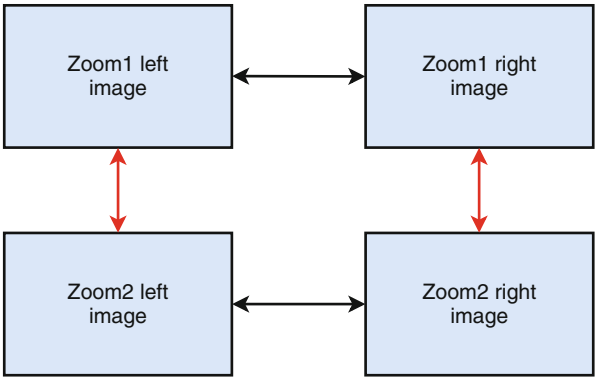


Fig. 8 The schematic diagram of the stereo with zooming framework. The matching cost consists of two sources for computation. One is from the conventional stereo matching of the image pairs (the horizontal direction), and the other is from the correspondence matching of the zoom image sequences (the vertical direction)

deal with this problem, the zoom-out image needs to be transformed to the zoom-in image for correspondence matching.

As shown in Fig. 8, the matching cost consists of two sources for computation. One is from the conventional stereo matching of the image pairs (the horizontal direction in the figure), and the other is from the correspondence matching of the zoom image sequences (the vertical direction in the figure). In the former case, the matching cost is evaluated by global or local stereo algorithms. As for the latter case, we utilize the local matching with adaptive support window to calculate the correspondence matching cost. The stereo and zoom matching costs are then combined for minimization.

In the stereo with zooming approach, the stereo matching algorithm is first used to derive the disparity maps for Zoom1 and Zoom2 image pairs. The point correspondences in Zoom1 and Zoom2 images are found with the zoom constraint. We define the correspondence matching cost as

$$C = Local_{ASW}(Z_{1L}(i, j), Z_{2L}(i', j')), \tag{6}$$

where $Z_{1L}(i, j)$ and $Z_{2L}(i', j')$ are the pixels (i, j) and (i', j') in Zoom1 left image and Zoom2 left image, respectively. The "winner-take-all" (WTA) is generally adopted by stereo matching algorithms for disparity computation. However, the best match in the ground-truth disparity map is not necessarily the pixel correspondence with the lowest cost. Thus, we select three pairs of correspondences with the lowest match costs among the unreliable points given by

$$C_n = cost(i, j, d_n(i, j)), 1 \leq n \leq 3 \tag{7}$$

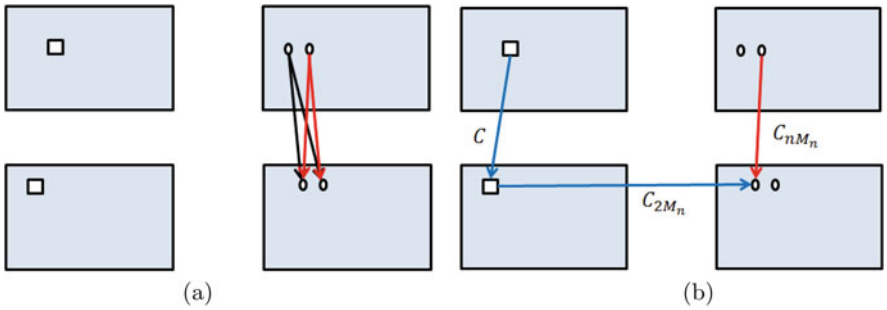


Fig. 9 Zoom1 and Zoom2 matching. (**a**) The circles in the right images represent the correspondence candidates of the squares in the left images. (**b**) The total cost is calculated by the combined matching cost of the left and right zoom images and the Zoom2 stereo image pair. (**a**) The candidate point matching. (**b**) The cost aggregation

as candidates for further evaluation, where d_n is the disparity of the candidate pixel (i, j). The matching cost between Zoom1 right and Zoom2 right images is then computed on the candidate correspondences using local matching methods with adaptive support window. As illustrated in Fig. 9a, the circles in the right images represent the correspondence candidates of the squares in the left images. The true correspondences are then derived by WTA using the cost given by

$$C_{nm} = Local_{ASW}(Z_{1R}(i + d_n(i, j), j), Z_{2R}(i' + d_m(i', j'), j')), \tag{8}$$

where $1 \leq n \leq 3$, $1 \leq m \leq 3$, and

$$M_n = \arg\min_m C_{nm}, \quad 1 \leq n \leq 3, 1 \leq m \leq 3. \tag{9}$$

Finally, the total cost is calculated by the combined matching cost of the left and right zoom images and the Zoom2 stereo image pair, as shown in Fig. 9b. The total cost function is defined as the weighted sum

$$TotalCost_n = \alpha C_{2M_n} + \beta(C + C_{nM_n}), 1 \leq n \leq 3, \tag{10}$$

where C, C_{2M_n}, and C_{nM_n} are given by Eqs. (6), (8), and (9), and the parameters α and β are empirically set as 1.0 and 1.0.

4 Experiments

Two zoom lens cameras as shown in Fig. 10 are used in our experiments to perform stereo with zooming. These two cameras are connected to a host computer via the GigE interface, and the dual image captures are synchronized. The computer is

Fig. 10 Two DFK Z12GP031 cameras used in the experiments. The pixel size and original image resolution are 2.2 μm and 2592 $\times$ 1944, respectively

Table 1 Middlebury BPR (%)

Method	Tsukuba	Venus	Teddy	Cones
Block Matching	9.48	**7.60**	**25.58**	**17.26**
Zoom-Stereo Block Matching	**9.26**	7.61	25.62	17.45
Belief Propagation	3.73	1.53	12.1	8.20
Zoom-Stereo Belief Propagation	**3.70**	**1.52**	**12.09**	**8.17**

Bold values illustrate the better results

equipped with an i7-4790 3.6 GHz CPU, 8 GB memory, and runs on Windows 10 and Visual Studio 2010. The original image resolution is set as 2048 $\times$ 1536 for acquisition and downsampled to 512 $\times$ 384 for processing. We set the focal length at the Zoom1 position as 4.8 mm.

The Middlebury datasets are not primarily used in our experiments since they do not provide the zoom images from the same viewpoints. Although a common practice is resizing the images to simulate the zooming effects, it is generally not realistic and without enough quality due to resampling and aliasing. The evaluation on the Middlebury datasets, "Tsukuba," "Venus," "Teddy," and "Cone" with the block matching and belief propagation algorithms, as well as incorporated with the stereo from zooming framework, is shown in Table 1. In the real scene experiments, the stereo image pairs are captured with two zoom cameras. Figure 11 shows the difference of the Zoom1 and Zoom2 images. We use the equation

$$d = \frac{bf}{z} \tag{11}$$

to limit the stereo correspondence search range, where d, f, b, and z are disparity, focal length, stereo baseline, and depth, respectively.

In the implementation, the mask size of 17 $\times$ 17 is adopted for stereo matching. The adaptive support window with a local matching method is used to compute the zoom image correspondences. For the local stereo matching methods, we use block

Fig. 11 Different zoom images in the experiment. (**a**) The Zoom1 image. (**b**) The Zoom2 image

matching (BM) and MC-CNN algorithms [29]. The mask size for block matching is 15×15, and the model for MC-CNN is the fast model trained by the Middlebury datasets. For the global stereo matching methods, we use belief propagation (BP) and semi-global matching (SGM) algorithms. The message in belief propagation is passed through 10 iterations to derive the cost. The mask size for semi-global matching is set as 7×7, and two other parameters P_1 and P_2 are given by 392 and 735, respectively. In addition, the parameters for zoom image matching are the same as those described previously.

To evaluate the performance, the bad pixel rate (BPR) is adopted as an objective measure. The bad pixel rate is defined as the percentage of incorrect matching in the disparity map and given by

$$BPR = \frac{1}{n} \sum_{i=1}^{N} \left(|d_i - d_i'| \geq \tau \right), \tag{12}$$

where τ denotes the disparity error threshold (set as 2 in this work), d_i is the value in the ground-truth disparity map, d_i' is the computed disparity using the conventional stereo matching techniques and the proposed stereo with zooming approach, and N is the total number of pixels in the image.

Figure 12 shows the real scene images and the disparity maps used in our experiments. The image "Real1" consists of three planar surfaces, and the image "Real2" contains a doll in the front, an inclined plane, and a planar surface parallel perpendicular to the camera viewpoint. The ground-truth disparity maps are shown in the second row of the figure, where the black regions indicate occlusions. Figure 13 shows the disparity maps of "Real2" derived by SGM (top row) and MC-CNN (bottom row). The stereo matching results obtained without and with zooming are shown in the left and right columns, respectively. Tables 2 and 3 tabulate the performance evaluation of "Real1" and "Real2." The results demonstrate that our stereo with zooming is able to further reduce the bad pixel rate using both the local

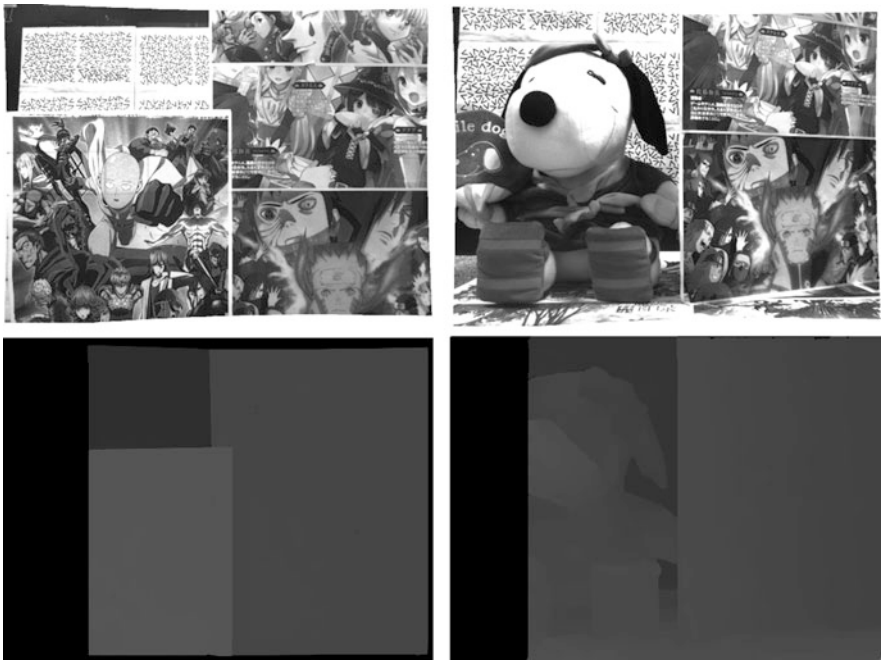


Fig. 12 The real scene images used in the experiments: Real1 (left) and Real2 (right)

Table 2 Bad pixel rate (BPR) of Real1 stereo pair (%)

Real1	Original	Out method
Block matching	3.927	**3.907**
Semi-global matching	4.551	**4.486**
Belief propagation	**2.173**	2.220
MC-CNN	3.412	**3.382**

Bold values illustrate the better results

Table 3 Bad pixel rate (BPR) of Real2 stereo pair (%)

Real2	Original	Out method
Block matching	15.884	**15.650**
Semi-global matching	11.719	**11.182**
Belief propagation	17.848	**17.846**
MC-CNN	12.130	**11.823**

Bold values illustrate the better results

and global stereo matching algorithms. It can also be seen that the proposed method derives better results using SGM even for cluttered scenes.

Since our candidates for zoom correspondence matching are determined by the surrounding disparities from stereo matching, the improvement looks like image erosion in SGM. In general, this process should be carried out until the disparities in the whole image are improved. However, not every pixel has suitable zoom image correspondences, and the disparity update will terminate. To deal with this problem,

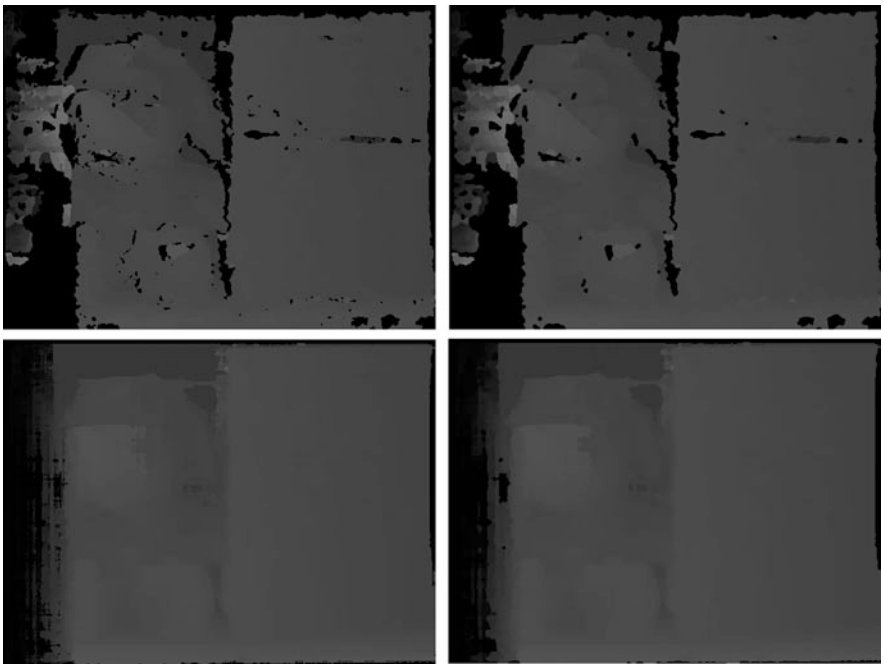


Fig. 13 The disparity maps of "Real2" derived by SGM (top row) and MC-CNN (bottom row). The left column shows the original stereo matching results and the right column shows the results using our stereo with zooming approach

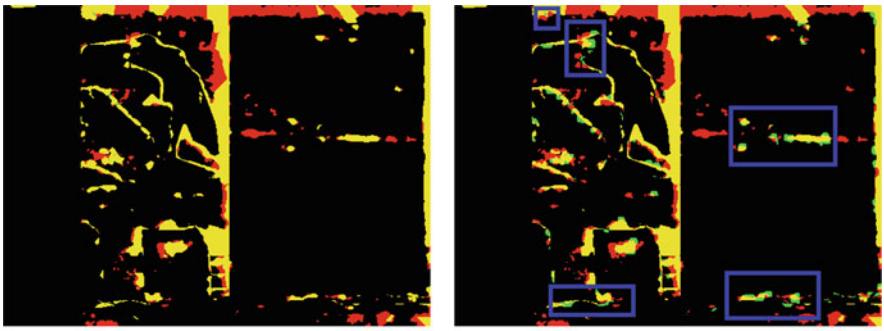


Fig. 14 The improvement with diffusion. If the diffusion effect is provided for the original disparity without zero, it could generate good or bad results simultaneously. (**a**) Diffusion to black. (**b**) Diffusion to disparity

we remove the unreliable pixels using diffusion with a region of 3×3, and the final BPR is improved at 7.71(%). If there are no constraints for the disparity diffusion with the surrounding pixels, we adopt a voting scheme to determine if they should be diffused. As shown in Fig. 14, if the diffusion effect is provided for the original

disparity without zero, it could generate good or bad results simultaneously (see the blue regions in Fig. 14b). Finally, the green and red pixels in the figure indicate the bad change and improvement, respectively.

5 Conclusion

In this chapter, a stereo with zooming framework for disparity computation is presented. With the multiple stereo image pairs acquired with different focal length settings, the correspondence mismatching can be reduced. The geometric image formation for stereo with zooming is presented with mathematical formulation. A cooperative cost aggregation with zoom and stereo is proposed to improve the matching results. The experiments using real scene zoom images, we have demonstrated that, in terms of bad pixel rate, the disparity maps obtained by the proposed approach provide better results upon the general stereo matching algorithms. In the future work, the estimation of more precise unreliable points and more efficient computation will be investigated.

Acknowledgments The support of this work in part by the Ministry of Science and Technology of Taiwan under Grant MOST 106-2221-E-194-004 is gratefully acknowledged.

References

1. Bouguet, J.Y.: Matlab camera calibration toolbox. Caltech Technical Report (2000)
2. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on pattern analysis and machine intelligence **23**(11), 1222–1239 (2001)
3. Buades, A., Facciolo, G.: Reliable multiscale and multiwindow stereo matching. SIAM Journal on Imaging Sciences **8**(2), 888–915 (2015)
4. Deng, Y., Yang, Q., Lin, X., Tang, X.: A symmetric patch-based correspondence model for occlusion handling. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. vol. 2, pp. 1316–1322. IEEE (2005)
5. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient belief propagation for early vision. International journal of computer vision **70**(1), 41–54 (2006)
6. Fusiello, A., Trucco, E., Verri, A.: A compact algorithm for rectification of stereo pairs. Machine Vision and Applications **12**(1), 16–22 (2000)
7. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, second edn. (2004)
8. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge university press (2003)
9. Hirschmuller, H.: Accurate and efficient stereo processing by semi-global matching and mutual information. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. vol. 2, pp. 807–814. IEEE (2005)
10. Hirschmuller, H.: Stereo processing by semiglobal matching and mutual information. IEEE Transactions on pattern analysis and machine intelligence **30**(2), 328–341 (2008)
11. Hirschmuller, H., Scharstein, D.: Evaluation of stereo matching costs on images with radiometric differences. IEEE transactions on pattern analysis and machine intelligence **31**(9), 1582–1599 (2009)

12. Hsiao, W.T., Leou, J.J., Hsiao, H.H.: Super-resolution reconstruction for binocular 3d data. In: Pattern Recognition (ICPR), 2014 22nd International Conference on. pp. 4206–4211. IEEE (2014)
13. Ivanov, M., Sergyienko, O., Tyrsa, V., Lindner, L., Flores-Fuentes, W., Rodríguez-Quiñonez, J.C., Hernandez, W., Mercorelli, P.: Influence of data clouds fusion from 3d real-time vision system on robotic group dead reckoning in unknown terrain. IEEE/CAA Journal of Automatica Sinica **7**(2), 368–385 (2020). https://doi.org/10.1109/JAS.2020.1003027
14. Kim, K.R., Kim, C.S.: Adaptive smoothness constraints for efficient stereo matching using texture and edge information. In: Image Processing (ICIP), 2016 IEEE International Conference on. pp. 3429–3433. IEEE (2016)
15. Lin, H.Y., Gu, K.D., Chang, C.H.: Photo-consistent synthesis of motion blur and depth-of-field effects with a real camera model. Image and Vision Computing **30**(9), 605–618 (2012)
16. Lin, H.Y., Lin, P.Z.: Hierarchical stereo matching with image bit-plane slicing. Machine Vision and Applications **24**(5) (2013)
17. Lin, H.Y., Tsai, C.L., Tran, V.L.: Depth measurement based on stereo vision with integrated camera rotation. IEEE Transactions on Instrumentation and Measurement **70**, 1–10 (2021). https://doi.org/10.1109/TIM.2021.3073687
18. Loop, C., Zhang, Z.: Computing rectifying homographies for stereo vision. In: Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. vol. 1, pp. 125–131. IEEE (1999)
19. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision **60**(2), 91–110 (2004)
20. Okutomi, M., Kanade, T.: A multiple-baseline stereo. IEEE Transactions on pattern analysis and machine intelligence **15**(4), 353–363 (1993)
21. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International journal of computer vision **47**(1–3), 7–42 (2002)
22. Sepulveda-Valdez, C., Sergiyenko, O., Tyrsa, V., Flores-Fuentes, W., Rodríguez-Quiñonez, J.C., Murrienta-Rico, F.N., Miranda-Vega, J.E., Hernandez-Balbuena, D., Mercorelli, P., Kolendovska, M.: Geometric analysis of a laser scanner functioning based on dynamic triangulation. In: 2020 IEEE 29th International Symposium on Industrial Electronics (ISIE). pp. 1398–1403 (2020). https://doi.org/10.1109/ISIE45063.2020.9152268
23. Sepúlveda-Valdez, C.A., Sergiyenko, O., Hernandez-Balbuena, D., Tyrsa, V., Mercorelli, P., Flores-Fuentes, W., Reyez-García, M., Lindner, L., Melnik, V.: Circular scanning resolution improvement by its velocity close loop control. In: 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE). pp. 244–249 (2019). https://doi.org/10.1109/ISIE.2019.8781135
24. Sergiyenko, O.Y., Ivanov, M.V., Tyrsa, V., Kartashov, V.M., Rivas-López, M., Hernández-Balbuena, D., Flores-Fuentes, W., Rodríguez-Quiñonez, J.C., Nieto-Hipólito, J.I., Hernandez, W., et al.: Data transferring model determination in robotic group. Robotics and Autonomous Systems **83**, 251–260 (2016)
25. Sergiyenko, O.Y., Tyrsa, V.V.: 3d optical machine vision sensors with intelligent data management for robotic swarm navigation improvement. IEEE Sensors Journal **21**(10), 11262–11274 (2021). https://doi.org/10.1109/JSEN.2020.3007856
26. Veksler, O.: Stereo matching by compact windows via minimum ratio cycle. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. vol. 1, pp. 540–547 vol.1 (2001). https://doi.org/10.1109/ICCV.2001.937563
27. Veksler, O.: Fast variable window for stereo correspondence using integral images. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. vol. 1, pp. I–I (2003). https://doi.org/10.1109/CVPR.2003.1211403
28. Yang, Q., Wang, L., Ahuja, N.: A constant-space belief propagation algorithm for stereo matching. In: Computer vision and pattern recognition (CVPR), 2010 IEEE Conference on. pp. 1458–1465. IEEE (2010)
29. Zbontar, J., LeCun, Y.: Stereo matching by training a convolutional neural network to compare image patches. Journal of Machine Learning Research **17**(1–32), 2 (2016)
30. Zhang, H., Kankanhalli, A., Smoliar, S.W.: Automatic partitioning of full-motion video. Multimedia systems **1**(1), 10–28 (1993)

ROS and Stereovision Collaborative System

Ruben Alaniz-Plata, Oleg Sergiyenko, Wendy Flores-Fuentes, Vera V. (Vira) Tyrsa, Julio Cesar Rodríguez-Quiñonez, Cesar Antonio Sepúlveda-Valdez, Humberto Andrade-Collazo, Paolo Mercorelli, and Lars Lindner

Abbreviations

CCD	Charge-coupled device
CMOS	Complementary metal-oxide-semiconductor
DC	Direct current
DLT	Direct linear transformation
FOV	Field of view
FPGA	Field-programmable gate array
NCC	Normalized cross-correlation
RADAR	Radio detection and ranging
ROI	Region of interest
ROS	Rotational optical scanner
SAD	Sum of absolute differences
SONAR	Sound navigation and ranging
SSD	Sum of squared differences
TOF	Time of flight

R. Alaniz-Plata (✉) · O. Sergiyenko · W. Flores-Fuentes · J. C. Rodríguez-Quiñonez
C. A. Sepúlveda-Valdez · H. Andrade-Collazo
Autonomous University of Baja California, Mexicali, Mexico
e-mail: ruben.alaniz@uabc.edu.mx

V. V. (Vira) Tyrsa
Polytechnic University of Baja California, Mexicali, Mexico

P. Mercorelli
Leuphana University of Lueneburg, Lüneburg, Germany

L. Lindner
Fraunhofer Institute for Energy Economics and Energy System Technology IEE, Kassel, Germany
e-mail: lars.lindner@iee.fraunhofer.de

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1_4

1 Introduction

A system with automatic decision capabilities taken according to collected data by sensors is considered as a vision system [1]. Generally, sensors acquire depth data for positioning static and dynamic objects in 3D coordinates [2, 3], path planning in robot navigation [4, 5], topography analysis [6], obstacle detection [7, 8], and so forth.

Vision techniques are defined by sensors arrays and algorithms used in vision systems, using measurement systems for depth estimation as laser scanning and image processing, with advantages strongly dependent on the specific use. Commonly, image processing is used in high-speed processing and data volume. Laser scanners are applied in high precision needs, where it is imperative to know with accuracy the position of one object.

Some depth estimation techniques used in previously mentioned applications are RADAR and SONAR with longer distance detection, LiDAR with higher data acquisition but lower distance detection, rotational optical scanner (ROS) with high precision and low volume data, monocular depth estimation with high complexity, binocular depth estimation with stereo system using triangulation, and multicamera depth estimation with movement detection capabilities and high computing cost [9].

For this project, considering a collaborative system where image processing and laser scanning can work together, binocular depth estimation using stereovision was selected taking account that is more robust than monocular due to the simplicity of theoretical models to estimate depth and simpler than multicamera depth estimation due to the lower computational cost, and ROS, for laser scanning, due to the high precision and lower cost and complexity than LiDAR and RADAR.

ROS is a system capable of localizing elements in a scene with rectangular coordinates. Nowadays, it has been used for robot navigation [10], structural health [11], body scanning [12], and vegetation measurement [13]. Although LiDAR and ROS are laser scanning systems, the path used to localize a point is different; LiDAR and RADAR localize a point using time-of-flight [14, 15] and ROS uses triangulation, so the subsystems required for its functioning have some differences; all laser techniques requires a sensitive receptor that discriminates the laser pulse bandwidth from the rest of the light sources than can generate noise in the signal, but LiDAR and RADAR also require an additional system for precise light time travel measurement, making them more complex.

On the image processing side, stereo systems are widely implemented in depth estimation using two parallel cameras or pitched cameras. For distance estimation with stereovision, a pair of images (or stereo pair) are captured and processed using correspondence algorithms. This correspondence is based on similarity models, so an object in a stereo pair is located in a group of pixels with similar features [16].

Viewing the general panorama, stereo system and ROS have the goal of getting 3D coordinates of elements in a scene taking the system itself as an origin for that coordinates [17–19], but each one achieves this goal using different basic principles, components, and algorithms. Giving as a result, the main differences between the

3D data obtained by these systems are the volume and accuracy, the set of these parameters variates according to instantly used system and the working conditions [20], which finally propose the option of smart combination of these two data sets in order to optimize the environment perception.

On the specific case of ROS and stereovision taken as the analyzed systems, the advantages of each one over the other are complementary. ROS has better precision even without signal processing and stereo systems get a considerably higher data volume in less time. So, collaborative work can reach an enriched depth map, with a large quantity of very accurate depth points of a specific object or even a whole scene.

To achieve the synchronization of the collaborative system are required, at first, an origin in a defined point for both techniques and an algorithm for the working sequence and data processing. Also, a high computing capacity system is preferable, considering the high number of iterations required in stereo systems' algorithms. The sequence used to get enriched depth map involves getting a first depth map with the stereo system, detection of regions where we could not get depth data, and analysis of those regions with ROS. In the same way, the ROS can be used to recalculate a given point, to ensure a precise measurement.

Missed depth data and recalculated points are defined as regions of interest (ROIs). To ensure that ROIs are scanned correctly, it is necessary to consider, in the processing algorithm, all the subsystems' specifications (cameras' focal length, photosensors' response, stepper motors' required compensations). With this consideration, it is possible a precise laser positioning in ROIs. According to the distribution of ROIs in a depth map, the ROS makes a scanning using three different scanning modes: single point scanning, line scanning sequence, and surface scanning sequence. Defining scanning modes leads to energy savings, because only the defined points are analyzed and the rest of movements are used for positioning.

In order to get a better understanding about how cameras and laser scanners can achieve the complementary enhancement and to enrich the fused information from both, let us consider more in detail their principles of functioning, with attention to their specific advantages and lacks. It can be useful for conclusion about the optimal way of their combination.

2 Background

2.1 Stereovision

A stereo system is a technical vision system based on human vision. It works using two cameras, installed with fixed distance between them, and a processing system (computer vision system) as imitation of human eyes and the brain, respectively. This technique is widely used for depth estimation and object detection in robot navigation, path planning, and ranging.

To localize objects in a scene, one image is taken from each camera (stereo pair), and both images are processed with a computer algorithm, where it is considered that images are formed from light intensity values which are dependent on camera specs like lens focal length, relative position of both cameras, sensor resolution, and environmental illumination.

2.1.1 Stereo System Array

Basic stereo system consists of a pair of cameras with known specs mounted at a specified distance, horizontally calibrated and with a parallel vision line. These initial characteristics are used for depth estimation with triangulation, shown in Fig. 1.

Only using the system array is not possible to obtain α and β, so it is necessary to capture and process images from scene to estimate them.

2.1.2 Image Capture

Cameras' main components are lenses, photosensors, and processors with the goal of capturing, measuring, and processing light intensities that are converted into signals and, afterward, in images. Complementary metal-oxide-semiconductor

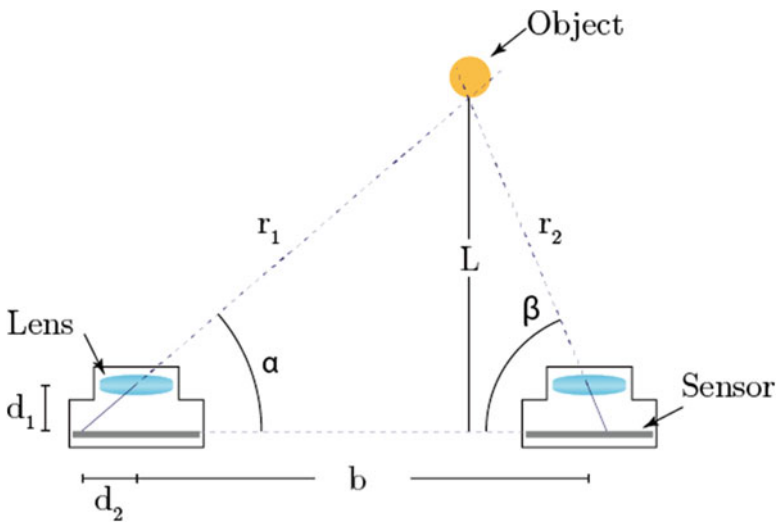


Fig. 1 Distribution and properties of a basic stereo system. Where b is the known distance between camera centers or baseline, d_1 is the distance from the center to the incidence point on the sensor, and d_2 is the distance from the sensor to the lens. With d_1 and d_2, α and β could be obtained to calculate L with trigonometrical relations

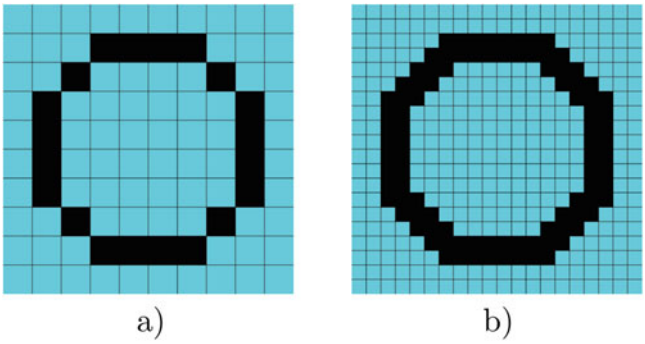


Fig. 2 (**a**) Shape formed with 100 pixels. (**b**) Shape formed with 400 pixels

(CMOS) and charge-coupled device (CCD) are camera sensors that have a matrix of photodiodes used for the measurement of different wavelength intensities for each pixel. So, the quantity of photodiodes in the sensor indicates how many pixels will give a form to each picture taken and, therefore, the data volume that should be processed for each image.

Pixel quantity is one of the variables related to picture resolution; at more pixels, more detailed is the picture. Figure 2 illustrates the difference between a shape formed with 100 pixels and 400 pixels, where it is shown that more pixels could give a sensation that details in pictures are more similar compared with a real circle.

Larger images require more processing capacity due to the amount of data. According to the application, it is important to analyze whether is necessary to have a detailed image with bigger amount of data or smaller image with less data and less processing requirements. For stereo systems, a common matrix size is 640 × 480 pixels, giving form to an image of 307,200 total pixels. Image size should be selected according to the computing capacity and specific application.

So, visually, each pixel represents a color in a picture, but depending on the color map used, this color could be obtained combining base colors or by a tonal variation of one color (monochrome). If the color map used is monochrome, each pixel is one value in a matrix, but if it is not, each base color is one value in a n dimensional matrix, where n is the quantity of base colors per pixel. As reference, color maps used to have the initials of base colors, i.e., RGB (red, green, and blue) and CMYK (cyan, magenta, yellow, and key) which determine n value.

RGB Color Map

RGB is the most common color map used for screens in cellphones, computers, tablets, and digital cameras. Each pixel is formed by a combination of tonal intensities of red, blue, and green. This intensity is commonly formed (but not limited) to an 8-bit value, so each color (or channel) has 255 tones of each base color, generating a three-dimensional matrix per image (Fig. 3a). With a specialized

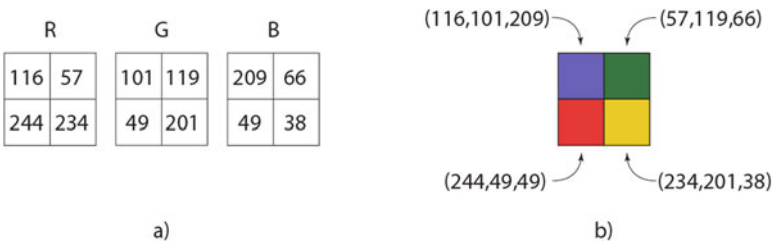


Fig. 3 RGB map. (**a**) 3D matrix with RGB values. (**b**) Four pixels formed by the combination of its respective RGB values

software, RGB matrix can be seen as an image where each pixel is a combination of one value from the three channels (Fig. 3b).

For stereo systems, a pair of RGB images could give the opportunity of a precise pattern match [21–23], but due to the high amount of data, processing time would be considerably large. This is the main reason why the color map used in stereo systems is gray scale. This circumstance will be considered later, in the optimization of ROS and stereovision co-functioning.

Grayscale Map

A grayscale map is a representation of black, white, and their middle tones, where black is the absence of light intensity, or a zero as value, and white is the maximum light intensity represented with the bigger number that could be represented in 8 bits, or 255, generating a one-dimensional matrix per image. Middle tones are different tones of gray and are considered as enough to get a detailed image that could be used to detect shapes and contrast with a stereo algorithm.

It is easy to visualize that a grayscale map could reach less details and contrast than RBG map, but has a third part of data, which is a parameter directly related to processing time.

In practice, commonly a conversion from RGB to gray scale is required, and, to achieve it, it is necessary to consider that a pixel in gray scale is a rounded average of RGB values of a pixel, expressed in Eq. (1):

$$\text{gray}_\text{p} = \frac{R_\text{p} + G_\text{p} + B_\text{p}}{3} \tag{1}$$

being R_p, G_p, and B_p the given values of a RGB pixel (p) and gray_p the rounded value of gray. In Fig. 4 it's possible to observe a conversion from RGB to gray scale and the considerable reduction of data amount from 12 values to 4.

The capture of images and their conversion to gray scale are considered as a data acquisition and preparation. Is from this point where the rest of the work requires pure data processing and takes the major part of its time cost.

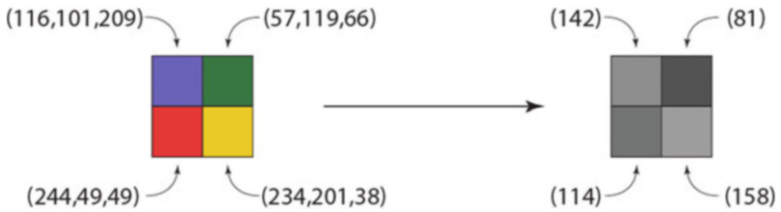


Fig. 4 Conversion from RGB to gray scale

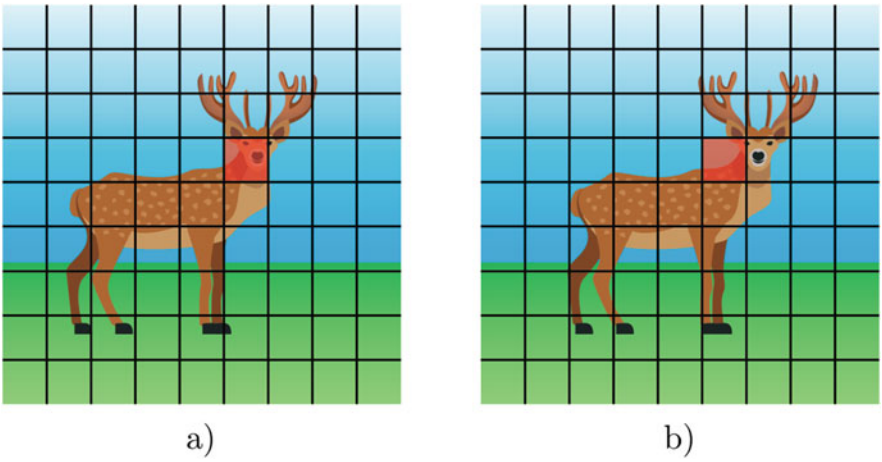


Fig. 5 (**a**, **b**) Correspondence error

2.1.3 Correspondence

Correspondence is the first data processing step and consists in finding the same group of pixels (or local region) which represents a real physical object [24] that can be confused on the stereo pair due to different sources of optical noise and processing algorithm imperfections. In order to localize any object in a stereo pair, it is required to run a process of correspondence using a matching algorithm that will divide both images and estimate the location of the object in both images. If, in the beginning, it is considered that an object is located in the same group of pixels for both images, then there will be a correspondence error [25], where a group of pixels represent a different object for each image and, if it is used for depth calculation, the measurement will be wrong. This correspondence error could be illustrated in Fig. 5, where it shows that part of the deer face is located in different pixel groups for each image.

Due to the physical distance between cameras, each pair of images has a different perspective of the scene, generating a displacement of the objects in an image relative to the other and making necessary a correspondence process, which consists in a recognition of a group of pixels in an image taking as template other group of

pixels of a second image. These images are taken with stereo system cameras at the same time under the same light conditions, ideally. Once the object is located, displacement value (d_2) is used for depth estimation with triangulation.

2.1.4 Matching Algorithms

To achieve the correspondence, a large quantity of matching algorithms has been developed and classified by feature matching, region matching, and semi-global matching. All stereo algorithms give as a result a disparity map.

Feature matching takes a portion of an image (template) or edges and search for it in the other image, a template and a portion of the second image with the best correlation have a correspondence and are used to depth estimation, and this method gives precise and high processing speed but sparse data results. Region matching considers intensities in a window of pixels, they have a correspondence when two templates have similar intensities, and with this method, it is possible to compare all the pixels in an image, throwing dense and precise data results but low-speed processing in counterbalance [26].

Semi-global matching is the technique with a more dense and precise output, but with a considerably low speed of processing. In this technique, minimization of energy is an interesting subject with the goal to achieve even real-time processing [27]; generally this can be expressed as Eq. (2):

$$E(d) = \sum_{\mathrm{p}} \left(C\left(p, d_{\mathrm{p}}\right) + \sum_{\mathrm{q} \epsilon N_{\mathrm{p}}} \mathrm{PT}\left[d_{\mathrm{p}} - d_{\mathrm{q}}\right| \geq 1 \Big] \right) \tag{2}$$

where the term $C(p, d_{\mathrm{p}})$ defines all the matching cost and $\mathrm{PT}[d_{\mathrm{p}} - d_{\mathrm{q}}| \geq 1]$ is a penalty for all the pixels (p) in which neighbors (q) have a different disparity.

All matching techniques require iterative comparisons, and, depending on the matching strategies, mathematical operations, and quantity of iterations, some techniques have more matching cost than others. In this work, region matching techniques are implemented considering that they have better balance between precision and computational cost than semi-global matching. Main region matching algorithms are listed below.

Sum of Absolute Differences

Sum of absolute differences (SAD) is also a theoretical tool for detection of the same point on different images obtained from different positions. It consists of a subtraction of intensity values of a template from another image. This subtraction is repeated for all the regions where could be the best correspondence. It is common to take a horizontal line (epipolar line) where could be aligned (if there is a good calibration) the searched region. This process is repeated for each template and its cost function is represented as Eq. (3):

$$SAD\,(x, y, d) = \sum_{(x,y)\epsilon W} |\, I_L\,(x, y) - I_R\,(x, y - d)\,| \tag{3}$$

All disparity maps obtained in this work are obtained using SAD, due to its low computational cost and robustness [28, 29].

Sum of Squared Differences

Sum of squared differences (SSD) algorithm is a variant of SAD, where all values of a window are squared. Its cost function is expressed as Eq. (4):

$$SSD\,(x, y, d) = \sum_{(x,y)\epsilon W} (I_L\,(x, y) - I_R\,(x, y - d))^2 \tag{4}$$

Due to the squared terms, SSD has higher computational cost than SAD but better performance in conditions with illumination changes.

Normalized Cross-Correlation

Normalized cross-correlation (NCC) is correspondence an algorithm whose stronger point is to be robust to illumination changes between cameras but with high computational cost, making it harder to implement it in real-time processing. Taking into consideration its advantage, there is some research related to time optimization [30]. Cost function of NCC is expressed as Eq. (5):

$$NCC\,(x, y, d) = \frac{\sum_{(x,y)\epsilon W} I_L\,(x, y) \cdot I_R\,(x, y - d)}{\sqrt{\sum_{x,y\epsilon W} I_L^2\,(x, y) \cdot \sum_{x,y\epsilon W} I_R^2\,(x, y - d)}} \tag{5}$$

where x and y are horizontal and vertical position of template central pixel that belongs to W (window or template) in the left and right image intensity values (I_L and I_R), respectively, and d is the disparity.

2.1.5 Disparity Map

A processed stereo pair with stereo matching algorithm gives, as a result, a matrix of disparity for each scene element. This matrix is the disparity map where, due to the parallax, the closer objects have higher displacement and further objects have lower displacement [31]. Visually, disparity maps can be shown as Fig. 6c obtained from disparity pair Fig. 6a, b using SAD algorithm. In a gray scale, closer objects are represented in white and further in black.

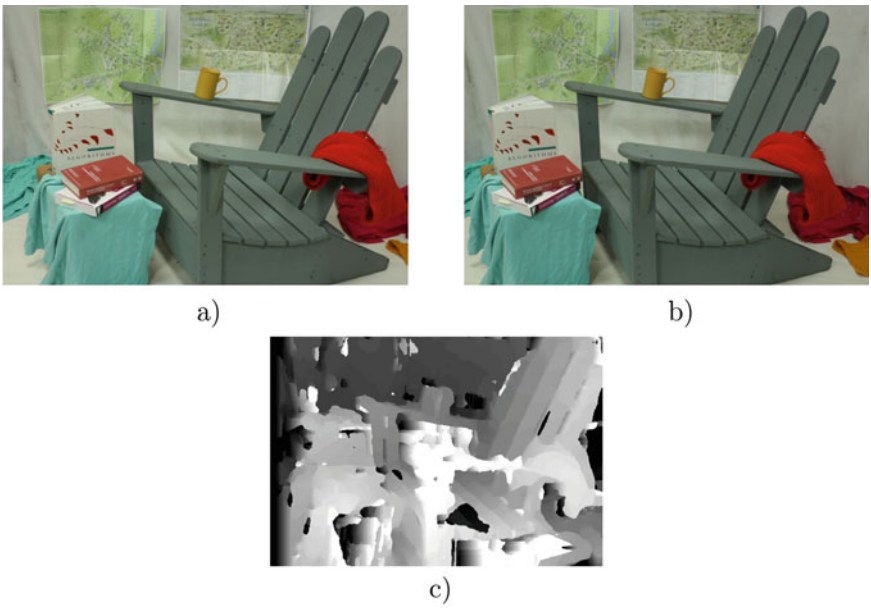


Fig. 6 (**a**) Left image of stereo pair, (**b**) right image of stereo pair, (**c**) disparity map [32]

2.1.6 Depth Map

Depth estimation is obtained from disparity map, using Eq. (6):

$$z = \frac{bf}{d} \tag{6}$$

where b is the baseline, f is the focal length, and d is the disparity; however, depth map is only considering distance from the system to the object in each pixel; to obtain x and y coordinates, Eqs. (7) and (8) are used:

$$x = \frac{zx_i}{f} \tag{7}$$

$$y = \frac{zy_j}{f} \tag{8}$$

x_i and y_j are the position of pixels in x and y, respectively.

Another way to get the (x, y, z) coordinates in a system with parallel cameras is calculating triangle angles (B, C, and β) for each pixel with Eqs. (9)–(13):

$$B_{(i,j)} = 90° + \left(\left| \frac{H}{2} - x_{\mathrm{L}} \right| \cdot R_{\mathrm{h}} \right) \text{ at } x \leq \frac{H}{2} \tag{9}$$

$$B_{(i,j)} = 90^\circ - \left(\left| \frac{H}{2} - x_\mathrm{L} \right| \cdot R_\mathrm{h} \right) \text{ at } x \geq \frac{H}{2} \tag{10}$$

$$C_{(i,j)} = 90^\circ + \left(\left| \frac{H}{2} - x_\mathrm{R} \right| \cdot R_\mathrm{h} \right) \text{ at } x \geq \frac{H}{2} \tag{11}$$

$$C_{(i,j)} = 90^\circ - \left(\left| \frac{H}{2} - x_\mathrm{R} \right| \cdot R_\mathrm{h} \right) \text{ at } x \leq \frac{H}{2} \tag{12}$$

$$\beta_{(x,y)} = \left(\frac{V}{2} - y_\mathrm{R} \right) \cdot R_\mathrm{V} \tag{13}$$

where x_L and y_L are the positions of the pixel taking the center of the image as origin in the left image and x_R and y_R in the right image, H and V the horizontal and vertical size of the camera sensor, and R_h and R_v the horizontal and vertical resolution.

Coordinates are obtained using triangulation, expressed in (15) to (19).

2.1.7 Stereovision for Collaborative System

These region matching methods used for object identification in a stereo pair are compared by matching precision, sensitivity to optical noise, and processing cost, where matching precision is slightly superior for SAD than SSD and NCC under [33, 34] conditions. In the case of Fig. 7, it also can be shown that SAD generates a disparity map with less noise and smother borders, but NCC is more precise.

Optical sensitivity is lower in NCC than SAD and SSD, making NCC more robust to different lighting conditions for each camera [30]. And SAD has better performance or lower processing cost with capabilities of real-time processing [35].

Considering a stereovision system used for collaborative work with ROS, the optimal matching algorithm is SAD, due to its simplicity and fast processing. Although NCC could be a good choice because of its robustness, time processing is considerably high with larger images and big window sizes, letting down the main advantage of stereovision over ROS: massive volume data acquisition at high speed.

So, SAD is the best choice to achieve a reliable combination with ROS, in which capabilities take weak points of stereovision and improve them.

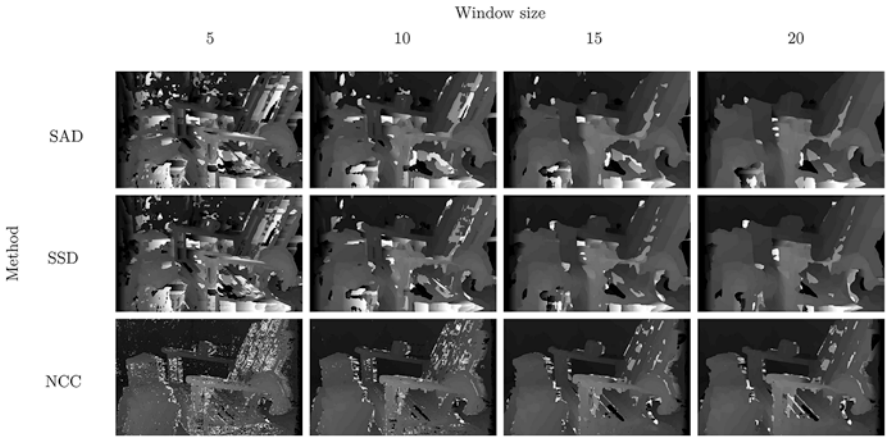


Fig. 7 SAD, SSD, and NCC disparity maps' comparison

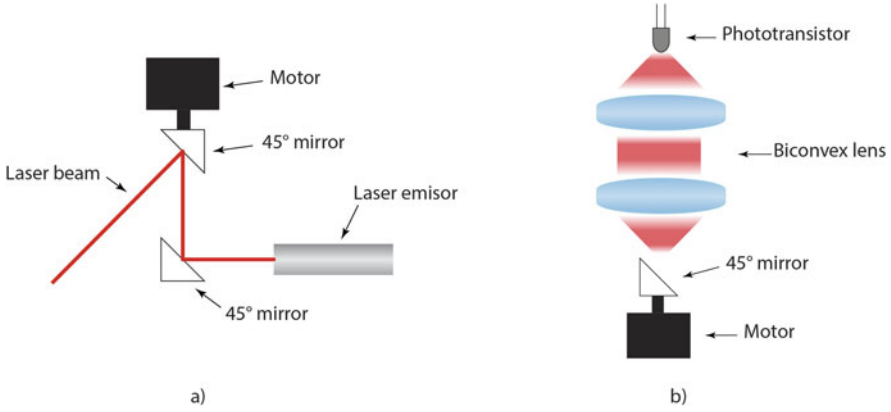


Fig. 8 (**a**) Laser positioner, (**b**) scanning aperture

2.2 Rotational Optical Scanner

A rotational optical scanner (ROS) is a laser scanning technique used to obtain 3D coordinates of an incident point. If the laser ray is able to move along the surface of scanned object, a digitalized map of the scanned surface is formed as a matrix of coordinates of n points where the laser incident point is fixed. To take every lecture ROS uses its two main subsystems: laser positioner (Fig. 8a) and scanning aperture (Fig. 8b).

Positioner and aperture are mounted in a cylindrical body (indicated in Fig. 9a) at known distance or baseline (a).

Vertical and horizontal positioner control is made with stepper motors in which the resolution is over 19,200 steps per revolution using a gear and controlled with

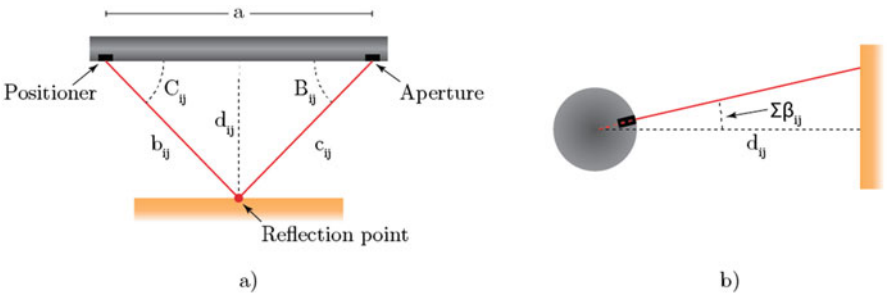


Fig. 9 ROS triangulation. (**a**) Top view, (**b**) lateral view

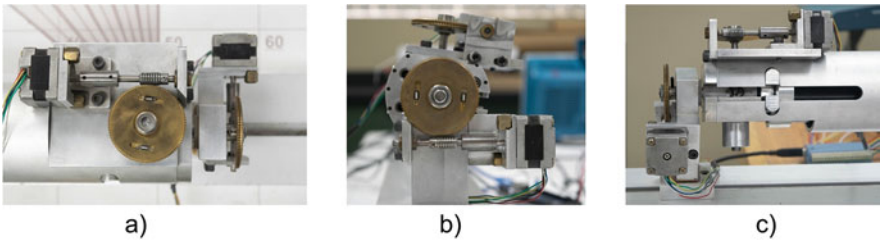


Fig. 10 Positioner components. (**a**) Top view with horizontal positioner, (**b**) lateral view with vertical positioner, and (**c**) front view with horizontal positioner, vertical positioner, and notch of laser beam

an Arduino Mega using code of Appendix A, which sends the movement sequence for each direction of both motors. In Fig. 10a, b, stepper motors and 96 teeth gears can be seen as a horizontal and vertical positioner. Figure 10c is a front view that shows both positioners and the notch for laser beam.

When laser is positioned in a C_{ij} known value, part of light is reflected on a surface to the scanning aperture, forming, by a brief time, a triangle.

2.2.1 Aperture

Aperture is capable of obtaining B_{ij} for each triangle formed using 45-degree mirror rotation and light reflection. Every 360-degree rotation is registered by an optocoupler (blue signal in Fig. 11b), while a photosensor is sending data of light intensity received during all the loop. With a photosensor capable of detecting laser bandwidth over the rest of the light, a notable curve is formed which indicates the moment in the loop in which reflected laser arrives at the photosensor (yellow signal in Fig. 11b).

Fig. 11 Optocoupler and photosensor signals. (**a**) Oscilloscope used for signal visualization, (**b**) closer view of signals

A programmed microprocessor saves both sensors' data and detects rise flanks in optocoupler signal and maximum point of each peak in photosensor signal. Using the location in time of rise flanks and peaks in (14), B_{ij} is obtained with the code of #define m1_1 22:

```
#define m1_2 23
#define m1_3 24
#define m1_4 25
#define m2_1 26
#define m2_2 27
#define m2_3 28
#define m2_4 29

int dir = 0, lastdirH = 0, lastdirV = 0,
steps = 0, binH = 1, binV = 1;
bool enc = false;
float vMotor = 0;

void setup() {
  pinMode(m1_1,OUTPUT);
  pinMode(m1_2,OUTPUT);
  pinMode(m1_3,OUTPUT);
  pinMode(m1_4,OUTPUT);
  pinMode(m2_1,OUTPUT);
  pinMode(m2_2,OUTPUT);
  pinMode(m2_3,OUTPUT);
  pinMode(m2_4,OUTPUT);

  digitalWrite(m1_1,0);
  digitalWrite(m1_2,0);
  digitalWrite(m1_3,0);
  digitalWrite(m1_4,0);
  digitalWrite(m2_1,0);
  digitalWrite(m2_2,0);
  digitalWrite(m2_3,0);
  digitalWrite(m2_4,0);

  Serial.begin(9600);
  }

void loop(){
```

```
  if (Serial.available() > 0){
    String dataSt = Serial.readString();
    int str_len = dataSt.length() + 1;
    char data[str_len];
    dataSt.toCharArray(data, str_len);
    sscanf(data,"D%d S%d",&dir,&steps);

if (lastdirH == 0 && dir < 3){
      lastdirH = dir;
      }
    if (lastdirV == 0 && dir > 2){
      lastdirV = dir;
      }
      if (dir == 1 || dir == 2){
        if (lastdirH != dir){
                binH = -(binH-5);
                lastdirH = dir;
          }
        }
      if (dir == 3 || dir == 4){
        if (lastdirV != dir){
                binV = -(binV-5);
                lastdirV = dir;
          }
        }
      int i = 0;
      for (i = 1;i <= steps;i++){
        delay(500);         if(dir == 2 || dir == 4){
          if (dir == 2){digitalWrite(m1_1+binH-1,0);}
          if (dir == 4){digitalWrite(m2_1+binV-1,0);}
        }
        if(dir == 1 || dir == 3){
          if (dir == 1){digitalWrite(m1_4-binH+1,0);}
          if (dir == 3){digitalWrite(m2_4-binV+1,0);}
        }         if (dir == 1 || dir == 2){
          if (binH == 4){binH = 0;}
          binH++;
        }else{
          if (binV == 4){binV = 0;}
          binV++;}
        if (dir == 1 || dir == 2){
          if (dir == 2){digitalWrite(m1_1+binH-1,1);}
          if (dir == 1){digitalWrite(m1_4-binH+1,1);}         }
        if (dir == 3 || dir == 4){
          if (dir == 4){digitalWrite(m2_1+binV-1,1);}
          if (dir == 3){digitalWrite(m2_4-binV+1,1);}
        }
      }
      Serial.println((String) 5);
    }
  }
```

Appendix B and all parameters are ready for dynamic triangulation:

$$B_{ij} = 360 * \left(\frac{t_{\mathrm{a}}}{t_{\mathrm{p2}} - t_{\mathrm{p1}}} \right) \tag{14}$$

taking account that t_{a} is the point in time where every photosensor signal peak is found and t_{p2} and t_{p1} are rise flanks of the optocoupler neighbor signal.

2.2.2 Dynamic Triangulation

Dynamic triangulation is due to the pulsed laser forming a unique short time light triangle. Every triangle is different from the previous one because of physical conditions like stability of the positioner and scanned surface.

Knowing C_{ij} and B_{ij}, 3D coordinates of the reflected point could be obtained with dynamic triangulation, described with Eqs. (15)–(19):

$$d_{ij} = a \frac{\sin\left(B_{ij}\right) * \sin\left(C_{ij}\right)}{\sin\left[180 - \left(B_{ij} + C_{ij}\right)\right]} \tag{15}$$

$$x_{ij} = a \frac{\sin\left(B_{ij}\right) * \sin\left(C_{ij}\right) * \cos \sum_{j=1}^{n} \beta_{ij}}{\sin\left[180 - \left(B_{ij} + C_{ij}\right)\right]} \tag{16}$$

$$y_{ij} = a \left(\frac{1}{2} - \frac{\sin\left(B_{ij}\right) * \sin\left(C_{ij}\right)}{\sin\left[180 - \left(B_{ij} + C_{ij}\right)\right]} \right) \text{ at } B_{ij} \leq 90^{\circ} \tag{17}$$

$$y_{ij} = a \left(\frac{1}{2} + \frac{\sin\left(B_{ij}\right) * \sin\left(C_{ij}\right)}{\sin\left[180 - \left(B_{ij} + C_{ij}\right)\right]} \right) \text{ at } B_{ij} \geq 90^{\circ} \tag{18}$$

$$z_{ij} = a \frac{\sin\left(B_{ij}\right) * \sin\left(C_{ij}\right) * \sin \sum_{j=1}^{n} \beta_{ij}}{\sin\left[180 - \left(B_{ij} + C_{ij}\right)\right]} \tag{19}$$

For all values, i and j describe the parameters of a point in a depth matrix.

3 Optimization of ROS and Stereovision Combined Use

3.1 Collaborative System

A mounted system using stereovision and ROS advantages is designed in order to complement strong points and enrich the final obtained information, as well as decrease the influence of weak points of both systems. It is not necessary the use of ROS for a whole scene scan, and stereovision doesn't use larger images to maximize precision with high processing cost. This even gives the opportunity of energy savings in both technical systems due to the reduction of processing time for stereovision, and the use of ROS is minimized, representing a lower use of energy for laser beam positioning and real-time data processing. In fact, all the sub-systems of ROS are set on standby when they are not used. A version of fused system is represented in Fig. 12, where a pair of cameras with a resolution of 640 × 480 pixels per image is mounted over a ROS system as shown in Fig. 13, both systems with a baseline of 20 centimeters.

Fused system with capabilities of generating a depth map with stereovision and ROS results in an enriched 3D data, linked through regions of interest (ROIs) that indicates which spaces are required to be enriched adding more data or recalculating depth. For this collaborative system, stereovision's goal is to obtain a first depth map and ROS will fill void data spaces and recalculate depths where stereo data

Fig. 12 Collaborative system of ROS and stereovision for robot navigation

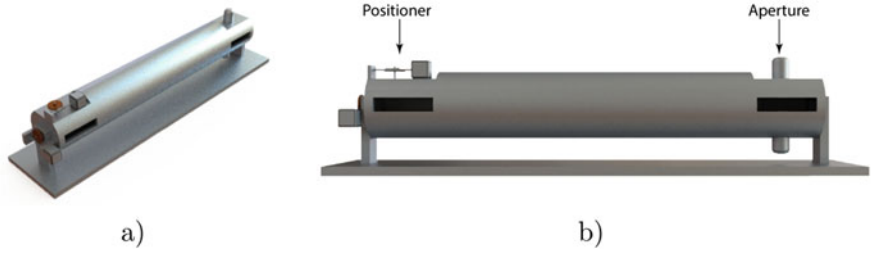


Fig. 13 ROS representation. (**a**) Isometric view, (**b**) front view

Fig. 14 Blockchain diagram of the general processing data flow

is not reliable. General work process is represented with the Fig. 14 blockchain. All subsystems processes belong to the stereo system and ROS already described functions.

So, as a result, enriched depth data takes the two main advantages of each system: high data volume and high precision.

3.1.1 High Data Volume

The amount of 3D points that ROS and stereo-vision are capable to save with a single capture in a unit of time is defined as the volume of data. In the stereo-vision case, it's directly related to the camera's image resolution (each pixel represents a depth point), for ROS, the volume depends on stepper motors and DC motor velocity, and processing capabilities of a microprocessor.

For ROS, using a MATLAB script as processing code running in a computer with 2.8GHz microprocessor, connected through serial protocol at 9600 bps, is possible to reach 6.131 points per second, considering acquisition, processing, and plotting in real time. SAD algorithm in MATLAB with the same computer is capable of processing a 640 $\times$ 480 pixel data set in 2.9 seconds.

Using a field-programmable gate array (FPGA) in both techniques, stereo system is capable of reaching 20 fps of 1024 $\times$ 768 pixels each [36–39], and ROS could reach ~20 lectures per second, giving to notice the difference in data volume of both systems. So it is easier to obtain high-volume data with a stereo system in less time than ROS.

During the definition of ROI, ROS volume data time effect could be less strong in a scene analysis if the definition of ROI is optimized (calculated for ROS scan in regions where it is truly necessary); however, only with the time for one long ROS positioning could be sufficient to process a whole depth map in stereovision. This is the reason why ROS is used for complementary and correction analysis, taking advantage of its high precision.

3.1.2 High Precision

On the other hand, the principal advantage of ROS over stereo system is the highly accurate measure of the distance of an object using dynamic triangulation; ROS can reach an uncertainty of ± 0.1 mm in x axis, ± 1.3 mm in y, and ± 0.0071 mm in z without post-processing, resulting in a distance of 1.303 mm from ideal point to

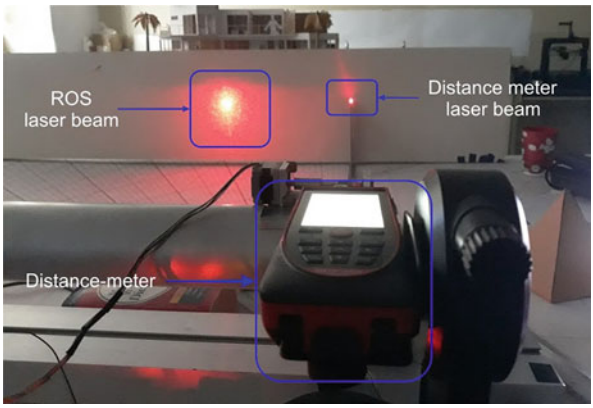


Fig. 15 Coordinate measurement with ROS and distance meter LEICA Disto D810

real point, equivalent to 0.1303% of error. Error data was confirmed comparing 3D coordinates obtained using ROS with a distance meter (Fig. 15). Even, an optimization of results can be achieved with [40] method.

Stereovision uncertainty is, under [41] conditions, ± 13.3 mm in x axis, ± 0.4 mm in y, and ± 0.3 mm in z (axis has been adjusted for direct comparison), giving a total distance of 13.309 mm, equivalent to an error of 1.3309%. Error difference of ~1.2% gives to notice the precision superior precision of ROS over stereo system and its relevance in combined use with stereovision.

Another aspect to consider is that, in stereo systems, depending on the lighting conditions and objects geometry and texture, some information could be lost and leave important spaces in depth matrix without information. These spaces are defined as a region of interest (ROI) and bring the opportunity to use the stereo system for a first depth analysis and ROS to fill void information spaces or take a second measure in critical regions which could have some characteristics as abrupt depth changes or elements that could collide with a robot if used for navigation. Assuming this, we can note that independent use of every single system for this purpose cannot complete the total requirement of the application. Only combined algorithm, uniting the strong skills of both, is only able to fulfill the application needs in the meaning of taking the decision right on time, not a little bit later.

3.1.3 Regions of Interest

ROI is a data bridge for ROS and stereo system communication. It's generated by output depth data of stereo system and is used for ROS control, where an algorithm establishes which regions are void of data or which ones had no reliable data. So, once these regions are identified, it is relevant to establish a standard and simplified format to achieve an optimal control and data transfer.

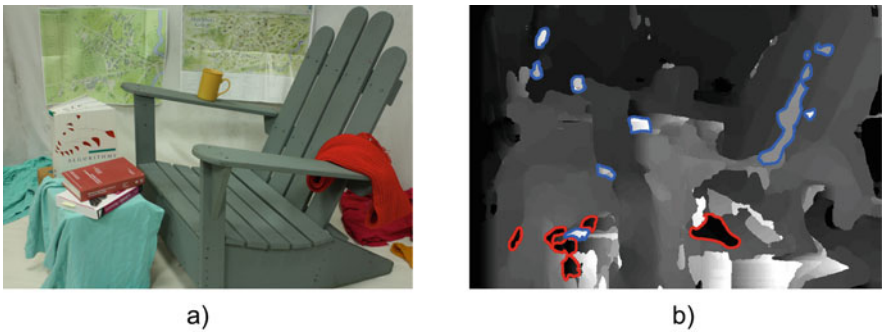


Fig. 16 Regions of interest in disparity map. (**a**) Left image of stereo data set, (**b**) disparity map with void spaces in red and wrong estimations in blue

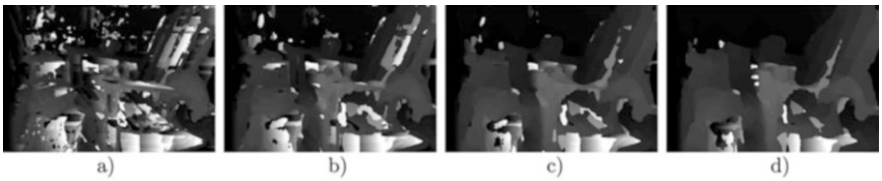


Fig. 17 Disparity map with different window sizes. (**a**) Window of 5 pixels, (**b**) window of 10 pixels, (**c**) window of 15 pixels, (**d**) window of 20 pixels

To get ROIs, stereo depth data is processed with the goal of identifying void spaces or unreliable data and classifying it for optimal laser scanning. In Fig. 16b ROIs are marked, void spaces in red and wrong measurements in blue.

ROIs that are defined as no reliable data need a recalculation when a surface has considerable variations that can be identified as abrupt changes in depth data. This can be obtained when SAD window size is small enough to generate noise in disparity map (Fig. 17).

Considering that window size is dependent on the number of details in a scene [42], taking the system from an environment to another may need an adjustment of the window size. So, it is hard to obtain a disparity map with good number of details and low amount of noise, making redundance a necessary process for most cases.

After ROIs are defined, data is classified in the four types of scanning modes of ROS and represented as Fig. 18 shows.

All ROS scanning modes share the following parameters: horizontal starting point (h_1), vertical starting point (v_1), horizontal ending point (h_2), vertical ending point (v_2), angle per step in degrees (χ), and averaged points per step.

Averaged and χ points per step are preestablished in the initial ROS working parameters and used in every displacement transition and direction change.

In terms of displacement, steps between each point coordinate measurement are obtained with (20), for horizontal displacement:

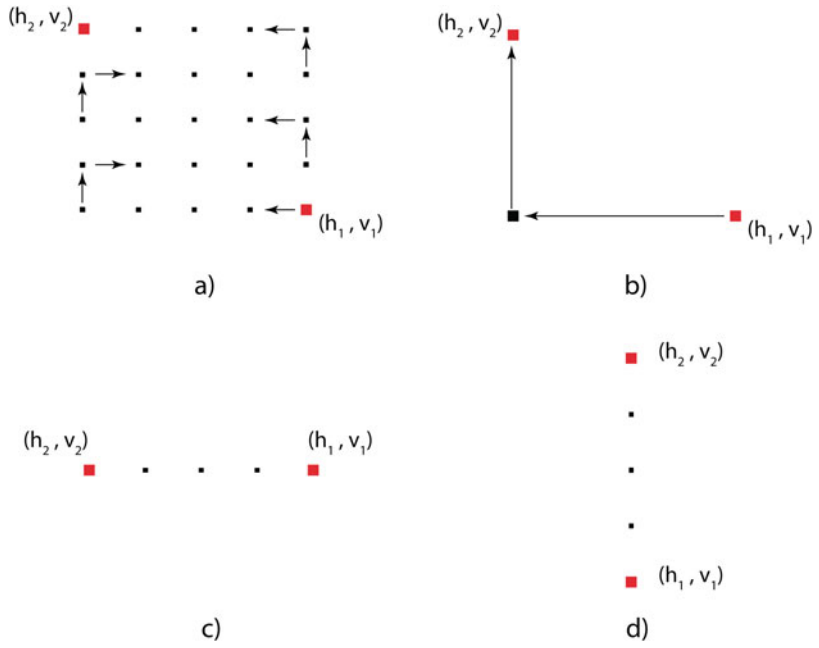


Fig. 18 Classification of ROIs. (**a**) Full area, (**b**) single point, (**c**) horizontal line, (**d**) vertical line

$$D_{\mathrm{h}} = \frac{h_2 - h_1}{\chi} \tag{20}$$

For vertical displacement, Eq. (21) is used:

$$D_{\mathrm{v}} = \frac{v_2 - v_1}{\chi} \tag{21}$$

For the ROS, the position of (h_1, v_1) and (h_2, v_2) points may be in any way, making faster the transition between ROIs' scanning and avoiding a readjustment for every scanned line.

Once the ROIs are formed and the starting and ending points defined, a vector is sent to the ROS control panel as a scanning sequence (SS) with the following format:

$$\mathrm{SS} = [\vartheta, h_1, v_1, h_2, v_2] \tag{22}$$

where ϑ defines the ROI typical types of scanning modes presented in Fig. 18, using the numbers from 1 to 4, respectively.

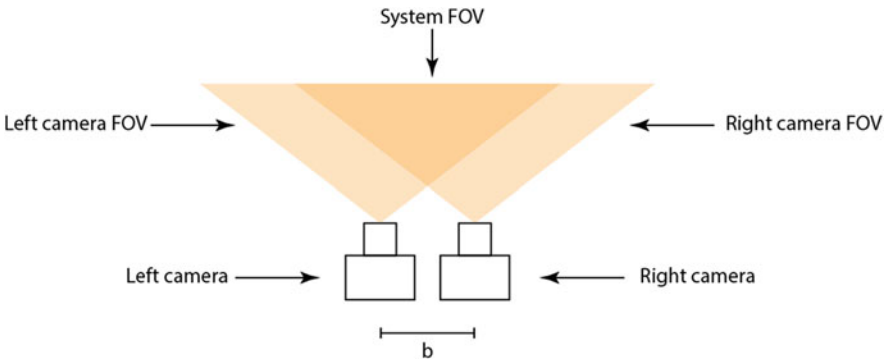


Fig. 19 Field of view of a stereo system

Displacement from (h_1, v_1) to (h_2, v_2) are limited by the field of view of combined system, where the vision system with less scope defines horizontal and vertical limits of vision of the whole combined system.

3.1.4 Field of View Synchronization

Ideally, displacement of ROS positioner and aperture scanner can cover all the ROIs generated with stereovision, but considering the physical construction, the field of view (FOV) of ROS could be bigger or smaller than stereovision when there are variations on parameters like focal length of cameras and baseline of ROS.

FOV of each system is restricted by very different subcomponents. For stereo system, FOV is given by the lens focal length (f) and cameras' baseline (b), and at more f and b value, FOV is tighter and vice versa. It can be shown in Fig. 19.

FOV angle (FA) for lenses can be obtained with Eq. (23):

$$\mathrm{FA} = 2 \arctan\left(\frac{L}{f}\right) \tag{23}$$

L is the CMOS or CCD camera sensor width and both L and f are expressed in millimeters and FA is obtained in degrees.

FOV of ROS is defined by the baseline and the notches of the aperture and positioner; with wider notches and shorter baseline (a) value, a wider FOV will be obtained, but it is important to consider that notches prevents light lacks that could generate noise in the aperture, so it cannot be totally open. This FOV can be represented in Fig. 20.

For fused system, it is more convenient that FOV of ROS is wider than stereovision, ensuring that all ROIs are covered in the redundance process where depths are recalculated or complemented.

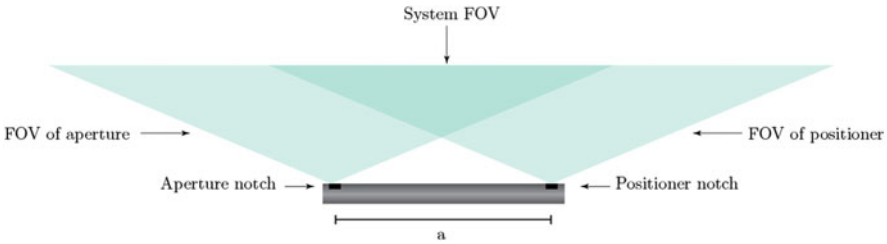
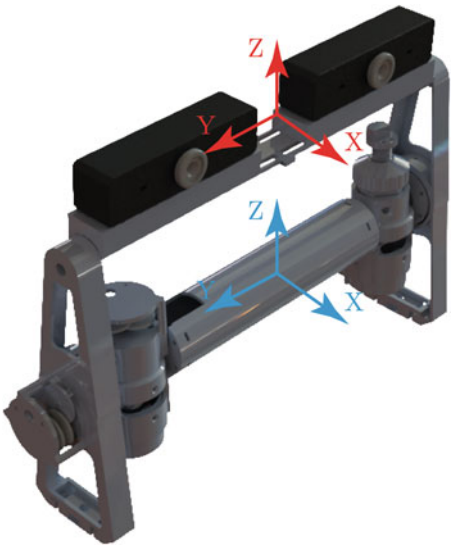


Fig. 20 Field of view of ROS

Fig. 21 Axis of stereovision (orange) and ROS (blue)



3.1.5 Data Link for Depth Enrichment

Synchronization of data between systems results in a ROS scanning of defined points where a re-estimation is necessary, because of lack of information in determined regions or low confidence in stereovision depth estimation in a specific area of the scene.

To achieve a precise scan of ROS in a region of interest, both systems share the same origin for rectangular coordinate estimation, so calibration has to be done in order to have the same center point. For this case, ROS origin is established as the collaborative system origin, due to the ease to keep fixed the position of positioner and aperture. In Fig. 21, orange coordinate system represents the stereovision axis, and blue coordinate system is the ROS and general axis at the same time.

Knowing the vertical and horizontal FOV of stereo system, pixels of images can be divided by their angular position which will stay constant despite the depth estimation for each pixel. This can be represented as Fig. 22a, b, where horizontal

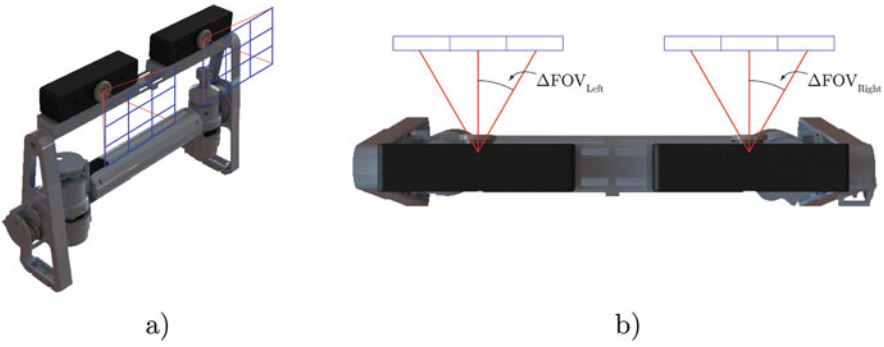


Fig. 22 (**a**) Vertical and horizontal FOV of each camera and (**b**) the differential FOV per pixel from top view

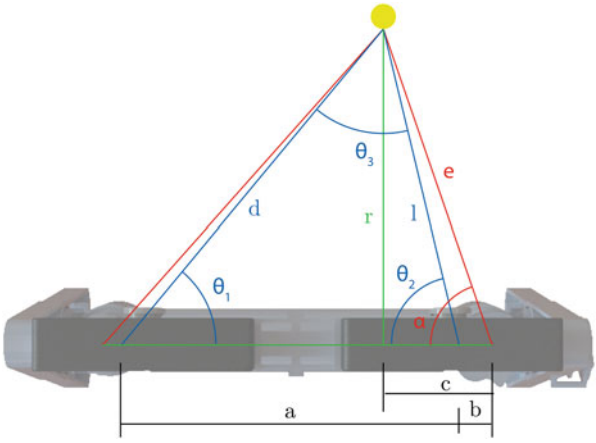


Fig. 23 Relation between positioning of and object with ROS and stereovision

and vertical FOV is divided by width and height of the matrix which contains image data.

Once the matching is done, the angular position of the correspondent pixels allows to point the laser in the same (x, y) coordinates to take measurements. This can be illustrated as shown in Fig. 23, where blue lines represent triangulation with stereovision and red lines represent triangulation with ROS.

So with the angular positions of pixels for left image (θ_1) and right image (θ_2), as shown in Fig. 23, r can be calculated using the Heron equation:

$$r = \frac{2}{a}\sqrt{s\,(s-a)\,(s-d)\,(s-l)} \tag{24}$$

The semi-perimeter (s) can be obtained with

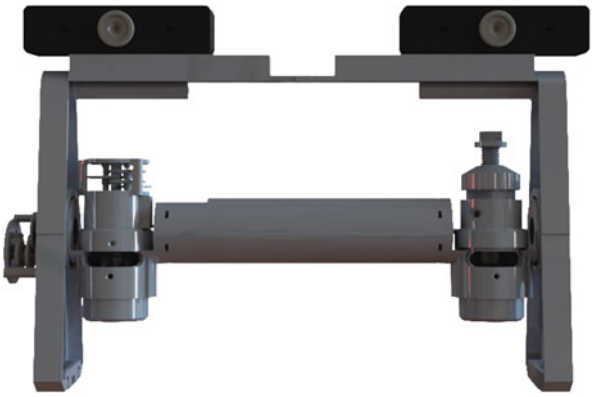


Fig. 24 Cameras vertically aligned with ROS positioner and aperture

$$s = \frac{a + d + l}{2} \tag{25}$$

Using the law of sines, d and l are calculated in the following way:

$$d = \frac{a\text{Sen}\,(\theta_2)}{\text{Sen}\,(\theta_3)} \tag{26}$$

$$l = \frac{a\text{Sen}\,(\theta_1)}{\text{Sen}\,(\theta_3)} \tag{27}$$

Finally, the laser positioner angle is calculated with the relation

$$\alpha = \tan^{-1}\left(\frac{r}{c}\right) \tag{28}$$

where

$$c = a + b - \frac{d}{\text{Cos}\,(\theta_1)} \tag{29}$$

This procedure allows a known displacement between cameras, laser positioner, and aperture. The alternative and low processing cost way is the installation of cameras directly over the ROS positioner and aperture (Fig. 24).

A shared vertical axis of each pair of components forms one same triangle for both systems; therefore, θ_2 and α are equal or just pitched in reference to the camera over the positioner. However, considering this setup, at bigger baseline, cameras should be more pitched to see the same scene from different perspective, and at smaller baseline, ROS will have lower depth scope. In conclusion, this alignment gives more limitations to the collaborative system when the main goal is to obtain the best results of each technique. So, in order to obtain better results of collaborative

system and minimize synchronization errors, it is better to achieve a good calibration of each subcomponent.

3.1.6 Calibration

Calibration of ROS and stereovision has an enormous importance to achieve minimal losses of information and, therefore, measurements which estimation have a small affectation due to a bad functioning of subsystems. These losses could be generated in any of both systems, so each of them has to be analyzed by its own sources of error.

Focal Point of ROS Aperture

ROS aperture uses two biconvex lenses for light concentration over a phototransistor (Fig. 8b), so if the photosensor has a deviation in its position with respect to the focal point, intensity of measurements will be lower than maximum, making a weaker signal and, in light conditions of over 150 lux, could represent between distinguished laser beam and unwanted external light sources.

To estimate the focal point and, perhaps, the photosensor position, lens calculations are used (Eq. (30):

$$\frac{1}{d_{\mathrm{i}}} + \frac{1}{f_{\mathrm{l}}} = \frac{1}{d_{\mathrm{O}}} \tag{30}$$

where f_{l} is the focal point, d_{O} is the distance from the object to the lens, and d_{i} is the distance from the lens to the image. To obtain d_{i} (considering that d_{O} is defined by the user), magnification (M_{l}) of lens is used:

$$M_{\mathrm{l}} = -\frac{d_{\mathrm{i}}}{d_{\mathrm{O}}} \tag{31}$$

For ROS aperture, these calculations have to be done twice, one for each lens. So, first all the parameters have to be settled, as Fig. 25 shows.

Considering that 45-degree-cut mirror is positioned in focal length f_0 and both lenses have the same optical and geometrical properties, then

$$f_{\mathrm{l}} = f_0 \tag{32}$$

To obtain f_{l} and f_0 Eq. (33) is used:

$$\frac{1}{f_1} = \frac{1}{f_0} = \frac{1}{d_{\mathrm{O}}} + \frac{1}{d_{\mathrm{O}} M} \tag{33}$$

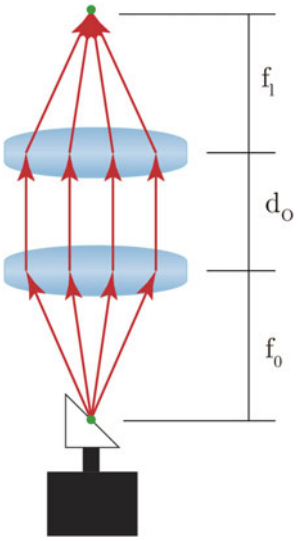


Fig. 25 Array of 45-degree mirror, lenses, and photosensor position

According to the inverse square law, all the components should be the closest possible to avoid light intensity losses; however, contact between elements could damage the lenses or limit a robust adjustment of all of them.

ROS Positioner Calibration

To reach a precise depth estimation with ROS, measurement of B_{ij} and C_{ij} angles must be as accurate as could be possible, but B_{ij} has the benefit of using the aperture as a dedicated subsystem for it, in which estimation is optimal under low light conditions, a microcontroller with high sampling, and a sensor with high sensitivity to the laser wavelength. However, C_{ij} does not have dedicated sensors to estimate its position; in fact, the horizontal and vertical position is inferred according to the number of steps moved by each stepper motor. So, considering that stepper motors have a fixed position for each step, one of them is considered as the starting point or the step which corresponds to 90 degrees counting in clockwise (Fig. 26), ensuring that a determinate number of steps will move the laser beam to an expected position.

The calibration of the positioner implies the use of external references to determine if laser is pointing to the desired spot. For ROS, a calibration sheet is used in order to achieve an initial positioning in 90 degrees, as Fig. 27 shows.

This calibration sheet is also useful for aperture lecture estimation; in other words, it shows an approximation of which could be the result of the aperture readings which gives a reference value during testing.

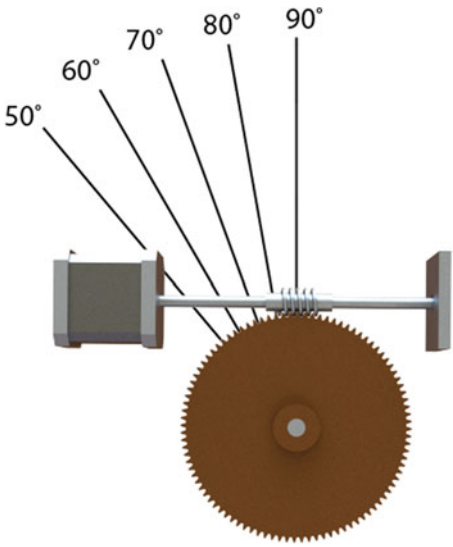


Fig. 26 Positioner degrees counted in clockwise

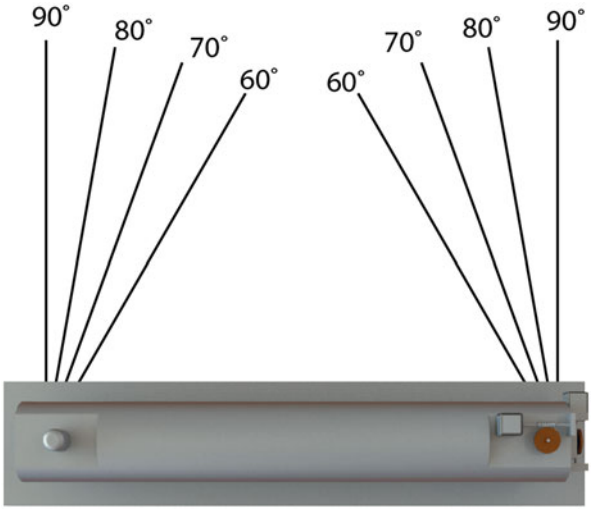


Fig. 27 Representation of ROS over calibration sheet

Stereo Cameras' Calibration

The most common problem before the capture of images in stereovision system is the lack of parallelism of cameras' axis and the deformation in images due to the use of spherical lens. Even if precision tools are used for mounting, it's improbable to achieve a perfect alignment of stereo pairs without rectification. So, calibration implies the use of an algorithm for rectification and reduction of deformation in images [43–45].

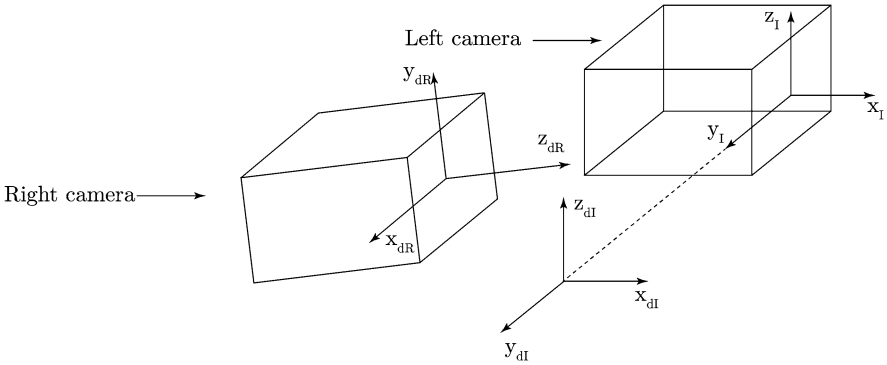
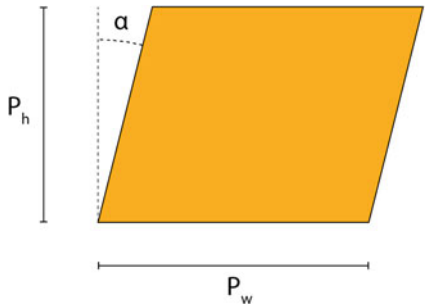


Fig. 28 Unaligned stereo cameras

Fig. 29 Skewed pixel



In the case of rectification, it is used to adjust the rotation and position of one camera with respect to the other [46]. In Fig. 28, left camera is taken as the reference camera and right camera is the one that will be rectified.

To rectify images, a relation between images and real scene has to be settled; this is achieved using intrinsic and extrinsic parameters. Intrinsic parameters are the camera specs compiled in a 3×3 matrix:

$$K = \begin{bmatrix} f_x & s & R_x \\ 0 & f_y & R_y \\ 0 & 0 & 1 \end{bmatrix} \tag{34}$$

where f_x and f_y are the horizontal and vertical focal lengths; these values can be different if pixels are not square. R_x and R_y are the positions of the central pixel and s is the skew factor that can be calculated with Eq. (35):

$$s = -fx \cdot \cot(\alpha) \tag{35}$$

being α the skew angle in a not square pixel (Fig. 29).

Extrinsic parameters establish the position of cameras in real world using rotation matrixes (R) and a position vector (t). Rotation matrixes for each camera are described in Eqs. (36) to (38):

$$R_{\mathrm{x}}(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\psi) & -\sin(\psi) \\ 0 & \sin(\psi) & \cos(\psi) \end{bmatrix} \tag{36}$$

$$R_{\mathrm{y}}(\theta) = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \tag{37}$$

$$R_{\mathrm{z}}(\theta) = \begin{bmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{38}$$

These matrixes are frequently simplified using Eq. (39):

$$R = R_x(\psi) \cdot R_y(\theta) \cdot R_z(\phi) \tag{39}$$

Resulting in one matrix R

$$R = \begin{bmatrix} \cos(\psi)\cos(\theta) & \sin(\psi)\cos(\theta) & -\sin(\theta) \\ \cos(\psi)\sin(\theta)\sin(\phi) + \cos(\psi)\cos(\phi) & \sin(\psi)\sin(\theta)\sin(\phi) - \sin(\psi)\cos(\phi) & \cos(\theta)\sin(\phi) \\ \cos(\psi)\sin(\theta)\cos(\phi) - \sin(\psi)\cos(\phi) & \sin(\psi)\sin(\theta)\cos(\phi) - \cos(\psi)\sin(\phi) & \cos(\theta)\cos(\phi) \end{bmatrix} \tag{40}$$

Position vector contains the offset of the camera respecting an ideal position:

$$t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \tag{41}$$

It can be graphically represented in Fig. 30.

Intrinsic and extrinsic matrixes can be computed using a variety of algorithms, like [47, 48]; one common algorithm is direct linear transformation (DLT) [49]. In MATLAB, obtaining parameters is simplified through the use of "Stereo Camera Calibrator App," which only requires 10–20 stereo pairs where a checkerboard pattern with known distance between rectangles can be seen (Fig. 31). Borders are detected and parameters are computed for parameter estimation.

Once all parameters are computed, images are rotated and cropped in order to obtain an aligned stereo pair. This can be done through all the variety of algorithms used in rectification [50–53].

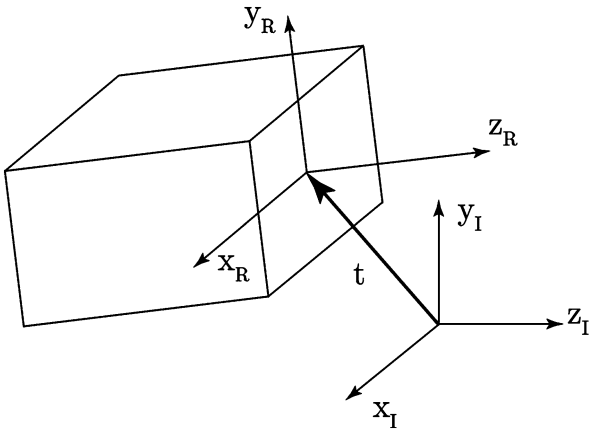


Fig. 30 Position vector (t) from ideal position (x_I, y_I, z_I) to real position (x_R, y_R, z_R) of camera

Fig. 31 Stereo pair with calibration checkerboard

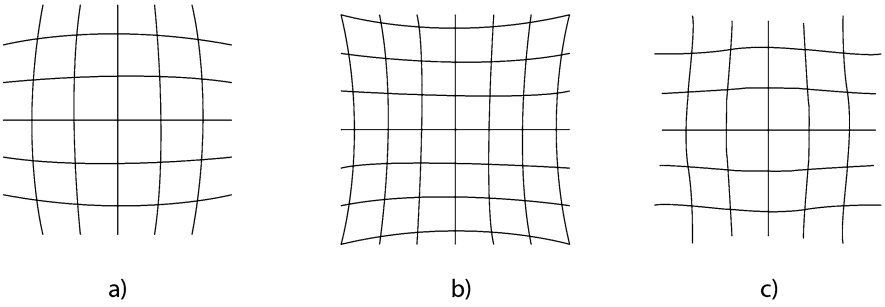


Fig. 32 Distortions in images. (**a**) Barrel, (**b**) pincushion, and (**c**) mustache

The next step in calibration is the reduction of distortion in images due to the adaptation of the light by spherical elements (lenses array) in a plane (CMOS or CCD sensors), creating curved lines where should be plane lines. These distortions are classified according to the deformation caused in the captured image; some of them are the barrel distortion, pincushion distortion, and a combination of both, mustache distortion (Fig. 32).

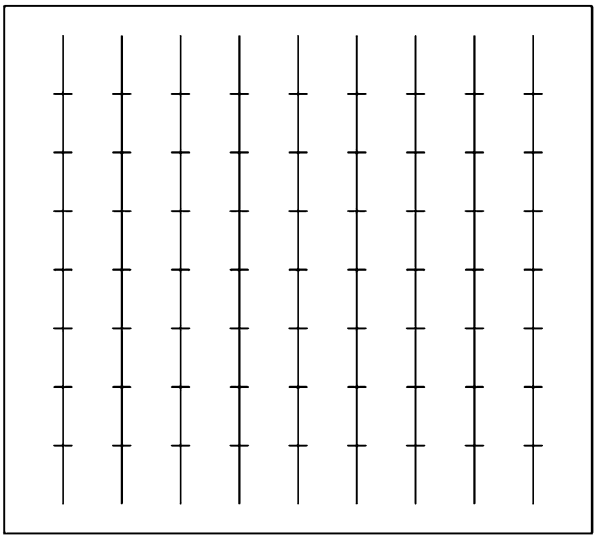


Fig. 33 Pattern used for distortion correction algorithm

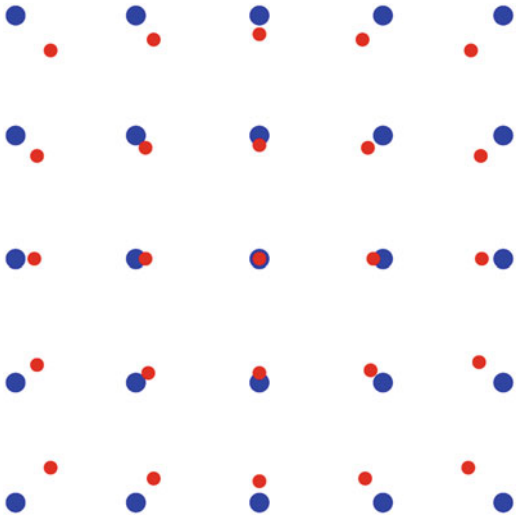


Fig. 34 Comparison of ideal position of points (blue) and real position (red)

Taking a capture of a calibration sheet (Fig. 33), computing the position of all the cross in the sheet, and comparing them with ideal points, distortion can be estimated for a specific lens, focal length, and focus. If one parameter changes, then there should be a recalibration.

A comparison of ideal (blue) and real (red) distribution of dots (blue) is represented in Fig. 34, where red dots have a barrel distortion which is lower in the center and more notable in borders.

Once the distortion is estimated with algorithms as used in [54] where a calibration sheet like that shown in Fig. 33 is used, these values can be saved in a file like a calibration profile. So, each captured image can be corrected using the profile without the requirement of repeating the calibration method.

3.1.7 Depth Estimation Time

The last topic to analyze is the processing time required to obtain enriched depth map, where, considering that it is not possible to make parallel process for the data of the collaborative systems due to the positioning of ROS which strongly depends on the definition of ROI gotten of the stereovision depth data (as Fig. 14 processing blockchain diagram describes), the processing time necessary to get a simple enriched depth matrix mainly depends on image resolution in megapixels taken from stereo cameras, processing capacity of the computer or embedded chip used for depth estimation, ROIs' establishment, the velocity of stepper motors for positioning the laser beam in a desired vertical and horizontal place, and velocity of DC motor in every loop of the 45-degree-cut mirror, considering that in movement transition, no reading is taken. And the last factor is the position point estimation time using an embedded chip:

$$t_{\mathrm{p}} = \frac{t_{\mathrm{ppx}}}{W_{\mathrm{sz}}} + \frac{I_{\mathrm{px}}}{\kappa} + (\kappa * t_{\mathrm{DCM}}) + \sum_{i=1}^{\kappa-1} \frac{V_{\mathrm{SM}}}{\mid p_{i+1} - p_i \mid} \tag{42}$$

where t_{ppx} is the processing time per pixel, W_{sz} is the window size, I_{px} is the quantity of pixels of the picture, κ are the number of interest pixels, t_{DCM} is the time required for a loop of the DC motor, V_{SM} is the velocity of the stepper motor, and p is the position of each i interest pixel.

For ROS is used an Arduino Mega 2560 with a real-time processing algorithm capable of reaching a sample rate of ~4 kHz, limiting to ~7 rps in the DC motor. In this stage, more processing capacity results in more position point estimation per second and more precision. Also, actual system achieves a displacement of 0.64 deg/seconds, being the most time cost process of both systems; however, the use of stepper motors ensures a more reliable position system.

Stereovision system gets a bigger FOV with short focal lengths and vice versa, but is widely known that short focal length generates deformations in images [55], so to avoid problematic deformations with ultra-wide camera lenses, 20–35 mm focal length is selected (equivalent in full frame cameras), giving as a result, a horizontal FOV of ~81–112 angular degrees. The time required to move ROS positioner between the furthest points in a horizontal 81-degree FOV and a vertical 54-degree FOV (3:2 ratio) is 3.51 minutes.

4 Summarized Description of Stereovision and ROS Cooperative Use

Fused system combines the best features of ROS and stereovision, achieving high data volume with over 179,200 3D points with a minimum uncertainty of 0.1303%, generating an enriched depth map which could not be obtained with each system on its own, remarking the benefits of combined use. On the other hand, processing time is the cost to pay, due to the slow movements of ROS stepper motors requiring 3.51 minutes for the furthest point movements and estimating 3D points with a rate of 6.131 points per second; however, system speed is open to improvements, thanks to new technologies as FPGAs for processing and precise servo motors for quick movements.

To obtain a 3D point matrix with data of both systems, a control panel developed in MATLAB is used; it captures a stereo pair with cameras, estimates depth of elements of a scene, determines ROIs according to data acquired, generates and sends a scanning routine which is used for ROS for void spaces' filling and depth re-estimation, and presents a point cloud. These routines are standardized in a simple and robust format for minimum delay in data handle, requiring only starting point, ending point, and scanning type.

For stereovision is used a region matching algorithm, specifically SAD cost function, capable of processing grayscale stereo pairs of 640 $\times$ 480 pixels in 2.9 seconds in a computer with a 2.8GHz microprocessor. Considering that has been used a medium to low computational capacity computer, disparity maps can be estimated in fractions of seconds using CPU with high frequency, GPUs, or FPGAs, giving more responsibility to ROS for high scanning time of each enriched depth map.

In stereovision and ROS cooperative use, FOV is also a determining factor for scanning time. While a wider FOV gives a very open view of a scene, ROS requires more time to analyze each ROI due to longer displacements between each region and point, and also stereo pair images present more deformation, requiring a better calibration. On the opposite, a narrow FOV presents the benefit of shorter displacements in ROS but a closer view of the scene. So, FOV was established in 112.61 degrees (focal length of 24 mm equivalent in a full sensor camera), being a good balance between an open view and a not too long distance between furthest points.

5 Solutions and Recommendations

As shown in Fig. 17, disparity map noise obtained from stereo data set generates dispersed and large quantity of ROIs that, in some cases, does not belong to important objects in a scene. This even leads to more corrections and complementary data acquisition from ROS, spending valuable time in not valuable data. A way to reduce this noise, which is already implemented in many depth estimation algorithms, is

the use of image filters to smooth borders and even disappear a good amount of noise. Median and bilateral filtering are examples of noise reduction filters that can help to save energy and time.

Another important aspect to consider is that scope of combined system is not defined, making necessary to test which baselines, focal lengths, lighting conditions, and sensors are suitable for short distance elements, which ones are for long distance elements, and which are the distance limits where precise measurements can be obtained.

6 Future Research Directions

Knowing that weakest point of combined system is the time cost of data acquisition due to the use of stepper motors for positioning in ROS, a solution for faster and continuous movements is presented in [56]; however, friction plays an important role when servo motors try to arrive in a very precise position in long transitions, decreasing reliability of the measurements. This maintains opening the opportunity for the optimization of subsystems capable of compensating the movement through monitoring with external sensors, strengthening ROS precise movements. Also, stereovision demands a high amount of processing cost for correspondence, so the implementation of systems as FPGAs leads to obtaining depth maps in a fraction of a second. For ROS, implementation of FPGA represents a bigger sampling rate which results in faster rotations in DC motors and more precise depth estimations.

Considering mentioned implementations, it is possible to reach fast depth estimation with enriched data for better path planning and quicker decision-making in robot navigation.

7 Conclusion

The modern robotics naturally itself starts a magisterial trend: presented robots of the most recognized laboratories and universities in the field [57–59] are using multiple devices for detection of the same characters. It means that a battle between the different theoretical approaches to machine vision is finished empirically. All invented machine vision methods always have unique advantages, which are impossible to reach for other methods. So, for better and more complete knowledge about the surrounding, robots preferably have to spend the information coming from physically different devices [60], and combine it in an enriched matrix of surrounding data. However, this implies another theoretical problem. The data coming from different informational channels no doubt bring more diverse, varied, complete, and comprehensive information about the surrounding state. But at the same time, it significantly raises the measurement noise, due to the same reason of different physical nature of measuring approaches. Sometimes, it is difficult to classify all physical sources and cross-talk effects within the limited time. Also, the use of several expensive technical systems onboard of robot for the same purpose

increases significantly the cost of such robot. This is a negative factor; obviously, it can be compensated with time by devices' price decrease, especially after massive production. However, it requires several years. Still, a more hard disadvantage is that combination of information from multiple independent channels and its ranking requires additional processing time. In conditions of limited time for decision, this lack becomes crucial.

So, the present chapter is an original proposition to solve this complex of problems. Herein above we are trying to establish a new approach to smart selections of technical systems from their big amount, and trying to establish the physical and mathematic bases for optimized combination of them. The criteria are the maximal uniqueness of data, maximal versatility of obtained data, and the most simple way to obtain the combined matrix of data. Fused system offers high-quality depth maps with possibilities of fast data enrichment using FPGAs and quick positioning ready for path planning and decision-making in robot navigation. These depth maps can reach a precision with a minimum error of ± 2 mm without processing, thanks to ROS. Also, in terms of data volume, with stereovision it is possible to obtain over 307,200 three-dimensional dots forming a precise and detailed cloud of points of the elements of a scene. So, combined system obtains improved features of ROS and stereovision compared when they work by separately.

In terms of synchronization, ROS can add or correct depth data knowing the relative position between cameras and ROS subsystems. To achieve it, ROIs from stereovision depth estimation are formed and converted to angular positions; a control panel generates a scan sequence and sends the order to the ROS controller for scan of enrichment of these ROIs. Taking conversion of stereo data to ROS scanning sequence into consideration, bad physical adjustment of stereovision and ROS components, as any measurement system, can lead to take bad lectures from both systems, so, although cooperative function between both techniques as separate devices is totally possible, it is recommended the collaborative work as a unique integrated device, with the purpose of keeping fixed the relative positions and reducing time between calibrations due to non-intentional displacements of components.

Besides production cost, a technical disadvantage of collaborative system is the scanning time due to the slow movements of ROS stepper motors, which is used due to its repeatability of movement and very precise positioning. However, rotary encoders and servo motors are taken into consideration to accelerate data acquisition in ROS. It is expected that these changes can be applied once a reliable way to obtain a precise positioning close to stepper motors can be found.

Collaborative work of stereovision and ROS is a reliable way to obtain dense and precise data for decision-making and object avoiding in robot navigation with wide improvement opportunities in positioning velocities and processing components for estimation time reduction, taking advantage of other systems like less complexity in its construction, algorithms, and theoretical and mathematical fundaments.

Acknowledgments Authors have special thanks to Mexican organization CONACyT for funding of PhD students participating as co-authors of this chapter, as well as our universities who provide our group with facilities, laboratory and equipment for herein presented research.

Appendix A

Code used to send stepper motors' sequence usable for Arduino. This code considers a serial input with format "DX SY," where *D* is a symbolic reference to the direction where *Y* can take a value from 1 to 4, being 1–2 both directions of one stepper motor (horizontal for ROS) and 3–4 for the second stepper motor (vertical for ROS). *S* refers to the number of steps *Y*.

```
#define m1_1 22
#define m1_2 23
#define m1_3 24
#define m1_4 25
#define m2_1 26
#define m2_2 27
#define m2_3 28
#define m2_4 29

int dir = 0, lastdirH = 0, lastdirV = 0,
steps = 0, binH = 1, binV = 1;
bool enc = false;
float vMotor = 0;

void setup() {
  pinMode(m1_1,OUTPUT);
  pinMode(m1_2,OUTPUT);
  pinMode(m1_3,OUTPUT);
  pinMode(m1_4,OUTPUT);
  pinMode(m2_1,OUTPUT);
  pinMode(m2_2,OUTPUT);
  pinMode(m2_3,OUTPUT);
  pinMode(m2_4,OUTPUT);

  digitalWrite(m1_1,0);
  digitalWrite(m1_2,0);
  digitalWrite(m1_3,0);
  digitalWrite(m1_4,0);
  digitalWrite(m2_1,0);
  digitalWrite(m2_2,0);
  digitalWrite(m2_3,0);
  digitalWrite(m2_4,0);

  Serial.begin(9600);
  }

void loop(){
  if (Serial.available() > 0){
    String dataSt = Serial.readString();
    int str_len = dataSt.length() + 1;
    char data[str_len];
    dataSt.toCharArray(data, str_len);
    sscanf(data,"D%d S%d",&dir,&steps);
```

```
    if (lastdirH == 0 && dir < 3){
      lastdirH = dir;
      }
    if (lastdirV == 0 && dir > 2){
      lastdirV = dir;
      }
      if (dir == 1 || dir == 2){
        if (lastdirH != dir){
                binH = -(binH-5);
                lastdirH = dir;
          }
        }
      if (dir == 3 || dir == 4){
        if (lastdirV != dir){
                binV = -(binV-5);
                lastdirV = dir;
          }
        }
      int i = 0;
      for (i = 1;i <= steps;i++){
        delay(500);

if(dir == 2 || dir == 4){
          if (dir == 2){digitalWrite(m1_1+binH-1,0);}
          if (dir == 4){digitalWrite(m2_1+binV-1,0);}
        }
        if(dir == 1 || dir == 3){
          if (dir == 1){digitalWrite(m1_4-binH+1,0);}
          if (dir == 3){digitalWrite(m2_4-binV+1,0);}
        }

if (dir == 1 || dir == 2){
          if (binH == 4){binH = 0;}
          binH++;
        }else{
          if (binV == 4){binV = 0;}
          binV++;}
        if (dir == 1 || dir == 2){
          if (dir == 2){digitalWrite(m1_1+binH-1,1);}
          if (dir == 1){digitalWrite(m1_4-binH+1,1);}

}
        if (dir == 3 || dir == 4){
          if (dir == 4){digitalWrite(m2_1+binV-1,1);}
          if (dir == 3){digitalWrite(m2_4-binV+1,1);}
        }
      }
      Serial.println((String) 5);
    }
  }
```

Appendix B

Code used for beta angle estimation in ROS usable for Arduino

```
float vmaxO = 0,vminO = 999,vrefO = 0; //Reference volages
int capture = 0,c = 0,pos = 0, pos_ant = 0,change = 0; //Flags
float opto = 0,foto = 0; //Data
int picoO[2] = {0},picoF = 0, maxF = -9999; //Peaks
int cpicosO = 0,diffO = 0,diffF = 0; //Peaks;void setup() {
    pinMode(A0,INPUT);
    pinMode(A1,INPUT);
    Serial.begin(9600);
}void loop() {
  if (c>50000){
    vmaxO = vminO * 2;
    }
  opto = analogRead(A0);
  foto = analogRead(A1);
  foto = - foto;  if (opto < vminO){vminO = opto;}
  if (opto > vmaxO || vrefO > vmaxO){vmaxO = opto;}
  vrefO = (vminO + vmaxO) / 2;
  if (opto<=vrefO){capture=1;}  if (capture==1){
    if (foto > maxF){
      maxF = foto;
      picoF = c;
      }
    c++;
    if (opto > vrefO * 1.1){
        pos = 1;
        if (pos_ant == 0){
            change = 1;
            if (cpicosO == 1){
              picoO[0] = c;
              }
            else{
              picoO[1] = c;
              }
          }
        pos_ant = 1;
      }
    else if (opto < vrefO * .9){
        pos = 0;
        if (pos_ant == 1){
            change = 2;
            if (cpicosO == 0){
              cpicosO++;
              }
            else{
              cpicosO=2;
              }
          }
        pos_ant = 0;
      }    if (cpicosO == 2 && change  == 1){
```

```
      diffO=picoO[1]-picoO[0];
      diffF=picoF-picoO[0];
      Serial.println((String) "O" + diffO + " F" + diffF);
      maxF = -9999;
      c = 0;
      cpicosO = 1;
      picoO[0] = 0;
      picoO[1] = 0;
      }
    }
  }
```

References

1. Vongbunyong S, Chen WH (2015) Vision System. Springer, Cham, pp. 55–93
2. Chen X (2007) Stereo vision based motion identification. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 575–586
3. He Z, Ren Q, Yang T, Li J, Zhang Y (2016) Multi-object detection based on binocular stereo vision. In: Communications in Computer and Information Science. Springer Verlag, pp. 114–121
4. Mo H, Luo C, Liu K (2016) Robot indoor navigation based on computer vision and machine learning. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag, pp. 528–534
5. Panigrahi PK, Tripathy HK (2015) Analysis on intelligent based navigation and path finding of autonomous mobile robot. In: Advances in Intelligent Systems and Computing. Springer Verlag, pp. 219–232
6. Hart S, Mikhailova E, Post C, McMillan P, Sharp J, Bridges W (2017) Spatio-temporal analysis of flowering using LiDAR topography. J Geogr Sci 27:62–78. https://doi.org/10.1007/s11442-017-1364-x
7. Broggi A, Medici P, Porta PP (2007) StereoBox: A robust and efficient solution for automotive short-range obstacle detection. Eurasip J Embed Syst 2007:1–7. https://doi.org/10.1155/2007/70256
8. Kim S, Kim H Bin (2010) High resolution mobile robot obstacle detection using low directivity ultrasonic sensor ring. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Berlin, Heidelberg, pp. 426–433
9. Zarandy A, Nagy Z, Vanek B, Zsedrovits T, Kiss A, Nemeth M (2013) A five-camera vision system for UAV visual attitude calculation and collision warning. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Berlin, Heidelberg, pp. 11–20
10. Sergiyenko O, Tyrsa V (2020) 3D optical machine vision sensors with intelligent data management for robotic swarm navigation improvement. IEEE Sens J. https://doi.org/10.1109/JSEN.2020.3007856
11. Flores-Fuentes W, Rivas-Lopez M, Sergiyenko O, Rodríguez-Quiñonez JC, Hernández-Balbuena D, Rivera-Castillo J (2014) Energy center detection in light scanning sensors for structural health monitoring accuracy enhancement. IEEE Sens J 14:2355–2361. https://doi.org/10.1109/JSEN.2014.2310224
12. C. J, Sergiyenko O, Tyrsa V, C. L, Rivas-Lopez M, Hernndez-Balbuena D, Pea-Cabrer M (2011) 3D Body & Medical Scanners' Technologies: Methodology and Spatial Discriminations. In: Optoelectronic Devices and Properties. InTech

13. Lindner L, Sergiyenko O, Rivas-López M, Hernández-Balbuena D, Flores-Fuentes W, Rodríguez-Quiñonez JC, Murrieta-Rico FN, Ivanov M, Tyrsa V, Básaca-Preciado LC (2017) Exact laser beam positioning for measurement of vegetation vitality. Ind Rob 44:532–541. https://doi.org/10.1108/IR-11-2016-0297
14. Chung SH, Lee SW, Lee SK, Park JH (2019) LIDAR system with electromagnetic two-axis scanning micromirror based on indirect time-of-flight method. Micro Nano Syst Lett 7:1–5. https://doi.org/10.1186/s40486-019-0082-9
15. Chung S-H, Lee S-W, Lee S-K, Park J-H (2019) LIDAR system with electromagnetic two-axis scanning micromirror based on indirect time-of-flight method. Micro Nano Syst Lett 2019 71 7:1–5. https://doi.org/10.1186/S40486-019-0082-9
16. Szeliski R (2011) Stereo correspondence. Springer, London, pp. 467–503
17. Vilaça JL, Fonseca JC, Pinho AM (2009) Non-contact 3D acquisition system based on stereo vision and laser triangulation. Mach Vis Appl 2008 213 21:341–350. https://doi.org/10.1007/S00138-008-0166-7
18. Denker K, Lehner B, Umlauf G (2010) Real-time triangulation of point streams. Eng with Comput 2010 271 27:67–80. https://doi.org/10.1007/S00366-010-0181-Y
19. Mikhaylichenko AA, Kleshchenkov AB (2018) Approach to Non-Contact Measurement of Geometric Parameters of Large-Sized Objects. Program Comput Softw 2018 444 44:271–277. https://doi.org/10.1134/S0361768818040096
20. Yoo H-S, Kim Y-S, Kwon S-W (2014) A comparative study of noise elimination algorithms for a 3D terrain model through object clustering and the differential method. KSCE J Civ Eng 2015 193 19:498–509. https://doi.org/10.1007/S12205-013-0370-5
21. Madeo S, Pelliccia R, Salvadori C, del Rincon JM, Nebel JC (2016) An optimized stereo vision implementation for embedded systems: application to RGB and infra-red images. In: Journal of Real-Time Image Processing. Springer Verlag, pp. 725–746
22. Lu R, Lai J, Xie X (2018) Asymmetric Two-Stream Networks for RGB-Disparity Based Object Detection. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 11259 LNCS:3–15. https://doi.org/10.1007/978-3-030-03341-5_1
23. Heng J, Xu Z, Zheng Y, Liu Y (2017) Disparity Refinement Using Merged Super-Pixels for Stereo Matching. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 10666 LNCS:295–305. https://doi.org/10.1007/978-3-319-71607-7_26
24. Popielski P, Wróbel Z, Koprowski R (2014) Object Detail Correspondence Problem in Stereovision. Adv Intell Syst Comput 283:209–222. https://doi.org/10.1007/978-3-319-06593-9_19
25. Yousfi J, Lahouar S, Ben Amara A (2017) Strategy of Image Capture and Its Impact on Correspondence Error in Reconstructed 3D-Images-Based Point. Lect Notes Mech Eng 115–126. https://doi.org/10.1007/978-3-319-66697-6_12
26. Lazaros N, Sirakoulis GC, Gasteratos A (2008) Review of stereo vision algorithms: From software to hardware. Int J Optomechatronics 2:435–462. https://doi.org/10.1080/15599610802438680
27. Hirschmüller H, Scharstein D (2007) Evaluation of cost functions for stereo matching. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. https://doi.org/10.1109/CVPR.2007.383248
28. Bae K ryeol, Moon B (2017) An accurate and cost-effective stereo matching algorithm and processor for real-time embedded multimedia systems. Multimed Tools Appl 76:17907–17922. https://doi.org/10.1007/s11042-016-3248-y
29. Patil S, Nadar JS, Gada J, Motghare S, Nair SS (2013) Comparison of Various Stereo Vision Cost Aggregation Methods. Int J Eng Innov Technol 2:222–226
30. Shen Y (2011) Efficient normalized cross correlation calculation method for stereo vision based robot navigation. Front Comput Sci China 5:227–235. https://doi.org/10.1007/s11704-011-9190-2
31. Dong Q, Feng J (2018) Adaptive disparity computation using local and non-local cost aggregations. Multimed Tools Appl 2018 7724 77:31647–31663. https://doi.org/10.1007/S11042-018-6236-6

32. Middlebury (2014) Adirondack. https://vision.middlebury.edu/stereo/data/scenes2014/
33. Hisham MB, Yaakob SN, Raof RAA, Nazren ABA, Embedded NMW (2015) Template Matching using Sum of Squared Difference and Normalized Cross Correlation. In: 2015 IEEE Student Conference on Research and Development, SCOReD 2015. Institute of Electrical and Electronics Engineers Inc., pp. 100–104
34. Ilmenau TU, Kuhl A (2005) Comparison of Stereo Matching Algorithms for Mobile Robots
35. Michalik S, Michalik S, Naghmouchi J, Berekovic M (2017) Real-Time smart stereo camera based on FPGA-SoC. In: IEEE-RAS International Conference on Humanoid Robots. IEEE Computer Society, pp. 311–317
36. Cambuim LFS, Oliveira LA, Barros ENS, Ferreira APA (2020) An FPGA-based real-time occlusion robust stereo vision system using semi-global matching. J Real-Time Image Process 17:1447–1468. https://doi.org/10.1007/s11554-019-00902-w
37. Masmoudi MBM, Jerad C, Attia R (2016) On-the-Fly Architecture Design and Implementation of a Real-Time Stereovision System. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 10016 LNCS:711–722. https://doi.org/10.1007/978-3-319-48680-2_62
38. Li Y, Huang K, Claesen L (2016) A Novel Hardware-Oriented Stereo Matching Algorithm and Its Architecture Design in FPGA. IFIP Adv Inf Commun Technol 508:213–232. https://doi.org/10.1007/978-3-319-67104-8_11
39. Jin S, Yuanzhi W, Yining S (2018) Design and implementation of wireless multimedia sensor network node based on FPGA and binocular vision. EURASIP J Wirel Commun Netw 2018 20181 2018:1–8. https://doi.org/10.1186/S13638-018-1172-8
40. Rodriguez-Quinonez JC, Sergiyenko O, Gonzalez-Navarro FF, Basaca-Preciado L, Tyrsa V (2013) Surface recognition improvement in 3D medical laser scanner using Levenberg-Marquardt method. Signal Processing 93:378–386. https://doi.org/10.1016/j.sigpro.2012.07.001
41. Sankowski W, Włodarczyk M, Kacperski D, Grabowski K (2017) Estimation of measurement uncertainty in stereo vision system. Image Vis Comput 61:70–81. https://doi.org/10.1016/j.imavis.2017.02.005
42. Rajeshkannan S, Korah R (2016) Improved CRC Based Disparity Estimation of Vision System Using Local Adaptive Hue Census and Mean Shift Clustering. Natl Acad Sci Lett 2016 391 39:35–38. https://doi.org/10.1007/S40009-015-0412-2
43. Pratt P, Bergeles C, Darzi A, Yang G-Z (2014) Practical Intraoperative Stereo Camera Calibration. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 8674 LNCS:667–675. https://doi.org/10.1007/978-3-319-10470-6_83
44. Yang S, Liu M, Song J, Yin S, Guo Y, Ren Y, Zhu J (2017) Projector calibration method based on stereo vision system. Opt Rev 2017 246 24:727–733. https://doi.org/10.1007/S10043-017-0370-7
45. Taryudi, Wang M-S (2017) Eye to hand calibration using ANFIS for stereo vision-based object manipulation system. Microsyst Technol 2017 241 24:305–317. https://doi.org/10.1007/S00542-017-3315-Y
46. Hyun J, Moon B (2016) A simplified rectification method and its hardware architecture for embedded multimedia systems. Multimed Tools Appl 2016 7619 76:19761–19779. https://doi.org/10.1007/S11042-016-3517-9
47. Hansen P, Alismail H, Rander P, Browning B (2012) Online continuous stereo extrinsic parameter estimation. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 1059–1066. https://doi.org/10.1109/CVPR.2012.6247784
48. Taketomi T, Okada K, Yamamoto G, Miyazaki J, Kato H (2014) Camera pose estimation under dynamic intrinsic parameter change for augmented reality. Comput Graph 44:11–19. https://doi.org/10.1016/j.cag.2014.07.003
49. Zhao Z, Ye D, Zhang X, Chen G, Zhang B (2016) Improved Direct Linear Transformation for Parameter Decoupling in Camera Calibration. Algorithms 9:31. https://doi.org/10.3390/a9020031
50. Monasse P, Morel JM, Tang Z (2010) Three-step image rectification. Br Mach Vis Conf BMVC 2010 – Proc. https://doi.org/10.5244/C.24.89

51. Dinh VQ, Nguyen TP, Jeon JW (2019) Rectification Using Different Types of Cameras Attached to a Vehicle. IEEE Trans Image Process 28:815–826. https://doi.org/10.1109/TIP.2018.2870930
52. Abraham S, Förstner W (2005) Fish-eye-stereo calibration and epipolar rectification. ISPRS J Photogramm Remote Sens 59:278–288. https://doi.org/10.1016/j.isprsjprs.2005.03.001
53. Kumar S, Micheloni C, Piciarelli C, Foresti GL (2010) Stereo rectification of uncalibrated and heterogeneous images. Pattern Recognit Lett 31:1445–1452. https://doi.org/10.1016/j.patrec.2010.03.019
54. Ramírez-Hernández LR, Rodríguez-Quiñonez JC, Castro-Toscano MJ, Hernández-Balbuena D, Flores-Fuentes W, Rascón-Carmona R, Lindner L, Sergiyenko O (2020) Improve three-dimensional point localization accuracy in stereo vision systems using a novel camera calibration method. Int J Adv Robot Syst 17:172988141989671. https://doi.org/10.1177/1729881419896717
55. Barrena N, Sánchez JR, Ugarte RJ, Alonso AG (2018) Proving the efficiency of template matching-based markerless tracking methods which consider the camera perspective deformations. Mach Vis Appl 29:573–584. https://doi.org/10.1007/s00138-018-0914-2
56. Lindner L, Sergiyenko O, Rodríguez-Quiñonez JC, Rivas-Lopez M, Hernandez-Balbuena D, Flores-Fuentes W, Murrieta-Rico FN, Tyrsa V (2016) Mobile robot vision system using continuous laser scanning for industrial application. Ind Rob 43:360–369. https://doi.org/10.1108/IR-01-2016-0048
57. Sohn K, Jang G (2019) Ground Vehicle Driving by Full Sized Humanoid. J Intell Robot Syst 2019 992 99:407–425. https://doi.org/10.1007/S10846-019-01130-X
58. Everett MF (2017) Robot designed for socially acceptable navigation
59. Belbachir A (2017) An embedded testbed architecture to evaluate autonomous car driving. Intell Serv Robot 2017 102 10:109–119. https://doi.org/10.1007/S11370-016-0213-6
60. Ellery A (2016) Autonomous Navigation—Self-localization and Mapping (SLAM). Planet Rovers 331–374. https://doi.org/10.1007/978-3-642-03259-2_9

Self-attention for 2D Hand Pose Estimation

Nicholas Santavas and Antonios Gasteratos

1 Robots and Humanity

In 1921, Karel Čapek's science fiction play *Rossumovi Univerzální Roboti* (R.U.R.) firstly introduced the word *robot*, and it provided the first concrete vision of how a robot should look, i.e., it should look like a human being. When science fiction author Isaac Asimov devised his Three Laws of Robotics, he was thinking about androids. He envisioned a world where humans would use these anthropomorphic machines as servants and they would need to be programmed with strict rules to prevent them causing us harm.

Throughout the twentieth century, science fiction pioneers long imagined that computers and robots would become a valuable part of modern society. Today, computers and robots serve a much wider range of functions. Technology is breaking down the barriers of the past, and the routine activities of humans and computers are becoming more and more dependent upon the interaction between them. As systems become more sophisticated, their interaction methods become more complex, making their use more cumbersome.

That being said, it is essential to make the interactions with machines as seamless and transparent as possible for the users, to be accessible from a larger part of the society, in terms of both cost and convenience. To fulfill these requirements, the design of human-in-the-loop systems should evolve and be developed side by side with technologies that maximize their social impact at the lowest possible cost (e.g., using established components concerning the current trend).

Developing intuitive ways of interacting with machines that mimic human-to-human communication is crucial to democratizing decision-making machines [1].

N. Santavas (✉) · A. Gasteratos
Department of Production and Management Engineering, Democritus University of Thrace, Xanthi, Greece
e-mail: nsantava@pme.duth.gr; agaster@pme.duth.gr

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1_5

The use of natural dialogues can be a convenient way for humans to communicate with machines. Virtual assistants that incorporate such examples are referred to as interactive conversational systems [2]. In this context, computer vision techniques have been widely applied, for example, to recognize faces and emotions [3, 4], to represent 3D facial meshes for augmented reality applications [5], to recognize actions [6], and to estimate body and hand poses [7–10]. Future seems quite promising, limited though only by science fiction.

2 Human Pose Estimation

Operating a robot or a machine using the intelligence of a human requires the availability of adequate human–machine interfaces. Using communication techniques such as natural language or gestures allows robots to be utilized in new work environments and in various scenarios, thereby expanding the scope and application of teleoperation. For decades, human pose estimation has been a challenging task, involving analysis of sensor inputs to determine the posture of parts of the body or the entire human body.

Using specialized equipment to control electromechanical devices is very common in areas such as civil engineering, assembly lines, or in parts, manipulation applications. In special cases in which the number of degrees of freedom grows, the use of more sophisticated devices becomes mandatory (e.g., exoskeletons, electromyography sensors, 3D space-aware wearables, flock of birds, etc.).

Having the same goal as the previously developed sensors and motivated by the avoidance of using complex mechanical structures, the most recent line of research is focused on designing computer vision algorithms to be utilized as a system's input device. Computer vision in industry and robotic applications has made outstanding progress, giving place to applications of detection of human movements and their gestures to interpret signals or orders. Therefore, the human operators can avoid the need to wear a physical device over their body.

A variety of applications such as machine control, augmented reality, and virtual reality can be achieved using human hand pose estimation, which is an established for decades research problem in the computer vision and graphics fields [11–13]. As a result of its significance, numerous solutions have been proposed in the literature. One of the most common methods involves spotting 2D keypoints accurately [14].

The field of deep neural networks has recently undergone substantial advances, and however this topic remains a challenging one, which is still far from being completely resolved. The properties of the hand, such as its morphology, the occlusion effects of objects interacting with it, the variety of appearances due to clothing and jewelry, as well as the varied lighting conditions and different backgrounds, add additional complexity to the problem.

It is nonetheless true that unlike the body or the face of a person, hands are almost uniform in shape and do not exhibit any local characteristics. Moreover, there are myriads of arbitrary poses that can be achieved by the few ten joints in a hand. Due

to these factors, it becomes crucial to precisely estimate the pose of each hand by localizing more than 20 keypoints [16, 17] and using the hand as a control device.

2.1 Notable Approaches

In recent years, vision-based hand pose estimation has made significant progress. A large number of approaches use Convolutional Neural Networks as a basis, owing to their potent ability to extract features from a given input.

2.1.1 Two-Stage Pipelines

With CNNs, 2D body pose estimation can be successfully performed by determining whether or not joints are present in each pixel [18, 19]. In the proposed methods, also called Convolutional Pose Machines (CPMs), a CNN is enforced in order to generate a set of heat maps, each of which is estimated to have its maximum activation value in the pixel containing the corresponding keypoint. Figure 1 illustrates a typical CPM architecture for pose estimation.

However, the heat map generation procedure is applied iteratively upon the generated heat maps to refine the outcome. Additionally, the majority of hand pose estimation methods also utilize this same methodology [16, 17], which leads to computationally expensive networks and complex system architectures.

2.1.2 Single-Stage Pipelines

The objective of another line of research concerns mapping an image directly to the coordinates of keypoints on the plane or to a specific frame of reference for the estimation of 2D and 3D poses, respectively, known as holistic regression [21, 22]. In the abovementioned procedure, no intermediate representations are generated (pixel-wise classification) while retaining all of the ability to understand

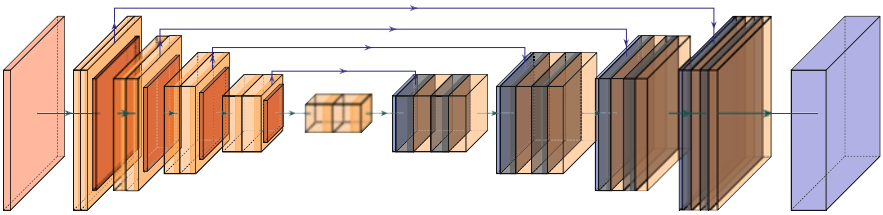


Fig. 1 A typical Convolutional Pose Machine building block. The output feature map represents the category of original image in pixel level. Pixels with higher activation values in each channel considered as detected joints

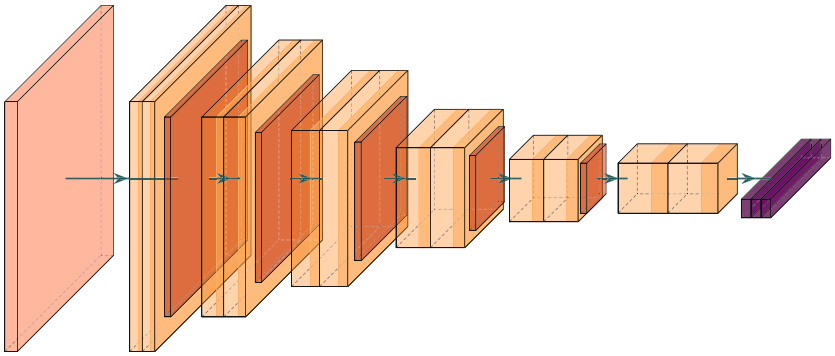


Fig. 2 A holistic regression architecture. The output of the network predicts the actual coordinates of each joint on the image plane

global constraints and correlations between keypoints. Yet, regression from highly disparate domains such as images and poses can be a particularly challenging task. Additionally, a simple holistic regression cannot generalize easily to combinations of local evidence such as the different poses of different fingers and is highly sensitive to hand locations and translational variance [23]. Figure 2 shows clearly the simplicity of the single-stage pipeline.

Despite these considerations, the proposed work is based on the holistic regression approach, validating its effectiveness when combined with a robust feature extraction algorithm and an appropriate anti-aliasing filter.

2.2 Methodology

Our objective in this section is to provide an overview of how we developed our concept and what elements we selected and modified to arrive at a compact and lightweight architecture able to estimate a hand's 2D keypoint coordinates from a single RGB image [15, 20, 24]. In order to address this challenge, we employ a feed-forward CNN architecture, which directly generates the coordinates in a single stage without intermediate supervision. Essentially, the network is divided into two parts, the *stem* and the remainder, for now referred to as the *tail*.

2.2.1 Proposed Architecture

Based upon the concept of *DenseNets* [15], the proposed architecture constructs a network where all subsequent layers receive additional input from all preceding ones, and each layer propagates its own feature maps to the subsequent layers

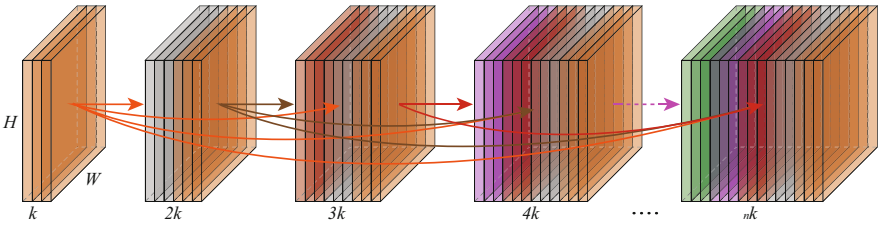


Fig. 3 Dense Block with growth rate k [15]

via channel-wise concatenation, as shown in Fig. 3, by leveraging the "collective knowledge" of all prior layers.

With the goal of keeping the number of parameters as low as possible, we have employed a highly effective feature extractor, viz the *inverted residual block* [24]. Using depthwise separable convolutional layers in place of standard ones reduces the computations by

$$k_f^2 \cdot d_o/(k^2 + d_o), \tag{1}$$

where k_f equals the kernel's size, and d_o equals the output depth size. The first convolutional layer expands the depth size by an e factor, while the last squeezes it by dividing the input's depth size by the same factor. Here, we have chosen $e = 4$.

2.2.2 Stem

Our *stem* design comprises a number of *dense blocks* that contain an inverted residual block in contrast to the original design. The authors of [15] found that architectures with concatenated skip connections retain more information since the concatenated connections permit subsequent layers to reuse intermediate representations, which results in an improved performance.

In regard to its nonlinearity, this block differs from the original in that we use the recently proposed *Mish* activation function [25]. Contrary to ReLU, *Mish* is non-monotonic and smooth, which is defined as

$$f(x) = x \cdot \tanh(\ln(1 + e^x)), \tag{2}$$

Mish exhibits better results than both *Swish* [26] and *ReLU* for classification tasks, as discussed in [25]. During extensive trials using both *Swish* and *ReLU*, we confirmed the above behavior for the regression task at hand.

2.2.3 Blur Pooling

Feature maps are typically subsampled between convolutional layers by using either a pooling or a striding operation. In [27], it is explicitly stated that, when a system combines convolution and subsampling operations, it lacks of translation invariance. When the translation is a multiple of each of the subsampling factors, the statement previously mentioned does not apply. Otherwise, subsampling generates aliases that detract from the quality of the output. As a result of this property, CNNs are also susceptible to error degradation, since small image transformations can adversely impact the accuracy of the network [28, 29]. As discussed in [30], a feature extractor function $F \in \mathbb{R}^{H \times W \times C}$ is shift-equivariant when shifting the input equally shifts the output, making shifting and feature extraction commutable:

$$Shift_{\Delta h, \Delta w}(F(X)) = F(Shift_{\Delta h, \Delta w}(X)) \qquad \forall(\Delta h, \Delta w). \tag{3}$$

In addition, an invariant representation is one in which shifting the inputs produces the same output regardless of the changed inputs:

$$F(X) = F(Shift_{\Delta h, \Delta w}(X)) \qquad \forall(\Delta h, \Delta w). \tag{4}$$

Shift-equivariance is broken by regular pooling methods. In order to mitigate this problem, we propose the adaptation of an anti-aliasing filter that is convolved with feature maps [30] and with stride 2 to reduce spatial resolution. The method provides the ability to choose between different sizes of kernels, producible by a box filter. An implementation of the anti-aliasing filter $Filt$ is shown below:

$$B_n[x] = \begin{cases} 1, \text{ for } 0 \leq x < n, \\ 0, \ otherwise, \end{cases} \tag{5}$$

$$Box_m = B_n * B_n, \tag{6}$$

$$Filt = Box_m \otimes Box_m, \tag{7}$$

where $\otimes$ denotes the outer product, $n \in \mathbb{N}^*$, $x \in \mathbb{Z}$, and $m = 2n - 1$. The utilized anti-aliasing filter $Filt$ $n = 2$.

2.2.4 Visual Attention

A brilliant aspect of the human brain is the fact that it is able to focus on various elements that are parts or details of its sensory input. This becomes possible, due to the already acquired knowledge and experience of the person from its environment, and in turn, from its expectations from it. Natural selection resulted in

the enhancement of human vision in several ways, firstly by creating space-variant distributions of the retinal receptive cells [57] and then by forcing the brain to create, maintain, and change the representation of reality, by ranking the importance of every component while scanning the visual scene [44, 45].

The visual attention that we have in the course of our daily lives is controlled both by cognitive factors (top-down), such as knowledge, expectations, and current goals, and by bottom-up factors that reflect sensory input. Another factor that affects attention is novelty and unexpectedness, which are caused by a combination of cognitive–sensorial influences. The research that has been conducted around the human visual system suggests the presence of an attention mechanism.

This hypothesis comes to the conclusion that biological vision, rather than compressing an image into a static representation, exploits an attention mechanism that allows for salient features to dynamically come to the forefront as needed. The intuition behind it can coherently be explained when our visual processing system tends to focus selectively on some parts of the field of view, while ignoring irrelevant or redundant information in a way that assists the perception [46].

In computational intelligence, the equivalent of the attention mechanism origins back in 1964, initially proposed by Nadaraya–Watson [47, 48]. Given a training data set of n samples comprising of features and their corresponding values $\{(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)\}$, a naive approach to estimate the value $\hat{y}$ for an arbitrary query x would simply to compute the average of target values of all training instances,

$$\hat{y} = \frac{1}{n} \sum_{i=0}^{n} y_i. \tag{8}$$

Instead, they introduced a more sophisticated approach, achieving superior results compared to the simple average, in which the estimator uses a weighted average, where weights correspond to relevance of the training instance to the query. The weighted version Eq. (8) is as follows:

$$\hat{y} = \frac{1}{n} \sum_{i=0}^{n} a(x, x_i) y_i, \tag{9}$$

where $a(x, x_i)$ is the weighting function that encodes the relevance of instance x_i to predict for x. The evolution of this idea has become the cornerstone of the state-of-the-art deep learning models for multiple use cases. In modern days, the first use of an attention model was introduced for Natural Language Processing tasks [49], which substantially improved the accuracy of Machine Translation.

The attention module as introduced by Bahdanau et al. [49] in its general form can be interpreted as a mapping of sequence of keys K to an attention distribution a according to query q. Thus, the attention distribution a emphasizes the keys which are relevant for the main task with respect to the query q. Usually, an additional input of values V is introduced, on which the attention distribution a is applied. Hence, a generalized form of the attention model A can be described as

$$A(q, K, V) = \sum_i p(a(k_i, q)) * u_i, \tag{10}$$

where K and V represent a set of key–value pairs and query q. To maintain the analogy of the original idea of the attention, one can look the instance x as the query and the training data points x_i as the keys and y_i as the values [50].

These improvements affected neural network-based computer vision tasks too. Visual attention has become the new de facto addition in the neural network architectures due to its ability to capture long-range dependencies between parts of the image. A number of variations have been introduced for different computer vision tasks, such as image classification [51], object detection [52], image generation [53], referring image segmentation [54], and a standard backbone for visual tasks [55, 56].

2.2.5 Attention Augmented Inverted Bottleneck Block

By using attention mechanisms, a CNN can be trained to focus more on relevant information than on irrelevant information. A prominent idea in the research field of deep learning is that of visual attention. Attention mechanisms and specifically self-attention serve as powerful tools for the processing of not only text but also images. A wide range of visual attention mechanisms have been proposed in order to improve the performance efficiency of CNNs while still utilizing a similar number of parameters [20, 31].

The general idea is that given a query and a set of key elements, the attention mechanism aggregates w.r.t. the trainable parameters, the resemblance between key–query pairs. Multiple attention functions provide the ability to attend multiple representation subspaces and spatial positions. The output of each head is then aggregated linearly with learnable weights [32]. Taking inspiration from a design proposed in [32], we enclose a standard residual block within a self-attention mechanism. To be more specific, we apply an Attention Augmented Convolutional layer [20], which embeds an inverted bottleneck block by combining its output with that of the Depthwise Separable Convolutional layer, as shown in Fig. 4.

The self-attention mechanism achieves better results when combined with convolutional layers [20]. Essentially, the self-attention module uses three sets of learnable parameters W^Q, W^K, W^V, where Q, K, V refer to *Query*, *Key*, and *Value*, respectively. According to [31], an input tensor $T \in \mathbb{R}^{H \times W \times F_{in}}$ is flattened to a matrix $X \in \mathbb{R}^{HW \times F_{in}}$ and then forwarded to the transformer attention architecture. In light of the fact that multiple applications of self-attention have been found to be beneficial, Eq. (11) is applied once for each attention head, producing $O_{[1,.,h]}$ outputs, where $h \in \mathbb{N}^*$.

$$O_h = Softmax\left(\frac{(XW^Q)(XW^K)^T}{\sqrt{d_k^h}}\right)(XW^V). \tag{11}$$

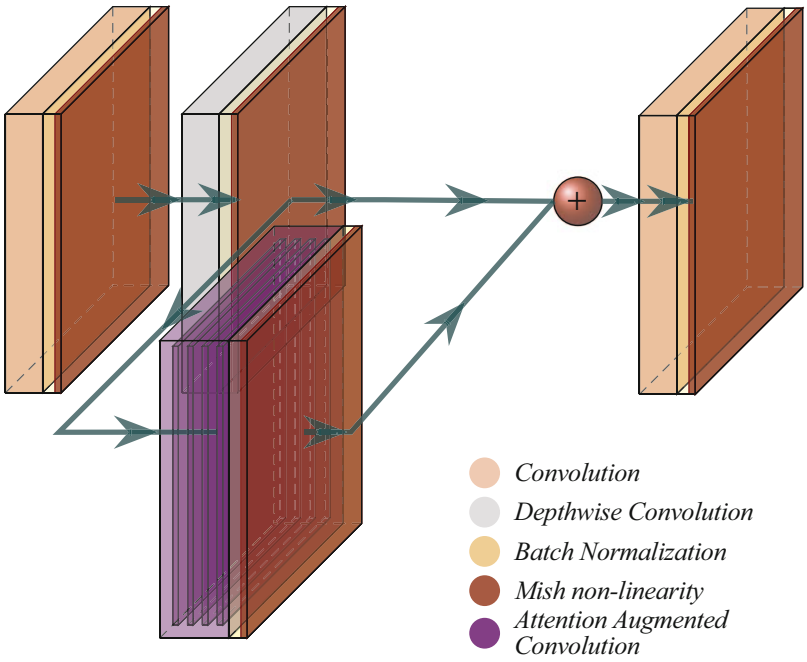


Fig. 4 Our proposed Attention Augmented Inverted Bottleneck Layer

$W^Q, W^K \in \mathbb{R}^{F_{in} \times d_k^h}$, and $W^V \in \mathbb{R}^{F_{in} \times d_v^h}$. A multihead attention mechanism is formed by concatenating the output of each head with that of the remaining heads.

$$MHA(X) = Concat\left[O_1, .., O_h\right] W^O, \tag{12}$$

where $W^O \in \mathbb{R}^{d_v \times d_x}$ is a trainable matrix which linearly transforms the aggregated output of each head. The *Values'* depth is d_v, *Queries'* depth d_k, and the number of heads N_h.

Self-attention has an inherent characteristic that it is equivariant to reordering of inputs. As a result, any spatial information is not maintained, which eliminates the possibility of performing vision tasks due to the structured nature of the images.

The transformer is incapable of making use of the sequence's order, without providing any information about each token's position in the sentence, as was originally discussed for NLP tasks [31]. In order to overcome this inability, the input embeddings were enhanced by adding extra information, in the form of *positional encodings* followed by a self-attention mechanism with *relative position embeddings*, enabling translation equivariance, while preventing permutation equivariance [33]. Two matrices are used to represent the relative position of each *Query* and *Key* pixel for both width and height, containing a relative embedding for each pair of pixels:

$$l_{i,j} = \frac{q_i^T}{\sqrt{d_k^h}} \left(k_j + r^W_{j_x - i_x} + r^H_{j_y - i_y} \right), \tag{13}$$

where q_i and k_j are the *Query* and *Key* vectors for pixels i, j, while $r^W_{j_x - i_x}$ and $r^H_{j_y - i_y}$ are learned embeddings for relative width and height, respectively.

Each attention head enhanced by *relative position embeddings* becomes

$$O_h = Softmax \left(\frac{(XW^Q)(XW^K)^T + S^{rel}_H + S^{rel}_W}{\sqrt{d_k^h}} \right) (XW^V), \tag{14}$$

where S^{rel}_H, $S^{rel}_H \in \mathbb{R}^{HW \times HW}$ are matrices of *relative positional embeddings* for every pixel pair.

As previously mentioned, visual attention of this type is capable of attending feature subspaces and spatial positions simultaneously, both thanks to the attention mechanism that introduces additional feature maps and the convolution operator. Concatenation of the convolutional operator and Multiheaded Attention's output is the final step in the Attention's Augmented Convolution integration.

$$AAC(X) = concat\left[Conv(X), MHA(X)\right]. \tag{15}$$

Denoted as $u = \frac{d_v}{F_{out}}$ is the ratio between attention depth size and the output depth size, while $\kappa = \frac{d_k}{F_{out}}$ is the ratio of key depth over the output depth.

In order to implement the Attention Augmented Inverted Bottleneck Layer for the *tail* of the network, we implement recurrently *dense blocks*, in a similar fashion to the *stem*.

2.2.6 Subsampling

Feature maps are downsampled between *dense blocks* using a *transition layer*, which consists of a pointwise convolutional layer to reduce depth, a Blur Pooling filter with stride 2, as well as batch normalization.

2.3 Training Settings

During training, we used *Cyclical Learning Rate* [34] with triangular policy and the *Stochastic Gradient Descent* for optimizer. Among the best hyper-parameters are the minimum learning rate of 10^{-4} and the maximum of 10^{-1}, as well as the $stepsize = 6$. The *batch size* equals 256, and the training was executed using *Tensor Processing Units* (TPUs) on the cloud, provided by Google. Last but not least, a mixed-precision

Table 1 Network's architecture. The growth rate is k=10

Layers	Output size	Architecture
Dense block (1)	224×224	[Inverted Bottleneck Layer] ×8
Transition layer	112 × 112	1 × 1 conv × 64 3 × 3 BlurPool, s2
Dense block (2)	112 × 112	[Inverted Bottleneck Layer] ×8
Transition layer	56 × 56	1 × 1 conv × 64 3 × 3 BlurPool, s2
Dense block (3)	56 × 56	[Attention Augmented Inverted Bottleneck Layer] ×6
Transition layer	28 × 28	1 × 1 conv × 64 3 × 3 BlurPool, s2
Dense block (4)	28 × 28	[Attention Augmented Inverted Bottleneck Layer] ×8
Transition layer	14 × 14	1 × 1 conv × 64 3 × 3 BlurPool, s2
Dense block (5)	14 × 14	[Attention Augmented Inverted Bottleneck Layer] ×10
Transition layer	7 × 7	1 × 1 conv × 64 3 × 3 BlurPool, s2
Dense block (6)	7 × 7	[Attention Augmented Inverted Bottleneck Layer] ×12
Transition layer	4 × 4	1 × 1 conv × 128 3 × 3 BlurPool, s2
Dense block (7)	4 × 4	[Attention Augmented Inverted Bottleneck Layer] ×14
Transition layer	2 × 2	1 × 1 conv × 128 3 × 3 BlurPool, s2
Dense block (8)	2 × 2	[Attention Augmented Inverted Bottleneck Layer] ×32
AA bottleneck	2 × 2	[Attention Augmented Inverted Bottleneck Layer] ×1
	1 × 1	2 × 2 average pooling, s2
	1 × 1 × 42	1 × 1 conv × 42

training policy was implemented by exploiting both 16-bit (bfloat16) and 32-bit (float32) floating point types [35]. In effect, this practice resulted in a memory gain, having a knock-on effect of allowing batches to be larger, models to be smaller, and execution times to be faster. In Table 1, we give a detailed description of the model's architecture, consisting of just *1.9 Million* parameters and *7.1 Million* FLOPs in terms of computational demand, which is built using the TensorFlow library [36].

3 Evaluation

A number of contemporary datasets have been used to evaluate our 2D pose estimation method in comparison with current state-of-the-art methods. Our method outperforms larger, more complex, more computationally expensive deep learning architectures with its lightweight and straightforward nature. Several experiments

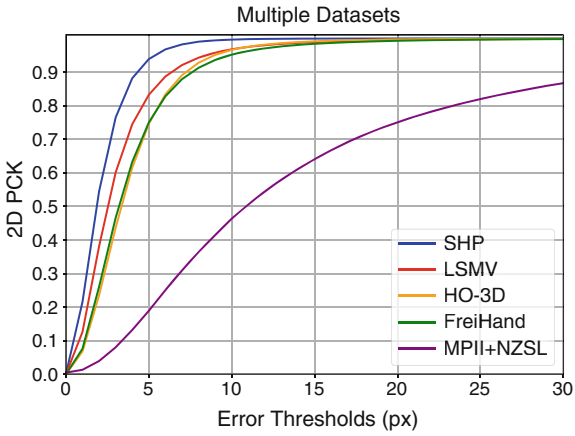


Fig. 5 Evaluation with the best setting architecture performed on the listed datasets

were conducted on five different datasets, the characteristics of which are described below. A qualitative evaluation compared to other methods over different datasets is shown in Figs. 5 and 6.

3.1 Datasets

PANOPTIC [37] is an accurate large-scale human posture dataset with many instances of occluded subjects. Our training set was composed of three dataset sessions: *office1, tools1*, and *cello3*. According to the literature [16], additional training images encompassing MPII+NZSL were included [37], resulting in a total of 165000 training images. The evaluation was conducted on the MPII+NZSL testing set.

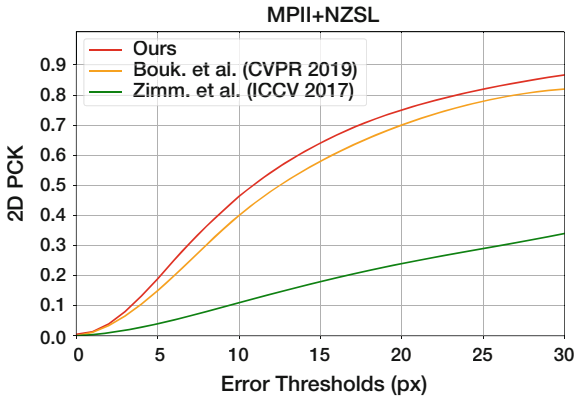
The **HO-3D** [38] is a recently released markerless dataset that contains 10505 images. The images in the dataset were flipped and rotated by 0–90–180 degrees.

The **FreiHAND** [39] provides a multiview hands dataset, recorded in front of a green screen and augmented with artificial background, resulting in a total of 130240 image instances.

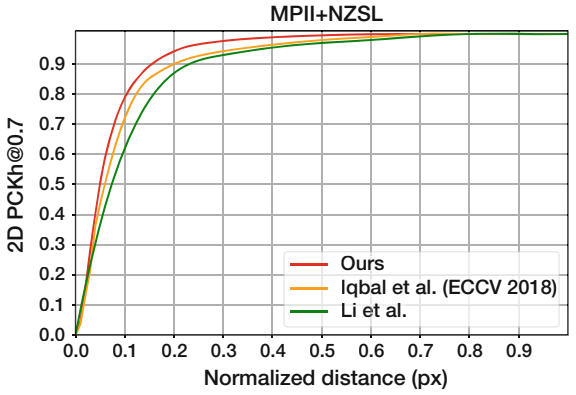
The **LSMV** [40] provides images of hands from multiple points of view. The total of frames is 80000.

The **SHP** [41] provides 3D pose annotations of a person's hand, performing various gestures in 18000 frames.

The datasets were evaluated separately and split according to a rule of 80%–10%–10% for training, validation, and testing. We cropped all images to a resolution of 224 × 224. The datasets selected cover a wide variety of lighting conditions, environments (e.g., green room, natural landscape, etc.), wearables (e.g., gloves), and hands that hold various objects.



(a)



(b)

Fig. 6 Performance Evaluation. (**a**) PCK curves on MPII+NZSL testing set. (**b**) PCKh curves on MPII+NZSL testing set

3.2 Ablation Studies

We would like to elaborate upon the architectural elements that we have incorporated into our system in this subsection. We evaluate the proposed method under different settings by (i) simply excluding the attention augmented convolution module, (ii) experimenting with different pooling methods, and (iii) varying the activation function, as presented in Table 3. We found that adding the attention augmented convolution module resulted in a slight increase in the overall FLOPs to *7.1 Million*. All twelve configurations have been tested and trained on the same four datasets, namely PANOPTIC [37], FreiHAND [39], LSMV [40], and SHP [41].

The results for each combination are shown in Fig. 8. *Architecture 1* demonstrates its superiority over all others in every radar chart since it occupies the smallest area overall. During evaluation only, we performed random shifts from vertical to horizontal in either axis with an interval of 20 pixels, in order to prove that the

Table 2 Quantitative performance results

	AUC	EPE(px)	
		Mean	Median
MPII+NZSL Dataset			
Zimm. et al. (ICCV 2017)	0.17	59.4	–
Bouk. et al. (CVPR 2019)	0.50	18.95	–
Ours	**0.55**	**16.1**	**11**
LSMV Dataset			
Gomez-Donoso et al.	–	10	–
Li et al.	–	8	–
Ours	0.89	**3.3**	2.5
Stereo Hand Pose Dataset			
Zimm et al. (ICCV 2017)	0.81	5	5.5
Ours	**0.92**	**2.2**	**1.8**
HO-3D Dataset			
Ours	0.87	3.9	3.3
FreiHAND Dataset			
Ours	0.87	4	3.1

Table 3 Different architectures utilized in the ablation study

	Attention module	Pooling method	Activation function
Architecture 1	✓	Blur	Mish
Architecture 2	–	Blur	Mish
Architecture 3	–	Average	Mish
Architecture 4	✓	Average	Mish
Architecture 5	✓	Blur	ReLU
Architecture 6	–	Average	ReLU
Architecture 7	✓	Average	ReLU
Architecture 8	–	Blur	ReLU
Architecture 9	–	Max	Mish
Architecture 10	✓	Max	Mish
Architecture 11	–	Max	ReLU
Architecture 12	✓	Max	ReLU

proposed method is robust to input's translations. As one can see from the results presented in Fig. 9, every tested architecture shows a degradation in accuracy due to the input's translation. However, *Architecture 1* exhibits a modest decrease, and therefore its use in our final architecture is justified.

It is thought that this behavior is caused by the *Blur Pooling layer*, which functions as an anti-aliasing filter during subsampling. Consequently, the network can propagate as much information as possible to the deeper layers, resulting in improved regression results. It is also worth noting that, as opposed to the typical nonlinearity of *ReLU*, the combination of *Blur Pooling* and *Mish* activation function leads to the best overall system performance.

3.3 Comparative Results

As shown in Fig. 5, we compare our results with those of other state-of-the-art methods, showing that our method performed better than the others, following the protocol proposed in [37]. Figure 6 illustrates the percentage of correct keypoints for different absolute thresholds and normalized thresholds and compares them with other techniques. We show the performance of our method on multiple datasets in Fig. 5. The abovementioned results are also summarized in Table 2. In Fig. 7, we present representative examples of the feature maps obtained from our *Attention Augmented Inverted Bottleneck Block* for a sample of images during various stages of the network structure. Despite the fact that the network was not specifically

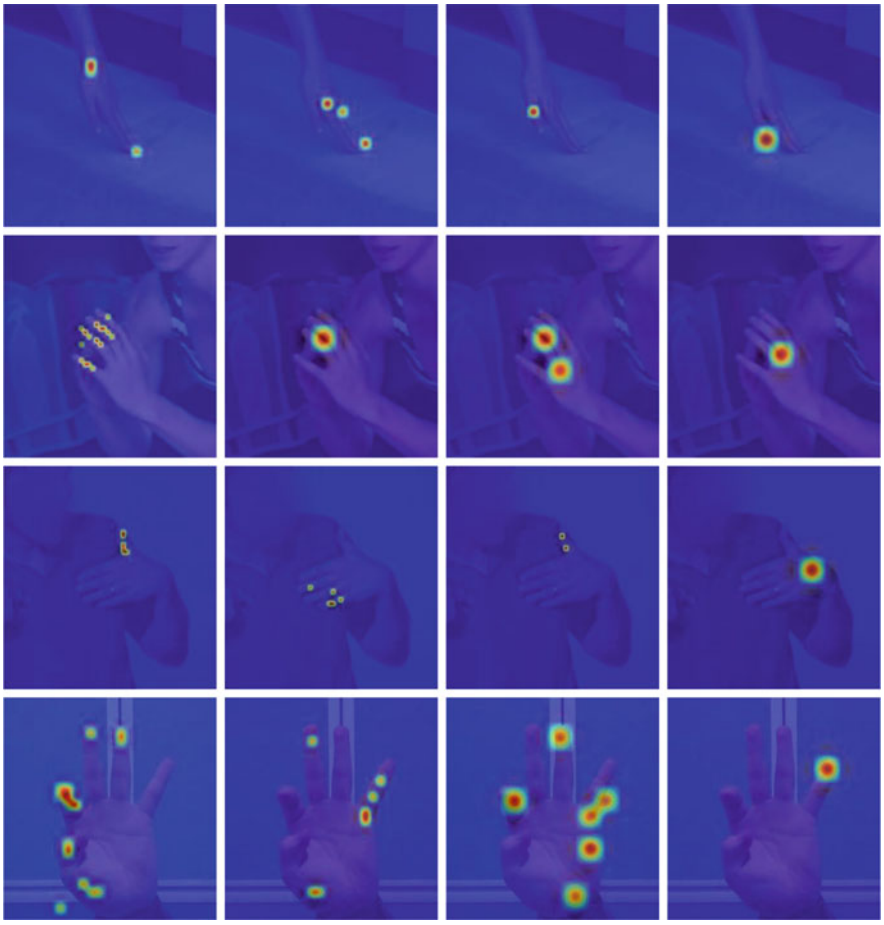


Fig. 7 Representative feature maps of our proposed Attention Augmented Inverted Bottleneck Layer

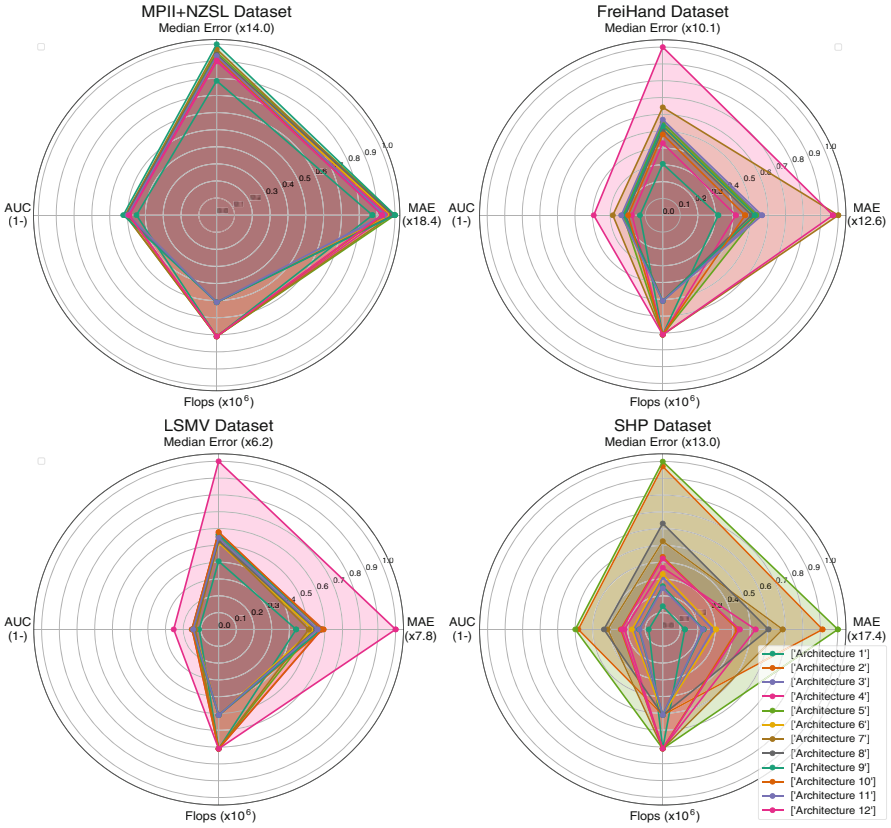


Fig. 8 Results under different architecture configurations

trained for the generation of a heat map, the effectiveness of the specific block in detecting the points of interest can easily be seen, in this case the hand's keypoints. Furthermore, we present some qualitative results of our method in Fig. 10 for different dataset instances.

4 Conclusions

Unlike the majority of pose estimation methods, we present an alternative approach which is simple, yet computationally efficient. We have demonstrated that our proposed single-stage end-to-end CNN model produces competitive results by

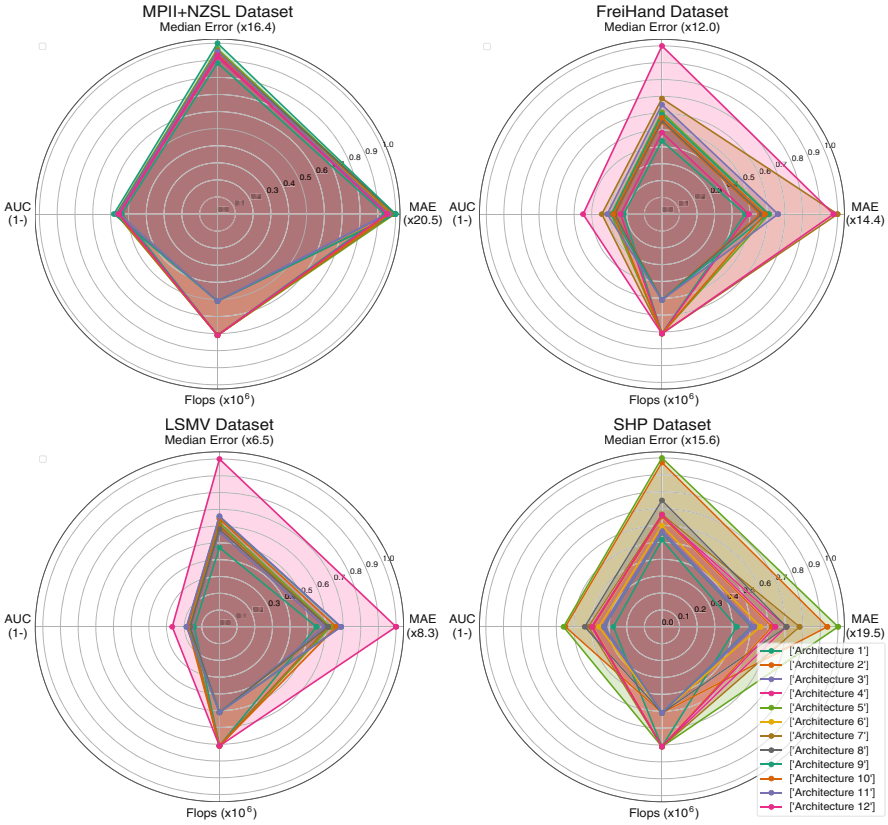


Fig. 9 Results under different architecture configurations with randomly shifted input

predicting joint coordinates directly with just *1.9M* parameters and a model's size of 11 Mbytes. Consequently, our system can be deployed on low-processing-power devices while meeting the operational guidelines for mobile devices.

The method's success can largely be attributed to the proposed *Attention Augmented Inverted Bottleneck Block* being effective in understanding global constraints and correlations between keypoints, as well as to the architecture's ability to share a "collective knowledge" among successive layers. Finally, our proposed approach could be adapted to other tasks as well, such as 3D hand, body, or object pose estimation, leveraging depth estimation methods [42, 43], human body pose estimation or classification, which we intend to investigate further in future work.

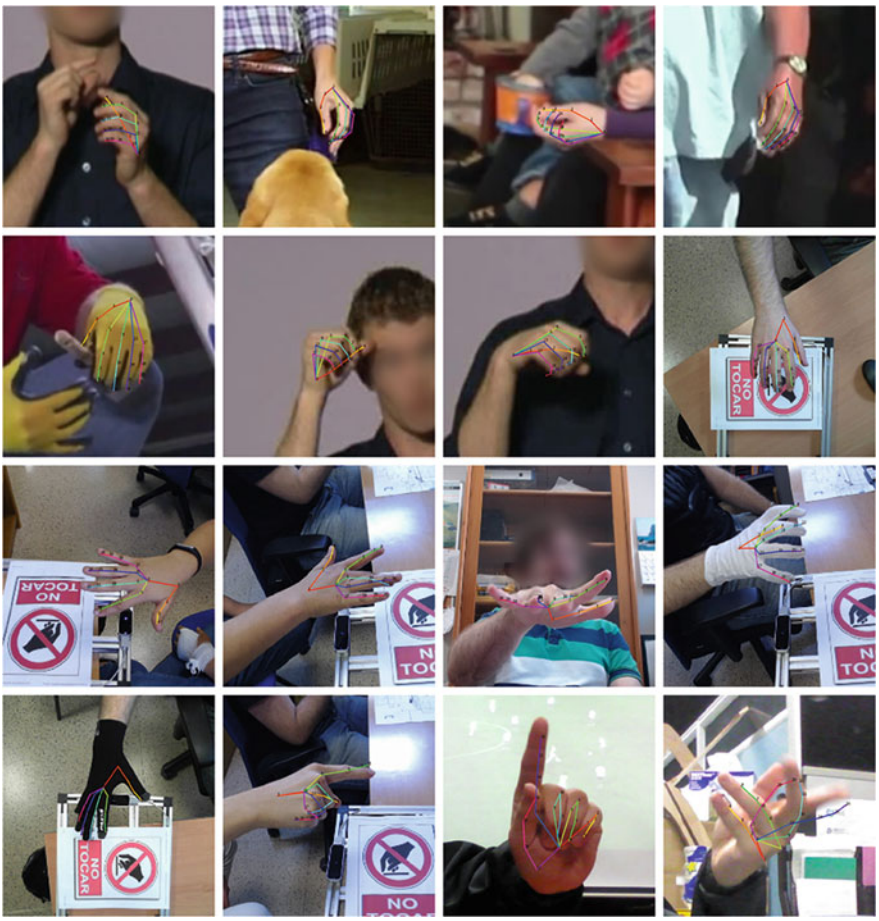


Fig. 10 Our 2D hand pose estimations on different testing sets

References

1. J. Złotowski, D. Proudfoot, K. Yogeeswaran, and C. Bartneck, "Anthropomorphism: opportunities and challenges in human–robot interaction," *International Journal of Social Robotics*, vol. 7, no. 3, pp. 347–360, 2015.
2. V. Kepuska and G. Bohouta, "Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home)," in *Proc. IEEE Computing and Communication Workshop and Conference*, 2018, pp. 99–103.
3. I. Kansizoglou, L. Bampis, and A. Gasteratos, "An active learning paradigm for online audio-visual emotion recognition," *IEEE Transactions on Affective Computing*, 2019.
4. N. Efremova, M. Patkin, and D. Sokolov, "Face and emotion recognition with neural networks on mobile devices: Practical implementation on different platforms," in *Proc. IEEE International Conference on Automatic Face & Gesture Recognition*, 2019, pp. 1–5.

5. Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, "Real-time facial surface geometry from monocular video on mobile GPUs," *arXiv preprint arXiv:1907.06724*, 2019.
6. R. Vemulapalli, F. Arrate, and R. Chellappa, "Human action recognition by representing 3d skeletons as points in a lie group," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 588–595.
7. Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: realtime multi-person 2d pose estimation using part affinity fields," *arXiv preprint arXiv:1812.08008*, 2018.
8. F. Zhang, X. Zhu, and M. Ye, "Fast human pose estimation," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3517–3526.
9. X. Chen, G. Wang, H. Guo, and C. Zhang, "Pose guided structured region ensemble network for cascaded hand pose estimation," *Neurocomputing*, 2019.
10. S. Yuan, G. Garcia-Hernando, B. Stenger, G. Moon, J. Yong Chang, K. Mu Lee, P. Molchanov, J. Kautz, S. Honari, L. Ge *et al.*, "Depth-based 3d hand pose estimation: From current achievements to future goals," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2636–2645.
11. Y. Jang, S.-T. Noh, H. J. Chang, T.-K. Kim, and W. Woo, "3D finger cape: Clicking action and position estimation under self-occlusions in egocentric viewpoint," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 4, pp. 501–510, 2015.
12. T. Piumsomboon, A. Clark, M. Billinghurst, and A. Cockburn, "User-defined gestures for augmented reality," in *Proc. IFIP Conference on Human-Computer Interaction*, 2013, pp. 282–299.
13. B. Fang, D. Guo, F. Sun, H. Liu, and Y. Wu, "A robotic hand-arm teleoperation system using human arm/hand with a novel data glove," in *Proc. IEEE International Conference on Robotics and Biomimetics*, 2015, pp. 2483–2488.
14. J. M. Rehg and T. Kanade, "Visual tracking of high DOF articulated structures: an application to human hand tracking," in *Proc. European Conference on Computer Vision*, 1994, pp. 35–46.
15. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
16. A. Boukhayma, R. d. Bem, and P. H. Torr, "3d hand shape and pose from images in the wild," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 843–10 852.
17. U. Iqbal, P. Molchanov, T. Breuel Juergen Gall, and J. Kautz, "Hand pose estimation via latent 2.5 d heatmap regression," in *Proc. European Conference on Computer Vision*, 2018, pp. 118–134.
18. A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proc. European Conference on Computer Vision*, 2016, pp. 483–499.
19. S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4724–4732.
20. I. Bello, B. Zoph, A. Vaswani, J. Shlens, and Q. V. Le, "Attention augmented convolutional networks," *arXiv preprint arXiv:1904.09925*, 2019.
21. B. Tekin, A. Rozantsev, V. Lepetit, and P. Fua, "Direct prediction of 3d body poses from motion compensated sequences," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 991–1000.
22. S. Li and A. B. Chan, "3d human pose estimation from monocular images with deep convolutional neural network," in *Proc. Asian Conference on Computer Vision*, 2014, pp. 332–347.
23. C. Wan, T. Probst, L. Van Gool, and A. Yao, "Dense 3D regression for hand pose estimation," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5147–5156.
24. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.
25. D. Misra, "Mish: A self-regularized non-monotonic neural activation function," *arXiv preprint arXiv:1908.08681*, 2019.

26. P. Ramachandran, B. Zoph, and Q. V. Le, "Swish: a self-gated activation function," *arXiv preprint arXiv:1710.05941*, vol. 7, 2017.
27. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, 1992.
28. L. Engstrom, B. Tran, D. Tsipras, L. Schmidt, and A. Madry, "Exploring the landscape of spatial robustness," in *Proc. International Conference on Machine Learning*, 2019, pp. 1802–1811.
29. A. Azulay and Y. Weiss, "Why do deep convolutional networks generalize so poorly to small image transformations?" *arXiv preprint arXiv:1805.12177*, 2018.
30. R. Zhang, "Making convolutional networks shift-invariant again," *arXiv preprint arXiv:1904.11486*, 2019.
31. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
32. X. Zhu, D. Cheng, Z. Zhang, S. Lin, and J. Dai, "An empirical study of spatial attention mechanisms in deep networks," *arXiv preprint arXiv:1904.05873*, 2019.
33. P. Shaw, J. Uszkoreit, and A. Vaswani, "Self-attention with relative position representations," *arXiv preprint arXiv:1803.02155*, 2018.
34. L. N. Smith, "Cyclical learning rates for training neural networks," in *Proc. IEEE Winter Conference on Applications of Computer Vision*, 2017, pp. 464–472.
35. P. Micikevicius, S. Narang, J. Alben, G. Diamos, E. Elsen, D. Garcia, B. Ginsburg, M. Houston, O. Kuchaiev, G. Venkatesh *et al.*, "Mixed precision training," *arXiv preprint arXiv:1710.03740*, 2017.
36. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "TensorFlow: A system for large-scale machine learning," in *Proc. {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 2016, pp. 265–283.
37. T. Simon, H. Joo, and Y. Sheikh, "Hand keypoint detection in single images using multiview bootstrapping," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
38. S. Hampali, M. Oberweger, M. Rad, and V. Lepetit, "HO-3D: A multi-user, multi-object dataset for joint 3D hand-object pose estimation," *arXiv preprint arXiv:1907.01481*, 2019.
39. C. Zimmermann, D. Ceylan, J. Yang, B. Russell, M. Argus, and T. Brox, "FreiHAND: A dataset for markerless capture of hand pose and shape from single RGB images," in *Proc. IEEE International Conference on Computer Vision*, 2019, pp. 813–822.
40. F. Gomez-Donoso, S. Orts-Escolano, and M. Cazorla, "Large-scale multiview 3D hand pose dataset," *Image and Vision Computing*, vol. 81, pp. 25–33, 2019.
41. J. Zhang, J. Jiao, M. Chen, L. Qu, X. Xu, and Q. Yang, "3D hand pose tracking and estimation using stereo matching," *arXiv preprint arXiv:1610.07214*, 2016.
42. L. R. Ramírez-Hernández, J. C. Rodríguez-Quiñonez, M. J. Castro-Toscano, D. Hernández-Balbuena, W. Flores-Fuentes, R. Rascón-Carmona, L. Lindner, and O. Sergiyenko, "Improve three-dimensional point localization accuracy in stereo vision systems using a novel camera calibration method," *International Journal of Advanced Robotic Systems*, vol. 17, no. 1, p. 1729881419896717, 2020.
43. Y. He, W. Sun, H. Huang, J. Liu, H. Fan, and J. Sun, "PVN3D: A deep point-wise 3D keypoints voting network for 6dof pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 632–11 641.
44. R. A. Rensink, "The dynamic representation of scenes," *Visual cognition*, vol. 7, no. 1–3, pp. 17–42, 2000.
45. M. Corbetta and G. L. Shulman, "Control of goal-directed and stimulus-driven attention in the brain," *Nature reviews neuroscience*, vol. 3, no. 3, pp. 201–215, 2002.
46. K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in *International conference on machine learning*. PMLR, 2015, pp. 2048–2057.

47. E. A. Nadaraya, "On estimating regression," *Theory of Probability & Its Applications*, vol. 9, no. 1, pp. 141–142, 1964.
48. G. S. Watson, "Smooth regression analysis," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 359–372, 1964.
49. D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
50. S. Chaudhari, V. Mithal, G. Polatkan, and R. Ramanath, "An attentive survey of attention models," *arXiv preprint arXiv:1904.02874*, 2019.
51. H. Zhao, J. Jia, and V. Koltun, "Exploring self-attention for image recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10 076–10 085.
52. P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-alone self-attention in vision models," *arXiv preprint arXiv:1906.05909*, 2019.
53. H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International conference on machine learning*. PMLR, 2019, pp. 7354–7363.
54. D.-J. Chen, S. Jia, Y.-C. Lo, H.-T. Chen, and T.-L. Liu, "See-through-text grouping for referring image segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 7454–7463.
55. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
56. B. Wu, C. Xu, X. Dai, A. Wan, P. Zhang, Z. Yan, M. Tomizuka, J. Gonzalez, K. Keutzer, and P. Vajda, "Visual transformers: Token-based image representation and processing for computer vision," *arXiv preprint arXiv:2006.03677*, 2020.
57. G. Sandini and V. Tagliasco, "An anthropomorphic retina-like structure for scene analysis," *Computer Graphics and Image Processing*, vol. 14, no. 4, pp. 365–372, 1980.

Visual-Inertial Navigation Systems and Technologies

Jorge Alejandro Valdez-Rodríguez, Julio César Rodríguez-Quiñonez, Wendy Flores-Fuentes, Luis Roberto Ramírez-Hernández, Gabriel Trujillo-Hernández, Oscar Real-Moreno, Moisés J. Castro-Toscano, Jesús Elías Miranda-Vega, and Paolo Mercorelli

Abbreviations

CCW	Counterclockwise
CW	Clockwise
DCM	Direction cosine matrix
FOV	Field of view
GNSS	Global navigation satellite system
IMU	Inertial measurement unit
INS	Inertial navigation system
LSS	Laser scanner system
NED	North, east, down
NEU	North, east, up
OSV	Omnidirectional stereo video
OVINS	Omnidirectional visual-inertial navigation system
RMIS	Robot-assisted minimally invasive surgery

J. A. Valdez-Rodríguez · J. C. Rodríguez-Quiñonez (✉) · W. Flores-Fuentes
L. R. Ramírez-Hernández · G. Trujillo-Hernández · O. Real-Moreno · M. J. Castro-Toscano
Facultad de Ingeniería Mexicali, Universidad Autónoma de Baja California, Mexicali, Baja California, México
e-mail: valdez.jorge71@uabc.edu.mx; julio.rodriguez81@uabc.edu.mx; flores.wendy@uabc.edu.mx; luis.ramirez16@uabc.edu.mx; gabriel.trujillo@uabc.edu.mx; oreal@uabc.edu.mx; moises.castro@uabc.edu.mx

J. E. Miranda-Vega
Tecnológico Nacional de México/IT de Mexicali, Mexicali, Baja California, México
e-mail: elias.miranda@itmexicali.edu.mx

P. Mercorelli
Leuphana University of Lueneburg, Lueneburg, Germany
e-mail: paolo.mercorelli@leuphana.de

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1_6

SNR	Surgical navigation robot
SVS	Stereoscopic vision system
VINS	Visual-inertial navigation system
VIO	Visual-inertial odometry

1 Introduction

Navigation is an ancient activity of the human history, originated from the necessity of travel from a region to another in the search for a better environment, better resources, or better opportunities [57]. It is a task that not everyone is able to achieve; even the most experienced person can get lost in a trip and never arrive at the desired location. Therefore, a diversity of tools and methods are developed as an aid for travelers; those elements are known in navigation as references.

The use of references as support in navigation could be interpreted as an external object used for orientation. Landscapes are some of the oldest and effective references, but as the task becomes more complicated, the instrumentation used in navigation grows in precision and complexity.

However, navigation is not exclusive to travel; in the present days, a robot performing the specific task of placing objects in a diversity of defined points is part of the navigation problem. The possibilities of following a track, evading objects, and mapping a room are elements of navigation; they require a robot which has the capability of orientating in unknown environments.

For example, a pipeline inspection gauge is a system which has the capacity of performing navigation to inspect a gas or oil pipeline to locate and detect critical deformations [6]. The system propelled possesses a navigation system conformed of inertial sensors and/or GPS signals, allowing to track and enlist the places where a failure exists. Other tasks demand low error orientation as a result of the consequences in a poor navigation. For instance, the field of medicine requires manipulators capable of performing meticulous operation procedures in humans [61]. Teleoperated robot-assisted minimally invasive surgery (RMIS) is more common and an important part of medical surgery. RMIS systems require precision in their movements, sometimes to compensate the inexperienced movements of novice surgeons or simply to balance between high clinical importance and technical complexity.

Different resources are implemented as reference for INS, although a popular and continuously growing solution is the addition of visual references as cameras and laser scanners [46, 71]. Vehicles, such as cars, planes, and teleoperated RMIS, are technologies taking advantage of INS with visual sensors. The chapter presented discusses how INS are aided by visual references and the benefits gained in modern technologies.

2 VINS

Navigation is the science of maneuvering from one point to another using references to know the current position [8]. Tools, as maps or compasses, are commonly used when someone has in mind the word navigation, but more sophisticated devices are applied nowadays and the integration of two or more forms a "navigation system." Popular gadgets as cameras, instrumentation as accelerometers and gyroscopes, or systems as laser scanners are elements included in navigation systems.

The combination of two or more navigation references with navigation techniques, physics, and mathematical analysis receives the name of "navigation system" [42]. Navigation references are classified according to the coordinate reference frame, where devices with a fixed origin are known as "absolute references" [81]; otherwise, instruments with a relative coordinate reference frame where the position and attitude require to be constantly calculated over time are called "inertial references" [20].

A widely implemented instrument in inertial navigation is the inertial measurement unit (IMU) which is conformed by inertial instruments as accelerometers and gyroscopes. The combination of IMUs and mathematical navigation calculus is known as inertial navigation system (INS) [11] Fig. 1.

As a result of the mathematical integration required to obtain position and attitude, inertial sensors have an inherent bias that reduces precision in the navigation calculation. A common effect in INS is the presence of "drift," a deviation in the estimated position and orientation. Drift is an accumulative multifactorial error which predicts a different location from the actual body position as time elapses.

In order to reduce the drift error, a variety of solutions are recommended in literature to assist INS. Kalman filters help to improve the computational efficiency and diminish the error in navigation applying two weighting factors to compute a new estimation: the previous estimation in accordance with the known equations of motion and the obtained measurement from the IMU [16, 76]. The new estimation calculated by the Kalman filter increments the accuracy but is not capable of reducing all the drift error of an inertial sensor and computational miscalculations. It is possible to enhance the data accuracy of Kalman filter by complementing it with a different type of filters to correct computation.

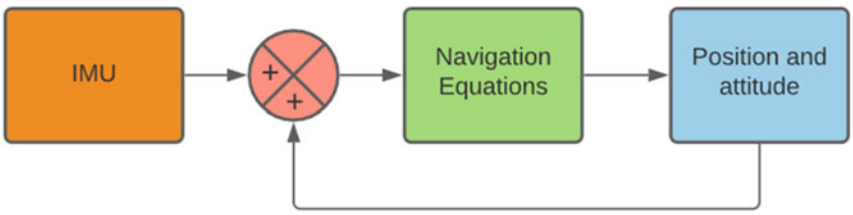


Fig. 1 Inertial navigation system

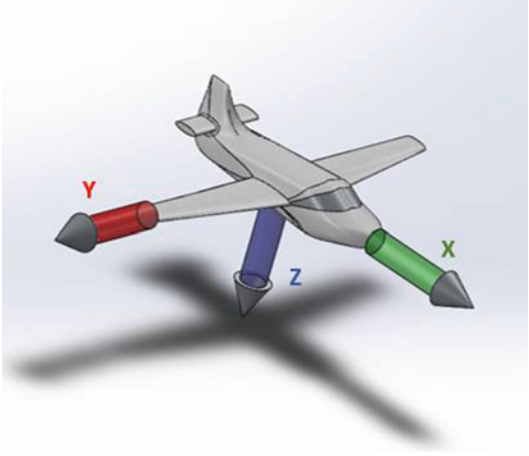
Another approach to reduce drift error in INS is the implementation of external navigation references such as GNSS (global navigation satellite systems), magnetometers (electronic compass), and visual systems as cameras to the INS. Visual sensors provide information about the environment where the body is located; however, they are sensitive to illumination conditions and motions [60]. An INS complemented with a visual system is called visual-inertial navigation system (VINS); the structure is an INS combined with one or more visual references as cameras, visual odometry [25], or laser scanner techniques.

The used approach in image-based navigation to complement INS is supplying and processing images using visual systems to provide accurate data of the surrounding environment or navigation object [59]. In addition, image-based navigation emulates the sense of orientation and navigation in human beings, allowing to determine object attitude and position in addition to object recognition [24].

In navigation is required a coordinate reference frame to express the position of a point in relation to some reference [43]. A coordinate reference frame is a Cartesian, right-handed axis set defined by a reference. Objects, the point of view, the Earth, and sensors are examples of references adopted to create a reference frame and with the aid of mathematical transformations are possible to translate from one coordinate reference frame to another.

One example of coordinate reference frame is the body frame. The body frame is the coordinate reference frame related to the vehicle or navigation object. The axes are related to the direction of the movements of the body, where **X** is the forward direction, **Y** is the right direction, and **Z** is related to the gravity pointing in the down direction (Fig. 2). In navigation, the body frame is assumed to have the axes in the same direction as the inertial frame to align the inertial sensors with the navigation vehicle. There is a wide variety of coordinate reference frames in navigation and

Fig. 2 Body frame



some of the related visual systems are going to be discussed further in the present chapter.

The translation from a coordinate reference frame to another is commonly solved through two different rotations in a plane, using methodologies such as the DCM and quaternion.

DCM is applied in a three-dimensional space R^3 where rotations are given through a coordinate angle called Euler angle [26, 62]. The rotations follow the right-handed coordinate frame rule where every rotation is defined, e.g., a CW rotation as positive and CCW rotation as negative.

A rotation in the Z axis is called yaw and the Euler angle is represented with ψ letter. The C_ψ rotation matrix is represented as follows:

$$C_\psi = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

The X axis rotation is represented by φ the Euler angle and is defined as roll. Eq. (2) shows the corresponding rotation matrix C_φ:

$$C_\varphi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi \\ 0 & \sin\varphi & \cos\varphi \end{bmatrix} \tag{2}$$

And for a rotation in the Y axis, the θ Euler angle is used and the rotation is named pitch. The C_θ rotation matrix is described below:

$$C_\theta = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \tag{3}$$

Therefore, for a succession of three rotations in each of the mentioned axis is created a DCM which represents a general translation from an A frame to a B frame:

$$C_{\mathrm{A}}^{\mathrm{B}} = \begin{bmatrix} \cos\theta\cos\psi & -\cos\varphi\sin\psi + \sin\varphi\sin\theta\cos\psi & \sin\varphi\sin\psi + \cos\varphi\sin\theta\cos\psi \\ \cos\theta\sin\psi & \cos\varphi\cos\psi + \sin\varphi\sin\theta\sin\psi & -\sin\varphi\cos\psi + \cos\varphi\sin\theta\sin\psi \\ -\sin\theta & \sin\varphi\cos\theta & \cos\varphi\cos\theta \end{bmatrix} \tag{4}$$

It is important to take into consideration that any translation between frames could perform two or more successive rotations involving Eqs. (1, 2, and 3). In this chapter are presented some examples applied for specific cases.

However, there exists another form to interpret navigation translations; this is by using quaternions [30, 41]. As the name suggests, a quaternion is formed by

four elements, an "s" scalar value and a vector $\overrightarrow{v}$ conformed of three scalars representing an axis x, an axis y, and an axis z as shown below:

$$q = \begin{bmatrix} s \\ \overrightarrow{v} \end{bmatrix} = \begin{bmatrix} s \\ v_x \\ v_y \\ v_z \end{bmatrix} = \begin{bmatrix} q_s \\ q_x \\ q_y \\ q_z \end{bmatrix} \tag{5}$$

Therefore, every rotation in a quaternion is governed by the following equation:

$$q_{\mathrm{B}\leftrightarrow A} = \begin{bmatrix} q_s \\ q_x \\ q_y \\ q_z \end{bmatrix} = \begin{bmatrix} \cos \theta/2 \\ \| \overrightarrow{e} \| \bullet \sin \theta/2 \end{bmatrix} \tag{6}$$

where:

- $q_{B \leftrightarrow A}$ defines a translation from an A reference frame to a B reference frame and backward. Furthermore, the expression could represent a one-sided rotation if it is necessary.
- $\| \overrightarrow{e} \|$ vector represents a rotation in the axis of interest, which could be described as one of the following vectors shown in Eq. (7):

$$\| \overrightarrow{x} \| = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}; \| \overrightarrow{y} \| = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}; \| \overrightarrow{z} \| = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{7}$$

Additionally, it is possible for mentioned vectors $\| \overrightarrow{e} \|$ to represent a combination of two or three simultaneous rotations.

Hence, to rotate θ degrees a three-dimensional vector $\overrightarrow{v_a}$ and finalize in a $\overrightarrow{v_b}$ position, the equation is described as quaternion multiplication $\bigotimes$ as shown in Eq. (8):

$$\overrightarrow{v_b} = q_{B \leftarrow A} \bigotimes \begin{bmatrix} s \\ \overrightarrow{v_a} \end{bmatrix} \bigotimes {q_{B \leftarrow A}}^{-1} \tag{8}$$

Therefore, for every reference frame, different sequences of rotations are realized according to the planes and elements involved. Consequently, it is possible to convert the information from a DCM structure to a quaternion representation.

A quaternion is described by the diagonal elements of a DCM, where the equations applied are:

$$q_s = \sqrt{\frac{1}{4} \bullet (1 + C_{11} + C_{22} + C_{33})} \tag{9}$$

$$q_x = \sqrt{\frac{1}{4} \bullet (1 + C_{11} - C_{22} - C_{33})} \tag{10}$$

$$q_y = \sqrt{\frac{1}{4} \bullet (1 - C_{11} + C_{22} - C_{33})} \tag{11}$$

$$q_z = \sqrt{\frac{1}{4} \bullet (1 - C_{11} - C_{22} + C_{33})} \tag{12}$$

Thus, some elements of the DCM matrix required for one of the quaternion values may be equal to zero. Consequently, it is necessary to involve the elements in the other subdiagonals of the matrix [5].

The process followed indicates the evaluation of the recently computed q_s, q_x, q_y, and q_z, where the one with the greatest absolute value is selected. The other elements are recalculated accordingly to the selected value.

For a selected q_s, the other elements are estimated as:

$$q_x = \frac{C_{32} - C_{23}}{4 \bullet q_s} \tag{13}$$

$$q_y = \frac{C_{13} - C_{31}}{4 \bullet q_s} \tag{14}$$

$$q_z = \frac{C_{21} - C_{12}}{4 \bullet q_s} \tag{15}$$

If the determined greatest absolute value is q_x, then the equations for the other elements are:

$$q_s = \frac{C_{32} - C_{23}}{4 \bullet q_x} \tag{16}$$

$$q_y = \frac{C_{21} + C_{12}}{4 \bullet q_x} \tag{17}$$

$$q_z = \frac{C_{13} + C_{31}}{4 \bullet q_x} \tag{18}$$

When the greatest absolute value is q_y, the equations for the other values are:

$$q_s = \frac{C_{13} - C_{31}}{4 \bullet q_y} \tag{19}$$

$$q_x = \frac{C_{21} + C_{12}}{4 \bullet q_y} \tag{20}$$

$$q_z = \frac{C_{32} + C_{23}}{4 \bullet q_y} \tag{21}$$

And finally, if the greatest absolute value is q_z, the equations are determined as follows:

$$q_s = \frac{C_{21} - C_{12}}{4 \bullet q_z} \tag{22}$$

$$q_x = \frac{C_{13} + C_{31}}{4 \bullet q_z} \tag{23}$$

$$q_y = \frac{C_{32} + C_{23}}{4 \bullet q_z} \tag{24}$$

Furthermore, the elements in a quaternion can describe a DCM matrix if necessary [22]. The equation is described as:

$$C_A^B = \begin{bmatrix} q_s^2 + q_x^2 - q_y^2 - q_z^2 & 2 \bullet (q_x \bullet q_y - q_y \bullet q_s) & 2 \bullet (q_x \bullet q_z + q_y \bullet q_s) \\ 2 \bullet (q_x \bullet q_y + q_z \bullet q_s) & q_s^2 - q_x^2 + q_y^2 - q_z^2 & 2 \bullet (q_y \bullet q_z - q_x \bullet q_s) \\ 2 \bullet (q_x \bullet q_z - q_y \bullet q_s) & 2 \bullet (q_y \bullet q_z + q_x \bullet q_s) & q_s^2 - q_x^2 - q_y^2 + q_z^2 \end{bmatrix} \tag{25}$$

It is important to indicate that the information presented in the chapter to translate from one reference frame to another is expressed in DCM, but as previously described, it is possible to handle the rotations with quaternions.

Stereoscopic vision systems and laser scanner systems are two examples of visual systems applied in VINS. They are widely used in navigation applications along with popular technologies as LIDAR and are part of current methodologies. The aim of both systems is to provide absolute references to diminish the inherent drift error of INS using visual sensors as cameras or photoelectric sensors.

3 Stereoscopic Vision Systems

The cameras in navigation attempt to recreate the eyes' function, giving information about the surrounding environment. One single camera will only offer some information of the environment, but for terms of depth, surface shape, and curvature, it is necessary to add two or more cameras.

Stereoscopic vision systems, or SVS, acquire visual information from two or more cameras to obtain features of a specific scene [50]. SVS are portable systems with a wide field of view (FOV), capable of obtaining distance and object information, and also, SVS have advantages over other navigation devices as sonar and radar, because they do not require mechanical components and attain the pixels of the image at the same point in time [39, 80] (Fig. 3).

On the other hand, if SVS loss information in the image digitalization, there is a distortion in the lens or the system cannot find the corresponding points in the two (or multiple) images; they will not be available to achieve the triangulation process [51].

The integration of SVS to a VINS states a system capable of performing visual odometry, where two or more cameras work in conjunction with inertial references as an IMU to navigate in real time [21]. The SVS acts as an absolute reference for the INS, helping the inertial references with information about the surroundings to diminish drift and reduce the error in position and attitude.

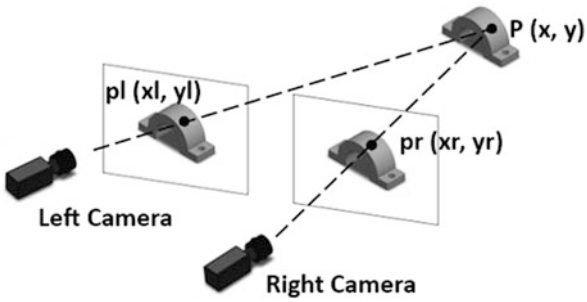


Fig. 3 Stereoscopic vision system

For a navigation environment, SVS systems are subject to constant motion, fast dynamics, limited computation on board (for small vehicles or bodies), and the constant necessity of odometry [1]. In VINS, a SVS system must operate in a self-positioning configuration, where a camera is part of the body hardware and provides images of the current environment where the body is navigating [72].

In order to navigate implementing a VINS arrangement with an SVS as the visual sensor, the system acquires a set of images according to the number of cameras available. Therefore, it initiates the detection of significant points and recognizes geometries in the set of images, gets information of the surrounding, and then finds the similarities. Afterward, a pattern match process begins where points are localized in each image to subsequently identify the same points but in the previous iteration of the set of images. Finally, the system identifies the variations of the pixels in the sequence of the images to perform the estimation of the motion [48, 56].

Figure 4 shows a block diagram where an SVS process is integrated in an INS, resulting in VINS. The SVS process is followed by a block transforming the information to body reference frame, allowing the data to properly be compared with the INS dead reckoning and feedback the system.

Figure 4 shows that it is necessary to transform the data coming from the SVS process to a body reference frame. As other references, the cameras in a SVS own a reference frame according to their properties. A camera coordinate reference frame is shown in Fig. 5; the axis is aligned following the right-hand rule. In the presented frame, Z axis is pointing to the object or the environment in direction of the depth, which is the view of the camera; X and Y axes, on the other hand, follow the image axis.

As noted, the camera frame is formed by a different coordinate reference frame in comparison with the body frame (Fig. 2), a common situation when diverse references are integrated in any navigation system. Therefore, the axis is aligned through a mathematical transformation matrix to avoid misinterpretations during the navigation. A particular solution for multiple coordinate reference frames is rotating through the coordinate frames until arriving at the navigation frame, where the navigation is interpreted and related to a remarkable amount of coordinate reference

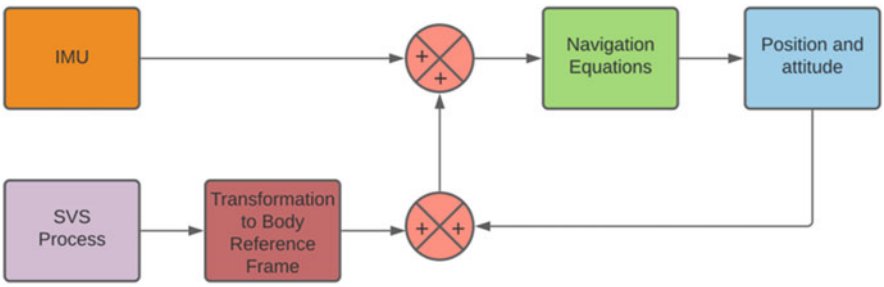


Fig. 4 SVS in an INS

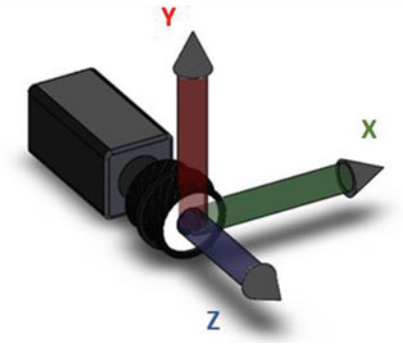


Fig. 5 Camera coordinate reference frame

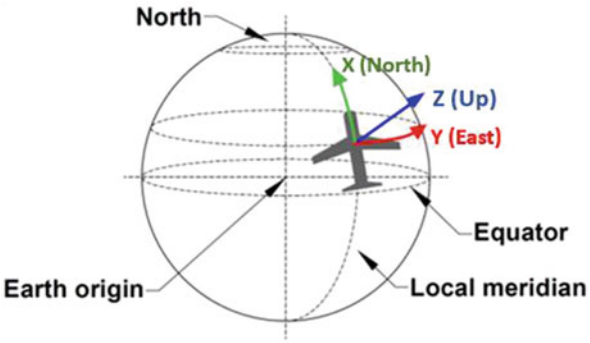


Fig. 6 Navigation frame

frames. The navigation frame is configurated with earth gravity and the cardinal points north and east. X axis is pointing to north; Y axis is aligned to east; and Z axis is pointing up or down, creating a NEU and a NED configuration. The arrangement is specified by the user (Fig. 6).

According to Wang et al. [73] and Veth [69], the camera is modeled after the camera perspective as showed in Fig. 7, where the presented frame shows the virtual frame in substitution of the focal plane to correct the inversion in the directions of the X and Y axes [74]. Also, Eq. (26) shows the line of sight vector from the camera pinhole in navigation coordinates (s^{n}):

$$s^{\mathrm{n}} = [x_i - x_c \;\; y_i - y_c \;\; -f] \tag{26}$$

The line of sight vector s^{n} is the difference between the image target location (i) and the camera position (c) in X, Y coordinates; x_i and y_i are the image coordinates and x_c and y_c are the camera coordinates. The vector also includes the camera's focal length f which is the distance between the camera and the image.

s_c^n, describes the cameras position transformed from camera frame to navigation frame. Therefore, the equation describes two DCM and considers the distance between the image target location and the camera position (27):

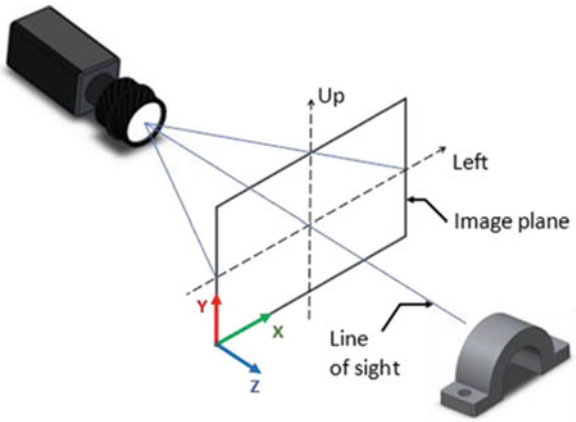


Fig. 7 Camera pinhole

$$s_{\mathrm{c}}^{\mathrm{n}} = C_{\mathrm{c}}^{\mathrm{b}} C_{\mathrm{b}}^{\mathrm{n}} s^{\mathrm{n}} \tag{27}$$

C_b^n is the DCM to transform from body frame to navigation frame. It is defined in Eqs. (28) and (29), where φ expresses the roll angle in X axis, θ is the pitch angle in Y axis, and ψ is the yaw angle in Z axis:

$$C_{\mathrm{b}}^{\mathrm{n}} = R_x(\varphi)\, R_y(\theta)\, R_z(\psi) \tag{28}$$

$$C_{\mathrm{b}}^{\mathrm{n}} = \begin{bmatrix} \cos\psi\cos\theta & \sin\psi\cos\theta & -\sin\theta \\ -\sin\psi\cos\phi + \cos\psi\sin\theta\sin\phi & \cos\psi\cos\phi + \sin\psi\sin\theta\sin\phi & \cos\theta\sin\phi \\ \sin\psi\sin\phi + \cos\psi\sin\theta\cos\phi & -\cos\psi\sin\phi + \sin\psi\sin\theta\cos\phi & \cos\theta\cos\phi \end{bmatrix} \tag{29}$$

$C_{\mathrm{c}}^{\mathrm{b}}$ DCM represent the camera to body transformation and involve rotations in Z and Y axes, transforming the attitude from the camera frame to the body frame (Eq. 30) [7]. The equation shows an azimuth angle expressed as α for Z axis and an elevation angle expressed as β for the rotation in Y axis:

$$C_{\mathrm{c}}^{\mathrm{b}} = \begin{bmatrix} \cos(\alpha) & \sin(\alpha) & 0 \\ -\sin(\alpha) & \cos(\alpha) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\beta) & 0 & -\sin(\beta) \\ 0 & 1 & 0 \\ \sin(\beta) & 0 & \cos(\beta) \end{bmatrix}$$

$$C_{\mathrm{c}}^{\mathrm{b}} = \begin{bmatrix} \cos(\beta)\cos(\alpha) & \sin(\alpha) & -\sin(\beta)\cos(\alpha) \\ -\sin(\alpha)\cos(\beta) & \cos(\alpha) & \sin(\alpha)\sin(\beta) \\ \sin(\beta) & 0 & \cos(\beta) \end{bmatrix} \tag{30}$$

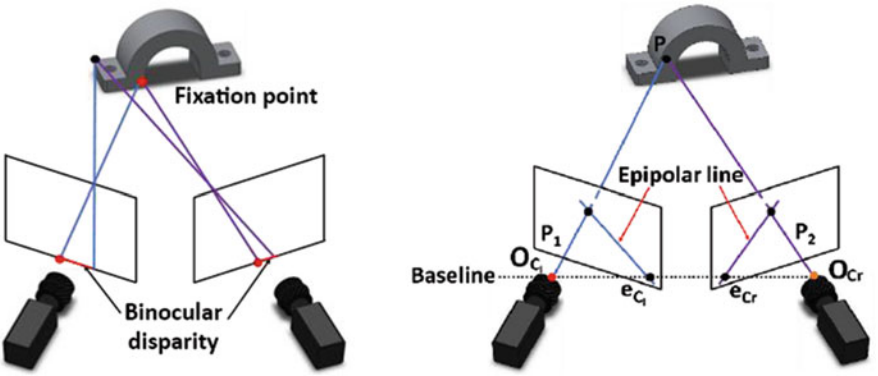


Fig. 8 Binocular disparity (**a**) and epipolar geometry (**b**)

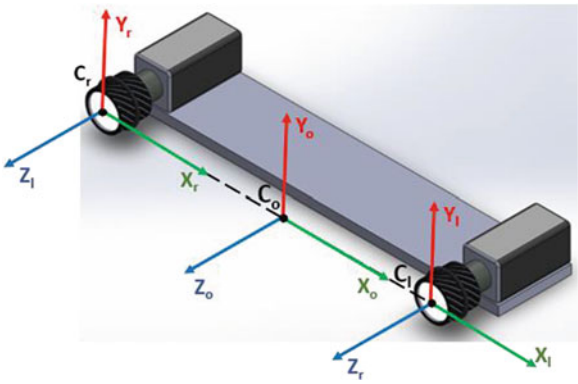


Fig. 9 Active stereo coordinate frame

As mentioned before, an SVS implements two or more cameras in order to estimate depth of objects and planes in a scene. The calculation of depth is produced by a binocular disparity created between the cameras [12]. Each camera possesses its own projection of the image, a triangulation between the points of each camera, and a point in the object of view creates an epipolar plane and determines a pair of epipolar lines in the two images, where the epipole point is the center of projection of the other camera [15, 29] (Fig. 8).

Consequently, to integrate a SVS in an INS, both cameras' frames are located in a midpoint frame where the two encounters. The active stereo coordinate reference frame C_{AS} describes the relation of left and right cameras, C_{l} and C_{r}, respectively; to an origin position, the calculation is made through a cross product as shown in Eq. (31) and Fig. 9 [31, 40].

$$C_{\mathrm{AS}} = (C_{\mathrm{r}} - C_{\mathrm{l}})\,(C_{\mathrm{o}} - C_{\mathrm{l}}) \tag{31}$$

The active stereo coordinate reference frame is located in the center and provides the data used in dead reckoning, where the C_{ASk} is compared with the C_{ASk-1} to compute a new attitude and position estimation.

4 Mobile Binocular Visual Inertial Odometry

Visual-inertial odometry (VIO) is a method employed in navigation to estimate motion using images acquired by camera sensors [64].

VIO systems are capable of functioning with a monocular camera; however, it is recommended to use a binocular camera to get better results in terms of environment recognition and motion due to its ability to perform depth determination [33]. Therefore, it is possible to implement a VIO system in vehicles or robotic manipulators to perform navigation tasks and work in conjunction with inertial instruments.

Vehicles as planes, drones, all-terrain mobile robots, or humanoid robots rely on binocular vision for navigation. For example, explorer vehicles take advantages of the cameras to travel across unknown environments and avoid collisions, or being positioned in a dangerous place [78].

All-terrain mobile robots are common in emergency situations in which there is no opportunity for persons to enter buildings. Binocular VIO aids not only to navigate through the place but also to provide opportunity to recognize and locate objects. A similar situation applies for drones where it is necessary to fly and identify landmarks [28] (Fig. 10).

However, VIO binocular mobile robots require obtaining non-blurry images to properly execute the process between frames and calculate the current distance and attitude [52]. Hence, VIO binocular mobile systems could be affected by the same surrounding environment they are exploring. Weather conditions, non-plain floors,

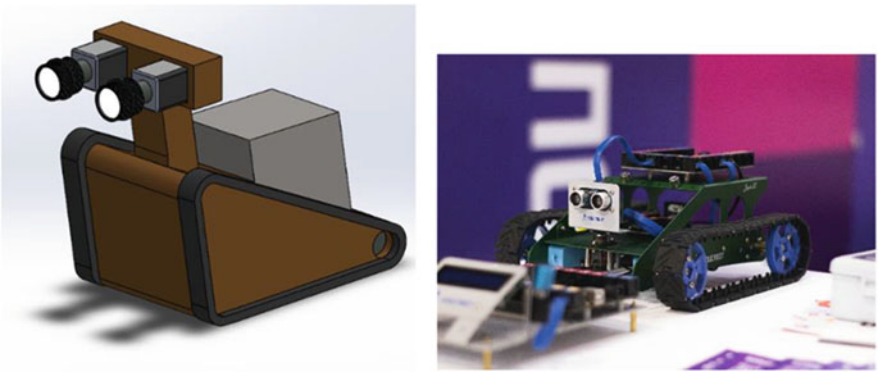


Fig. 10 SVS binocular rescue mobile vehicle (left) and binocular educational mobile vehicle (right) [10]

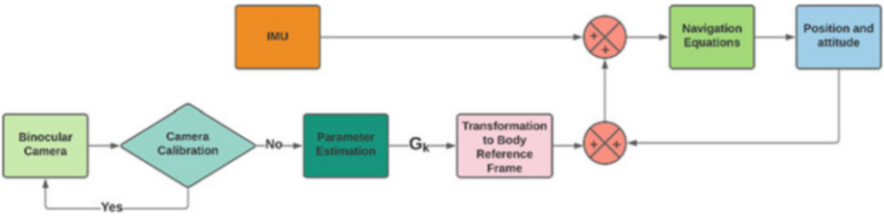


Fig. 11 Calibration and correction for blurry images

absence of light, or the cameras being directly affected by light are some of the drawbacks impeding the cameras to accurately get images.

VIO binocular mobile robots follow the majority of the steps in the process of VINS. However, in some systems, at the beginning a camera calibration phase and in a subsequent state a learning technique for parameter estimation through random samples are added [23, 66]. The methodologies mentioned work as a help to diminish the errors for the image taken by the binocular cameras.

Besides the feature point extraction of the image, the parameter estimation compares the current image with the previous one in order to obtain the current position and attitude of the navigation body or the relative motion T_k of the camera [4, 77]. Relative motion T_k is the computation of the current position and attitude of the navigation body, where every new data k obtained from the SVS process is concatenated with the previous data $k-1$. T_k is expressed as:

$$T_{k,k-1} = \begin{bmatrix} R_{k,k-1} & t_{k,k-1} \\ 0 & 1 \end{bmatrix} \tag{32}$$

where $R_{k,k-1}$ is a rotation matrix. T_k is a translation vector between frames taken at timesteps. Therefore, T_k is employed to determine a global estimate G_k to transform the information to the body reference frame in the particular case of VINS, as shown in Fig. 11 and afterward, compare the information with inertial instruments as IMU. G_k is obtained with the previous G_{k-1} and the relative motion T_k referenced at the initial frame G_0 at $k = 0$:

$$G_k = G_{k-1} T_k \tag{33}$$

5 Omnidirectional Visual-Inertial Navigation Systems

OVINS are navigation systems build with two omnidirectional cameras or two rotating cameras; their purpose is to measure depth in the horizontal plane of the cameras additionally to horizontal and vertical distances. In navigation, the device

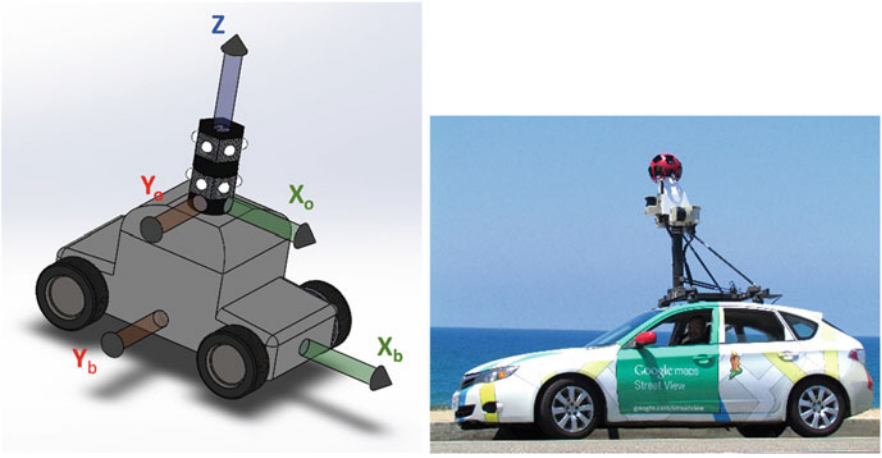


Fig. 12 OVINS: Two omnidirectional cameras in a vehicle to calculate depth (left) and Google Maps car (right) [9]

Fig. 13 OVINS detecting points in two omni-images

employed must be able to record omnidirectional stereo video (OSV) and create panoramic images to estimate distance and attitude in real time (Fig. 12)

In a city, OVINS allow INS to be aware of the different obstacles a mobile vehicle may encounter. From small places as malls or amusement parks to a bigger environment as the street, OVINS have the possibility to determine the distance between the mobile vehicle and the object (Fig. 13).

In OVINS, the FOV they possess is bigger than the one proportioned in conventional cameras; it allows the possibility to perform navigation tasks in a large environment. UAV as drones perform specific tasks such as recognition and reconstruction of the surroundings where the mobile vehicle is navigating. The

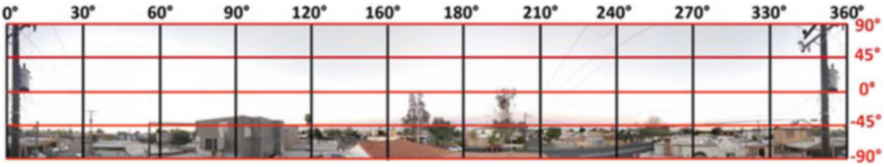


Fig. 14 Panoramic image and axis

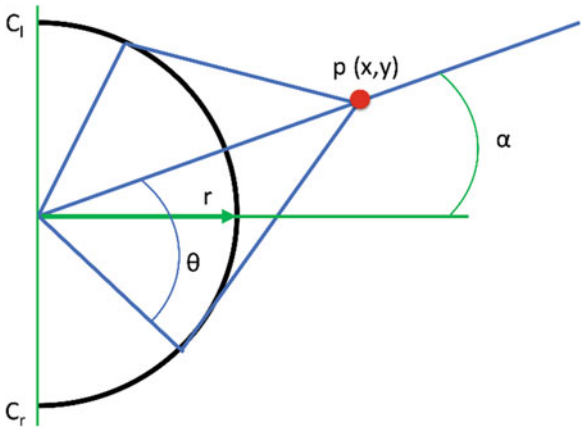


Fig. 15 Left and right camera FOV over a point $p(x, y)$

OVINS also realize odometry work with the images, defining coordinates for the points demonstrated in the captured spherical images [49].

OVINS takes panoramic images of the surroundings, where the image shows the horizontal and vertical axis through the pixels. A panoramic image is able to display between 0° and 360° in the horizontal axis and from −90° to 90° in the vertical axis [32] (Figs. 14 and 15).

There are various methods to create panoramic images as mentioned by Peleg et al. [47]. Circular projections or rotating cameras are some of the most common and still, as mentioned SVS, it exists a vertical disparity. The vertical disparity is created when the pixel is not founded in the same location of the vertical axis. The points are computed from Eq. (34):

$$\theta = \cos^{-1}\left(\frac{r}{p_x}\right) \quad \phi = \tan^{-1}\left(\frac{p_y}{\sqrt{p_x^2 - r^2}}\right) \tag{34}$$

where a point $(p_x, p_y, 0)$ is projected in the image. The panorama possesses a radius r, and due to the rotational capacities of the cameras, a pair of perspective images is generated from the left and right cameras, C_l and C_r, respectively. Also, an α angle represents the direction of view from the camera [3, 75]:

$$C_{\mathrm{l}} = \begin{bmatrix} \cos\phi \sin\left(\frac{\pi}{2} - \theta - \alpha\right) \\ \sin\left(\phi\right) \\ \cos\phi \cos\left(\frac{\pi}{2} - \theta - \alpha\right) \end{bmatrix} \quad C_{\mathrm{r}} = \begin{bmatrix} \cos\phi \sin\left(\theta - \frac{\pi}{2} - \alpha\right) \\ \sin\left(\phi\right) \\ \cos\phi \cos\left(\theta - \frac{\pi}{2} - \alpha\right) \end{bmatrix} \tag{35}$$

6 Laser Scanner Systems

The object recognition through laser scanner systems (LSS) is a methodology employing photoelectronic instruments capable of detecting the light emitted by a laser. There exist different types of LSS and a diversity of photoelectronic sensors that detect light at different speeds. Therefore, it is important to note the impact of the methodologies in the precision of the point estimation; the different strategies improve the resolution incrementing the number of detected points during the scanning.

One approach to compute measurements is with the help of triangulation methods. Laser triangulation can be static, where the method requires an adjusted and fixed laser and camera to capture the light. Hence, a dynamic laser triangulation method requires a moving laser until the point of light goes through the selected area or object, and then the triangulation is calculated and the measurement is obtained [53, 54].

The LSS process produces a point cloud representing the object surface, where the points generated must present a low dispersion to reflect the true form of the scanned object. Therefore, despite mainstream methods for registering the measurements as triangulation laser scanner, an error in the estimation of the object's shape is still present. Methodologies as artificial neural networks (ANN) are a helpful solution to LSS, thanks to their capacity to detect and predict patterns also applied in image classification [58, 67] (Fig. 16).

In a navigation system, LSS provides information about the surrounding environment, shapes, size, and depth of objects. As discussed previously, the photoelectronic sensors and laser are key factors in distance measurement, where some lasers are designed to measure large distances. Aerial vehicles as drones require LSS to create maps and determine features for the study of ecologic areas. On the other hand, terrestrial vehicles as cars implement LSS to recognize the road, helping self-driving vehicles avoid collision and measure distances (Fig. 17).

Nevertheless, there exist some lasers in LSS for short distances, applied in the high-precision detection, where they help in differentiation of near objects with detail and provide support for navigation body until they find the desired point. Close-range navigation with laser is applied for microsurgery systems where the level of precision required is high to properly complete the procedure on people. Mobile robots in pipelines realize a mapping of structures to extract features to detect damaged elements such as elbows, T-junctions, or corrosion in the pipeline.

For an LSS working with an INS, the inertial navigation process gets feedback from the last iteration of position and attitude in conjunction with the data coming

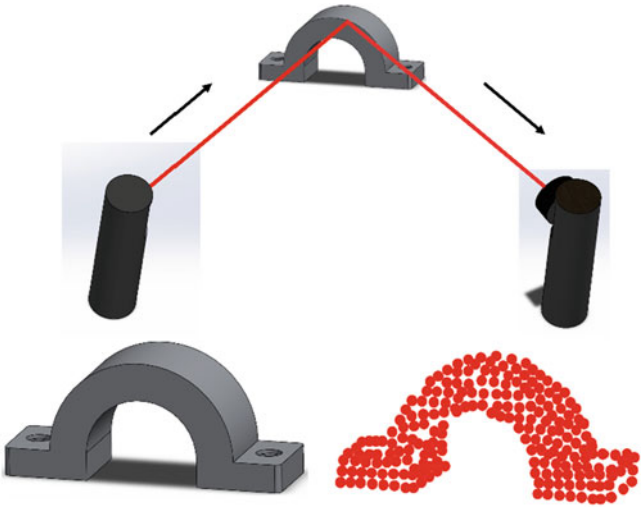


Fig. 16 Laser scanning (upper image) and point cloud (below image)

Fig. 17 Aerial mapping with laser scanners [55]

from the LSS. The LSS system performs a point triangulation to measure the distances and then proceed to apply a transformation of the data in LSS to the body reference frame (Fig. 18).

LSS coordinate reference frame as other reference frames follows the right-hand rule, where Z axis is parallel to the scanning aperture of a laser scanner system and is pointing up, Y axis is the pointing direction of the scanning aperture laser, and X axis is orthogonal to Z axis and Y axis and pointing to the left [68] (Fig. 19).

And to transform the LSS coordinate reference frame to body coordinate reference frame, the system as other references can employ a quaternion transformation matrix or a direction cosine matrix. The present chapter shows a direction cosine matrix required to transform the LSS frame. However, the LSS frame requires

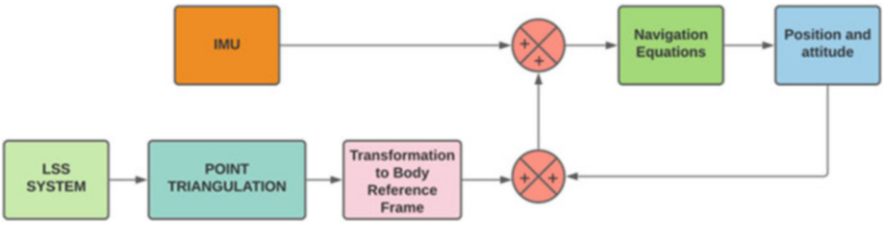


Fig. 18 LSS and INS integration

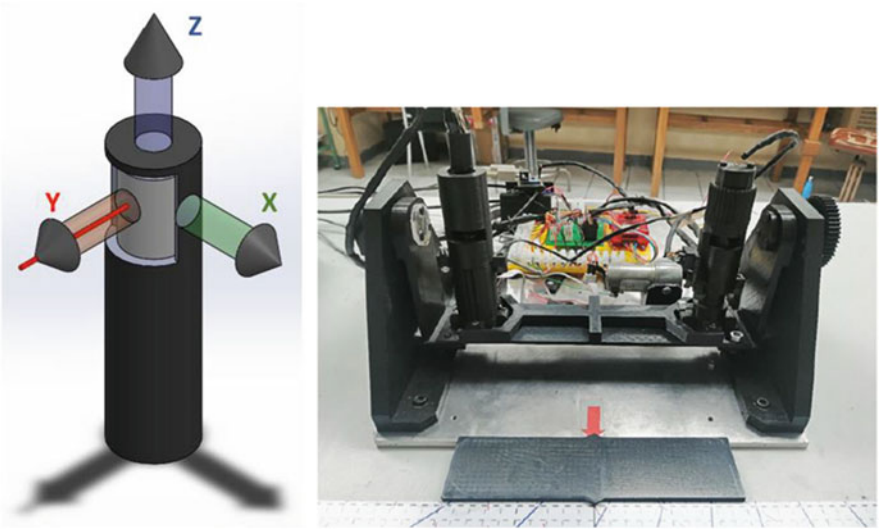


Fig. 19 LSS coordinate reference frame

aligning the laser scanner; the process presents the following equations to properly center the scanner [36, 65]:

$$P^{\mathrm{L}} = \begin{bmatrix} p_i \sin\phi_i \cos\theta_i \\ p_i \sin\phi_i \sin\theta_i \\ p_i \cos\phi_i \end{bmatrix} \tag{36}$$

For Eq. (6), P^{L} is the offset to properly align the laser scanner to the system, ϕ is the mirror angle, θ is the laser scanning axis, and p_i is an offset of the pulse on the distance measurement. The DCM complements the aligning process. As a result, to transform the LSS measurement to body reference frame, the equation requires the rotation in X and Y axis:

$$C_{\mathrm{la}}^{\mathrm{b}} = C_{\mathrm{l}}^{\mathrm{b}} P^{\mathrm{L}} \tag{37}$$

where $C_{\mathrm{la}}^{\mathrm{b}}$ is the DCM of the aligned laser in the body reference frame and $C_{\mathrm{l}}^{\mathrm{b}}$ represents only the rotation from laser frame to body frame. The transformation performs a rotation in X, Y, and Z axis as shown before in Eq. (38):

$$C_{\mathrm{l}}^{\mathrm{b}} = \begin{bmatrix} \cos\psi\cos\theta & \sin\psi\cos\theta & -\sin\theta \\ -\sin\psi\cos\phi + \cos\psi\sin\theta\sin\phi & \cos\psi\cos\phi + \sin\psi\sin\theta\sin\phi & \cos\theta\sin\phi \\ \sin\psi\sin\phi + \cos\psi\sin\theta\cos\phi & -\cos\psi\sin\phi + \sin\psi\sin\theta\cos\phi & \cos\theta\cos\phi \end{bmatrix} \tag{38}$$

7 LIDAR Odometry and Mapping

The motion estimation is a task discussed in different parts of the present chapter. INS has the capability to perform odometry through different means and work in real time. LSS systems are capable of executing the task, and due to their capacity to characterize forms and their measurements, mapping in real-time using LSS is still a popular technology.

On the other hand, light detection and ranging (LIDAR) is a measurement technology based on laser technology. The system principle consists of a transmitter and a receiver; LIDAR measures the time it takes for the laser to travel to a point and go back to the receptor. Therefore, a common practice for the improvement of precision is comparing the measurements to other instruments' data. Consequently, LIDAR technology is used for mapping in topography and exploration [70] (Fig. 20).

The mapping process requires the LIDAR system in movement; as a consequence, there are a series of problems affecting the precision in the measurements. In order to implement a LIDAR in an INS for a mapping process in real time, it is

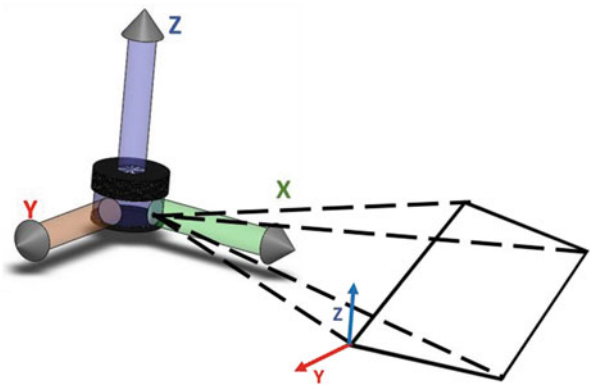


Fig. 20 LIDAR system and FOV

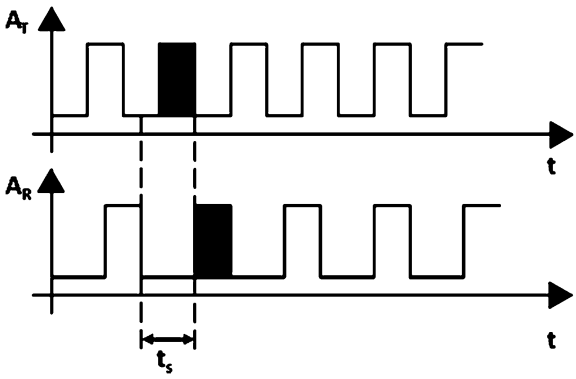


Fig. 21 Traveling time t_s of a laser pulse

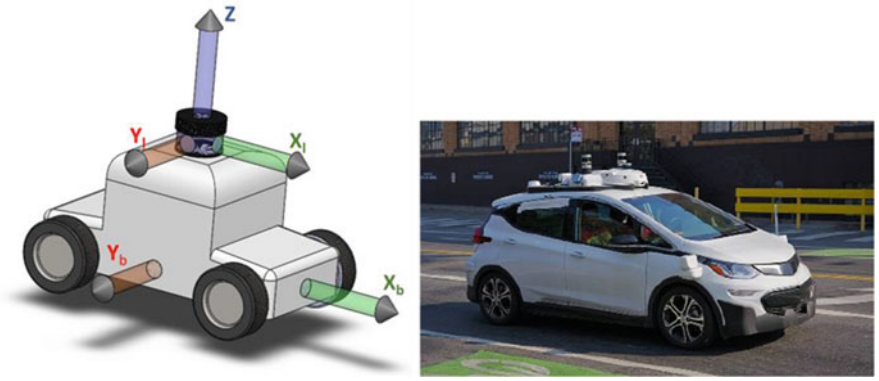


Fig. 22 Body and LIDAR coordinate reference frame and car with a LIDAR [18]

necessary to consider the synchronization of the data received. LIDAR measures the timestep a light pulse takes to reach the target and back to the receiver [44, 79].

The range of LIDAR pulse of light is calculated as follows:

$$R = \frac{1}{2} c \bullet t_s \tag{39}$$

where R is the range or distance between the laser transmitter and the surface object, c is the speed of light, and t_s is the traveling time of the laser pulse (Fig. 21). The amplitudes of the laser pulse in Fig. 22 demonstrate the traveling time t_s [17]. A_T and A_R are the amplitude transmitted and the amplitude received, respectively.

The information proportioned by LIDAR is matched with the inertial instruments as in other INS. Besides the synchronization in time of the data samples, the LIDAR information must be in the body reference [34]. Henceforth, due to LIDAR scanning laser process, the system can share the same coordinate reference frame for X and Y axis. Z axis is pointing in the opposite direction.

After LIDAR initiates the scanning, the data is stored in submaps to form the map of the scanned zone. A variety of methodologies [17, 45, 63] are well known to proceed with the mapping, but in an INS, the information is complemented with inertial references as IMU. Hence, the map created and later stored is an accompaniment for the navigation process; it is still necessary for the technology to properly calculate and process the information coming from the inertial sensor to afterward receive feedback from the mapping process.

Therefore, for the integration of LIDAR with an INS, Su et al. [63] propose to employ the timestep i difference as a reference coordinate, where in a next timestep, j is describing the system trajectory c_{ij} according to the IMU data. Then, it is calculated a component vector of the c_{ij} followed by the body and the current pitch and yaw variation, represented with $d\theta$ and $d\varphi$, respectively. In the particular case of ground vehicles, c is considered as the ground roughness. The component vector possesses a chord length of the trajectory:

$$l_{ij} = c_{ij} \cos(c * d\varphi) \tag{40}$$

Thus, a motion vector p_{ij} is determined with the chord length l_{ij} to obtain the variation between the i and j timesteps:

$$p_{ij} = \begin{bmatrix} c_{ij} \cos(c * d\varphi) \cos(d\theta) \cos(d\varphi) \\ c_{ij} \cos(c * d\varphi) \cos(d\theta) \sin(d\varphi) \\ -c_{ij} \cos(c * d\varphi) \sin(d\theta) \end{bmatrix} \tag{41}$$

8 Surgical Navigation Robots

To perform a surgery, accuracy is a key element for surgeons. It takes years of practice to meticulously realize a complex surgery. Thus, new technologies to perform surgeries with precision are now part of some hospitals and are a helpful tool for inexperienced surgeons. Surgical navigation robots (SNR) are manipulated by trained surgeons and built with inertial sensors as accelerometers or IMU to support the movements and improve the precision.

Surgical navigation robots (SNR) are structures composed of robotic arms, INS system, and complementary sensors to increase the precision. A SNR in conjunction with a LSS allows the systems to perform scans and obtain detailed information of the body part in depth and mapping. Thus, the LSS is capable of realizing a mapping process to create a mesh and to compute finite element calculation [19] (Fig. 23).

LSS is implemented at the end of the link of SNR where the surgical tool is located; in some systems, the laser points directly to the position where it is going to be performed the surgical work. The laser position is corrected through inertial sensors and with the help of cameras where the image is displayed to the surgeon.

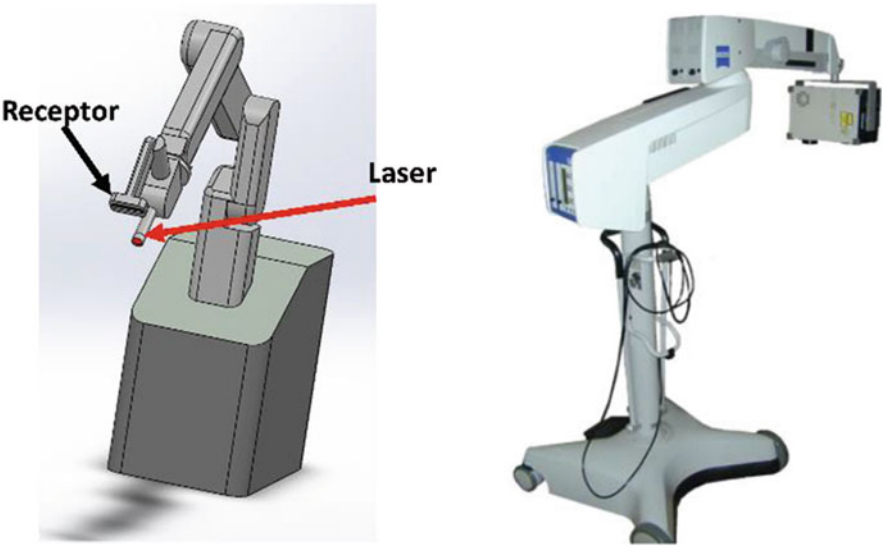


Fig. 23 Laser surgical navigation robot and the Minolta VI-900 class I laser scanner [38]

Robust SNR applies additional optical tracking systems to adjust the laser position [14] [37].

SNR are robots with a variety of configurations depending on the surgery to be performed. Specifications as the number of links for the robotic arm, coupling with other robotic manipulators to improve the precision in the surgery, are elements to consider where the mathematical model is described. For the purposes of the present chapter, the following Eqs. (42, 43, and 44) correspond to the final link of a robotic arm where an LSS could be found. Thus, elements as the body coordinate reference frame and distances for offset laser must be considered.

For a mapping coordinate frame m, Liao et al. [35], Jerbić et al. [27], and Al-Durgham et al. [2] suggest:

$$r_p^m(t) = r_b^m(t) + C_b^m(t)\left\{a_{\mathrm{IMU}/s}^b + C_s^b r_p^s(t)\right\} \tag{42}$$

where ${r_p}^m$, ${r_b}^m$, and ${r_p}^s$ are the positions of the body frame, point p at the end of the arm and a function of the observed range, respectively. ${C_b}^m$ and ${C_s}^b$ are DCM for the rotations; thus, the equations can vary depending on the arm configuration. And $a_{\mathrm{IMU}/s}^b$ is the lever arm offset (if exists) between the laser and the body frame.

Finally, for a validation in the accuracy of measurements Chen et al. [13] propose a pivot P and axis A calibration, where the actual i points collected are compared to the estimated pivotal points P^*, to obtain a distance error:

$$P_{i\text{err}} = \sqrt{(P_{ix} - P*_{ix})^2 + \left(P_{iy} - P*_{iy}\right)^2 + (P_{iz} - P*_{iz})^2} \tag{43}$$

And for the angle error are included the actual angles A and the calculated angle $A*$:

$$A_{i\text{err}} = \cos^{-1}\left[\frac{A_{ix} - A*_{ix} + A_{iy} - A*_{iy} + A_{iz} - A*_{iz}}{\sqrt{{A_{ix}}^2 + {A_{iy}}^2 + {A_{iz}}^2} \times \sqrt{{A*_{ix}}^2 + {A*_{iy}}^2 + {A*_{iz}}^2}}\right] \tag{44}$$

9 Conclusions

For the integration of a reference to INS, even for similar types of instrumentation, proper identification of the coordinate reference frame and interpretation of the data expressed are required. The data from both SVS and LSS must be transformed to the navigation frame or the measurement frame where it is needed for analysis. All the transformation matrices in this chapter are DCM, but the described methodologies could also be applied for quaternions if the reader is more familiar with them.

When a SVS is incorporated into an INS, it enhances the possibility to attach two or more cameras in a different set of configurations for the system. For a proper interpretation of the data obtained from the image, it is necessary to define where the epipolar line is located with their epipolar points and consider the line of sight and FOV demanded for every camera at the moment to fix them in their location.

Besides, when more cameras are added to the system, a robust calibration is demanded in order to diminish errors in the measurements.

SVS are more common to be implemented with INS or navigation systems; the reason is their low complexity in data interpretation, and they are fast systems. Thus, SVS have the property to determine the objects in the environment even without the involvement of a methodology; the user can simply get the image and correct the trajectory if the system is sending information in real time.

LSS, on the other hand, is more useful for mapping in navigation. The mapping possibility helps the system to recognize a familiar environment, avoid collisions, and improve the navigation. Two of the mentioned LSS own different methodologies to perform measurements during navigation were triangulation and time of flight.

LSS systems could require more time than SVS to perform their scanning and be able to recognize the objects or the structure in front of the body. But a properly scanned object can provide useful information and help the system to perform precise movements as required in medical surgeries.

For both systems, it is necessary to remember that the body could make movements that can distort the camera image and the laser reception, generating measurement errors and increments in the INS drift. The absolute references aid the IMU of the INS, but in other situations, the IMU is helping the vision systems in the navigation, as in the LIDAR case. During the integration of the vision system with

INS, it is desirable to define which system is the one receiving feedback through the navigation process or if both systems are going to work parallel to each other.

References

1. Achtelik M, Bachrach A, He R, et al (2009) Stereo vision and laser odometry for autonomous helicopters in GPS-denied indoor environments. In: Gerhart GR, Gage DW, Shoemaker CM (eds) SPIE Defense, Security, and Sensing. p 733219
2. Al-Durgham K, Lichti DD, Kwak E, Dixon R (2021) Automated Accuracy Assessment of a Mobile Mapping System with Lightweight Laser Scanning and MEMS Sensors. Applied Sciences 11:1007. https://doi.org/10.3390/app11031007
3. Anderson R, Gallup D, Barron JT, et al (2016) JUMP: Virtual reality video. ACM Transactions on Graphics 35:1–13. https://doi.org/10.1145/2980179.2980257
4. Antigny N, Uchiyama H, Servières M, et al (2019) Solving Monocular Visual Odometry Scale Factor with Adaptive Step Length Estimates for Pedestrians Using Handheld Devices. Sensors 19:953. https://doi.org/10.3390/s19040953
5. Bekir, E. (2007). Introduction to modern navigation systems. Introduction to Modern Navigation Systems, 1–235. https://doi.org/10.1142/6481
6. Ben Y, Yang J, Yin D, Li Q (2014) System reset of strapdown INS for pipeline inspection gauge. Ocean Engineering 88:357–365. https://doi.org/10.1016/j.oceaneng.2014.07.004
7. Börjesson, N. (2005). Transformations between Camera Images and Map Coordinates with Applications. Institutionen för systemteknik.
8. Bose A, Bhat KN, Kurian T (2013) Fundamentals of Navigation and Inertial Sensors by Amitava Bose, K.N. Bhat, Thomas Kurian: New Softcover (2013) First edition. | BookVistas. PHI Learning
9. Brooks, S., 2021. https://unsplash.com/photos/jRFJ28gq8ME. [image]. [Accessed 28 March 2021].
10. Castelli, M., 2021. https://unsplash.com/photos/l74VaCZon7I. [image]. [Accessed 28 March 2021].
11. Castro-Toscano MJ, Rodríguez-Quiñonez JC, Hernández-Balbuena D, et al (2018) Obtención de Trayectorias Empleando el Marco Strapdown INS/KF: Propuesta Metodológica. Revista Iberoamericana de Automática e Informática industrial 15:391–403. https://doi.org/10.4995/riai.2018.8660
12. Chauhan T, Héjja-Brichard Y, Cottereau BR (2020) Modelling binocular disparity processing from statistics in natural scenes. Vision Research 176:27–39. https://doi.org/10.1016/j.visres.2020.07.009
13. Chen X, Xu L, Wang H, et al (2017) Development of a surgical navigation system based on 3D Slicer for intraoperative implant placement surgery. Medical Engineering and Physics 41:81–89. https://doi.org/10.1016/j.medengphy.2017.01.005
14. Choi J, Kim J, Hwang JY, et al (2017) A novel smart navigation system for intramedullary nailing in orthopedic surgery. PLOS ONE 12:e0174407. https://doi.org/10.1371/journal.pone.0174407
15. Distante A, Distante C (2020) Paradigms for 3D Vision. In: Handbook of Image Processing and Computer Vision. Springer International Publishing, Cham, pp 315–411
16. Dong Y, Wang D, Zhang L, et al (2020) Tightly Coupled GNSS/INS Integration with Robust Sequential Kalman Filter for Accurate Vehicular Navigation. Sensors 20:561. https://doi.org/10.3390/s20020561
17. Dong P, Chen Q (2017) LiDAR Remote Sensing and Applications. CRC Press
18. Dllu, 2021. A Cruise Automation Chevrolet Bolt. [image] https://commons.wikimedia.org/wiki/File:Cruise_Automation_Bolt_EV_third_generation_in_San_Francisco.jpg [Accessed 1 April 2021].

19. Esslinger D, Rapp P, Knödler L, et al (2019) A novel finite element model–based navigation system–supported workflow for breast tumor excision. Medical and Biological Engineering and Computing 57:1537–1552. https://doi.org/10.1007/s11517-019-01977-0
20. Grewal M, L.~R.~Weill, Andrews A (2007) Global Positioning Systems, Inertial Navigation, and Integration
21. Griesbach D, Baumbach D, Zuev S (2014) Stereo-vision-aided inertial navigation for unknown indoor and outdoor environments. In: 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN). IEEE, pp 709–716
22. Hanson, A. Andrew J. (2006). *Visualizing quaternions*. 498.
23. He M, Zhu C, Huang Q, et al (2020) A review of monocular visual odometry. Visual Computer 36:1053–1065. https://doi.org/10.1007/s00371-019-01714-6
24. Hofmann-Wellenhof B, Legat K, Wieser M (2011) Navigation: Principles of Positioning and Guidance. Springer Science & Business Media
25. Huang G (2019) Visual-Inertial Navigation: A Concise Review. In: 2019 International Conference on Robotics and Automation (ICRA). pp 9572–9582
26. Jazar, R. N. (2010). *Theory of Applied Robotics Kinematics, Dynamics, and Control (2nd Edition)*.
27. Jerbić B, Švaco M, Chudy D, et al (2020) RONNA G4—Robotic Neuronavigation: A Novel Robotic Navigation Device for Stereotactic Neurosurgery. In: Handbook of Robotic and Image-Guided Surgery. Elsevier, pp 599–625
28. Kelly J, Sukhatme GS (2007) An experimental study of aerial stereo visual odometry. In: IFAC Proceedings Volumes (IFAC-PapersOnline). IFAC Secretariat, pp 197–202
29. Kim H, Lee S (2012) Simultaneous line matching and epipolar geometry estimation based on the intersection context of coplanar line pairs. Pattern Recognition Letters 33:1349–1363. https://doi.org/10.1016/j.patrec.2012.03.014
30. Kuipers, J. B. (2020). *Quaternions and rotation sequences : a primer with applications to orbits, aerospace, and virtual reality*.
31. Kwon H, Park J, Kak AC (2007) A New Approach for Active Stereo Camera Calibration. In: Proceedings 2007 IEEE International Conference on Robotics and Automation. pp 3180–3185
32. Lai PK, Xie S, Lang J, Laqaruere R (2019) Real-time panoramic depth maps from omni-directional stereo images for 6 dof videos in virtual reality. In: 26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 – Proceedings. Institute of Electrical and Electronics Engineers Inc., pp 405–412
33. Li G, Yu L, Fei S (2020) A Binocular MSCKF-Based Visual Inertial Odometry System Using LK Optical Flow. Journal of Intelligent and Robotic Systems: Theory and Applications 100:1179–1194. https://doi.org/10.1007/s10846-020-01222-z
34. Liang S, Cao Z, Guan P, et al (2020) A Novel Sparse Geometric 3-D LiDAR Odometry Approach. IEEE Systems Journal 1–11. https://doi.org/10.1109/jsyst.2020.2995727
35. Liao H, Ishihara H, Tran HH, et al (2010) Precision-guided surgical navigation system using laser guidance and 3D autostereoscopic image overlay. Computerized Medical Imaging and Graphics 34:46–54. https://doi.org/10.1016/j.compmedimag.2009.07.003
36. Liu W, Li Z, Sun S, et al (2019) Improving Positioning Accuracy of the Mobile Laser Scanning in GPS-Denied Environments: An Experimental Case Study. IEEE Sensors Journal 19:10753–10763. https://doi.org/10.1109/JSEN.2019.2929142
37. Luo X, Mori K, Peters TM (2018) Advanced Endoscopic Navigation: Surgical Big Data, Methodology, and Applications. Annual Review of Biomedical Engineering 20:221–251
38. Marmulla R, Hassfeld S, Lüth T, Mühling J (2003) Laser-scan-based navigation in cranio-maxillofacial surgery. Journal of Cranio-Maxillofacial Surgery 31:267–277. https://doi.org/10.1016/S1010-5182(03)00056-8
39. Matthies LH (1989) Dynamic stereo vision. Carnegie Mellon University
40. Mohamed A, Culverhouse P, Cangelosi A, Yang C (2018) Active stereo platform: online epipolar geometry update. EURASIP Journal on Image and Video Processing 2018:54. https://doi.org/10.1186/s13640-018-0292-8
41. Morais, J. P., Georgiev, S., & Sprößig, W. (2014). *Real quaternionic calculus handbook*. 216.

42. Noureldin A, Karamat TB, Georgy J (2013a) Fundamentals of Inertial Navigation, Satellite-based Positioning and their Integration. Springer-Verlag, Berlin Heidelberg
43. Noureldin A, Karamat TB, Georgy J (2013b) Basic Navigational Mathematics, Reference Frames and the Earth's Geometry. In: Noureldin A, Karamat TB, Georgy J (eds) Fundamentals of Inertial Navigation, Satellite-based Positioning and their Integration. Springer, Berlin, Heidelberg, pp 21–63
44. Palieri M, Morrell B, Thakur A, et al (2021) LOCUS: A Multi-Sensor LIDAR-Centric Solution for High-Precision Odometry and 3D Mapping in Real-Time. IEEE Robotics and Automation Letters 6:421–428. https://doi.org/10.1109/LRA.2020.3044864
45. Pan Y, Han Y, Wang L, Chen J, Meng H, Wang G, Zhang Z, Wang S, (2019) 3D Reconstruction of Ground Crops Based on Airborne LiDAR Technology. In: IFAC-PapersOnLine. 52(24):35–40, ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2019.12.376
46. Pasqualetto Cassinis, L., Fonod, R., & Gill, E. (2019). Review of the robustness and applicability of monocular pose estimation systems for relative navigation with an uncooperative spacecraft. Progress in Aerospace Sciences, 110, 100548. https://doi.org/10.1016/J.PAEROSCI.2019.05.008
47. Peleg S, Ben-Ezra M, Pritch Y (2001) Omnistereo: Panoramic stereo imaging. IEEE Transactions on Pattern Analysis and Machine Intelligence 23:279–290. https://doi.org/10.1109/34.910880
48. Pesquet-Popescu B, Cagnazzo M, Dufaux F (2014) Chapter 2 – Motion Estimation—A Video Coding Viewpoint. In: Theodoridis S, Chellappa R (eds) Academic Press Library in Signal Processing. Elsevier, pp 27–92
49. Ramezani M, Khoshelham K, Fraser C (2018) Pose estimation by Omnidirectional Visual-Inertial Odometry. Robotics and Autonomous Systems 105:26–37. https://doi.org/10.1016/j.robot.2018.03.007
50. Ramírez-Hernández LR, Rodríguez-Quiñonez JC, Castro-Toscano MJ, et al (2020a) Improve three-dimensional point localization accuracy in stereo vision systems using a novel camera calibration method. International Journal of Advanced Robotic Systems 17:1729881419896717. https://doi.org/10.1177/1729881419896717
51. Ramírez-Hernández LR, Rodríguez-Quiñonez JC, Castro-Toscano MJ, et al (2020b) Stereoscopic Vision Systems in Machine Vision, Models, and Applications. In: Sergiyenko O, Flores-Fuentes W, Mercorelli P (eds) Machine Vision and Navigation. Springer International Publishing, Cham, pp 241–265
52. Rani P, Jangid A, Namboodiri VP, Venkatesh KS (2018) Visual odometry based omnidirectional hyperlapse. In: Communications in Computer and Information Science. Springer Verlag, pp 3–13
53. Real-Moreno O, Castro-Toscano MJ, Cesar Rodríguez-Quiñonez J, et al (2019) Surface Measurement Techniques in Machine Vision: Operation, Applications, and Trends. https://doi.org/10.4018/978-1-5225-5751-7.ch004
54. Real-Moreno O, Castro-Toscano MJ, Rodríguez-Quiñonez JC, et al (2018) Implementing k-nearest neighbor algorithm on scanning aperture for accuracy improvement. In: Proceedings: IECON 2018 – 44th Annual Conference of the IEEE Industrial Electronics Society. Institute of Electrical and Electronics Engineers Inc., pp 3182–3186
55. Roca David, Martínez-Sánchez Joaquín, Lagüela Susana, Arias Pedro, CC BY 4.0 <https://creativecommons.org/licenses/by/4.0>, via Wikimedia Commons. [image] [Accessed 28 March 2021].
56. Roma N, Sousa L (2005) Least squares motion estimation algorithm in the compressed DCT domain for H.26x/MPEG-x video sequences. In: IEEE Conference on Advanced Video and Signal Based Surveillance, 2005. pp 576–581
57. Samama N (2008) A Brief History of Navigation and Positioning. In: Global Positioning: Technologies and Performance. John Wiley & Sons, pp 1–27
58. Sanchez-Castro JJ, Rodriguez-Quinonez JC, Ramirez-Hernandez LR, et al (2020) A Lean Convolutional Neural Network for Vehicle Classification. In: IEEE International Symposium on Industrial Electronics. Institute of Electrical and Electronics Engineers Inc., pp 1365–1369

59. Santoso F, Garratt M, Anavatti S (2017) Visual–Inertial Navigation Systems for Aerial Robotics: Sensor Fusion and Technology. IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING 14:. https://doi.org/10.1109/TASE.2016.2582752
60. Seok H, Lim J (2020) ROVINS: Robust Omnidirectional Visual Inertial Navigation System. IEEE Robotics and Automation Letters 5:6225–6232. https://doi.org/10.1109/LRA.2020.3010457
61. Sharon Y, Lendvay TS, Nisky I (2017) Instrument Orientation-Based Metrics for Surgical Skill Evaluation in Robot-Assisted and Open Needle Driving. arXiv:170909452 [cs]
62. Spong, M. W., & Vidyasagar, M. (Mathukumalli). (2004). Robot dynamics and control.
63. Su Y, Wang T, Shao S, et al (2021) GR-LOAM: LiDAR-based sensor fusion SLAM for ground robots on complex terrain. Robotics and Autonomous Systems 140:103759. https://doi.org/10.1016/j.robot.2021.103759
64. Sun K, Mohta K, Pfrommer B, et al (2018) Robust Stereo Visual Inertial Odometry for Fast Autonomous Flight. IEEE Robotics and Automation Letters 3:965–972. https://doi.org/10.1109/LRA.2018.2793349
65. Talaya J, Alamus R, Bosch E, et al (2004) INTEGRATION OF A TERRESTRIAL LASER SCANNER WITH GPS/IMU ORIENTATION SENSORS
66. Tomažič S, Škrjanc I (2015) Fusion of visual odometry and inertial navigation system on a smartphone. Computers in Industry 74:119–134. https://doi.org/10.1016/j.compind.2015.05.003
67. Trujillo-Hernández G, Rodríguez-Quiñonez JC, Ramírez-Hernández LR, et al (2019) Accuracy Improvement by Artificial Neural Networks in Technical Vision System. In: IECON Proceedings (Industrial Electronics Conference). IEEE Computer Society, pp 5572–5577
68. Van Dang CB, Takahashi K, Phan ATT (2020) Accuracy assessment of 3d point clouds collected by a low cost uav-based laser scanner system. In: Lecture Notes in Civil Engineering. Springer, pp 815–823
69. Veth M (2006) Fusion of Imaging and Inertial Sensors for Navigation. Theses and Dissertations
70. Wandinger U (2006) Introduction to LIDAR. In: LIDAR. Springer-Verlag, pp 1–18
71. Wang, C., Wen, C., Dai, Y., Yu, S., & Liu, M. (2020). Urban 3D modeling with mobile laser scanning: a review. Virtual Reality & Intelligent Hardware, 2(3), 175–212. https://doi.org/10.1016/J.VRIH.2020.05.003
72. Wang D, Li M, Huang X, Zhang X (2021) Optical Autonomous Navigation Technology. Spacecraft Autonomous Navigation Technologies Based on Multi-source Information Fusion 163–210. https://doi.org/10.1007/978-981-15-4879-6_8
73. Wang L, Niu X, Zhang Q, et al (2013) A camera/IMU tightly- coupled navigation algorithm and verification by hybrid simulation. Journal of Harbin Institute of Technology (New Series) 20:84–90. https://doi.org/10.11916/j.issn.1005-9113.2013.06.012
74. Xu G, Zhang Z (1996) Epipolar Geometry in Stereo, Motion and Object Recognition: A Unified Approach. Springer Netherlands
75. Yan J, Kong L, Diao Z, et al (2018) Panoramic stereo imaging system for efficient mosaicking: parallax analyses and system design. Applied Optics 57:396. https://doi.org/10.1364/ao.57.000396
76. Yazdkhasti S, Sasiadek JZ, Ulrich S (2016) Performance enhancement for GPS/INS fusion by using a fuzzy adaptive unscented Kalman filter. In: 2016 21st International Conference on Methods and Models in Automation and Robotics (MMAR). IEEE, pp 1194–1199
77. Yousif K, Bab-Hadiashar A, Hoseinnezhad R (2015) An Overview to Visual Odometry and Visual SLAM: Applications to Mobile Robotics. Intelligent Industrial Systems 1:289–311. https://doi.org/10.1007/s40903-015-0032-7
78. Zhai G, Zhang W, Hu W, Ji Z (2020) Coal Mine Rescue Robots Based on Binocular Vision: A Review of the State of the Art. IEEE Access 8:130561–130575. https://doi.org/10.1109/ACCESS.2020.3009387
79. Zhang J, Singh S (2017) Low-drift and real-time lidar odometry and mapping. Autonomous Robots 41:401–416. https://doi.org/10.1007/s10514-016-9548-2

80. Zhang Y-J (2021) Handbook of Image Engineering. Springer Singapore
81. Cova TJ, Miller HJ, Beard K, Frank AU, Goodchild MF (eds) (2008) Geographic Information Science: 5th International Conference, GIScience 2008, Park City, UT, USA, September 23–26, 2008, Proceedings. Springer-Verlag, Berlin Heidelberg

Development of a Doppler Anemometry Method for Diagnosing Two-Phase Flows in a Liquid Metal Medium

Alexey Strelnik, Sergey Dvoynishnikov, Vladimir Meledin, and Ivan Kabardin

Abbreviations

ADC Analog digital converter
EMF Electro motive force
HSC Hardware-software complex
PC Personal computer

1 Introduction

The use of modern noncontact precision diagnostic tools significantly increases the efficiency and safety of technological processes in the power industry and industry [1]. Modern technologies for studying three-dimensional structural and dynamic parameters of multiphase optically opaque flows are the most complex and insufficiently developed for industrial use. This is due to the need to solve a set of complex multidisciplinary scientific and technical problems.

At present, the development of the atomic energy of the future is associated with the use of fast nuclear reactors [2]. The design characteristics of these reactors are associated with the need to remove a large amount of heat from the core. Water heat carriers have a few significant structural and operational limitations in terms of the amount of heat removed and the efficiency of heat transfer. The main solution for the intensification of heat transfer was the use of liquid metal coolants with significant thermal conductivity. Intensive heat removal from the core prevents the temperature

A. Strelnik · S. Dvoynishnikov · V. Meledin · I. Kabardin (✉)
Kutateladze Institute of Thermophysics, Siberian Branch, Russian Academy of Sciences, Novosibirsk, Russian Federation

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1_7

inside the reactor from rising above the permissible values of structural materials and makes nuclear power safer and more efficient. For the correct optimization of the parameters of the coolant flow (optimization of hydrodynamics and thermal physics of processes), experimental information on the local velocities and modes of the coolant flow is required. Due to high temperatures, pressure, chemical activity, and difficult access to the liquid metal coolant, the diagnostics of flow parameters is usually limited to measurements of temperature and pressure, which does not allow correct optimization of the coolant flows.

Nevertheless, it is especially important to obtain experimental data on the structural parameters of the flow. The problem of diagnosing the structural and dynamic parameters of the flow is especially acute in cases of the occurrence of a gas component caused by an emergency or local overheating. The use of complex diagnostics tools will make it possible to determine the nature and cause of an emergency and to perform the minimum necessary actions to prevent the reactor from failing [3].

For solving the problem of diagnostics of multiphase liquid metal media, devices operating based on optical methods are inapplicable, since the medium is opaque.

For noncontact diagnostics of structural and dynamic parameters of opaque multiphase media, there are a few methods [4]. The potentiometric method is based on measuring the conductivity of the "pipeline-heat carrier" system [5]. It has a low sensitivity (5–10% of the volumetric gas content) and gives information only about the integral gas content in the measured area. The vortex-current method [5] has design features that cannot be considered in all cases. The location of the pipeline must ensure that gas bubbles pass directly under the coils, which requires a horizontal or sloped pipeline. In addition, this method has a very limited service life of the sensors and fixes gas bubbles mainly in the volume of the coolant adjacent to the area where the coils are located.

X-ray methods [6] have a relatively low acquisition rate, which results in non-real-time meter operation. In addition, these meters are difficult to implement and are extremely expensive.

Methods using a magnetic flow meter [5] record the presence of gas by pulsations of the EMF value between the electrodes. They do not allow the detection of small gas bubbles, and do not work with a large number of bubbles.

For noncontact diagnostics of structural and dynamic parameters of opaque multiphase media, it is promising to use ultrasonic methods [7, 8]. The complex configuration and sophisticated data processing algorithms based on synchronous detection, correlation, and time-frequency analysis will make it possible to create a tool for three-dimensional diagnostics of geometric parameters and the velocity field of the investigated two-phase flow.

The specificity and difficulty of the task require a comprehensive research work in this area. The aim of this work is to develop a software and hardware complex for ultrasonic diagnostics of two-phase flows in a liquid metal medium. To achieve this goal, the following tasks were set:

1. Propose a method for diagnosing the structural characteristics of the liquid metal-gas medium.
2. Propose a method for diagnosing the dynamic characteristics of the liquid metal-gas medium.
3. Develop the measuring and hardware parts of the complex.
4. Develop the software part of the complex.

This chapter is divided into the following subsections:

- Review of Existing Diagnostic Methods
- Development of a Method for Ultrasonic Diagnostics of Two-Phase Flows in a Liquid Metal Medium
- Development of Software and Hardware Complex of the Method
- Experiments. This part is devoted to the description of the experiments carried out on water and liquid metal. The chapter also includes a description of the developed and manufactured ultrasonic sensors and experimental stands
- Conclusion

2 Overview of Existing Methods

2.1 Patent Review

A patent search was carried out for methods and devices for ultrasonic diagnostics of multiphase flows to analyze the distinctive features and advantages of the proposed development. Most of the known methods of ultrasonic diagnostics of multiphase flows are intended either for studying gas-liquid flows with a significantly higher concentration of the gas phase or do not allow spatially localizing the gas phase. Nevertheless, a number of ways are known for diagnosing structural or kinematic parameters of multiphase flows by ultrasonic methods.

There is a known method for monitoring flow discontinuities [9], which consists in the fact that the monitored medium is sounded by ultrasonic pulses in two mutually perpendicular directions, pulses passed through the medium are recorded, their amplitudes are recorded, pulses reflected from the phase boundary are also taken, their amplitudes and time are measured run, and the data obtained are used to judge the concentration of the gas phase and the flow regime. The disadvantage of this method is its low information content of the measurement results, high error, and low degree of automation.

The known method of ultrasonic monitoring of a multiphase flow [10], which consists in sounding a multiphase flow with ultrasonic vibrations, registers the signals passed through the flow and echo pulses reflected from the discontinuities of the multiphase flow. The disadvantage of this method is a high error, since the method does not consider many parameters that affect the determination of the flow regime. Known methods for determining the flow rate of two-phase flow

components and systems for their implementation [11, 12] include a measuring section with two diametrically located electroacoustic transducers installed on its surface. The disadvantages of this system for measuring the components of a two-phase flow are the complexity of the design and the integral nature of the gas content information.

Known devices for monitoring gas in a liquid metal coolant [13, 14] are based on eddy current and inductive methods. These devices do not allow continuous monitoring of the dynamic parameters of single pop-up bubbles and have a narrow range of applications. The disadvantage of these devices is their relatively low sensitivity.

The main distinguishing features of the method and hardware and software complex developed within the framework of this work for ultrasonic diagnostics of multiphase flow are in the original spatial configuration of the location of ultrasonic transducers in the investigated volume and the signal processing algorithm. This will provide a complete spatial diagnosis of the position of the gas phase in a liquid medium. Ultrasonic transducers operating at different frequencies will be used. This will provide simultaneous multichannel measurement of the structural and dynamic parameters of the investigated medium. Complex schemes for measuring the structural parameters of a two-phase medium based on complex time-frequency modulation of ultrasonic transducers will be created.

Thus, the developed technology of ultrasonic diagnostics of two-phase optically opaque media is patent-free.

2.2 Methods of Ultrasonic Diagnostics

Now, a huge number of different methods of ultrasonic diagnostics have been developed to solve a wide range of problems: from nondestructive testing of solid products to multiphase flow metering. In the context of the problem posed, these methods can be divided into two significantly different groups: methods that perform integral diagnostics of the environment and methods that allow localizing phase inhomogeneities in the environment and performing their spatial diagnostics.

The methods of the first group underlie the operation of ultrasonic flow meters [4] (phase, frequency, time-pulse, Doppler, etc.), which solve related problems, making integral diagnostics of different phases in motion, but cannot cope with the task of spatial local diagnostics of a multiphase medium. Although such methods do not provide knowledge about the local parameters of the flow, they provide important experimental information for the quantitative assessment of the motion of matter in volume.

The methods of the second group are of the greatest interest for the task at hand. It is these methods that underlie the developed hardware and software complex. Let's dwell on them in more detail.

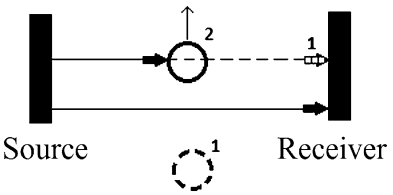


Fig. 1 Schematic of the amplitude-shadow method

2.2.1 Amplitude-Shadow Method

The shadow (or amplitude-shadow) method is based on registering a decrease in the amplitude of the transmitted wave (through signal) under the influence of phase inhomogeneity [15]. Figure 1 shows a diagram of this method.

In the absence of phase inhomogeneities in the measuring volume (bubble position 1), ultrasound is transmitted without losses between the source and the receiver. The gas bubble, being in position 2, screens part of the ultrasonic wave generated by the source from the receiver. As a result, a decrease in the amplitude of the receiving signal is observed, which can be used to judge the presence and size of a gas bubble in the measuring volume.

This method allows detecting the presence of a phase inhomogeneity, evaluating its position and size in the measuring volume at a certain moment in time. The method is practically inapplicable for extremely small volumes of inhomogeneity, and it is very sensitive to the position of the receiver and transmitter, and also does not allow one to accurately determine the velocity and position of the inhomogeneity in the medium. The amplitude-shadow method has several modifications and diagnostic methods similar in principle of operation: mirror-shadow, echo-shadow, and many others.

2.2.2 Time-of-Flight Method (Echo Method)

In this method, sounding of the measuring volume with short pulses and reception of echo signals reflected from inhomogeneities is carried out. Based on the time of detection of the echo signal and its shape, it is possible to calculate the position and size of the inhomogeneity [16].

Figure 2 shows a diagram of the time-of-flight method used to diagnose the flow of bubbles in a liquid. The combined transducer emits short pulses of ultrasonic waves. Ultrasonic waves are reflected from the floating bubble and the back wall of the tank and return to the transducer, which is already operating in the receiver mode. As a result, we obtain the characteristic echogram shown in the lower part of Fig. 2. By detecting on the echogram, the position in which the bubble echo reaches its maximum, we obtain the propagation time t_B, during which the ultrasound reaches the bubble and returns back to the transducer. Knowing the time t_B, the position of the bubble X_B is calculated as follows:

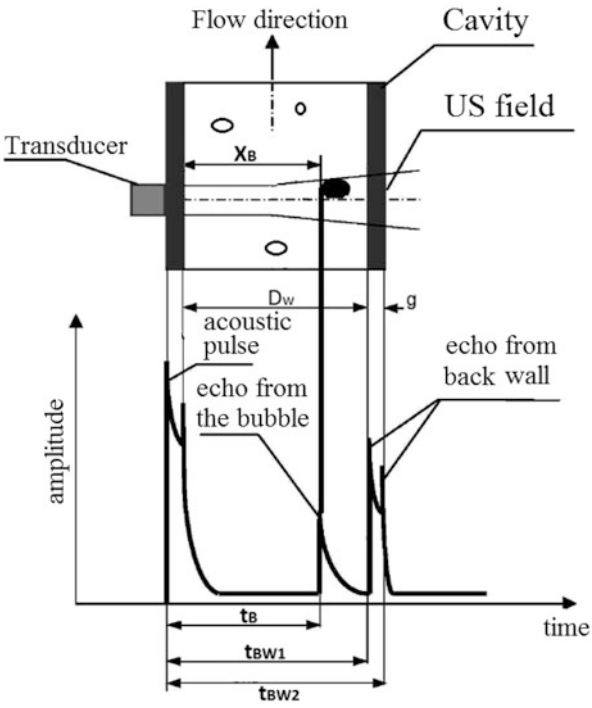


Fig. 2 Schematic diagram of the time-of-flight method

$$X_B = c \cdot t_B/2,$$

where c is the speed of ultrasound in the liquid, which can be calculated using the value t_{BW1}, the propagation time of the signal from the transducer to the back wall of the tank and back.

Having the dependence $X_B(t)$,it is possible to calculate the trajectory and speed of the bubbling bubble. This requires the use of several sensors installed in a certain way in a row, in combination with nontrivial algorithms for signal interpretation. Also, analyzing the echo signal from the floating bubble, and using transducers installed opposite each other at the same height, it is possible to obtain the dimensions of the bubble [17].

The described method is extremely difficult to implement and requires time-consuming calibration and adjustment. It is inappropriate to use this method for measuring the bubble velocity, since many sensors will be required, and the hardware part of the system will become much more complicated. As in the case of the amplitude-shadow method, the echo method has several configurations of the geometric arrangement of the transducers.

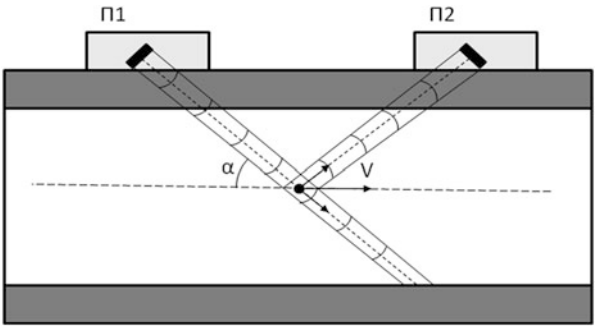


Fig. 3 Diagram of the Doppler method

2.2.3 Doppler Method

The Doppler method is based on the effect of the frequency shift of an ultrasonic signal scattered by phase inhomogeneities relative to the value of the reference frequency. The difference $f_1 - f_2$ depends on the speed of the scattering particle/inhomogeneity, V, and the speed of sound in the medium, c. The following dependence takes place:

$$f_1 - f_2 = 2 f_1 \cos(\alpha) \frac{V}{c},$$

where f_1 and f_2 are the frequencies of the reference (generated by the source) and reflected (recorded by the receiver) acoustic wave, respectively, and α is the angle between the direction of wave propagation and the vector of the scatterer velocity. The measured frequency difference can be used to determine the speed of motion of a phase inhomogeneity, in particular a floating gas bubble [18]. Figure 3 shows a diagram of the method: P1, source-transducer; P2, receiver- transducer.

Since the ultrasonic wave is scattered in absolutely all directions from the scatterer, the Doppler method can be implemented in many ways with different positions of the receiver and source. Thus, it is possible to determine different components of the scatterer velocity.

The Doppler method is most convenient for measuring the velocity of an inhomogeneity in a medium, but practically unsuitable for localizing the inhomogeneity and determining its dimensions [18].

2.3 Piezoceramic Transducers

The source (generator, emitter) and receiver of ultrasonic vibrations are the main functional elements of any ultrasonic diagnostic method. In many cases, a sensor,

a receiver, and a source combined into one is used, that is, the sensor at different points in time performs the functions of either a source or a receiver.

One of the most widespread and effective methods of creating ultrasonic vibrations in a medium is the use of piezoceramic transducers. There are crystals (piezoelectric elements) that, when deformed relative to the selected directions, form electric charges on the surface (direct piezoelectric effect), and, conversely, when an external electric field is applied, they are deformed (reverse piezoelectric effect). The operation of piezoceramic transducers is based on the described effect: the source (emitter) converts the alternating electrical signal applied to it into an acoustic wave of the same frequency, and the receiver converts the vibrations of the medium into an electrical voltage.

The parameters of acoustic sources are described by a piezoelectric module:

$$d = \frac{S}{E}(\text{M/B}),$$

where S is the size of the elastic deformation and E is the electrical tension on the piezoelectric element.

A similar parameter for the receiver – a piezo constant for deformation:

$$h = \frac{E}{S}(\text{B/M}),$$

These two parameters have the following dependency:

$$h = \frac{Ey}{\varepsilon\varepsilon_0}d,$$

where Ey is Young's modulus and ε and ε_0 are the dielectric constants of the crystal and vacuum, respectively. Consequently, with an increase in the piezoelectric modulus, the piezo constant for deformation also increases.

Historically, the piezoelectric effect was first observed in natural quartz, but today, as a rule, piezoelectric materials are used as transducers. The most common piezoceramic material is barium titanate ($BaTiO_3$) or lead titanate zirconate ($PbZ_2O_3 + PbTiO_3$). The use of these crystals is determined by their high piezomodulus d and high dielectric constant. Such piezoceramic transducers are used to create vibrations in the region of ultrasonic frequencies [19].

When generating powerful acoustic vibrations, it is necessary to observe the operating conditions at the resonant frequency f_p of the piezoelectric element:

$$f_p = \frac{C_n}{2\sigma},$$

where C_n is the speed of sound in the piezoelectric element and σ is the thickness of the piezoelectric element.

Taking Poisson's ratio for a piezoelectric crystal to be close to zero, the speed of sound is calculated by the following formula:

$$c_n = \sqrt{\frac{Ey}{\rho}}(\mathrm{M/C}),$$

where ρ is the density of the material and Ey is Young's modulus.

Also considering that

$$c_n = f\lambda\left(\frac{\mathrm{M}}{\mathrm{C}}\right),$$

where λ is the acoustic wavelength, we find that the resonant oscillation frequency of the piezoelectric element will be achieved under the following condition:

$$\sigma = \frac{\lambda}{2}(M),$$

that is, the length of the sound wave in the piezoelectric crystal at the emitted vibration frequency should be twice the thickness of the piezoelectric element.

When choosing a frequency for ultrasound studies, it is necessary to consider the dissipation and scattering of waves on gas bubbles, impurities, or liquid inhomogeneities, since the greatest interaction occurs when the orders of magnitude of the inhomogeneities and the acoustic wavelength are equal.

In this work, we used disk piezoelectric transducers made of barium titanate with a diameter of ≈6 mm and a thickness of ≈0.6 mm, operating at a resonance frequency of 4 MHz and giving directional radiation.

3 Development of the Method of Ultrasonic Diagnostics of Two-Phase Flows in a Liquid Metal Medium

The developed method of ultrasonic diagnostics of two-phase flows in a liquid metal medium is based on amplitude-shadow, time-of-flight, and Doppler methods. The scheme of the method is shown in Fig. 4. The method is very flexible in its application. It does not regulate the strict interposition of the receiver and the source, that is, different configurations of the positions of the source and receiver allow you to measure different parameters of inhomogeneity in the medium.

The signal generator sends a sinusoidal signal with a frequency f_1 to a piezoelectric element (source), which creates a sound wave with a frequency equal to the frequency of the applied electrical signal and the resonant frequency of the piezoelectric crystal. An acoustic wave is emitted into the medium at an angle α, after which it is scattered by phase inhomogeneities contained in the flow. In this

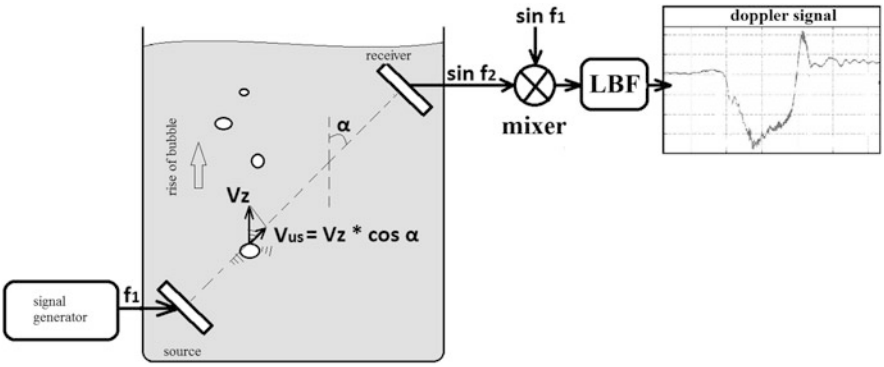


Fig. 4 Scheme of the proposed method for ultrasound diagnostics of the environment

case, the phase inhomogeneity means a bubble with air floating up in the measuring volume with a certain velocity V_z.

The scattered sound wave hits the receiving piezoelectric element and is converted back into a sinusoidal electrical signal of frequency $f_2 = f_1 + \varepsilon$, where ε is the shift of the reference frequency caused by the Doppler effect due to scattering on moving phase inhomogeneities.

It should be noted that only the projection of the bubble velocity on the direction of propagation of the sound wave influences the frequency shift:

$$f_1 - f_2 = 2f_1 \frac{Vus}{c},$$

where $Vus = Vz \cos(\alpha)$.

Thus, the projection of the speed on the selected direction is measured.

Next, a quadrature mixing operation is performed, that is, the operations of multiplying the recorded signal containing the Doppler frequency shift by a reference signal with a frequency f_1 and an orthogonal signal of the same frequency (shifted relative to the reference by $\pi/2$):

$$A \sin(f_1) * B \sin(f_1 + \varepsilon) = \frac{A * B}{2} (\cos(f_1 + \varepsilon - f_1) - \cos(f_1 + \varepsilon + f_1))$$
$$= \frac{A * B}{2} (\cos(\varepsilon) - \cos(2f_1 + \varepsilon)),$$

$$A \cos(f_1) * B \sin(f_1 + \varepsilon) = \frac{A * B}{2} (\sin(f_1 + \varepsilon - f_1) - \sin(f_1 + \varepsilon + f_1))$$
$$= \frac{A * B}{2} (\sin(\varepsilon) - \sin(2f_1 + \varepsilon)),$$

where A is the amplitude of the reference signal and B is the amplitude of the receiving signal after scattering by an air bubble. The value of B depends on the number and size of scatterers (bubbles) in the flow and will be used to implement the amplitude-shadow method.

This mixing operation is performed in order to obtain a low-frequency signal that carries information about the Doppler frequency shift. In fact, from the above trigonometric formulas shown that after multiplication, a two-component signal is formed, composed of a high-frequency ($2f_1 + \varepsilon$) and a low-frequency (ε) sinusoid. By applying a low-pass filter, that is, by attenuating the high-frequency signal, we can isolate the low-frequency Doppler signal we need. Further processing of this signal using a sliding Fourier transform makes it possible to obtain the dependence of the Doppler frequency shift on time and, accordingly, the dependence of the bubble velocity on time:

$$Vus\ (t) = c \cdot (f_1 - f_2)\ /2f_1 = c \cdot \varepsilon(t)/2f_1.$$

Thus, the described method makes it possible to track the dynamics of the bubble ascent, to distinguish the stages of its ascent, and, using the principle of the amplitude-shadow method, to estimate the size of the bubble and its position.

At the first stage of the operation, the mixing and filtering operations were implemented in hardware on analog signals, and the Doppler signal was already digitized. This approach has a few disadvantages: hardware-fixed frequency of the generated signal and filtering parameters, incompleteness of the extracted data from the receiver, etc. Therefore, for the flexibility and expandability of the capabilities of the complex, as well as to obtain more information about the receiving signal, it was decided to digitize the receiving high-frequency signal and carry out mixing operations and filtering programmatically.

Within the framework of this work, a laboratory sample of a software and hardware complex (HSC) was created, which implements the method described in the previous chapter. The scheme of the complex is shown in Fig. 5. Figure 6 shows a photo of the complex.

The following functional elements of the complex can be distinguished: hardware, software, and piezo transducers (source and receiver), the design and properties of which are described in the chapter "Experiments."

To excite the oscillations of the piezoelectric element (source), a high-frequency generator with high-frequency stability is required. The generator brand used is Rigol DG1062Z. Using a coaxial cable, the piezoelectric element (source) and the signal output of the generator were connected. The receiving piezoelectric element is connected by a coaxial cable to a 14-bit L-Card E20-10-D-1 ADC with a typical signal-to-noise ratio of 73 dB. This ADC does not have input filters, which is essential for digitizing high-frequency signals. The sampling rate used is 1 MHz and the oscillator signal frequency is 4.02 MHz.

The processing of the data coming from the ADC was carried out using a program developed in the MATLAB package.

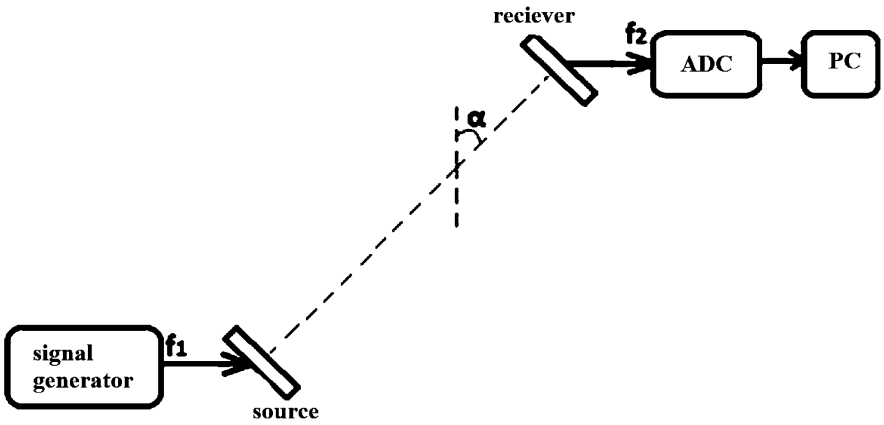


Fig. 5 Diagram of the implemented hardware and software complex

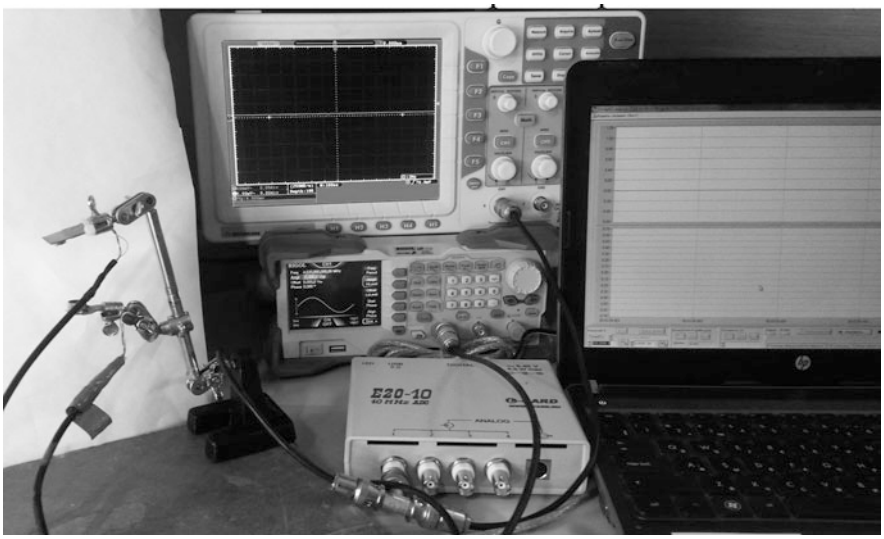


Fig. 6 Photo of the implemented hardware and software complex

Analysis of Error Sources

During experiments and data processing, several technical difficulties arise, as a result of which the measured parameters slightly differ from the real picture of what is happening.

Firstly, the wave emitted by the source is not ideally harmonic, monochrome, and directional, since the design of the ultrasonic sensor is imperfect, and the generator frequency is relatively unstable. Contacts soldered to the plates of the piezoelectric element add additional mass to the plates and affect the character and form of vibrations of the piezoelectric element. The situation is similar with the

receiving sensor. Thus, the readings of the generator connected to the source and the oscilloscope connected to the receiver are not completely accurate. Let us estimate the measurement error of the Doppler frequency shift (~2000 Hz) associated with the generator and source error of 10 Hz:

$$\sigma_g = \frac{\Delta_g}{f_{\text{Doppler}}} \approx 10\ \text{Hz}/2000\ \text{Hz} = 0.5\%$$

Secondly, ADC noise occurs during digitization. Since the amplitude of the Doppler signal is ≈0.001 V, and the signal-to-noise ratio of the ADC is 73 dB, the error in measuring the Doppler frequency shift associated with the noise of the ADC is estimated as follows:

$$73\ (dB) = 20 \cdot \log_{10} \frac{A_{\text{signal}}}{A_{\text{noise}}} => A_{\text{noise}} = \frac{A_{\text{signal}}}{10^{\frac{73}{20}}} \approx \frac{0.4\ \text{B}}{4467} \approx 9 \cdot 10^{-5}\text{B}$$

$$\sigma_{ADC} = \frac{A_{\text{noise}}}{A_{\text{Doppler}}} = \frac{9 \cdot 10^{-5}\text{B}}{0.001\ \text{B}} = 0.09\%.$$

Thirdly, there is an inaccuracy in determining the Doppler frequency shift (Doppler frequency) in the spectral analysis of the Doppler signal, since the Doppler signal frequency "floats" in a certain range. This range depends on the width of the sliding window, that is, with a small window, the Doppler frequency range in the signal spectrum is small and the desired Doppler frequency shift is easily determined, and with a large window, this frequency range is larger and there is an inaccuracy in determining the Doppler frequency shift.

The small window is too sensitive to small changes in the Doppler frequency caused by a change in the projection of the velocity on the direction of wave propagation and oscillations of the bubble. In this work, a wider window is used to measure the speed of the bubble as a whole and to exclude the influence of minor vibrations of the walls of the bubble and the flows of the environment.

Let us estimate the described inaccuracy as the ratio of the range of frequencies present in the signal to the maximum value in the spectrum:

$$\sigma_{FFT} = \frac{\Delta_{\text{freq}}}{f_{\text{Doppler}}} \approx 200\ \text{Hz}\ 2000\ \text{Hz} = 10\%.$$

Thus, we can make a rough upper estimate of the total error in measuring the velocity of the bubbling bubble:

$$\sigma \cong \sigma_g + \sigma_{ADC} + \sigma_{FFT} = 10.59\%.$$

4 Development of Sensors Based on Piezoelectric Transducers

The source and receiver of ultrasonic waves are the main working elements in any method of ultrasound diagnostics. To carry out experimental studies within the framework of this work, it was decided to develop and manufacture sensors independently, adapting them to the task at hand and eliminating the presence of "black boxes" in the measuring system.

Thus, a few sensors based on piezoceramic disk transducers (Fig. 7) made of barium titanate with a diameter of ≈6 mm and a thickness of ≈0.6 mm were developed.

During the development of the sensor construct, it was necessary to solve two main problems:

1. Ensure complete isolation of the sensor contacts, because measurements are made in an electrically conductive environment
2. To create a structure for shielding the sensor from electromagnetic waves, thereby excluding the influence of interference and other fields on measurements

The first task was solved by multilayer varnishing of the transducer and contacts (Fig. 16a–c). In one of the versions (Fig. 16d), the transducer was poured with epoxy resin into the sensor body, but this approach had a negative impact on the amplitude of ultrasonic vibrations.

For shielding, several versions of the copper body were created (Fig. 16a–c). The body consists of two parts: a flat substrate and a cylindrical part, in which the piezoelectric element was attached. Plasticine (Fig. 16b, c) and sealant (Fig. 16a) were used as a damper between the substrate and the piezoelectric element. The braid of the coaxial cable was soldered to the body, which ensured its grounding.

Special attention was paid to the development of a submersible sensor for experiments in liquid metal media since the metal has a high electrical conductivity. Figure 8e and d shows such a sensor. Its structure is like other sensors, except for the mount and the addition of an additional insulating rubber housing cover.

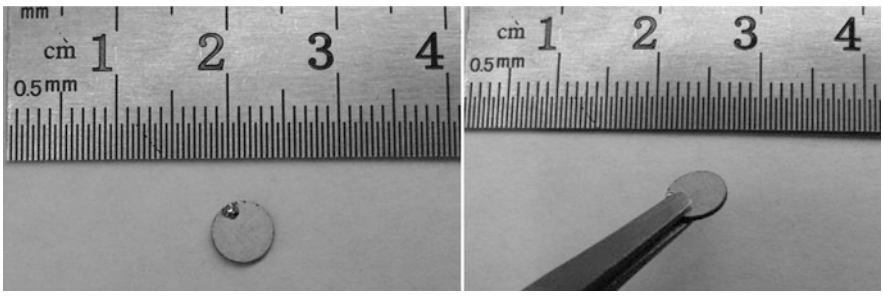


Fig. 7 Photo of piezoceramic transducers

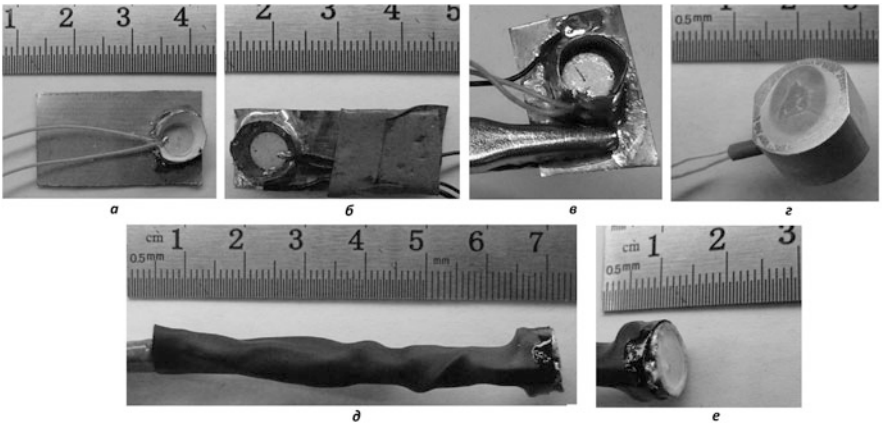


Fig. 8 Photo of the developed ultrasonic sensors

The developed sensors were tested in various operating modes (continuous, pulsed, mixed). The frequency response of the sensors was measured, as a result of which a resonant frequency of 4 MHz was revealed. The directivity of the radiation of the sensors was also experimentally investigated, and it was found that the sensors have an extremely small angle of divergence of the ultrasonic wave at a distance from 5 mm to 30 cm, that is, give directional "narrow" radiation in the specified range of distances.

5 Experiments

Ultrasonic Diagnostics of the Hydraulic Media

In experimental studies, at the first stage of operation, the liquid metal medium is modeled by the hydraulic medium. This greatly simplifies the experiment process and allows for alternative control using optical methods. At the second stage of the work, experiments are carried out in liquid metal, considering the peculiarities of such a medium.

As a phase inhomogeneity, air bubbles are injected into the medium and float to the surface of the liquid.

In experimental studies, the hardware-software complex described in the previous chapter is used. The data recording time from the transducers is 2 s, that is, a series of single short experiments of 2 s each are carried out.

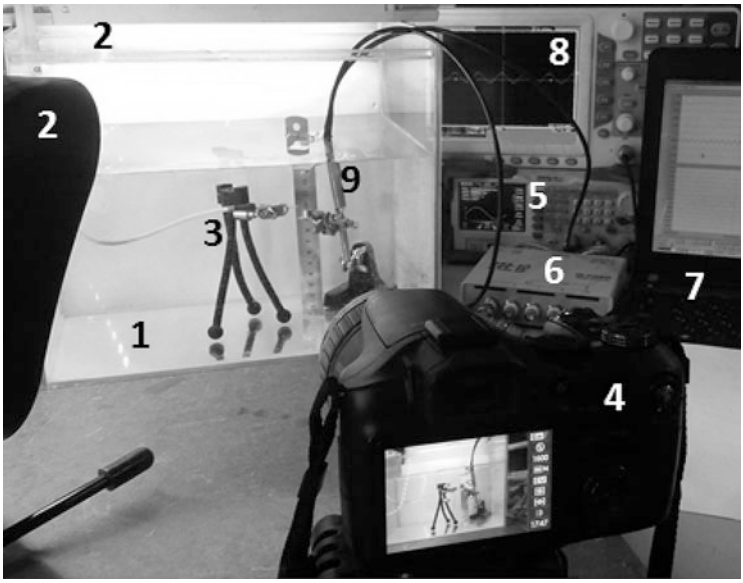


Fig. 9 Photo of the experimental stand with a software and hardware complex. (1) Aquarium with water, (2) illuminators, (3) air supply system, (4) video camera, (5) generator, (6) ADC, (7) PC, (8) oscilloscope, (9) measuring part (source-receiver)

Experimental Stand

To carry out experimental studies in a hydraulic environment, a compact stand is required, which makes it possible to quickly and conveniently align ultrasonic sensors, carry out alternative control using optical methods, and change the configuration of the measuring part of the complex. The developed stand solves the listed tasks and contains the following elements: glass aquarium ($30 \times 25 \times 20$ cm^3), filled with water, illuminators, high-speed (1200 frames per second) camera of the Casio EXILIM EX-F1 brand, air supply system for creating small (0.01–0.05 cm^3) bubbles, and oscilloscope brand Aktakom ADS-20131V (bandwidth, 30 MHz; sampling rate, 250 MSa/s) for signal visualization and measurement control.

Figure 9 shows a photo of the experimental stand with a software and hardware complex, and Fig. 10 – its measuring zone.

The experimental stand in conjunction with the software and hardware complex allows measurements with any angle of inclination of the transducers α and with a distance between the transducers in the range from 5 mm to 30 cm.

The descriptions and results of experiments with different configurations of the measuring part of the complex are given below: vertical, inclined, and horizontal configurations, as well as the reverse configuration with reflection of ultrasonic waves from the bubble.

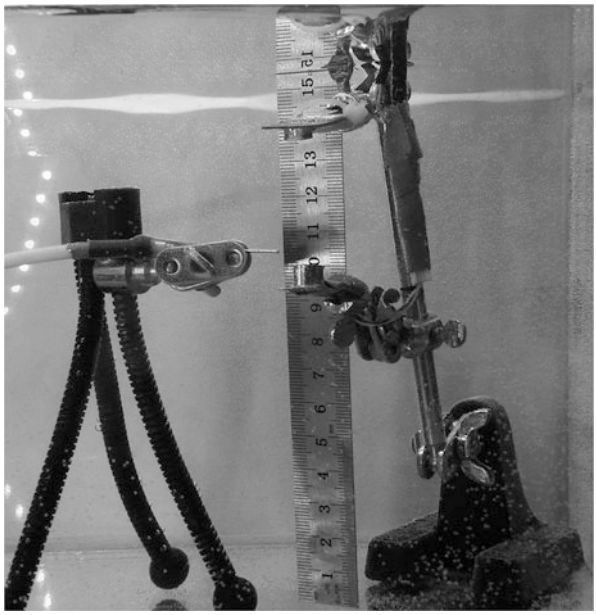


Fig. 10 Photo of the measuring area of the experimental stand

Vertical Configuration

In this configuration, the angle $\alpha = 0$ °. Figure 11 shows a photo of a bubble rising from a high-speed camera.

As a result, the following Doppler signal was obtained (Fig. 12), in which it is possible to clearly distinguish between the various stages of the bubble ascent: inflation, separation, ascent, and drift along the surface of the liquid.

After filtering from the low-frequency component (Fig. 13) and the sliding Fourier transform, the dependence of the Doppler frequency shift on time was obtained (Fig. 21, left Y axis) and, accordingly, the dependence of the bubble velocity on time (Fig. 21, right Y axis and Fig. 14).

Thus, with the help of the developed hardware and software complex, the dynamics of bubble ascent in a vertical configuration was established. The average bubble ascent velocity is 0.24 m/s, which is confirmed by the analysis of video recording from a high-speed camera.

Also, for the integral analysis of ascent, the spectrum of the "ascent" section of the Doppler signal was built (Fig. 15).

The presence of a relatively wide range of frequencies in the spectrum is due to the following factors: the emerging bubble is accelerating; the projection of the bubble velocity on the direction of propagation of the ultrasonic wave changes with time; in addition to the main movement of the bubble, there are vibrations of the walls of the bubble and the environment.

The Inclined Configuration

In an inclined configuration, the angle is $\alpha \approx 30°$ (Figs. 16 and 17).

Fig. 11 Photo of the bubble rising

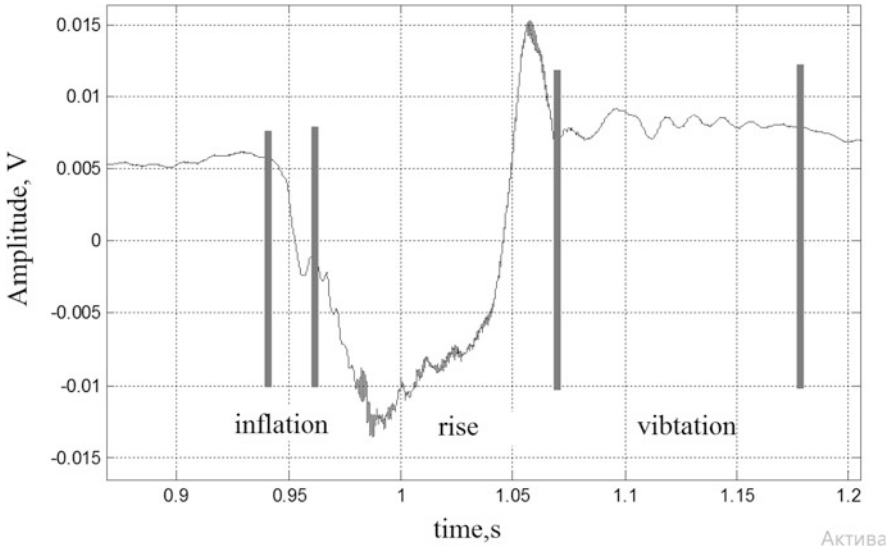


Fig. 12 Doppler signal

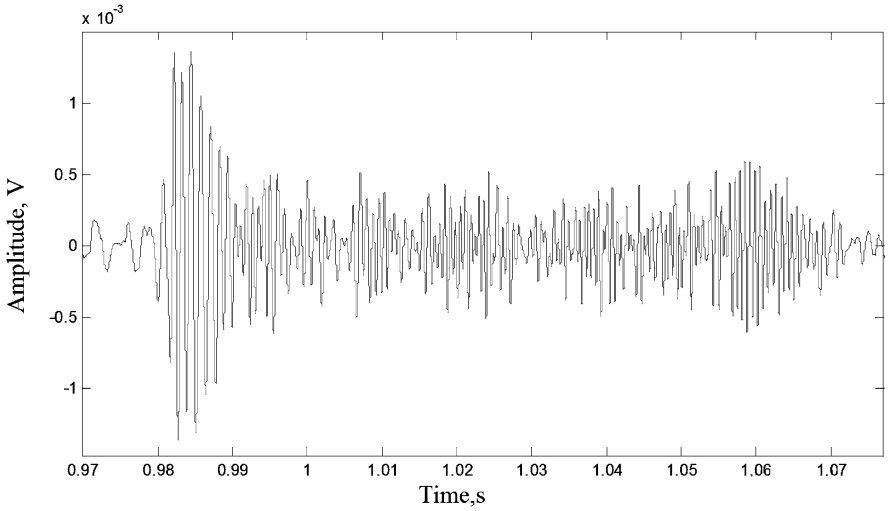


Fig. 13 Doppler signal without low-frequency component

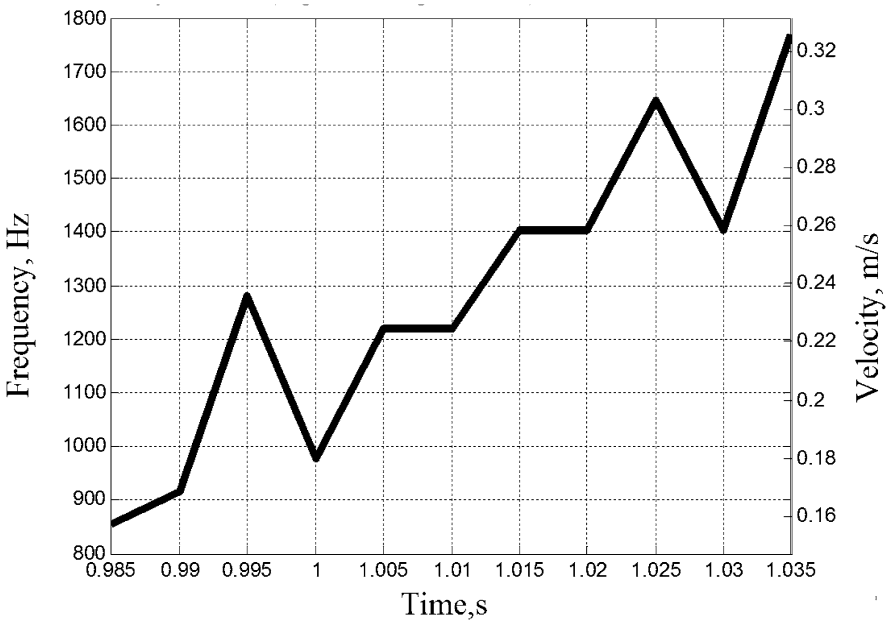


Fig. 14 Doppler frequency shift and rising rate versus time

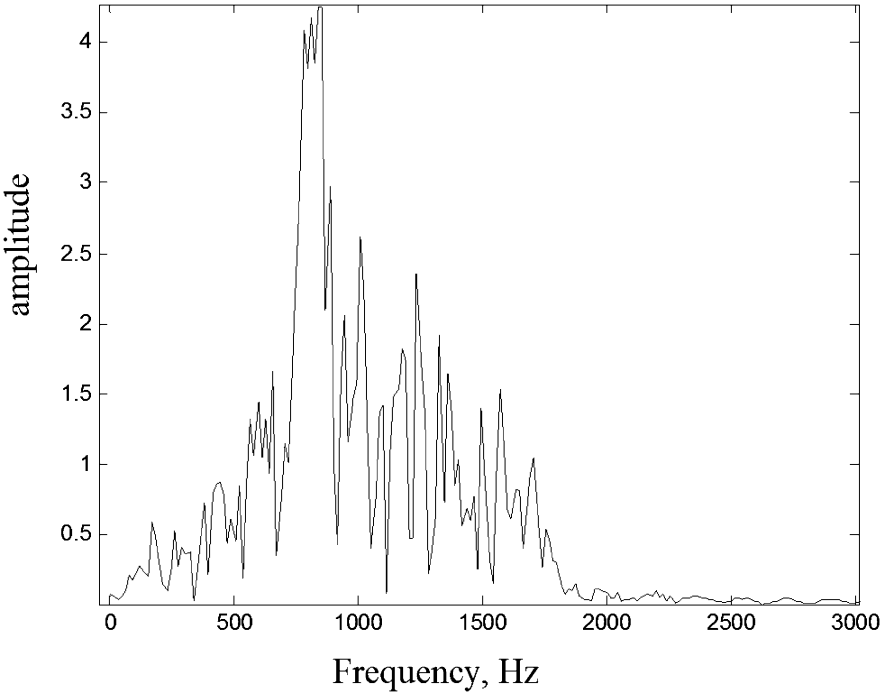


Fig. 15 Doppler signal spectrum

Similarly, to the vertical configuration, the time dependence of the Doppler frequency shift (ascent rate) was obtained, shown in Fig. 18.

The observed decrease in speed is explained as follows. When the bubble rises, it has a non-rectilinear trajectory, as a result of which, flying through the measurement zone, the projection of the velocity on the direction of ultrasound propagation can change, which we observe in Fig. 18. Figure 19 shows a series of frames from a high-speed camera explaining the images obtained using the software and hardware complex bubble ascent data.

Horizontal Configuration

In this configuration, $\alpha \approx 90°$. Figure 20 shows the bubbling of bubbles in a horizontal configuration.

This configuration does not allow the full use of the Doppler method, because the projection of the bubble velocity on the direction of ultrasound propagation is small. But the amplitude-shadow method is very well applicable. As a result of the operation of the software and hardware complex, the signal shown in Fig. 21 was received.

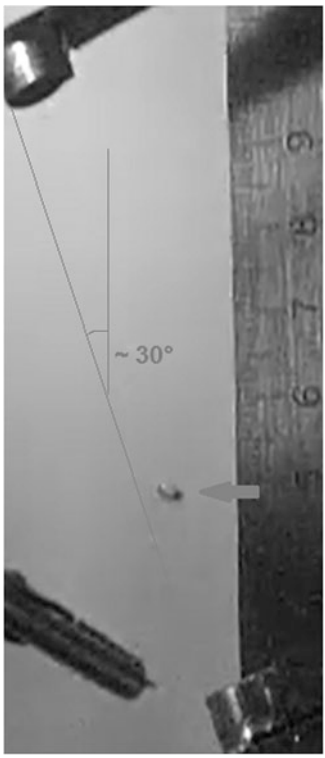


Fig. 16 Photo of a pop-up bubble

This dependence makes it possible to detect bubbles and estimate their size and even velocity, knowing the duration of one amplitude burst and the width of the measurement zone in the medium.

Reverse Configuration

This configuration is based on the reflection of the ultrasonic wave from the walls of the bubble. The source was positioned so that the wave reflected from the bubbling bubble and hit the receiver. Figure 22 shows the bubble rising in the reverse configuration.

The received signal from the receiver has amplitude flashes due to the reflection of the ultrasonic wave or bubbles. Figure 23 shows a graph of the signal from the receiver with flashes from two successively emerging bubbles. Based on this signal, it is possible to judge the presence of bubbles in the measurement zone and their size. The Doppler method is also applicable in this configuration.

Ultrasonic Diagnostics of Liquid Metal Media

For experimental studies in a liquid metal environment, an alloy of the following metals was prepared in the indicated proportion: gallium (62%), indium (25%), and tin (13%). The resulting alloy has a melting point of 4.85 °C. This alloy has low

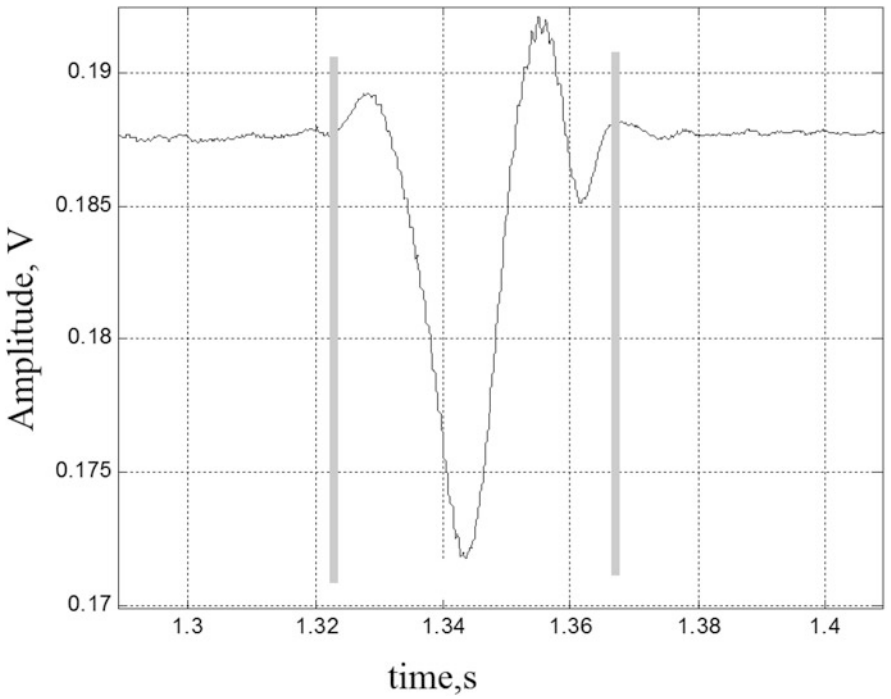


Fig. 17 Doppler signal

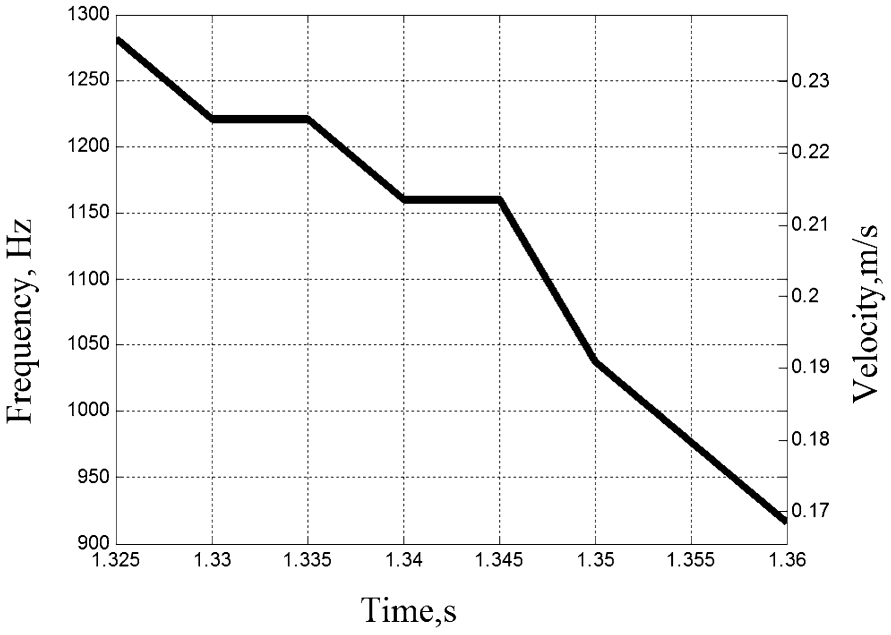


Fig. 18 Time dependence of the Doppler frequency and velocity shift

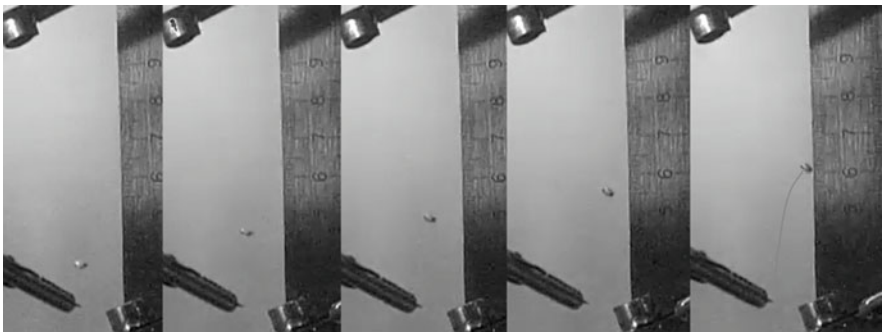


Fig. 19 A series of frames from a high-speed camera

Fig. 20 Photo of bubbles rising in a horizontal configuration

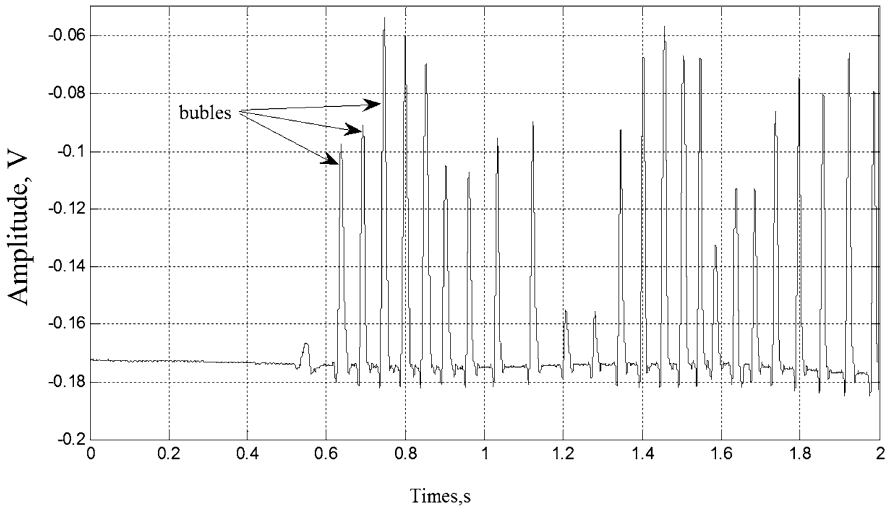


Fig. 21 Bubbles rising. Doppler signal

Fig. 22 Photo of bubbles rising in reverse configuration

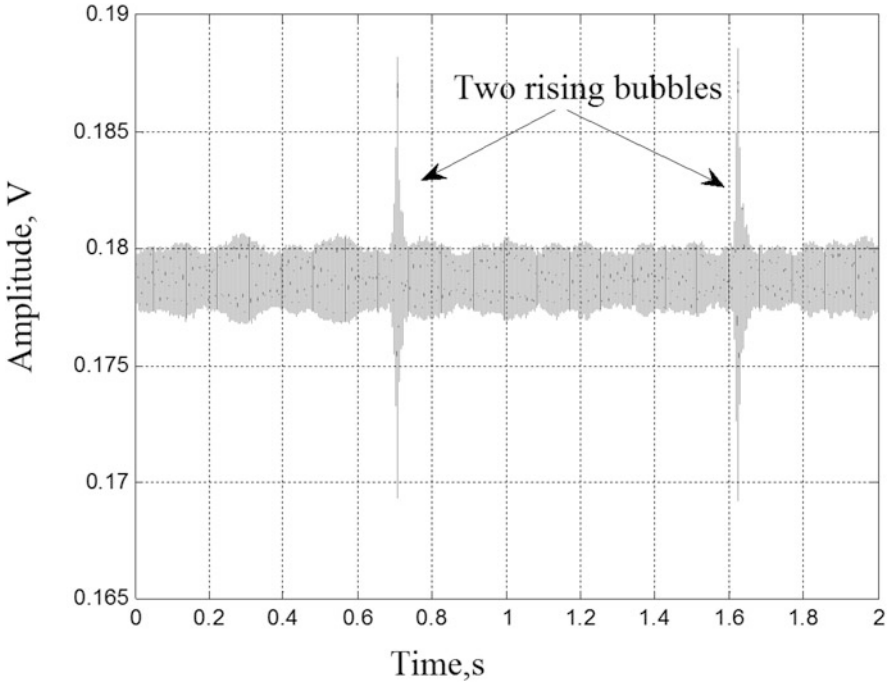


Fig. 23 Signal from the receiver. Reverse configuration

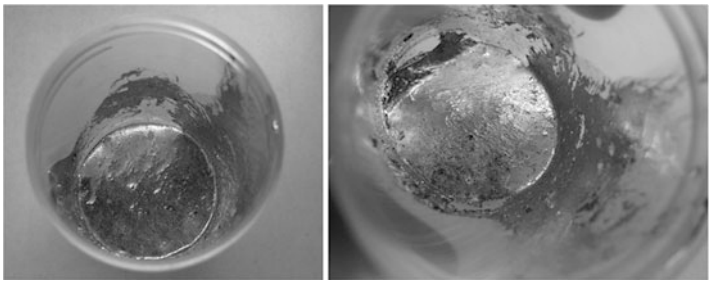


Fig. 24 Alloy of gallium, indium, and tin at room temperature

Fig. 25 Stand for experiments in a liquid metal environment

toxicity, is often used as a coolant, and is also convenient for laboratory experiments. A photo of the alloy is shown in Fig. 24.

Experimental Stand

The experimental stand used for research in a hydraulic environment is too large and ineffective for research in a liquid metal medium, because metal alloy is expensive. Therefore, a stand was created (Fig. 25), adapted for the liquid metal environment. The stand design has not fundamentally changed. In the stand, the measuring volume is minimized to 20 cm^3 in order to save metal and the system for supplying air to the medium. A specially developed submersible ultrasonic sensor was used in the hardware-software complex.

The Inclined Configuration

In a liquid metal medium, the following feature of the experiment occurred: a larger air bubble was formed than in water, due to a different value of viscosity. This

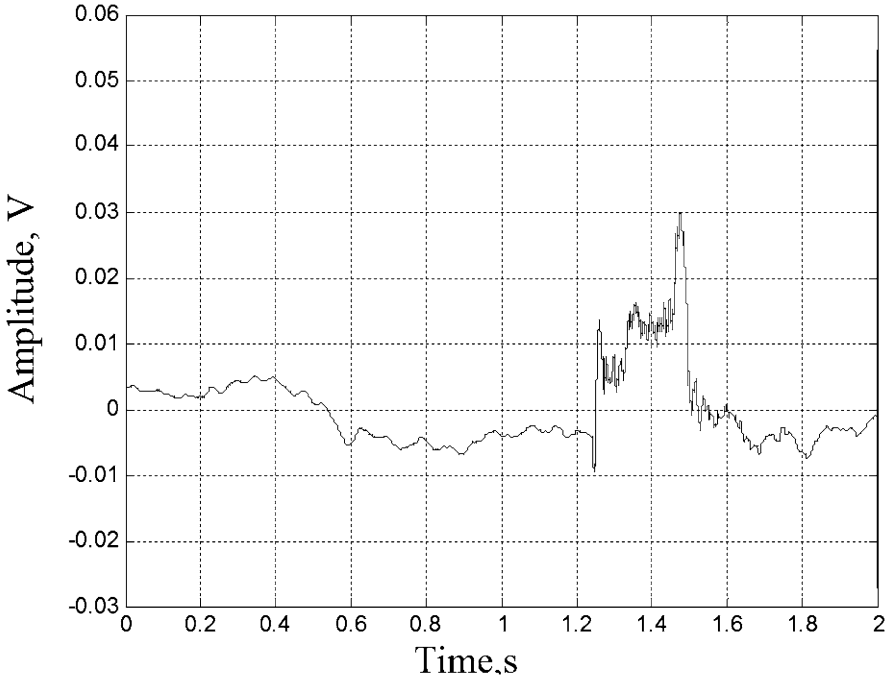


Fig. 26 The Doppler signal. Rise of a bubble in a liquid metal medium

bubble was squeezed out to the wall of the container with the metal, as a result of which the bubble floated up more slowly due to friction with the wall. Despite all this, the experiments gave good results, because it was possible to track the dynamics of the bubble ascent like the experiment in water. Figure 26 shows the Doppler signal from a single bubble ascent, and Fig. 27 shows the time dependence of the Doppler frequency (velocity) shift. This dependence confirms that the average bubble rise rate in the liquid metal is lower. The form of the dependence indicates that the projection of the velocity on the direction of wave propagation is constantly changing. This is due to the peculiarities of the geometry of the experimental stand and the hardware and software complex.

6 Conclusion

Within the framework of this work, the following results have been achieved:

1. A complex method for ultrasonic diagnostics of two-phase flows in a liquid metal medium is proposed. The method allows to measure the dynamic and structural characteristics of a floating inhomogeneity (gas bubble).

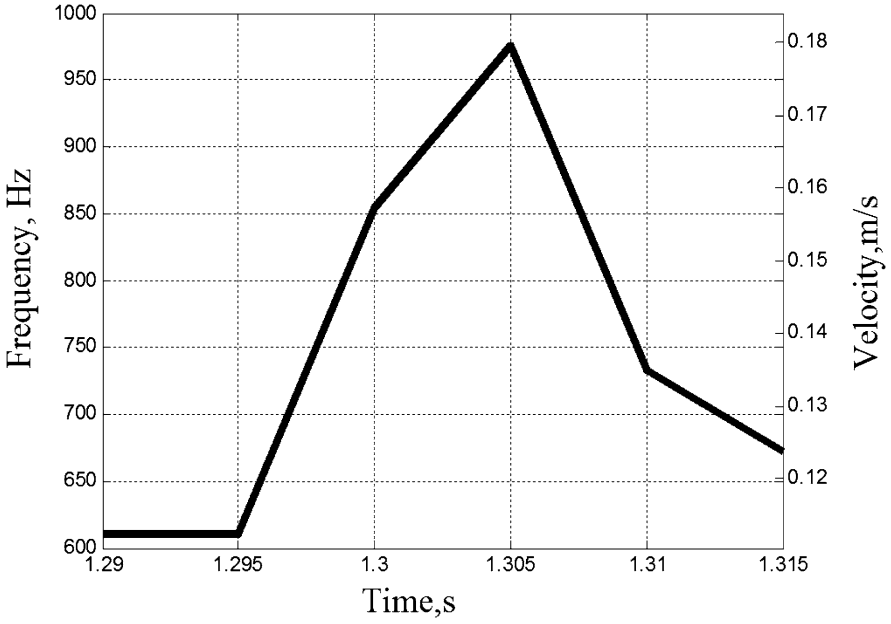


Fig. 27 The time dependence of frequency and velocity Doppler shift

2. A laboratory sample of a software and hardware complex has been developed. The complex is simple and flexible to use and allows measurements in different configurations.
3. Based on piezoelectric transducers, a series of ultrasonic submersible sensors has been developed and implemented, which allow working in a fluid and liquid metal. The sensors are universal and can work both as a receiver and as a transmitter.
4. Compact experimental stands have been developed and implemented for research in a hydraulic medium and in a liquid metal and for testing the developed complex.
5. The developed complex was tested in a liquid metal environment and its performance was shown.
6. The results of experimental studies show that the proposed method makes it possible to detect floating gas bubbles in a liquid medium, track their dynamics, and localize them in a volume of 5–50 cm^3.

Acknowledgments The study of the flow was carried out within the framework of the IT SB RAS state assignment (project 121032200034-4), the development of the software and hardware complex was carried out at the expense of the grant MD-1767.2020.8, and the method development was carried out within the framework of the IT SB RAS state assignment (AAAA-A19-119052190039-8).

References

1. Meledin V.G. Informatics of optoelectronic measurements: science and innovative industrial technologies / Novosibirsk: Publishing house of IT SB RAS, 2008. - 75 p.
2. Akatov A.A., Koryakovskiy Yu. S. Future of nuclear power engineering. Fast reactors. Atomic energy information centers. Moscow, 2012 - 36 p.
3. Sahai and Guthrie. Hydrodynamics of gas stirred melts: Part I and II. Metallurgical and Materials Transactions B, 1982.
4. Kremlevsky P.P. Flowmeters and counters of the amount of substances: Handbook: Book. 2 / P.P. Kremlin - SPb.: Polytechnic, 2004.
5. A. I. Mogilner, S. A. Morozov, S. O. Zakharov, A. Yu. Uralets. Detection of gas bubbles in a liquid metal coolant using magnetic flowmeters // IPPE, Obinsk - 1986.
6. I. Kokoreva, G. Shchelkunov. X-ray methods of non-destructive sensing. Electronics: Science, Technology, Business 5/2007.
7. K. V. Santhosh, B. K. Roy. An Intelligent Flow Measurement Technique using Ultrasonic Flow Meter with Optimized Neural Network // International Journal of Control and Automation, Vol. 5, No. 4, December, 2012
8. Andruszkiewicz et al. Ultrasonic measurements of flow in two-phase liquid gas systems I - III. Chemical and Process Engineering, 2008
9. E.C. Chistyakov and Yu. I. Dyshleva. A.S. 1631401 USSR - Method for monitoring fluid flow discontinuities in a pipeline (prototype).
10. Kuzyakov ON; Dudko S.A. Patent No. 2198397 Method for multiphase flow control in a pipeline.
11. Kosarev V. I., Mukhin L. N., Shternov A. A. Patent No. 2339915. Method for determining the flow rate of two-phase flow components and a system for its implementation
12. Dobrynin V. V., Shemetun G. K., Kosarev V. I. Patent No. 2425332. A method for the simultaneous determination of the flow rates of the liquid and gas phases of the flow of a gas-liquid mixture.
13. Leshkov VV, Shkolyarenko VV, Taranin VD Gas control device in liquid metal heat carrier. Patent No. 2426111.
14. Leshkov V.V. Gas analyzer. Patent No. 2249203.
15. I.N. Ermolov, M.I. Ermolov. Ultrasonic testing. A textbook for specialists of the first and second qualification levels. Fifth edition, stereotyped. –M: 2006 - 208 s: 77 ill.
16. Andruszkiewicz A., Eckert K., Eckert S., Odenbach S., 2013. Gas bubble detection in liquid metals by means of the ultrasound transit-time-technique. Eur. Phys. J. Spec. Top. 220 (1), 53–62.
17. T. Richter, K. Eckert, X. Yang, S. Odenbach, 2015. Measuring the diameter of rising gas bubbles by means of the ultrasound transit time technique. Nuclear Engineering and Design 291 (2015) 64–70.
18. T. Vogt, A. Andruszkiewicz, K. Eckert, S. Odenbach, S. Eckert, G. Gerbeth. Ultrasonic flow measurements and bubble detection in gas-stirred metallic melts. 8th International Symposium on Ultrasonic Doppler Methods for Fluid Mechanics and Fluid Engineering (2012).
19. Mozhaev VG Features of propagation and nonlinear interaction of acoustic waves in piezocrystals with plane and slightly curved boundaries: dis. ... Cand. phys.-mat. Sciences: 01.04.06 / Vladimir Gennadevich Mozhaev; Acoustic Institute named after Ak. N.N. Andreeva. - Moscow time, 1984. -192 p.

3D Reconstruction of Human Body Biometry

Gabriel Trujillo-Hernández, Wendy Flores-Fuentes, Julio Cesar Rodríguez-Quiñonez, Daniel Hernández-Balbuena, Oscar Real-Moreno, Jesús Elías Miranda-Vega, and Vikrant Bhateja

1 State of the Art

Nowadays, several diseases have affected the lifestyle of persons due to deformities in different areas of the body such as the feet [1], spinal [2], and chest [3]. Therefore, researchers and medical professionals have worked together to develop new systems that assess detailed measurements of the body's size and shape to provide effective diagnoses and treatment. However, there are diverse technologies to measure these deformities such as photogrammetry [4], 3D laser scanning [5], and structured light [6]. 3D laser scanner has demonstrated satisfactory results in medical applications in the creation of point clouds to represent the area of interest on the human body enabling medical professionals to detect deformities and analyze treatment's effectiveness.

3D scanning systems can be classified into measurements (geometry) and visualization (texture). According to Treleaven and Wells [7], also, healthcare application measurements can be divided into four groups: epidemiology, diagnosis, treatment, and monitoring. Following, we present several diagnostic applications, where the 3D laser scanning is used to detect deformities in areas such as the feet, spinal, head, face, and chest.

Feet deformities cause pain and injuries in persons that suffer congenital diseases; therefore, it is necessary to develop new systems, which enable the detection of deformities on the feet such as flat feet, cavus feet, valgus feet, and diabetic feet.

G. Trujillo-Hernández · W. Flores-Fuentes (✉) · J. C. Rodríguez-Quiñonez
D. Hernández-Balbuena · O. Real-Moreno · J. E. Miranda-Vega · V. Bhateja
Universidad Autónoma de Baja California, Mexicali, México

Tecnológico Nacional de México/IT de Mexicali, Mexicali, México

SRMGPC, Lucknow, (U.P.), India
e-mail: flores.wendy@uabc.edu.mx

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1_8

A novel technic employed to detect deformities on the feet is the 3D laser scanning; this technology captures the shape of the feet for creating a 3D point cloud, which provides detailed information that enables the development of effective treatments such as orthopedic insole. According to Pfeiffer et al. [8], a 3D laser scanner is employed to digitize the shape of the feet for establishing a prevalence of flat feet in a population of 3–6-year-old children; the obtained 3D point cloud enables to characterize the foot arch shape and determine the foot type. Savic et al. [9] uses a 3D laser scanner to capture the sole of the feet for manufacturing a customized insole. Subsequently, the insole is used to perform a stress comparison between the uses of a customized insole and when a customized insole is not used.

Serious health problems can be caused by chest deformities in persons that suffer congenital diseases such as pectus excavatum (PE) and pectus carinatum (PC). Pectus excavatum is associated with depression of the sternum; this defect could affect cardiac and pulmonary function causing pain in the chest, fast heart rate, and tiredness in physical activities. Unlike pectus excavatum, pectus carinatum consists of excessive growth of the costal cartilage, which typically results in protrusion of the sternum that does not affect the body function, and it is only a cosmetic problem on the chest [10, 11]. Lain et al. [12] employed a 3D laser scanner for creating a tridimensional representation of the affected area by pectus excavatum disease, and with this, a 3D model can reconstruct the area of interest to evaluate the severity of the deformity. In addition, 3D laser scanning applications enable creating a 3D model before or after treatment like surgery to evaluate the changes.

Face deformities are defects caused by a group of conditions such as chin deviation, cleft lip, cleft palate, and nose deformity. Several research highlights compare the efficacy of botulinum toxin in masseter muscle depending on the chin deviation. Therefore, a 3D laser scanner is used to measure the volume and bulkiest height of the lower face [13]. Otherwise, Verzé, Bianchi, and Ramieri [14] employed a 3D laser scanner to perform a study that compares 12 patients with maxillary retrognathic dysplasia and nose deformity.

On the other hand, spinal deformities are caused by diseases such as scoliosis and kyphosis. Scoliosis is a multifactorial three-dimensional disease that involves three main planes: lateral curvature in the frontal, anteroposterior deviation in sagittal, and vertebral axial rotation in horizontal plane. Otherwise, kyphosis is a posterior deviation of the spine in the sagittal plane beyond normal limits; this disease may result in pain, cardiopulmonary failure, and paraplegia. 3D laser scanners can be employed to detect deformations on the spinal. According to Rodríguez et al. [15], a 3D medical laser scanner based on the principle of dynamic triangulation is employed to analyze the method of operation, medical application, and orthopedical diseases as scoliosis. Also, the tridimensional model enables to illustrate the effects of the treatments.

Following the topic of human body deformities is important to mention the nonsynostotic head deformities, which occur more often in newborn babies. These deformities not only are a cosmetic problem but also can increase the risk of devel-

opmental delays. Therefore, several methods have been implemented to analyze the defect in the area. One of these methods is the 3D laser scan, which is used to obtain quantitative data on the shape of the head. In order to determine the prevalence of symmetrical and asymmetrical head deformities, with this information, possible risk factors can be identified [16].

As previously mentioned, 3D laser scanning has a wide variety of applications related to deformity detection on the body. In each application, it is necessary to scan accurate surfaces in order to provide a reliable assessment diagnosis. However, the performance of the 3D laser scanner can be affected by different factors such as ambient light, noise, and waste energy due to the reflectance of the laser. The effects cause distortion on the 3D point cloud preventing to provide reliable analysis and diagnostics. Therefore, it is necessary to employ methods that attenuate the distortion on the 3D point cloud.

The evolution of informatics systems and technologies has enabled the development of 3D reconstruction algorithms and mesh methods, which enable to perform the relation of representative points of the object to create surfaces using a triangle, squares, or any geometric form. 3D laser scanner improves when 3D reconstruction algorithms are employed, because it is possible to apply two kinds of reconstruction known as geometric and volumetric that enable to organize and optimize point clouds in order to improve the surface visualization. Even atypical measurements due to scanning systems and uncertainties that could distort the point cloud imaging during the scanning can be eliminated or attenuated as part of the reconstruction process.

This chapter provides a broad knowledge of different methods applied in 3D reconstruction, which could be applied in the medical area using reconstruction software. To perform a practical application of the presented methods and concepts, a biometric shape dataset previously obtained by a 3D laser scanner will be analyzed applying geometric reconstruction (polygon mesh (PM), Bezier curves) and volumetric reconstruction methods (Voronoi diagrams, Delaunay triangulations) to improve the 3D model shape and reduce the required number of points to perform a reliable characterization of the measured biometric shapes.

2 Geometric Reconstruction

Geometric reconstruction applies methods for analyzing and manipulating point clouds in order to obtain a defined representation of the scanned object and efficient processing. There are several applied methods to perform a geometric reconstruction of a point cloud obtained by laser scanning or stereo vision [17]. This section presents the polygon mesh and Bezier curve methods, which are employed in the majority of the reconstruction software.

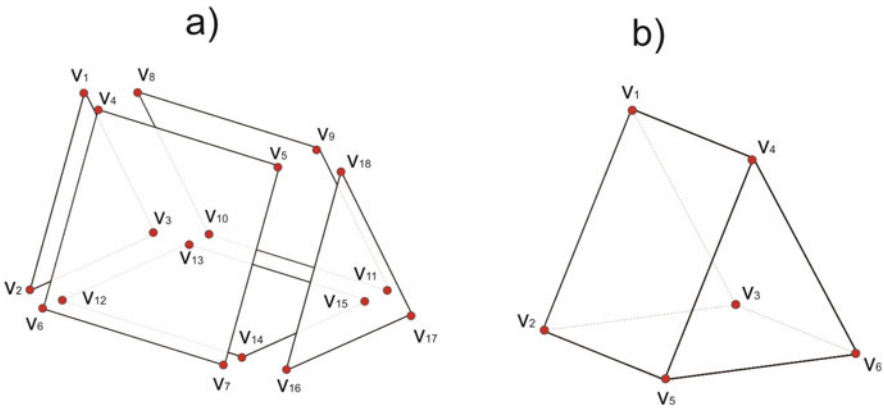


Fig. 1 (**a**) Prism exploded with 18 vertices. (**b**) Prism joined six vertices

2.1 Polygon Mesh

A widely employed method in reconstruction is polygon mesh (PM), which is used to connect vertices and create shapes in a point cloud. Generally, the shapes can be represented according to vertices, edges, and faces for creating shapes such as triangles, quadrilateral, and other polygons. A triangle is formed when three vertices are joined, whether four points are bounded a quadrilateral, but whether more than four vertices are related is formed a simple polygon. In each case, a face is obtained; hence, 3D shapes are formed when faces are joined [18].

The vertices are joined to form faces to reduce points on the 3D model to create a defined representation and highly efficient processing in the handle of the point clouds. Whether a triangular prism is exploded will be divided into two triangular faces and three rectangular faces for each one is required three and four vertices, respectively, i.e., the sum of all vertex's faces will be 18. Otherwise, whether faces are joined only will be necessary six vertices, to create a triangular prism [19]. Figure 1 shows a triangular prism and vertices with exploded faces and joined.

2.1.1 Representation of Meshes

Each triangle, rectangle, or simple polygon formed on the point cloud can be represented by strips or meshes; some alternatives employed are separate triangle, indexed triangle, rectangular, fan, strip triangles, and strip rectangles; and each one presents different patterns to create points that will be shown below.

Strip triangles consist of a sequence of triangles formed by three vertices; for the first triangle are necessary three vertices; however, for the subsequent triangles, only one vertex is required. A variant of the strip triangles is strip fan, which presents the same pattern for creating strip triangles [20]. However, this follows the trajectory of

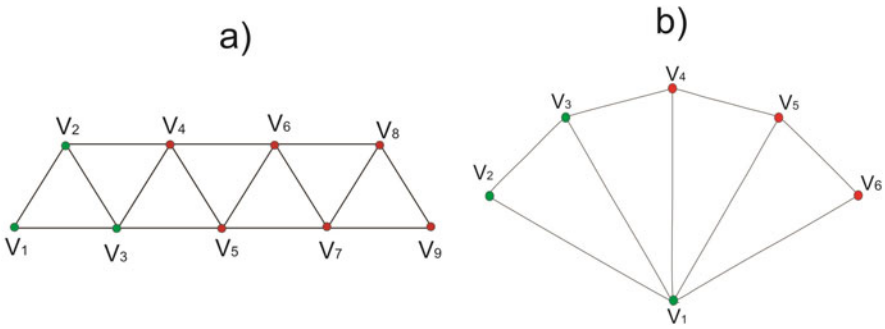


Fig. 2 Strip triangle (**a**) and strip fan (**b**) vertices V_1, V_2, V_3 (color green) represent the first triangle; subsequently only one point will be necessary

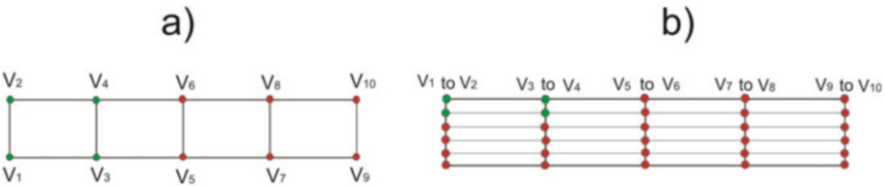


Fig. 3 (**a**) Strip rectangular (**b**) Rectangular mesh

a fan and allows the creation of shapes that would not be possible to get by the use of the strip triangle mesh type. Figure 2 shows the patterns to form the two types of strips.

Strip rectangles use four vertices to create the first rectangles; afterward, for the next rectangles, only two vertices are required. Also, rectangular mesh presents the same pattern as strip triangles. However, rectangular mesh consists in a matrix $m \times n$; moreover, for each first rectangle of the rows are necessary four points; afterward, for the next rectangles, only one point is required. A disadvantage of rectangular mesh is the ineffectiveness to represent different surfaces such as the sphere [21]. Figure 3 shows the patterns to form the two types of strips.

There are some occasions where it is impossible to represent an object using polygons; therefore, it is necessary to apply methods like separate triangle. This method consists of extracting topological information of separate polygons such as coordinates and vertices that are related to them. The topologic information is divided using different techniques, which enable separation of the vertices and polygons in individual lists. Whether the information of each polygon's vertex is stored, the same vertex can be used to represent different polygons and can be called to create different shapes. This technique, known as the indexed triangle set, enables to save storage space [23]. In Fig. 4, if the polygon T_1 is analyzed, we can note that it is formed by V_0, V_1, and V_2, and this information is stored in Table 1; afterward, to form the polygon T_2, it will only be necessary to call the index of V_0, V_1 and to call V_3 of the polygon T_3. Table 2 shows the indexes necessary to form T_1 and T_2.

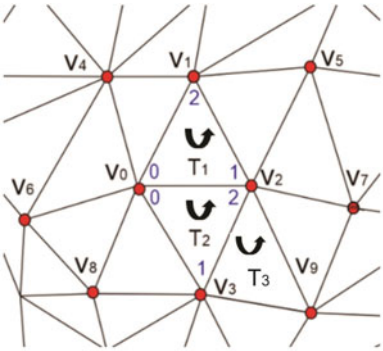


Fig. 4 The information of the point cloud is separated into two lists vertices: (V_n) and polygons (T_n)

Table 1 Values of T1 and T2

Indexes	Vertices	Coordinates
0	V_0	x_0, y_0, z_0
1	V_1	x_1, y_1, z_1
2	V_2	x_2, y_2, z_2
3	V_3	x_3, y_3, z_3

Table 2 Polygons of T1 and T2

Polygons	Indexes
T_1	0, 2, 1
T_2	0, 3, 2

2.2 Bezier Curves

Polygon mesh allows to create objects using triangles and rectangles; however, another method widely applied in geometric reconstruction is Bezier curves. These are polynomic curves based on Bernstein polynomial due to both having the same extreme control vertices and coincident tangents. Therefore, a polynomic curve has a Bezier representation and is defined as a polynomial contour. In addition, a Bezier curve is classified depending on the number of control vertices [24]. Following, some typical curves are presented: liner curve, quadratic, and cubic. Bezier curves can be implemented using Bezier and Casteljau algorithm.

2.2.1 Bezier Algorithm

Linear curve, quadratic, and cubic can be represented easily replacing linear equation in the general equation; however, every time the degree of the equation increases, it is more difficult to develop the equation of the Bezier curve, because it will be necessary to perform a greater number of replacements. Therefore, Bernstein polynomial is used to reduce the number of replacements, because it is only necessary to replace each control vertex from i to n [25]. According to Song et al. [26], the equation for a Bezier curve of order n can be expressed by (1):

$$B(t) = \sum_{i=0}^{n} b_{i,n}(t) P_i \tag{1}$$

where $t \epsilon [0,1]$. P_0 is the start point of the curve, P_n is the end point, and $P_i (0 < i < n)$ is the control vertex. $b_{i,n}$ is given by Eq. (2):

$$b_{i,n}(t) = \binom{n}{i} t^i (1-t)^{n-i} \quad i = 1, 2, 3 \ldots n \tag{2}$$

2.2.2 Casteljau Algorithm

We have seen that a Bezier curve can easily be calculated using polynomial Bernstein; however, sometimes it is difficult to represent them. Therefore, another broadly applied method is Casteljau algorithm, which is the most efficient way to evaluate Bezier curves. This algorithm is equivalent to Bernstein polynomials and consists of a successive linear interpolation between two points.

Linear Curve

A linear curve (u_0) is a first-degree equation based on two control vertices; to describe the trajectory of the Bezier curve, it is necessary to vary the t parameter from the first control vertex (a_0) to end vertex (b_0) [27]. The equation u_0 is given by (3):

$$u_o = a_0 \ (1-t) + a_1 t \quad \text{for } a \neq b \tag{3}$$

Quadratic Curve

A quadratic curve (q) is formed by two linear curve equations u_1 and u_2 (4, 5); therefore, they are second-degree equations based on three control vertices. The curve equation q is given by (6):

$$u_0 = a_0 \ (1-t) + a_1 t \ \text{ for } a_0 \neq b_0 \tag{4}$$

$$u_1 = a_1 \ (1-t) + a_2 t \quad \text{for } b_0 \neq b_1 \tag{5}$$

$$q = u_0 (1-t) + u_1 t \ \text{ for } u_0 \neq u_1 \tag{6}$$

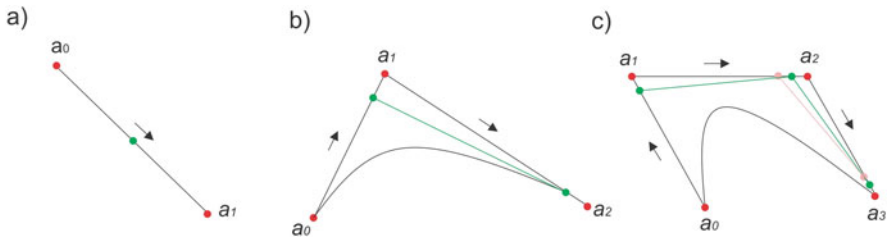


Fig. 5 The red points represent the control vertices in each curve; varying parameter t (green point), it describes the Bezier curves. (**a**) Linear curve. (**b**) Quadratic curve. (**c**) Cubic curve

Cubic Curve

A cubic curve (k) is formed by two quadratic Eqs. (7, 8), thereby they are third-degree equations based on four control vertices. The curve equation k is given by (9):

$$q_0 = u_0 \ (1-t) + u_1 t \quad u_0 \neq u_1 \tag{7}$$

$$q_1 = u_1 \ (1-t) + u_3 t \quad u_1 \neq u_3 \tag{8}$$

$$k = q_0 \, (1-t) + q_1 (t) \tag{9}$$

The number of control vertices defines shapes of the curve, i.e., whether there are more vertex complex curves can be created; therefore, it will be possible to perform a better reconstruction of the scanned object. Figure 5 shows linear, quadratic, and cubic curves.

Below is an example where can be shown linear Eqs. (10), (11) in quadratic curve q (12):

$$u_0 = a_0 \ (1-t) + a_1 t \ \text{ for } \ a_0 \neq b_0 \tag{10}$$

$$u_1 = a_1 \ (1-t) + a_2 t \quad \text{for } \ a_1 \neq b_1 \tag{11}$$

$$q = u_0 \, (1-t) + u_1 t \tag{12}$$

Replacing the linear curves u_0 and u_1 in q:

$$q = (1-t) \, (a_0 \ (1-t) + a_1 t) + t \, (a_1 \ (1-t) + a_2 t) \tag{13}$$

$$q = a_0\,(1-t)\,2 + a_1 t\,(1-t) + a_1 t\;\;(1-t) + a_2 t2 \tag{14}$$

$$q = a_0(1-t)^2 + 2a_1 t\,(1-t) + a_2 t2 \tag{15}$$

Depending on the variation of parameter t, Eq. (15) defines the trajectory of the Bezier curve, which is related to the vertices a_0, $a_{1,}$ and a_2.

3 Volumetric Reconstruction

Volumetric reconstruction consists of the representation of a point cloud as volume; this technology can be applied in a variety of applications especially in medical area because it enables rebuilding the point cloud of human body captured by laser scanning or photometry. A method widely used to transform a point cloud in a volume is voxel [28]. However, according to May [29], also Voronoi diagram can be applied to create a volumetric reconstruction of the material lithology. On the other hand, the application of volumetric reconstruction is highly applied to measure densities in a specific object.

3.1 Voronoi Diagram

A method widely applied in reconstruction of point clouds is Voronoi diagram, which is an efficient method that divides the point cloud in a particular Euclidean space, i.e., Voronoi diagram is the join of the regions (Voronoi cell) of each point of the cloud. This method is employed in reconstruction, thereby reducing the number of points in the scanned cloud, obtaining elasticity in structures, and enabling connecting nearby points and avoiding joining points not nearby [30]. Figure 6 shows the structure of a Voronoi diagram.

A Voronoi diagram is implemented applying some properties that depend on the edges, vertices, region, and Euclidean distances [31]. The following presents some of those properties:

- In a Voronoi diagram, each edge is located in the mediatrix of each line (p_i, p_j) that connect two points (neighbors) in the cloud.
- The neighbors nearest to each point p_i define one edge of the Voronoi diagram of the same point.
- In a Voronoi diagram, each vertex is equidistant with the points, i.e., the distance from V_n to g_k is the same. In addition, each vertex is the intersection between three edges.

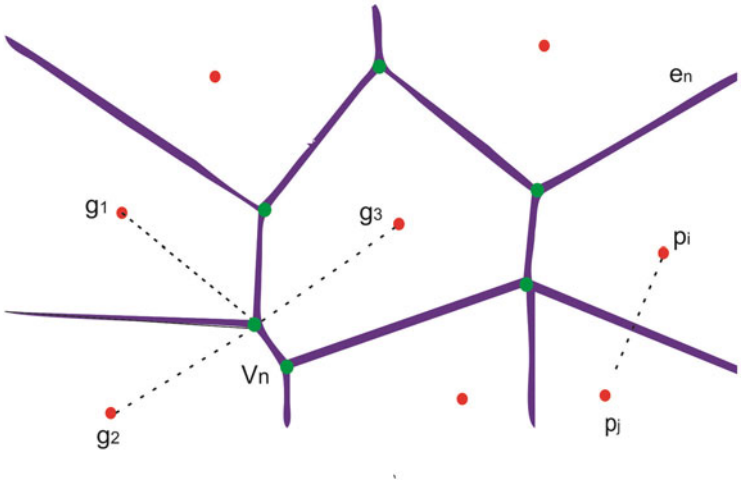


Fig. 6 e_n represents each edge of the Voronoi diagram; V_n indicates the vertices; g_k are the points used to create the vertices; and p_i and p_j are the points start and end of the imaginary segments used to draw the edges in the Voronoi diagram

- The regions in a Voronoi diagram can be classified as polygon convex or not convex; this classification depends on the localization of the points of each region, i.e., if the point is located into convex envelope is considered as polygon convex. Otherwise, whether the point is outside of the convex envelope is considered as not convex.
- An edge Vor(s) is formed, whether a point (o) is on the mediatrix of the point $p_i\ p_j$.
- There are several algorithms applied to create Voronoi diagram such as divide and conquer, incremental algorithm, and Fortune [32, 35]. In this section, divide and conquer and incremental algorithm are presented.

3.1.1 Divide and Conquer

Divide and conquer is an algorithm used to create Voronoi diagrams in a point cloud or set of point. This algorithm is considered faster and complex [33]. This algorithm can be applied using the following steps:

Step 1: Split using a line (l) the point cloud into two subsets (S_1 and S_2).
Step 2: Calculate the Voronoi diagram of $Y(S_1)$ and $Y(S_2)$ applying recursion.
Step 3: Calculate $X(S)$ using the diagram obtained in step 2:

- Assign a point in the extreme of l.
- Ubicate the points p and q of S_1 and S_2, respectively, and those points are nearest regarding the extreme point of l.

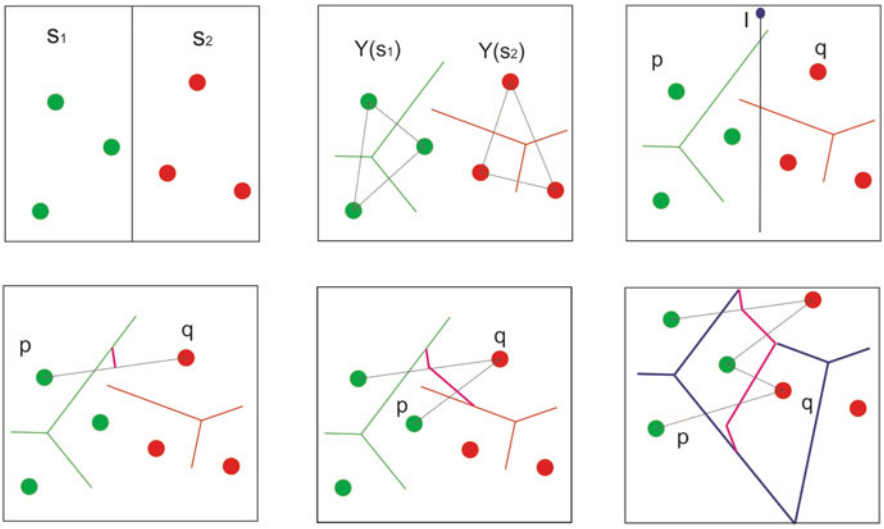


Fig. 7 Implementation of a Voronoi diagram using divide and conquer

- Draw a mediatrix between the lines p and q until intersecting with an edge of $Y(S_1)$ or $Y(S_2)$.
- If the mediatrix intersect with $Y(S_1)$ to assign p to the point that is located into the edge. Otherwise, if the mediatrix intersect with $Y(S_2)$ must be assigned q.
- Return to step 3 until completing the polygonal.

To detail each step, Fig. 7 shows the application of the method using a point cloud of six points.

3.1.2 Incremental Algorithm

Incremental algorithm consists of adding points ($\boldsymbol{p_i}$) in the point cloud to determine a new region $\boldsymbol{p_j}$; this algorithm only modifies the regions affected by the new points. The incremental algorithm is considered slow because it is necessary to recalculate the point cloud every time it adds a new point. However, has been an algorithm highly employed to build Voronoi diagrams [34]. For creating a Voronoi diagram of $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p_n}\}$ using this algorithm, it is necessary to apply the following steps:

Step 1: Add a new point p_{i+1}.
Step 2: Explore all points to find the region p_j closest to p_{i+1}.
Step 3: Draw a mediatrix in the points nearest to p_{k_n} to p_{i+1}.
Step 4: Eliminate the edges and vertices, that is, into of p_j.

Figure 8 shows an example to visually understand the application of the incremental algorithm.

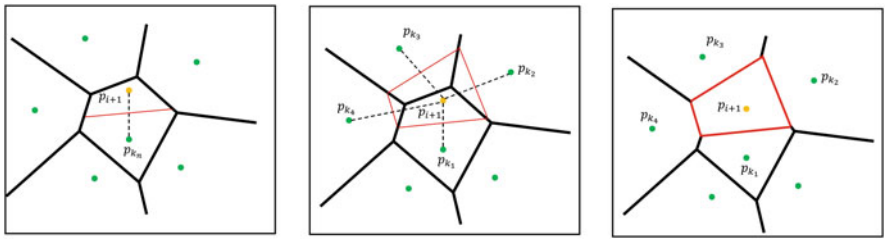


Fig. 8 Implementation of a Voronoi diagram applying incremental algorithm; the red region indicated the new region p_j

3.2 Voxel Algorithm

Voxel method is used to transform a point cloud in a 3D mesh, where the cells are small cubes called *voxel*. A voxel is defined as the cubic unit that forms a tridimensional object and the minimum unit of processing of a tridimensional matrix. Voxel algorithm presents advantages such as accurate 3D and quickly to represent a model. They can be employed in reconstruction in areas such as medical, architectonic, and animation [36].

3.2.1 Voxelization

The geometry of a voxel is defined by length *(l)*, height *(h)*, and width *(w)* and its localization in a 3D grid is indexed by columns *(i)*, rows *(j)*, and layers *(k)*. Therefore, a voxel can be represented by *v (i, j, k)* and *k* defines the voxels in layer *k* (M_k) [37]. Figure 9 shows the representation of a voxel into a specific space called voxel grid.

According to Wu et al. [22], considering a whole voxel grid V that has t voxel layers, the voxel grid can be defined as the join of all the layers, and the equation is given by (16):

$$V = \bigcup_{k=0}^{t-1} M_k \tag{16}$$

3.2.2 Voxel-Based on a Neighborhood

Voxel-based on a neighborhood is a method widely applied for solving region growing problem, which consists in to search and identify a voxel's neighborhood region in a voxel layer. In a vertical direction, a 3D voxel grid is represented by three types of neighbors: face, edge, and vertex.

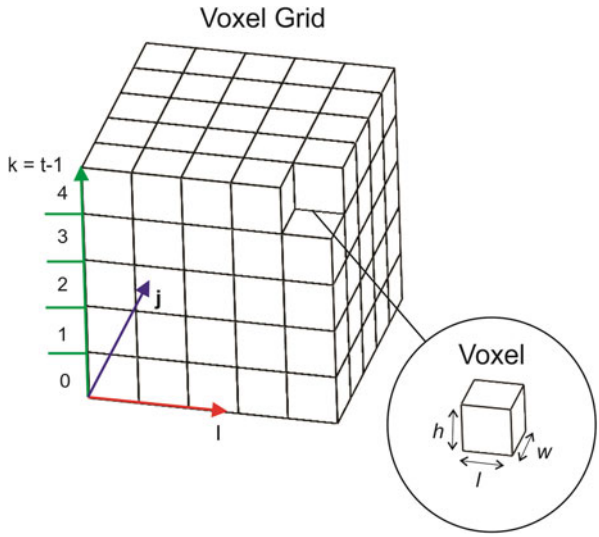


Fig. 9 Representation of voxel grid and voxel

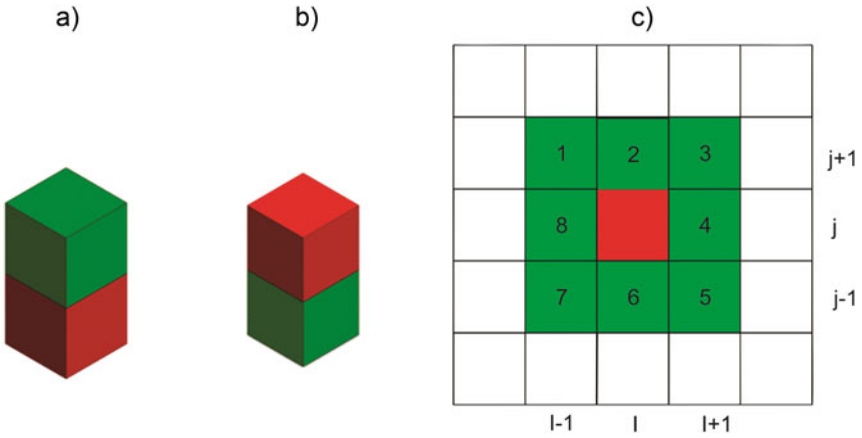


Fig. 10 The red cube and square represent a voxel *v (i, j, k)*, and the greens represent the neighbors nearest. (**a**) Top face *v (i, j, k + 1)*, (**b**) bottom face *v (i, j, k–1)*, (**c**) eight neighbors in the same layer

According to Wu [38], a 3D voxel grid has two neighbors (top face and bottom face), while in a voxel layer, each voxel has eight neighbors $v\{i–l, j–l, k\}$, $v\{i–1, j, k\}$, $v\{i–1, j + 1, k\}$, $v\{i, j–1, k\}$, $v\{i, j + 1, k\}$, $v\{i + 1, j–1, k\}$, $v\{i + 1, j, k\}$, and $v\{i + 1, j + 1, k\}$. The neighbors presented enable for creating a 3D neighborhood in the voxel grid. Figure 10 shows the three types of voxel neighborhood used in the research.

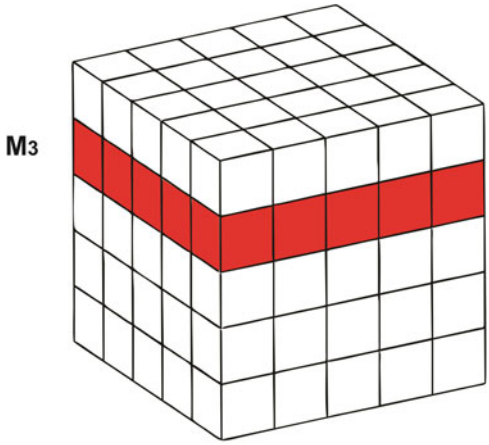


Fig. 11 The red cubes represent the layer M_3

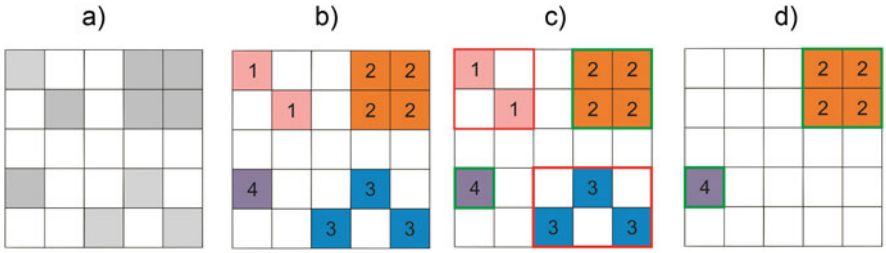


Fig. 12 (**a**) Fourth layer in the voxel grid. (**b**) Identification, grouping, and marking. (c) Removing voxel grid depending on the number and compactness index of voxel. (**b**) Seed voxel group [22]

3.2.3 Select Seed Voxel Groups and Neighborhood

According to Wu et al. [38], seed voxel consists in to select a layer ($\boldsymbol{M_k}$), which contains all the points of a specific layer of a point cloud (i.e., $\boldsymbol{M}_3$ indicates all the points that contain the fourth row) as shown in Fig. 11. All the voxel layers are scanned in row-wise and a seed voxel is settled, afterward seed voxel is expanded to include all voxels, and this process is repeated until all connected voxels are included and grouped.

The voxels in the same group are indexed and marked by groups using a different integer number. In addition, a false seed voxel will be removed in the following cases: if the count of voxel in the same group is larger than a threshold value or the extent of the group voxels is larger than a threshold value. In Fig. 12 is shown the application of the method.

Once obtained seed voxel group, the neighbors are defined in each layer of the voxel grid, i.e., in our example, the groups are in M_3; then in order to identify the neighbors nearest, it is necessary to compare our seed voxel group obtained in M_3 in regard to all layers M_k.

On the other hand, according to Torres et al. [39], an algorithm based on *voxels* can be implemented applying the following steps:

To each voxel v into of the volume V:

Step 1: To find the normal distance de v to all the planes of the images, obtaining a distance vector D

Step 2: To obtain the maximum distance and the two minimum distances ($d_{\max}$, d_1 *y* d_2). Also the plane of the minimum distances I_1 *y* I_2

Step 3: To each plane I_n, where $n = 1, 2$

- To calculate the intersection point between the normal that connect v and I_n
- To perform a dual bilinear interpolation (p_n) of the four pixels nearest the intersection point
- To obtain the weight of each plane I_n:

$$w_{n=}\left(1-\frac{d_n}{d_{\max}}\right)$$

- The final value v is given by (17):

$$v = \left(\frac{(w_1)\,(p_1) + (w_2)\,(p_2)}{2}\right) \tag{17}$$

4 Reconstruction of Human Body Surfaces

Nowadays, there are numerous systems such as laser scanner, photometry, and structured light, which are applied in reconstruction of human body surfaces in the medical area. These systems enable to capture surfaces of the human body to create point clouds. However, factors such as environment light and noise cause distortion avoiding reliable diagnostics of deformities in the body. Therefore, in order to define the geometry and optimize the measurements, and human body surface characterization, the point cloud methods of reconstruction are applied. This section presents previous methods used to reconstruct a cloud in different areas of the human body.

4.1 3D Spine Reconstruction

Section 1 presented several diseases associated with spine deformities such as scoliosis and kyphosis; using a laser scanning or photometry, it is possible to represent the geometry of the spine to provide diagnostics and orthopedic treatments

that enable to correct the curvature of the spine. To define the geometry of the point cloud acquired by laser scanning or photometry, it is necessary to apply a method of 3D reconstruction. According to Aroeira et al. [40], the highest incidence of vertebral deformity in adolescents occurs between the fifth and decimal vertebrae (T5 to T10). Therefore, the geometry of the T5 to T10 is obtained using images. Once the geometric model is acquired, the HyperMesh and Meshmixer software are used to create a 3D finite element model employing a meshing procedure. The meshing procedure enables to create a 3D model of the thoracic spine with the proper orientation of the joint faces, edges, and geometry. A similar analysis is performed using a laser scanner to create two-point clouds of eight vertebras of the spine; to join the two-point clouds is applied an iterative closest point (ICP) algorithm in order to obtain a distance lower than 0.8 mm and make the density uniform. Also, a Tessellation algorithm (Voronoi diagram) is employed to define the solid geometric model of the vertebra [41].

4.2 3D Chest Reconstruction

Pectus excavatum (*PE*) and pectus carinatum (*PC*) are chest deformities that cause depression sternum and excessive growth of the costal cartilage, respectively. *PE* and *PC* are considered as esthetics problems; however, PE could affect cardiac and pulmonary function causing pain in the chest. Therefore, systems as laser scanner and photometry have been employed in order to create a 3D point cloud, which along with reconstruction methods enable to create a precision digital 3D model. According to Bellia et al. [42], 3D computed tomography scan is developed to obtain images of patients in the chest area; subsequently, the images are reconstructed to 1.0-mm-thick sections with a soft-tissue algorithm. The application of the algorithm enables to characterize the deformity over the chest caused by *PE*, in order to develop custom implants enabling to give a solution to esthetic issues. Another research related to 3D chest reconstruction uses a laser scanner to capture the surfaces over the chest. The laser position is located at 1355 mm; the chest was rotted in front of the scanner to obtain 25 scans. Subsequently, a voxel algorithm is applied utilizing a 3D grid system set at $0.8 \times 0.8 \times 0.8\ m^3$ [43].

4.3 3D Face Reconstruction

A common problem in the population is congenital diseases related to facial deformities such as deviation, cleft lip, cleft palate, and nose deformity. For each deformity can be applied plastic surgeries in order to correct the anomalies. Laser scanning and photometry have enabled the 3D model representations, which are used by medical professionals to analyze the results of the surgeries. According to Matsushima, Nishi, and Nakahara [44], polygon mesh processing is applied to

define the geometry of a girl's face based on a point could obtained by laser scanning (Konica Minolta Vivid 910); the 3D model obtained with mesh polygon could be employed to characterize deformities over the face. A similar analysis employs a laser scanner to capture a point cloud that enables to represent the human face into cylindric spatial, which enable to efficiently process the 3D triangular mesh by applying 2D filters [45]. On the other hand, there are some researches based on photometry, which uses multiples images to represent the tridimensional shape of the face; however, there is a research area that intends to build a 3D facial model from a single image; this is performed by analyzing and classifying from a database the depth to apply a triangle mesh, which enables to connect the point of the cloud to create a 3D model [46]. The research presented in this section employ polygon meshes due to the advantages they presented in Sect. 2. Moreover, there is a wide variety of software that have this tool.

4.4 3D Feet Reconstruction

Another common problem in the population is foot deformities; the deformities are related to congenital diseases causing pain and fatigue to patients that suffer them. There are different types of deformities such as flat, cavus, valgus, and diabetic feet. For each one, it is necessary to employ an especial treatment that enables to correct the deformity such as orthopedic soles, physiotherapy, or surgeries. Therefore, medical professionals use systems to capture the low foot surfaces or anthropometric parameters in order to represent 3D model that provides detailed diagnostics. According to Lee, Lin, and Wang [47], a 3D laser scanner (INFOOT USB scanning system, IFU-S-01, I-Ware Laboratory Co., Ltd., Japan), force foot imprinted (Harris mat, Inc., USA), digital caliper, and ink footprint are employed to capture foot dimensions. However, they can be used to collect volumetric and surface data to provide more detailed size and shape (i.e., anthropometric parameters) in order to apply methods of biomechanical analysis of the foot. On the other hand, *INFOOT USB* and *Harris mat* systems employ the method of polygon to define the geometry of the obtained point cloud. The results demonstrate a better accuracy and higher efficiency utilizing 3D laser scanner when polygon mesh is applied in the cloud. However, several laser scanners have a distortion in the point cloud when it is necessary to capture the wrinkles of the fingers. Therefore, Bezier curve is applied using a quadratic function to define the curve of the wrinkles in the fingers in zones such as feet or hands [48].

4.5 3D Head Reconstruction

Craniosynostosis is a birth deformity in which the bones in a baby's skull join together too early. Meanwhile, nonsynostotic deformational plagiocephaly refers

to asymmetry of the infant's skull characterized by flattening or deformation. These deformities not only are cosmetic problems but also can increase the risk of developmental delays. According to Yuan, B. [49], a photometry system (3dMD) is used to scan a skull. Afterward, 3D triangle mesh is employed to represent the shape of a head for evaluating and analyzing results of infant cranial surgeries. Each 3D mesh contains approximately 10^5 vertices and 2×10^6 faces. Therefore, to avoid computation, extremely time-consuming mesh simplification algorithm is employed. This algorithm (known as quadric error metric decimation) consists of edge collapse, which merges two vertices into one, removes the edges connected to these two vertices, and adds new edges connected to the new vertex. On the other hand, a similar analysis applies photometry utilizing cranial focal point. This method can be calculated to determine the mean virtual intersection points of all normal surfaces to create a point cloud in the cranium with a center point and spread. Afterward, was implemented a polygon mesh to define the geometry of the point cloud [50].

5 Human Body Modeling Approaches' Comparison

In previous sections are presented 3D reconstruction methods applied in point clouds, which can create defined surfaces. The proposal of this section is to apply 3D reconstruction methods and representation in order to provide defined geometries that enable to detect deformities in the human body. To show the application of the methods, two sample datasets foot and chest downloaded from ([51, 52] webpages [http://www.i-ware.co.jp/content/infoot2.html *and* https://www.artec3d.com/3d-models] are used. The datasets can be used as point clouds to build mesh surfaces applying geometric and volumetric reconstruction methods in MeshLab software.

For the demonstration of the triangular mesh method are used foot and chest datasets. *Foot dataset* is formed by 70,000 points, while *chest dataset* is formed by 50,000 points. However, considering the capabilities of the computer system, only samples of 1000 points are used. Both point clouds are built applying the algorithm of screened Poisson algorithm (SPA), which enables to transform the point cloud to surfaces. Subsequently, a triangular mesh is employed to simplify the number faces and vertices of the obtained model in order to optimize its processing. Figures 13 and 14 show the point cloud and different reconstructions employing triangular mesh varying the number faces.

Once geometric reconstruction is generated, methods of remesh are used to analyze the volume in Blender software. The first method employed is voxel remesh tool setting up a voxel size to 1.5 m and adaptative to 1 m as is shown in Fig. 15a). The second method used is block remesh tool setting up the octree depth (it is a tree data structure in which each internal node has exactly children) to 7 and scale and 0.8 as is shown in Fig. 15b).

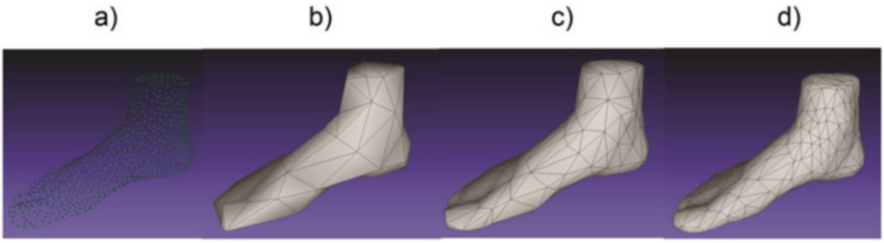


Fig. 13 (**a**) Point cloud formed by 1000 points, (**b**) 100 faces, (**c**) 300 faces, (**d**) 600 faces

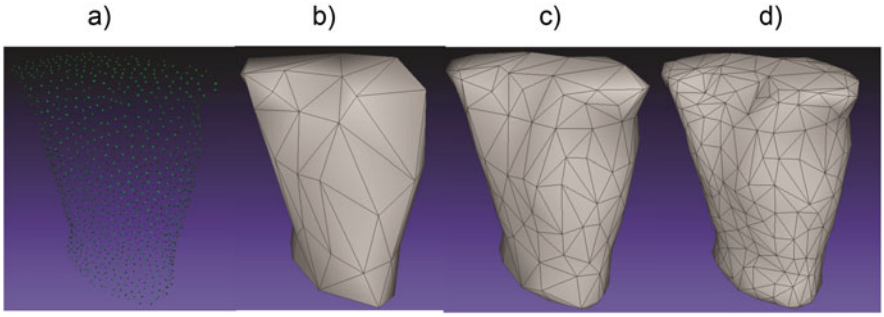


Fig. 14 (**a**) Point cloud formed by 100 points, (**b**) 100 faces, (**c**) 300 faces, (**d**) 600 faces

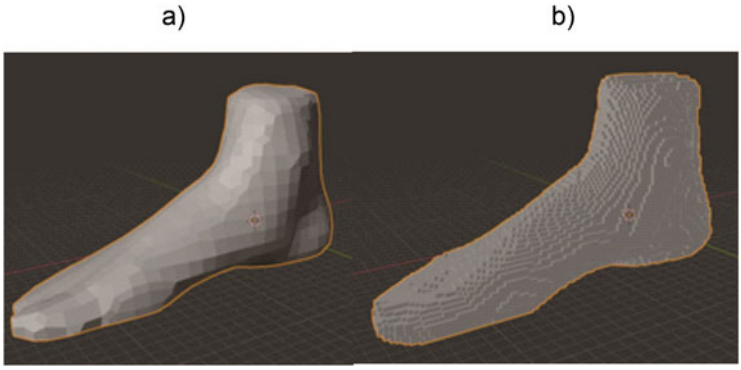


Fig. 15 (**a**) Remesh voxel setup to voxel size 1.5 m, (**b**) remesh block setup to octree depth 7 and scale 0.8

On the other hand, also Voronoi method is applied to reduce the volume of the point cloud and create a structure that enables to easily represent the biometric areas. To transform the reconstrued model to Voronoi diagram are used different type filters in MeshLab software; the first filter is *Voronoi vertex coloring*, which given a mesh (M) and a pointset (P) to project each vertex of P over M and color M according to the distance of these projected points. Afterward, a *select by vertex quality filter* is applied to select all the vertex within the specified vertex quality

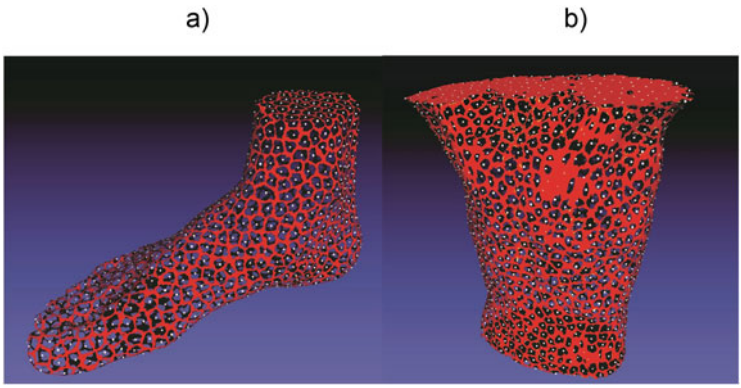


Fig. 16 (**a**) Foot model utilizing Voronoi method. (**b**) Chest model applying Voronoi method

range. Figure 16 shows the Voronoi method setting up the high and low quality to 3.5 and 1, respectively, in both models.

5.1 Measurements of Biometric Parameters

In previous sections are presented common deformities over the human body related to different diseases that the population is suffering. Also, the main 3D reconstruction methods used to build point clouds are presented. This section presents biometric technics such as longitudinal arch angle (LAA) and Hebal-Malas index (HMI) applied to detect and measure deformities in the foot and chest, respectively. To perform the measurements is employed SolidWorks software, which enables creating a view section on different surfaces in 3D models.

5.1.1 Deformities' Analysis in Foot

The first analysis is performed to classify the type of foot as normal, cavus, or flat applying anthropometric method as navicular drop and longitudinal arch angle. For this case, longitudinal arch angle (LAA) is used, which consists in determining the angle formed by first metatarsal head ($\boldsymbol{MT}_1$), navicular tuberosity (NT), and medial malleolus (MM). The mesh model created in MeshLab is exported to an *stl* extension file. Afterward, SolidWorks software is used because it enables easily to measure 3D model and to perform view-section in multiples planes. Figure 17a) shows the 3D foot model utilizing the view-section plane directed orthogonally toward the internal structure of the foot. When the plane of cross section is moved, the first contact point is the $\boldsymbol{MT}_1$; whether the view-section follow the same trajectory, the second and third contact points are *MM* and *NT*, respectively, as are

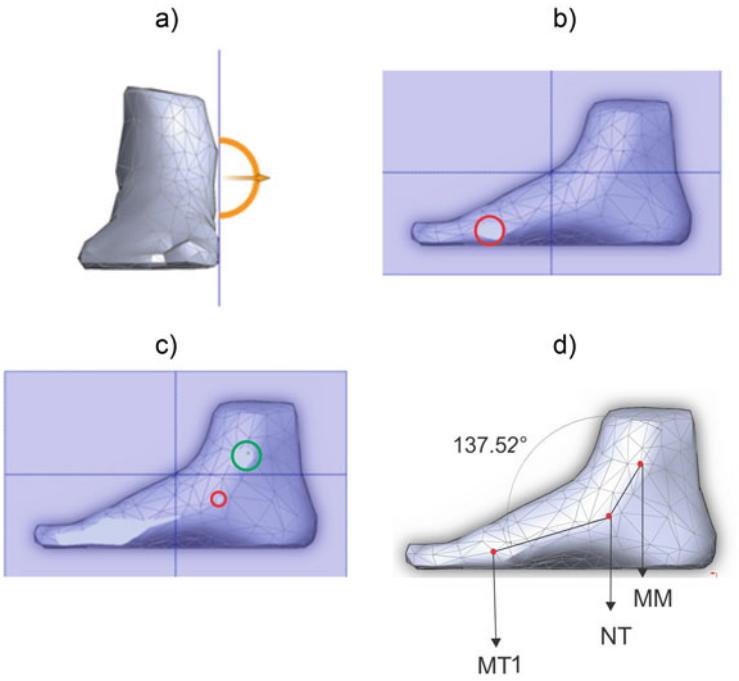


Fig. 17 (**a**) Plane view-section on the internal foot. (**b**) Red circle indicates the MT_1. (**c**) The red and green circles indicate the *NT* and *MM*. (**d**) LAA method applied on the internal foot obtaining as result 137.52°

shown in Fig. 17b, c). At least, Fig. 17d) shows three anthropometric point positions and the angle measured applying the LAA method in SolidWorks software**.**

According to Nilson et al. [53] and Dahle et al. [54], utilizing LAA method, a foot **can be classified as:**

- Flat: 121°–131°
- Normal: 131°–152°
- Cavus: 153°–162°
- Extreme cavus: more than 162°

Therefore, the result obtained enables to classify the 3D foot model as normal, because the angle obtained (137.52°) is between 131° and 152° range. This study can be used to provide diagnostics in a static state to patients that suffer these deformities. However, a medical professional must consider applying a dynamic state method to provide reliable diagnostics. On the other hand, also the obtained model could be used to perform measurements in the foot over the areas such as foot length, foot width, ball of foot length, and outside ball of foot length. Figure 18 shows some measurements obtained utilizing SolidWorks software.

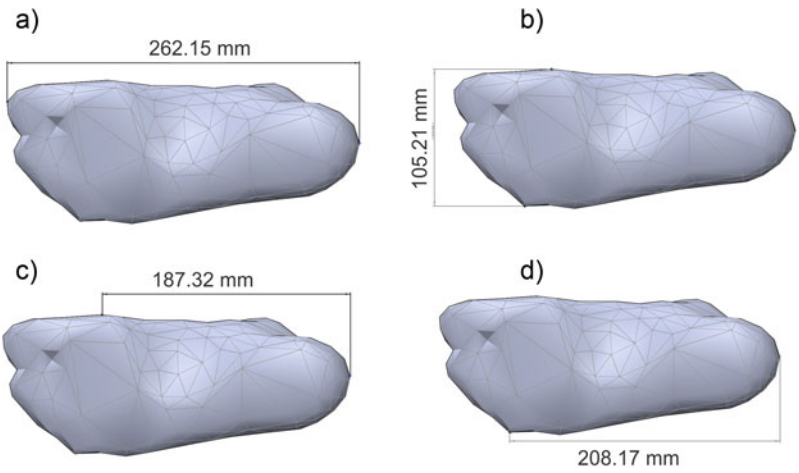


Fig. 18 (**a**) Foot length, (**b**) foot width, (**c**) ball of foot length, (**d**) outside ball of foot length

5.1.2 Deformities' Analysis in Chest

The second analysis is performed to verify whether it is possible to measure the depressed sternum caused by pectus excavatum (PE) deformity. A common technique uses a ruler as measurement instrument to measure the depth into the depression at its deepest point and recorded the distance from the deepest point to the surface of the depression. However, this procedure can cause uncertainties in measurements. Therefore, the analysis presented below provides a method based on 3D model to measure the depressed sternum attenuating uncertainness. According to Hebal et al. [55], a 3D model chest can be used to determine Hebal-Malas index (HMI), which consists in to assess PE deformity measurements. HMI is calculated dividing the mediolateral diameter (ML) by the anteroposterior diameter (AD) of the chest depth; the equation is given by (18):

$$\text{HMI} = \frac{\text{ML}}{\text{AD}} \tag{18}$$

The triangular mesh model obtained in Fig. 15 is imported to SolidWorks software to apply the view-section tool, which enables varying the top plane until reaching the maximum depression of the sternum as is shown in Fig. 19a). Once the top plane position in the maximum depression of the sternum is defined, this view is collocated from normal to a top plane as is shown in Fig. 19b).

Afterward, the ML and AD are measured using the reference point and smart dimension tool as is shown in Fig. 20.

These values enable to calculate HMI applying Eq. (18) obtaining as result 1.71 mm; therefore, according to Hebal et al. [55], the result obtained indicates that HMI is 0.09 below the mean. Thereby, the 3D model obtained does not present

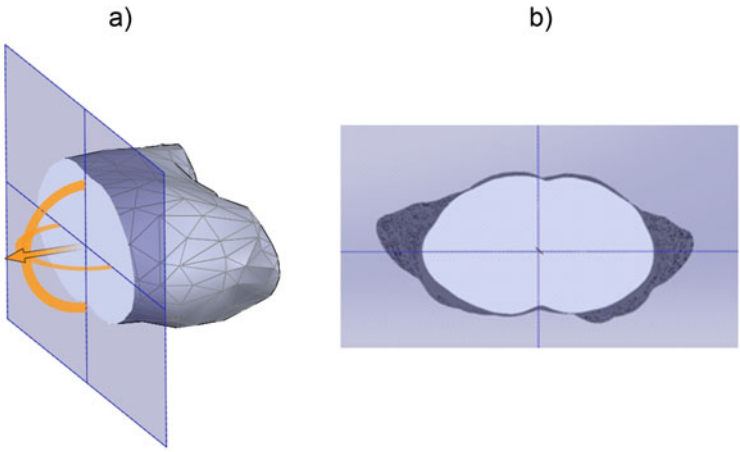


Fig. 19 (**a**) View-section collocated in maximum depressed sternum. (**b**) View-section normal to top planes

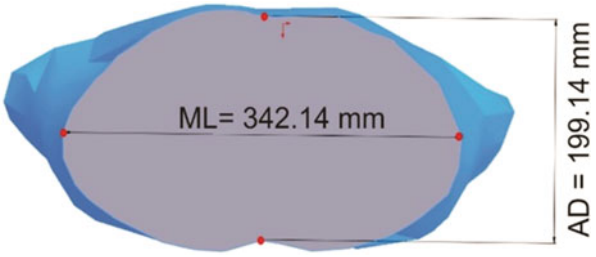


Fig. 20 ML and AD measurements determine utilizing view-section in SolidWorks software

PE deformity; however, the previous analysis demonstrates that it is possible to measure the depression of the sternum. Also, it can be noted that the accuracy of the 3D model depends on numerous faces, i.e., whether numerous faces increase in the triangle mesh model, the accuracy is enhanced, obtaining defined surfaces and reliable diagnostics.

6 Technical Vision System for 3D Human Body Measurements

Technical vision system (TVS) is a 3D laser scanner based on dynamic triangulation principle and developed at "Universidad Autónoma de Baja California." The system has provided satisfactory results in application such as autonomous mobile robot, structure health monitoring, and monitoring of scoliosis. This section presents the TVS operation principle and the surface reconstruction from point clouds in order to analyze biometric parameters in the human body.

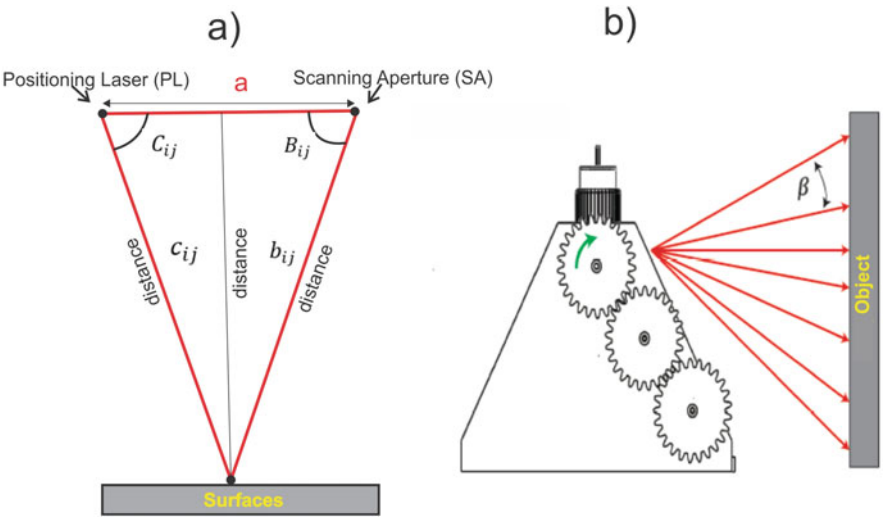


Fig. 21 (**a**) Principle of dynamic triangulation. (**b**) Developed mechanism for positioning the angle β

6.1 Dynamic Triangulation Principle

The principle of dynamic triangulation consists in the analysis of a triangle formed in short time; this is formed by three straight lines, which connect three points: emitter, receiver, and the reflection point of the laser. In TVS based on dynamic triangulation, the positioning laser (PL) projects a laser beam over the surface of an object. Afterward, the laser beam is reflected toward the scanning aperture (SA), as the fixed distance (a) between *PL* and *SA* is known a triangle is formed as is shown in Fig. 21a). Through this information and the application of trigonometric functions can be calculated the coordinates, where the laser beam gets in contact with the surfaces. The angles formed in dynamic triangulation are C_{ij} and B_{ij}. C_{ij} angle is delimited by *PL* and *a*, while B_{ij} angle is formed between *SA* and the distance a [56, 58].

To calculate the coordinates (x, y, z) in each point over the surfaces of the object, it is necessary to apply Eqs. (19), (20), and (21) implemented in previous research, where the value β shown in Fig. 21b) is used to calculate the coordinates of the z axis [57]:

$$x = a\left(\frac{\sin\left(B_{ij}\right)\ \sin\left(C_{ij}\right)}{\sin\left(B_{ij}+C_{ij}\right)}\right) \tag{19}$$

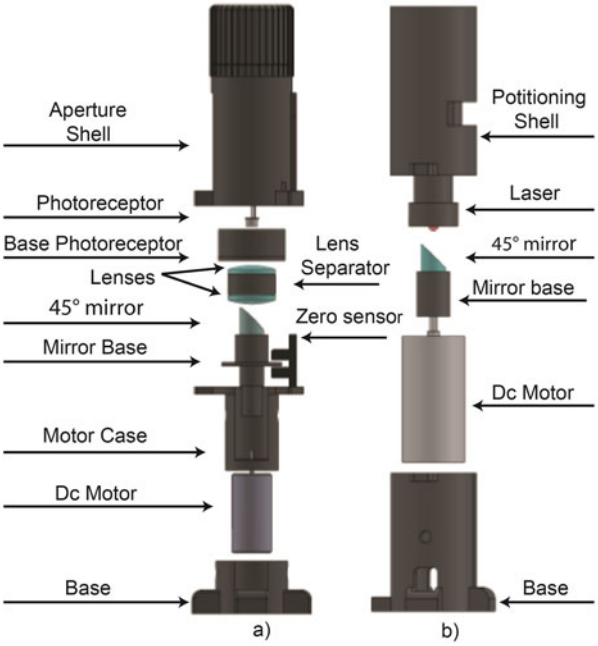


Fig. 22 (**a**, **b**) Accuracy improvement by artificial neural networks in technical vision system. Trujillo [58]

$$y = a\left(\frac{1}{2} - \frac{\cos\left(B_{ij}\right)\ \sin\left(C_{ij}\right)}{\sin\left(B_{ij} + C_{ij}\right)}\right) \tag{20}$$

$$z = a\left(\frac{\sin\left(B_{ij}\right)\sin\left(C_{ij}\right)\tan\left(\beta_{ij}\right)}{\sin\left(B_{ij} + C_{ij}\right)}\right) \tag{21}$$

6.2 Positioning Laser and Scanning Aperture

SA is an essential element that enables the dynamic triangulation system reception, which is formed by a 45° mirror, two biconvex convergent glasses, a zero sensor, and a photoreceptor as is shown in Fig. 22a). The PL is composed by a laser, which emits parallel beams toward 45° mirror to redirect them over to the object's surfaces to scan. Then, the lines are reflected to scanning aperture, where another 45° mirror redirects the lines to the biconvex convergent glasses, which focus the parallel beams to a single point (receptor) located at the top of the scanning aperture. On the other hand, zero sensor indicates when the motor pass 0°.

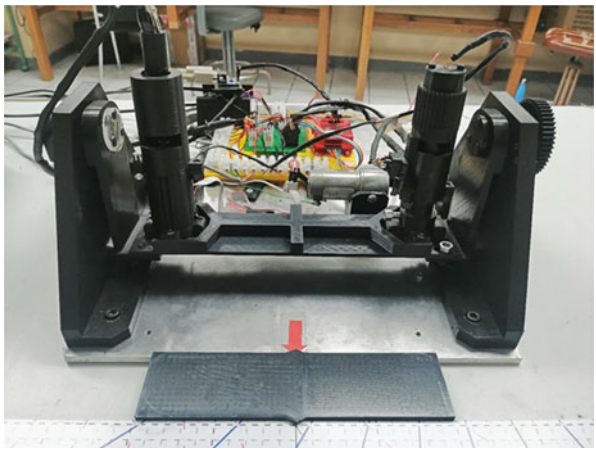


Fig. 23 Technical vision system (TVS)

PL is another essential element that enables the operation of the TVS, which is composed by 45° mirror, dc motor, and encoder as is shown in Fig. 22b). The laser beam is directed to the 45° mirror, which redirected the beam laser to 90° over the surface to scan. In addition, PL projects the laser beam moved along the object until finishing the scanning process; during this time is measured a laser beam point by each revolution of the SA, i.e., each detection of the zero sensor. The TVS design enables to capture tridimensional point clouds in order to characterize surfaces. Figure 23 shows the TVS assembly.

6.3 3D Reconstruction in Point Cloud Captured by the TVS

In order to determine the capacities of the TVS and the reconstruction methods presented throughout this chapter, it is performed a scanning over a hand surface to convert a point cloud to 3D polygon mesh. For this scanning, the TVS control software is set up with a laser beam speed to 5 $\frac{^\circ}{\text{seg}}$ and go forward in $\boldsymbol{\beta}$ angle at 0.7°. Once parameters are set up, the hand is painted with acrylic paint to obtain an adequate reflection of the laser. Afterward, the hand is vertically collocated in front of the TVS to perform the scanning. During the scanning, it is necessary to keep the hand relaxed until completing the scanning time of 3 minutes to avoid measurement errors. Afterward, a point cloud is obtained and presented in 3D coordinates as shown in Fig. 24a).

From the obtained point cloud, it is possible to note some measurement errors on the fingers caused by the reflection of the beam laser, involuntary movements, and calibration, but even with measurement errors, the hand can be characterized. However, the approach of this chapter is to apply methods of 3D reconstruction that enable to measure deformities over the human body. Therefore, to analyze the surface, MeshLab and Blender software are used, because they provide a

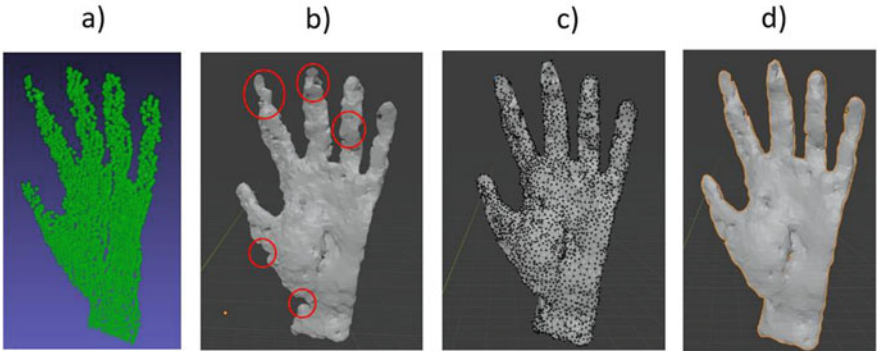


Fig. 24 (**a**) Point cloud formed by 6000 points. (**b**) 3D reconstruction using ball pivoting algorithm. (**c**) Adjustment of meshes utilizing the polygon mesh tool. (**d**) Model obtained applying a smoothing filter

wide variety of reconstruction algorithms, simplification, remeshing, sculpting, and smoothing. Following it is presented the procedure applied to rebuild the point cloud.

Using MeshLab software, the noise over the point cloud is removed; afterward, *compute normals for points set* tool is applied to create a normal in each vertex to apply a *ball pivoting algorithm*, which enables given a point cloud with normals to create mesh surfaces. However, the measurement error causes a poor reconstruction on the fingers as shown in Fig. 24b). Therefore, the model obtained is exported to Blender software to apply the *polygon mesh* tool, which enables to repair the geometry of the hand surface adjusting the meshes especially over the index and ring finger as shown in Fig. 24c). Once meshes are adjusted, *draw sharp* tool and *thumb* are employed to attenuate the noise in 3D model and flatten in the specific area. Afterward, to improve the quality of 3D model is used a *filter smoothing* applying a factor to 0.5 and repeat to 3 the final model is shown in Fig. 24d).

The procedure applied on the point cloud enabled to build a 3D mesh model and attenuate some measurement errors. However, it is still possible to note the errors over the index and thumb fingers; this can be corrected by a specific scan in the area of interest. Another alternative to enhance the measurements is the implementation of a fixture that assures to keep the hand relaxed during the scanning time.

7 Conclusion

3D reconstruction can be used to provide a more natural representation of a scanned model object and be able to visualize information that would go unnoticed if only a point cloud were used. The 3D reconstruction employs the methods of geometric and volumetric reconstruction in the point cloud to create a well-defined

and optimized 3D model, which can be processed by computer systems. In addition, the obtained 3D model can be employed to detect deformities over the human body to provide reliable diagnostics and treatments to patients.

This chapter provides information about the state of the art of the principal diseases that causes deformities over the human body and the 3D reconstruction methods applied in different research, which have demonstrated satisfactory results in deformities' detection. Also, some methods such as polygon mesh, voxel, and Voronoi diagram are applied on two-point clouds (*chest* and *foot*) utilizing *MeshLab* and *Blender* software. The screened Poisson algorithm enabled to build a 3D model of the feet; afterward, this model was simplified varying the number of meshes and can be noted an enhanced resolution when more meshes are applied. The 3D model was exported to SolidWorks to identify anthropometric parameters using the view-section tool. This tool enabled the identification of the internal foot of the metatarsal head, navicular tuberosity, and medial malleolus to calculate the *longitudinal arch angle* and determine the foot type. As well, the same procedure is applied to measure the depression of the sternum, and using the view-section tool was possible to calculate the Hebal-Malas index to determine the severity of the deformity.

On the other hand, the TVS is used to scan a human hand; the application of the *ball pivoting algorithm* enabled to build the point cloud to surface meshes. However, the thumb and index fingers presented issues. Therefore, polygon mesh tool, draw sharp, thumb, and smoothing were applied obtaining satisfactory results on the finger 3D reconstruction.

References

1. Puszczalowska-Lizis, E., Bujas, P., Omorczyk, J., Jandzis, S., & Zak, M. (2017). Feet deformities are correlated with impaired balance and postural stability in seniors over 75. *PloS one*, *12*(9), e0183227.
2. Laratta, J. L., Ha, A., Shillingford, J. N., Makhni, M. C., Lombardi, J. M., Thuet, E., ... & Lenke, L. G. (2018). Neuromonitoring in spinal deformity surgery: a multimodality approach. *Global spine journal*, *8*(1), 68–77.
3. Ors, S. (2017). Incidence and classification of chest wall deformities in breast augmentation patients. *Aesthetic Plastic Surgery*, *41*(6), 1280–1290.
4. Garza, E. M., Cortés, H. G. C., Godoy, B. D., & Campos, H. A. C. (2020). Assistive Technology for Patient with Congenital Foot Deformity. In *SHS Web of Conferences* (Vol. 77, p. 05002). EDP Sciences.
5. Rothstock, S., Weiss, H. R., Krueger, D., & Paul, L. (2020). Clinical classification of scoliosis patients using machine learning and markerless 3D surface trunk data. *Medical & Biological Engineering & Computing*, *58*(12), 2953–2962.
6. Szafer, D., Taylor, J. S., Pei, A., de Ruijter, V., Hosseini, H., Chao, S., & Wall, J. (2019). A simplified method for three-dimensional optical imaging and measurement of patients with chest wall deformities. *Journal of Laparoendoscopic & Advanced Surgical Techniques*, *29*(2), 267–271.
7. Treleaven, P., & Wells, J. (2007). 3D body scanning and healthcare applications. *Computer*, *40*(7), 28– 34.

8. Pfeiffer, M., Kotz, R., Ledl, T., Hauser, G., & Sluga, M. (2006). Prevalence of flat foot in preschool-aged children. *Pediatrics*, *118*(2), 634–639.
9. Savic, S. P., Lukic, N., Prodanovic, N., & Devedzic, G. (2019, June). A Comparative FE Analysis of a Flat Foot with and without a Custom-made Orthotic Insole. In *2019 8th Mediterranean Conference on Embedded Computing (MECO)* (pp. 1–4). IEEE.
10. López, M. E. S., Cano, R., Vera, C. S. C., Avendaño, L., Zavaleta, N. E., & García, C. (2018). Pectus excavatum y carinatum en el síndrome de Marfan y síndromes similares: prevalencia e impacto clínico pulmonar y cardiovascular. *Gaceta médica de México*, *154*(2), 67-78.
11. Kuru, P., Cakiroglu, A., Er, A., Ozbakir, H., Cinel, A. E., Cangut, B., ... & Yuksel, M. (2016). Pectus excavatum and pectus carinatum: associated conditions, family history, and postoperative patient satisfaction. *The Korean Journal of Thoracic and Cardiovascular Surgery*, *49*(1), 29.
12. Lain, A., Garcia, L., Gine, C., Tiffet, O., & Lopez, M. (2017). New methods for imaging evaluation of chest wall deformities. *Frontiers in pediatrics*, *5*, 257.
13. Kim, D., Park, J. H., Favero, V., Mah, J., Jung, Y. S., & Kim, S. T. (2020). Effect of botulinum toxin injection on asymmetric lower face with chin deviation. *Toxins*, *12*(7), 456
14. Verzé, L., Bianchi, F. A., & Ramieri, G. (2014). Three-dimensional laser scanner evaluation of facial soft tissue changes after LeFort I advancement and rhinoplasty surgery: patients with cleft lip and palate vs patients with nonclefted maxillary retrognathic dysplasia (control group). *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, *117*(4), 416–423.
15. Rodríguez-Quiñonez, J. C., Sergiyenko, O. Y., Preciado, L. C. B., Tyrsa, V. V., Gurko, A. G., Podrygalo, M. A., ... & Balbuena, D. H. (2014). Optical monitoring of scoliosis by 3D medical laser scanner. *Optics and Lasers in Engineering*, *54*, 175-186.
16. Ifflaender, S., Rüdiger, M., Konstantelos, D., Wahls, K., & Burkhardt, W. (2013). Prevalence of head deformities in preterm infants at term equivalent age. *Early human development*, *89*(12), 1041–1047.
17. Pitzer, B., Kammel, S., DuHadway, C., & Becker, J. (2010, May). Automatic reconstruction of textured 3D models. In *2010 IEEE International Conference on Robotics and Automation* (pp. 3486–3493). IEEE.
18. Godse, A. P. (2009). Computer Graphics. Technical Publications Pune. First edition (pp. 564)
19. Botsch, M., Kobbelt, L., Pauly, M., Alliez, P., & Lévy, B. (2010). *Polygon mesh processing*. CRC press.
20. Peng, J., Kim, C. S., & Kuo, C. C. J. (2005). Technologies for 3D mesh compression: A survey. *Journal of Visual Communication and Image Representation*, *16*(6), 688–733.
21. Wu, G., Yang, Z., Zhang, K., Dong, P., & Lin, Y. T. (2018). A non-equilibrium sediment transport model for dam break flow over moveable bed based on non-uniform rectangular mesh. *Water*, *10*(5), 616.
22. Wu, B., Yu, B., Yue, W., Shu, S., Tan, W., Hu, C., ... & Liu, H. (2013). A voxel-based method for automated identification and morphological parameters estimation of individual street trees from mobile laser scanning data. *Remote Sensing*, *5*(2), 584–611.
23. Van Verth, J. M., & Bishop, L. M. (2015). *Essential mathematics for games and interactive applications*. CRC Press.
24. Paluszny, M., Prautzsch, H., & Boehm, W. (2005). *Métodos de Bézier y B-splines*.
25. Kanjanasurat, I., Chutchavong, V., Pirajnanchai, V., & Janchitrapongvej, K. (2016, November). Bernstein polynomial and rational Bézier curve for blood pressure simulation. In *2016 IEEE Region 10 Conference (TENCON)* (pp. 1737–1741). IEEE.
26. Song, S., Li, Z., Meng, M. Q. H., Yu, H., & Ren, H. (2015). Real-time shape estimation for wire-driven flexible robots with multiple bending sections based on quadratic Bézier curves. *IEEE Sensors Journal*, *15*(11), 6326–6334.
27. Jafari, M., & Molaei, H. (2014). Spherical linear interpolation and Bézier curves. *General Scientific Researches*, *2*(1), 13–17.
28. Lee, K. T., Tsai, W. H., Liao, H. Y. M., Chen, T., Hsieh, J. W., & Tseng, C. C. (Eds.). (2011). *Advances in Multimedia Modeling: 17th International Multimedia Modeling Conference, MMM 2011, Taipei, Taiwan, January 5–7, 2011, Proceedings, Part I* (Vol. 6523). Springer.

29. May, D. A. (2012). Volume reconstruction of point cloud data sets derived from computational geodynamic simulations. *Geochemistry, Geophysics, Geosystems*, *13*(5).
30. Tse, R., Gold, C., & Kidner, D. (2007, July). Using the delaunay triangulation/voronoi diagram to extract building information from raw lidar data. In *4th International Symposium on Voronoi Diagrams in Science and Engineering (ISVD 2007)* (pp. 222–229). IEEE.
31. Shenmaier, V. V. (2016). Solving some vector subset problems by Voronoi diagrams. *Journal of Applied and Industrial Mathematics*, *10*(4), 560–566.
32. Smith, E., Trefftz, C., & DeVries, B. (2020, July). A Divide-and-Conquer Algorithm for Computing Voronoi Diagrams. In *2020 IEEE International Conference on Electro Information Technology (EIT)* (pp. 495–499). IEEE.
33. Jida, S., Aksasse, B., & Ouanan, M. (2018, July). Porosity estimation in carbonate rock based on Voronoi diagram and 2D histogram segmentation in HSV color space. In *International Conference on Advanced Intelligent Systems for Sustainable Development* (pp. 727–735). Springer, Cham.
34. Bonivardo, A. (2017). Geometría Computacional.
35. Peethambaran, J., Parakkat, A. D., Tagliasacchi, A., Wang, R., & Muthuganapathy, R. (2019, February). Incremental labelling of Voronoi vertices for shape reconstruction. In *Computer Graphics Forum* (Vol. 38, No. 1, pp. 521–536).
36. Boyle, R., Parvin, B., Koracin, D., Kuno, Y., Wang, J., Renato, P., ... & Coming, D. (Eds.). (2009). *Advances in Visual Computing: 5th International Symposium, ISVC 2009, Las Vegas, NV, USA, November 30-December 2, 2009, Proceedings, Part II* (Vol. 5876). Springer.
37. Patrikalakis, N. M. (Ed.). (2012). *Scientific visualization of physical phenomena*. Springer Science & Business Media.
38. Wu, Bin, et al. "Voxel-based marked neighborhood searching method for identifying street trees using vehicle-borne laser scanning data." *2012 Second International Workshop on Earth Observation and Remote Sensing Applications*. IEEE, 2012.
39. Torres, F., Fanti, Z., Lira, E., García-Segundo, C., Reyes-Ramírez, B., Hazan, E. J., ... & Arámbula-Cosío, F. (2012). Rastreo de imágenes y reconstrucción de volúmenes de ultrasonido médico. *Revista mexicana de ingeniería biomédica*, *33*(2), 101–115.
40. Aroeira, R. M. C., Pertence, A. E. D. M., Kemmoku, D. T., & Greco, M. (2017). Three-dimensional geometric model of the middle segment of the thoracic spine based on graphical images for finite element analysis. *Research on Biomedical Engineering*, *33*(2), 97–104.
41. Di Angelo, L., & Di Stefano, P. (2015). A new method for the automatic identification of the dimensional features of vertebrae. *Computer methods and programs in biomedicine*, *121*(1), 36–48.
42. Bellia-Munzon, G., Martinez, J., Toselli, L., Peirano, M. N., Sanjurjo, D., Vallee, M., & Martinez-Ferro, M. (2020). From bench to bedside: 3D reconstruction and printing as a valuable tool for the chest wall surgeon. *Journal of Pediatric Surgery*, *55*(12), 2703–2709.
43. Redaelli, D. F., Barsanti, S. G., Fraschini, P., Biffi, E., & Colombo, G. (2018). LOW-COST 3D DEVICES AND LASER SCANNERS COMPARISON FOR THE APPLICATION IN ORTHOPEDIC CENTRES. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *42*(2).
44. Matsushima, K., Nishi, H., & Nakahara, S. (2012). Simple wave-field rendering for photorealistic reconstruction in polygon-based high-definition computer holography. *Journal of Electronic Imaging*, *21*(2), 023002.
45. Hernandez, M., Choi, J., & Medioni, G. (2015). Near laser-scan quality 3-D face reconstruction from a low-quality depth stream. *Image and Vision Computing*, *36*, 61–69.
46. Savvides, M., & Heo, J. (2014). *U.S. Patent No. 8,861,800*. Washington, DC: U.S. Patent and Trademark.
47. Lee, Y. C., Lin, G., & Wang, M. J. J. (2014). Comparing 3D foot scanning with conventional measurement methods. *Journal of foot and ankle research*, *7*(1), 44.
48. Li, Y. B., Xiao, H., & Zhang, S. Y. (2007, August). The wrinkle generation method for facial reconstruction based on extraction of partition wrinkle line features and fractal interpolation. In *Fourth International Conference on Image and Graphics (ICIG 2007)* (pp. 933–937). IEEE.

49. Yuan, B. (2016). *A computational framework for evaluating outcomes in infant craniosynostosis reconstruction* (Doctoral dissertation).
50. De Jong, G. A., Maal, T. J., & Delye, H. (2015). The computed cranial focal point. *Journal of Cranio-Maxillofacial Surgery*, *43*(9), 1737–1742.
51. I-Ware Laboratory Co., L. (2020, 01 01). Downloads. Retrieved from Infoot Standard Sample Date: http://www.i-ware.co.jp/content/infoot.html
52. Artec3D, A. (2013). Eva 3D. Retrieved from Downloads: https://www.artec3d.com/3d-models
53. Nilsson, M. K., Friis, R., Michaelsen, M. S., Jakobsen, P. A., & Nielsen, R. O. (2012). Classification of the height and flexibility of the medial longitudinal arch of the foot. *Journal of foot and ankle research*, *5*(1).
54. Dahle LK, Mueller MJ, Delitto A, Diamond JE: Visual assessment of foot type and relationship of foot type to lower extremity injury. JOrthop Sports Phys Ther. 1991, 14: 70–74.
55. Hebal, F., Port, E., Hunter, C. J., Malas, B., Green, J., & Reynolds, M. (2019). A novel technique to measure severity of pediatric pectus excavatum using white light scanning. *Journal of pediatric surgery*, *54*(4), 656–662.
56. Real-Moreno, O., Castro-Toscano, M. J., Rodrízuez-Ouiñonez, J. C., Hernández-Balbuena, D., Flores-Fuentes, W., & Rivas-Lopez, M. (2018, October). Implementing k-nearest neighbor algorithm on scanning aperture for accuracy improvement. In *IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society* (pp. 3182–3186). IEEE.
57. Ramírez-Hernández, L. R., Rodríguez-Quiñonez, J. C., Castro-Toscano, M. J., Hernández-Balbuena, D., Flores-Fuentes, W., Rascón-Carmona, R., ... & Sergiyenko, O. (2020). Improve three-dimensional point localization accuracy in stereo vision systems using a novel camera calibration method. *International Journal of Advanced Robotic Systems*, *17*(1), 1729881419896717.
58. Trujillo, G. (2019). Accuracy improvement by artificial neural networks in technical vision system. [Fig. 22]. Recovered from https://ieeexplore.ieee.org/abstract/document/8927596?casa_token=dZLVXSG iRIAAAAA:ERdQY9gryR_hyEEAczLhor7RLx3l9LjZCUdnTfPwUJ1S6bjdiZwNZqIfo22mO6PpYm58JeHJL0M067c

Fuzzy Decision-Making for Intelligent Robotic System

Igor Nevliudov, Oleksandr Tsymbal, and Artem Bronnikov

Abbreviations

ACS	Automated control system
ANN	Artificial neural network
FIS	Flexible integrated systems
FWFF	Fuzzy well-formed formula
ODM	Original design manufacturer
WFF	Well-formed formula

1 Introduction

The research, development, and application of flexible integrated systems (FIS), is one of the feature entrances to the postindustrial stage of development. Application of FIS must provide the quick and low-cost transition to the new production type output, especially for conditions of low-series production. The efficiency of FIS is determined by optimal organization of technological equipment usage, supplied by robotized and transport systems, delivering bars, details, and instruments, making the required service and check of technological processes.

The mass use of FIS becomes effective only then, with the society demands for their application and the level and culture of production reach of a certain quality level. Commercially viable development, introduction, and exploitation of FISs and robots are still quite expensive and not possible for every factory in any country.

I. Nevliudov · O. Tsymbal (✉) · A. Bronnikov
Department of Computer-Integrated Technologies, Automation and Mechatronics of Kharkiv National University of Radioelectronics, Kharkiv Oblast, Ukraine
e-mail: igor.nevliudov@nure.ua; oleksandr.tsymbal@nure.ua; artem.bronnikov@nure.ua

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1_9

Recent industrial and natural disasters have shown that even the most developed countries have a level of robotics, which doesn't correspond to the modern current needs of technical tool application in dangerous human conditions.

A possible solution is in the introduction of intelligent mobile robotic systems able to combine manipulation properties of industrial robots, mobility of transport systems, and intelligence of human knowledge.

Intelligent control belongs to a class of technologies which use different AI methods, including control with application of artificial neural network (ANN), Bayesian control, fuzzy control, expert systems, genetic control, and intellectual agents (cognitive control).

The proposed article considers application of fuzzy methods for decision-making and control of robotic systems.

2 Fuzzy Expressions and Their Description in Automated Control Systems (ACS)

Classical approach to ACS strategy planning considers certainty of all events on the binary level. So, the event exists or not. The necessity to consider the risks connected with action implementation and robot state causes other approaches used for strategy planning process modeling in particular based on probability theory and theory of fuzzy sets. Fuzzy sets used for strategy planning mostly are fuzzy evaluations (fuzzification) of classic decision-making theories.

Similar to the clear expressions' logic [1], let's introduce the concept of atomic fuzzy expressions (or fuzzy atoms) $\tilde{A}$, $\tilde{B}$, $\tilde{C}$ etc.

Atomic expressions are such expressions that cannot be further broken down into more fundamental particles.

Each fuzzy atomic expression has truth value in the interval [0, 1]. Complicated formulas in the logic are built by symbols that are atomic expressions and logic connectors' combination:

$\neg$ negation
$\&$ logical multiplication or conjunction
$\vee$ logical sum or disjunction
$\rightarrow$ follow, implication
$\leftrightarrow$ identity

A complicated formula that contains logical connectors is called well-formed formula (WFF). Usually for expressions logic description symbols $\tilde{p}$, $\tilde{q}$, $\tilde{r}$ are used. They are called propositional variables. If propositional variables are fuzzy, WFF is determined in the next way.

Fuzzy atomic expression symbol is fuzzy WFF.

If $\tilde{p}$ and $\tilde{q}$ are fuzzy WFF, then the next formulas are also fuzzy WFF:

$$\neg\tilde{p} \quad - \quad \text{not} \quad \tilde{p}; \quad \tilde{p}\&\tilde{q}; \quad \tilde{p} \vee \tilde{q}; \quad \tilde{p} \to \tilde{q}; \quad \tilde{p} \leftrightarrow \tilde{q}. \tag{1}$$

Real fuzzy WFF can be written in the next way:

$$\bar{A}\&\left(\bar{B} \vee \bar{C}\right), \quad \left(\bar{A}\&\bar{B}\right) \vee \bar{C} \to \bar{D}, \quad \text{etc.} \tag{2}$$

The form in which WFF is written or the method by which it is constructed from atomic expressions is called WFF syntax. Truth values are WFF semantics. As expressions have regular truth values from the basis of them, we can determine truth of every regular or fuzzy WFF (FWFF) according to the laws (rules).

While truth counting of FWFF is constructed by atomic expressions and logical connectors, the main difference in connectors' interpretation appears. If atomic expressions $\tilde{p}$ and $\tilde{q}$ exist, and $\tilde{p} = \langle \mu_p/p \rangle$, $\tilde{q} = \langle \mu_p/p \rangle$, we can consider a table of FWFF truth verification (valence table appellation is used, for example, in [1, 2]) (Table 1):

If we consider fuzzy atoms $\tilde{A} = \{0.7/A\}$ and $\tilde{B} = \{0.2/B\}$, we can get the next truth table (Table 2):

Table 1 Truth table for conditional fuzzy expression

p	q	$\tilde{p}$	$\tilde{q}$	$\tilde{p}\&\tilde{q}$
t	t	$\langle \mu_p/t \rangle$	$\langle \mu_p/t \rangle$	$\langle \min(\mu_p, \mu_q)/t \rangle$
t	f	$\langle \mu_p/t \rangle$	$\langle \mu_p/f \rangle$	$\langle \min(\mu_p, \mu_q)/f \rangle$
f	t	$\langle \mu_p/f \rangle$	$\langle \mu_p/f \rangle$	$\langle \min(\mu_p, \mu_q)/f \rangle$
f	f	$\langle \mu_p/f \rangle$	$\langle \mu_p/t \rangle$	$\langle \min(\mu_p, \mu_q)/f \rangle$
p	q	$\tilde{p}\&\tilde{q}$	$\neg\tilde{p}$	$\neg\tilde{p} \vee \tilde{q}$
t	t	$\langle \max(\mu_p, \mu_q)/t \rangle$	$\langle \mu_p/f \rangle$	$\langle \max(\mu_p, \mu_q)/t \rangle$
t	f	$\langle \max(\mu_p, \mu_q)/f \rangle$	$\langle \mu_p/f \rangle$	$\langle \max(\mu_p, \mu_q)/f \rangle$
f	t	$\langle \max(\mu_p, \mu_q)/f \rangle$	$\langle \mu_p/t \rangle$	$\langle \max(\mu_p, \mu_q)/t \rangle$
f	f	$\langle \max(\mu_p, \mu_q)/f \rangle$	$\langle \mu_p/t \rangle$	$\langle \max(\mu_p, \mu_q)/t \rangle$

Table 2 Truth table for conditional fuzzy expression

A	B	$\tilde{A}$	$\tilde{B}$	$\tilde{A}\&\tilde{B}$	$\tilde{A} \vee \tilde{B}$	$\neg\tilde{A}$	$\neg\tilde{A} \vee \tilde{B}$
t	t	<0.7/t>	<0.2/t>	<0.2/t>	<0.7/t>	<0.7/f>	<0.7/t>
t	f	<0.7/t>	<0.2/f>	<0.2/f>	<0.7/t>	<0.7/f>	<0.7/f>
f	t	<0.7/f>	<0.2/t>	<0.2/f>	<0.7/t>	<0.7/t>	<0.7/t>
f	f	<0.7/f>	<0.2/f>	<0.2/f>	<0.7/f>	<0.7/t>	<0.7/t>

Let's consider objection operation for fuzzy atom $\tilde{A}$.

If $\tilde{A} = \{0.7/A\}$, then $\neg\tilde{A} = \{1 - 0.7/A\} = 0.3/A$. Similarly, we can write $\neg\tilde{A} = \{0.7/\neg A\} = 0.7/\neg A$. So we get $0.3/A = 0.7/\neg A$. These records are identical. So, on the one hand, objection is provided by difference calculation between 1 and certainty coefficient $1 - \mu_a$ and, on the other hand, by the denial of a regular base of the fuzzy atom. Moreover, function value doesn't change.

It seems that regular base of fuzzy atom is error, but in practice in fuzzy sets, this base contains an element of fuzzy sets, for example, set of "small numbers" [3]:

$$\tilde{A} = \{1/0, \quad 1/1, \quad 0.8/2, \quad 0.5/3, \quad 0.1/4\}.$$

Another ambiguous moment is implication consideration, i.e., such expressions as "*If A, then B*." According to classical logic, implication is written in such way: $\neg A \vee B$. This expression in fuzzy logic corresponds to record $\neg\tilde{A} \vee \tilde{B}$, which for certainty coefficient calculation gives such expression as:

$$\mu_{a\to b} = (1 - \mu_a) \vee \mu_a \Rightarrow \max(1 - \mu_a, \mu_b). \tag{3}$$

In [4] it is indicated that calculation results are not accurate enough, if they are obtained this way. It is proposed to use implication formula from multivalued Lukasevich logic:

$$\mu_{a\to b} = (1 - \mu_a + \mu_b)\,\&1 \Rightarrow \min(1 - \mu_a + \mu_b, 1). \tag{4}$$

For fuzzy sets and for appropriate fuzzy logic, it is possible to prove main logic laws, for example, idempotency, commutativity, associativity, twice objection, and De Morgan's laws. Complementarity law is not performed. Specified laws are shown below:

(a) Idempotence

$$\text{For regular logic, it is written this way}: A\&A = A, \quad A \vee A = A.$$

$$\text{For fuzzy logic, it can be written this way}: \tilde{A}\&\tilde{A} = \tilde{A}, \quad \tilde{A} \vee \tilde{A} = \tilde{A}.$$

$$\text{Coefficients}'\text{check gives the following}: \min(\mu_a, \mu_a) = \mu_a, \max(\mu_a, \mu_a) = \mu_a.$$

(b) Commutative property

$$\text{For regular logic}: A\&B = B\&A, \quad A \vee B = B \vee A$$

$$\text{For fuzzy logic}: \tilde{A}\&\tilde{B} = \tilde{B}\&\tilde{A}, \quad \tilde{A} \vee \tilde{B} = \tilde{B} \vee \tilde{A}$$

$$\text{Really}, \min(\mu_a, \mu_b) = \min(\mu_b, \mu_a), \quad \max(\mu_a, \mu_b) = \max(\mu_b, \mu_a).$$

(c) Associative property

For regular logic:

$$A\&\,(B\&C) = (A\&B)\,\&C, \qquad A \vee (B \vee C) = (A \vee B) \vee C$$

For fuzzy logic:

$$\tilde{A}\&\left(\tilde{B}\&\tilde{C}\right) = \left(\tilde{A}\&\tilde{B}\right)\&\tilde{C}, \qquad \tilde{A} \vee \left(\tilde{B} \vee \tilde{C}\right) = \left(\tilde{A} \vee \tilde{B}\right) \vee \tilde{C}$$

Really, coefficients' check gives the following:

$$\min\left(\mu_a, \min\left(\mu_b, \mu_c\right)\right) = \min\left(\min\left(\mu_a, \mu_b\right), \mu_c\right)$$

$$\max\left(\mu_a, \max\left(\mu_b, \mu_c\right)\right) = \max\left(\max\left(\mu_a, \mu_b\right), \mu_c\right)$$

(d) Distributive property

For regular logic:

$$A\&\,(B\&C) = (A\&B)\,\&C$$

$$A \vee (B \vee C) = (A \vee B) \vee C$$

For fuzzy logic:

$$\tilde{A}\&\left(\tilde{B}\&\tilde{C}\right) = \left(\tilde{A}\&\tilde{B}\right)\&\tilde{C}$$

$$\tilde{A} \vee \left(\tilde{B} \vee \tilde{C}\right) = \left(\tilde{A} \vee \tilde{B}\right) \vee \tilde{C}$$

Really, coefficients' check gives the following:

$$\min\left(\mu_a, \max\left(\mu_b, \mu_c\right)\right) = \max\left(\min\left(\mu_a, \mu_b\right), \min\left(\mu_a, \mu_c\right)\right)$$

$$\min\left(\mu_a, \min\left(\mu_b, \mu_c\right)\right) = \min\left(\max\left(\mu_a, \mu_b\right), \min\left(\mu_a, \mu_c\right)\right)$$

Thus, for systems in which fuzziness takes place, description is necessary to introduce the concept of fuzzy logic which is an extension of classical mathematical logic. The next steps must be fuzzy predicates' theory development [5–9].

3 Individual Strategy Planning in ACS

As usual several classes are switched among strategy planning tasks [10, 11]. According to one criterion, strategy planning tasks are classified by one or several subjects providing strategy planning availability. Such tasks belong to one or several persons' (collective) strategy planning. According to other criteria, those tasks are switched that require simple utility function optimization, optimization with some limitations, with several goals. Moreover, strategy planning may be performed in one step or iteratively as a sequence of steps (stages).

To the existing models, fuzziness is introduced in different ways; moreover, strategy planning task is characterized by the next component existence:

1. Set A of possible actions
2. Set of goals G_i ($i \in N_n$), each of which is written as a fuzzy set determined on A
3. Limitation set C_j ($j \in N_m$), by similar record conditions

Let $G_i^{'}$ and $C_j^{'}$ be fuzzy sets, determined on sets X_i and Y_j, likewise, where $i \in N_n$ and $j \in N_m$. Let's say that these fuzzy sets represent goals and limitations, expressed by the strategy planning system. Then for each $i \in N_n$ and $j \in N_m$, let's describe action values on set A in terms of sets X and Y:

$$g_i : A \to X_i, \quad c_j : A \to Y_j, \tag{5}$$

and display goals G_i and C_j by compositions g_i from $G_i^{'}$ and c_i from $C_i^{'}$, i.e.:

$$G_i(a) = G_i{}'(g_i(a)), \quad C_j(a) = C_j{}'\left(c_j(a)\right). \tag{6}$$

Fuzzy decision D is considered on set A and must correspond to specified goals G_i and conditions C_j at the same time, i.e.:

$$D(a) = \min\left[\inf_{i \in N_n} G_i(a), \quad \inf_{j \in N_m} C_j(a)\right], \quad \text{for all} \quad a \in A \tag{7}$$

Let's consider the example of a robot strategy planning system. The robot is in simple space with dimensions 4×4 with fuzzy characteristics (shown in Fig. 1).

Let's say for specified robot's world such section states are determined:

(a) Sections *c(0, 0), c(1, 0), c(2, 0), c(3, 0), c(0, 1), c(1, 1), c(0, 2), c(0, 3), c(3, 1), c(3, 2), c(2, 2)* are absolutely passable and have good cover state that allows to move quickly.
(b) Sections *c(2, 1), c(1, 2)* are occupied and can't be passed.
(c) Sections *c(1, 3), c(2, 3), c(3, 3)* are not occupied, but have bad cover state.

According to fuzzy set theory, let's evaluate quality and occupation of working space sections:

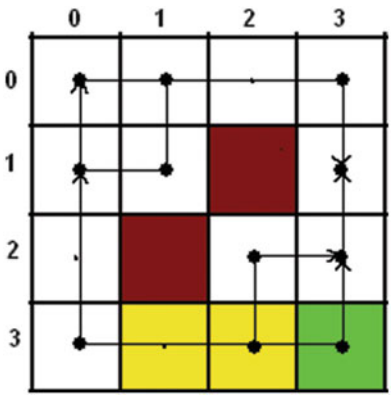


Fig. 1 Working space example with different degrees of passage section certainty (brown, impassable; white, absolutely passable; yellow and green, mid-level passability)

- Class A: K(c(0, 0), c(1, 0), c(2, 0), c(3, 0), c(0, 1), c(1, 1), c(0, 2), c(0, 3), c(3, 1), c(3, 2), c(2, 2)) =1
- Class B: K(c(2, 1), c(1, 2))=0
- Class C: K(c(1, 3), c(2, 3), c(3, 3))=0.4

Now let's consider potential ways in space. Let's assume that only horizontal and vertical paths are possible. Moreover, it is not allowed to visit sections repeatedly during route execution.

Thus, we'll have such routes:

$$(0, 3) \rightarrow (0, 2) \rightarrow (0, 1) \rightarrow (0, 0) \rightarrow (1, 0) \rightarrow (2, 0) \rightarrow (3, 0) \rightarrow (3, 1)$$

$$(0, 3) \rightarrow (0, 2) \rightarrow (0, 1) \rightarrow (1, 1) \rightarrow (1, 0) \rightarrow (2, 0) \rightarrow (3, 0) \rightarrow (3, 1)$$

$$(0, 3) \rightarrow (1, 3) \rightarrow (2, 3) \rightarrow (3, 3) \rightarrow (3, 2) \rightarrow (3, 1)$$

$$(0, 3) \rightarrow (1, 3) \rightarrow (2, 3) \rightarrow (2, 2) \rightarrow (3, 2) \rightarrow (3, 1)$$

After route inspection, let's evaluate them. The first two routes (1) and (2) are the longest – 7 conventional units. Two others are shorter – 5 units. So, considering the current situation, 7 is the maximum value but it isn't endless. That's why at the level of the "minimum route" concepts corresponding to (3) and (4), which is associated with certainty coefficient 1, let's evaluate first two routes as $5/7 \approx 0.7$.

So, certainty coefficients will get the next values:

$$K_{p1}{}' = 0.7; \quad K_{p2}{}' = 0.7; \quad K_{p3}{}' = 1.0; \quad K_{p4}{}' = 1.0.$$

Considering routes' pass ability, it's necessary to evaluate a common value characteristic for the entire route including coefficients corresponding to every section:

$$K_1''(rout1) = \min\left(K_{p1}, K_{p2}, K_{p3} \ldots, K_{p8}\right) = 1;$$

$$K_2''(rout2) = \min\left(K_{p1}, K_{p2}, K_{p3} \ldots, K_{p8}\right) = 1;$$

$$K_3''(rout3) = \min\left(K_{p1}, K_{p2}, K_{p3} \ldots, K_{p6}\right) = 0.4;$$

$$K_4''(rout4) = \min\left(K_{p1}, K_{p2}, K_{p3} \ldots, K_{p4}\right) = 0.4.$$

As is known from [12, 13], in expression that determines the decision, it is necessary to consider both goals (the shortest route here) and limitations (pavement quality). That's why described ways will formulate fuzzy goal set in next way:

$$G' = 0.7/r_1 + 0.7/r_2 + 1.0/r_3 + 1.0/r_4.$$

Fuzzy set corresponding to cover quality will have similar view:

$$C_1' = 1.0/r_1 + 1.0/r_2 + 0.4/r_3 + 0.4/r_4.$$

In addition, let's introduce one more limitation on the number of the route turns. It'll depend on the requirement of minimum turns provided so the speed will be constantly maximum:

$$C_2' = 0.8/r_1 + 0.6/r_2 + 1.0/r_3 + 0.7/r_4.$$

According to the expression about fuzzy decision record, let's write next:

$$D(r) = \min\left[infG_i(r), infC_j(r)\right],$$

i.e., for our occasion, we'll get next:

$$D(r) = 0.7/r_1 + 0.6/r_2 + 0.4/r_3 + 0.4/r_4.$$

As a result, in the form fuzzy set, we'll select the route with the maximum value of certainty coefficient, i.e., route r_1.

Proposed fuzzy model allows strategy planning system to describe goals and conditions in terms that have to display strategy planning situations more accurately. The fuzzy goals' membership function in this model is for the same goals as the utility function and target function in the classical strategy planning model

that regulates the output values according to preferences. Contrarily, the classical planning strategy theory within the limitations of symmetry between the goals and conditions in the fuzzy model allows to treat them in the same way.

Fuzzy model further can be extended to take into account the relative importance of different goals and limitations by using weighting coefficients. In this case, fuzzy decision D should be displayed by a combination of n weighted goals and m weighted limitations in the next form:

$$D(a) = \sum_{i=1}^{n} u_i G_i(a) + \sum_{j=1}^{m} v_j C_j(a), \tag{8}$$

for all $a \in A$, where u_i and v_j are more than zero weighting coefficients that are joined to each fuzzy goal $G_i^{'}$ $(i \in N_n)$ and each fuzzy limitation $C_j^{'}$ $(j \in N_m)$ accordingly, i.e.:

$$\sum_{i=1}^{n} u_i + \sum_{j=1}^{m} v_j = 1, \tag{9}$$

but direct formula expansion may be also used:

$$D(a) = \min \left[\inf_{i \in N_n} G_i^{u_i}(a), \quad \inf_{j \in N_m} C_j^{v_j}(a) \right], \tag{10}$$

where u_i and v_j have the above properties.

4 Multistep Strategy Planning

The strategy planning task in terms of fuzzy dynamic programming is considered as a problem of fuzzy finite automata [14, 15]. However, the theory limitation is a problem of the transition function describing, characterized by analog from the classical theory of automata. Another limitation is the lack of output values, i.e., the next internal state is used as output value and, therefore, it is necessary to distinguish the two states.

These limits define the problem formulation: automatic A, used in fuzzy dynamic programming, determined by the triplet:

$$A = \langle X, Z, f \rangle, \tag{11}$$

where X and Z are the sets of input and output states, respectively, so:

$$f : Z \times X \to Z. \tag{12}$$

It is a transition function, which automate following internal state z^{t+1} determined for each discrete time point $t \in N$ in terms of the following internal z^t and x^t current input states, i.e.:

$$z^{t+1} = f\left(z^t, x^t\right). \tag{13}$$

This automate type is used in classic dynamic programming. In the case of fuzzy dynamic programming, data must be fuzzificated using the expansion principle. In the fuzzy version scheme, automates A^t and C^t correspond to fuzzy input state and fuzzy internal state at the moment of time t, and C^{t+1} is a fuzzy internal state at the moment of time $t+1$. A^t corresponds to a fuzzy set at X, and C^t and C^{t+1} are fuzzy sets at Z.

Fuzzy automate using can provide a fuzzy dynamic programming description. In considering the concept, a necessary goal is expressed in terms of fuzzy set C^N (fuzzy internal state A at time N), where N is the time of the planning strategy end. The value N also indicates the number of planning strategy stages. It is assumed that the input value A is expressed in every moment of time t by fuzzy state at, and initial state z^0 is also set.

Considering the fuzzy input states $A^0, A^1, \ldots, A^{N-1}$ as limitations and fuzzy internal state C^N as a fuzzy goal strategy planning, now it is possible to start a fuzzy decision as fuzzy set X^N, which is defined in next way:

$$D = \tilde{A}^0 \cap \tilde{A}^1 \cap \cdots \cap \tilde{A}^{N-1} \cap \tilde{C}^N, \tag{14}$$

where $\tilde{A}^t$ is a cylindrical extension $\tilde{A}^t$ from X to X^N for each $t = 0, 1, \ldots, N-1$ i and $\tilde{C}^N$ is a fuzzy set at X^N that provides C^N at Z. That is, for each sequence $x^0, x^1, \ldots, x^{N-1}$, considered a decision sequence, the belonging degree to D is expressed in the next way:

$$D\left(x^0, \mathrm{x}^1, \ldots, \mathrm{x}^{\mathrm{N}-1}\right) = \min\left[A^0\left(x^0\right), A^1\left(x^1\right), \ldots, A^{N-1}\left(x^{\mathrm{N}-1}\right), C^N\left(z^N\right)\right], \tag{15}$$

where z^N is determined uniquely on the basis of the information about $x^0, x^1, \ldots, x^{N-1}$ and z^0. This definition provides a standard intersection operator availability. Strategy planning is to find the initial states' sequence $\tilde{x}^0, \tilde{x}^1, \ldots, \tilde{x}^{N-1}$ such that:

$$D\left(\tilde{x}^0, \tilde{\mathrm{x}}^1, \ldots, \tilde{\mathrm{x}}^{\mathrm{N}-1}\right) = \max_{x^0, \ldots, x^{N-1}} D\left(x^0, x^1, \ldots, x^{\mathrm{N}-1}\right). \tag{16}$$

To solve such a task by the fuzzy dynamic programming tools, it is necessary to use Bellman's principle of optimality: the optimal sequence of decisions have the next feature – for any initial states and initial decisions, other decisions must correspond to the optimal strategy respectively of the state got from the first decision [16–20].

Using the principle of optimality and substituting the expression for D, we can write next:

$$D\left(\tilde{x}^0, \tilde{\mathrm{x}}^1, \ldots, \tilde{\mathrm{x}}^{\mathrm{N}-1}\right) = \max_{x^0,\ldots,x^{N-2}} \left\{ \max_{x^{N-1}} \min\left[A^0\left(x^0\right), A^1\left(x^1\right), \ldots, A^{N-1}\left(x^{\mathrm{N}-1}\right), C^N\left(f\left(z^{\mathrm{N}-1}, x^{\mathrm{N}-1}\right)\right)\right]\right\}. \tag{17}$$

This expression can be rewritten:

$$\begin{aligned} D\left(\tilde{x}^0, \tilde{\mathrm{x}}^1, \ldots, \tilde{\mathrm{x}}^{\mathrm{N}-1}\right) &= \max_{x^0,\ldots,x^{N-2}} \left\{ \min\left[A^0\left(x^0\right), A^1\left(x^1\right), \ldots, A^{N-2}\left(x^{\mathrm{N}-2}\right), \right.\right. \\ &\left.\left. \max_{x^{N-1}} \min\left[A^{N-1}\left(x^{\mathrm{N}-1}\right), C^N\left(f\left(z^{\mathrm{N}-1}, x^{\mathrm{N}-1}\right)\right)\right]\right]\right\} \\ &= \max_{x^0,\ldots,x^{N-2}} \left\{\min\left[A^0\left(x^0\right), A^1\left(x^1\right), \ldots, A^{N-2}\left(x^{\mathrm{N}-2}\right), \right.\right. \\ &\left.\left. \max_{x^{N-1}} \min\left[A^{N-1}\left(x^{\mathrm{N}-1}\right), C^N\left(z^{\mathrm{N}-1}\right)\right]\right]\right\} \\ &= \max_{x^0,\ldots,x^{N-2}} \left\{\min\left[A^0\left(x^0\right), A^1\left(x^1\right), \ldots, A^{N-2}\left(x^{\mathrm{N}-2}\right), C^{N-1}\left(z^{N-1}\right)\right]\right\} \end{aligned}$$

where $C^{N-1}\left(z^{N-1}\right) = \max_{,x^{N-1}} \min\left[A^{N-1}\left(x^{N-1}\right), C^N\left(z^N\right)\right]$.

Repeating inverse iteration, we get a set of recurrence equations:

$$C^{N-k}\left(z^{N-k}\right) = \max_{,x^{N-k}} \min\left[A^{N-k}\left(x^{\mathrm{N}-\mathrm{k}}\right), C^{N-k+1}\left(z^{N-k+1}\right)\right] \tag{18}$$

for $k = 1, 2, \ldots, N$, where

$$z^{N-k+1} = f\left(z^{N-k}, x^{N-k}\right) \tag{19}$$

Decision optimal sequence $\tilde{x}^0, \tilde{x}^1, \ldots, \tilde{x}^{N-1}$ can be obtained by the values x^{N-k} maximization for $k = 1, 2, \ldots, N$, which leads to results $\tilde{x}^{N-1}, \tilde{x}^{N-2}, \ldots, \tilde{x}^0$.

Considering robotic system functioning strategy planning fuzzy model requires fuzzy sets $\tilde{X}, \tilde{D}, \tilde{S}$ introducing, which describe robotic systems states, their decisions, WS states, so according to this:

$$\tilde{x}_i \in \tilde{X}, \tilde{X} = \{\mu_0/x_0, \mu_1/x_1, \ldots, \mu_{n-1}/x_{n-1}\}, \tag{20}$$

$$\tilde{d}_i \in \tilde{D}, \tilde{D} = \{\mu_0/d_0, \mu_1/d_1, \ldots, \mu_{n-1}/d_{n-1}\},$$

$$\tilde{s}_i \in \tilde{S}, \tilde{S} = \{\mu_0/s_0, \mu_1/s_1, \ldots, \mu_{n-1}/s_{n-1}\}.$$

Fuzziness (or certainty) is determined by information about robotic system state fuzziness (it's difficult to determine the suitability of robot gripper to capture details, wheel adhesion to the surface and their condition, battery status, etc.). Space fuzziness is determined by incomplete information about its observation (errors of sensors and technical vision system) and space change fuzziness.

Discrete space state determines the space in a matrix:

$$S = \begin{bmatrix} \mu_0^0/S_0^0 & \mu_0^1/S_0^1 & \ldots \mu_0^{m-1}/S_0^{m-1} \\ \mu_1^0/S_1^0 & \mu_1^1/S_1^1 & \ldots \mu_1^{m-1}/S_1^{m-1} \\ \ldots & \ldots & \ldots \quad \ldots \\ \mu_{n-1}^0/S_{n-1}^0 & \mu_{n-1}^1/S_{n-1}^1 & \ldots \mu_{n-1}^{m-1}/S_{n-1}^{m-1} \end{bmatrix}$$

In papers about technical vision systems the trajectory setting, based on Freeman chain code, is widely used [21].

Similarly, space section states directly surrounding current robot position are described by the vector of dimension 9:

$$\begin{bmatrix} 4\ 3\ 2 \\ 5\ 0\ 1 \\ 6\ 7\ 8 \end{bmatrix}, \tilde{S} = \{\mu_0/s_0, \mu_1/s_1, \ldots, \mu_8/s_8\}.$$

Similarly, robot movements are determined:

$$\tilde{M}v = \{\mu_0/mv_0, \mu_1/mv_1, \ldots, \mu_8/mv_8\}, \tag{21}$$

Manipulations with the objects:

$$\tilde{M}p = \{\mu_0/mp_0, \mu_1/mp_1, \ldots, \mu_{n-1}/mp_{n-1}\}. \tag{22}$$

Robotic system functioning goal is the achievement of the target (goal) states $\tilde{Y}$ or $\tilde{S}^g$ that can be obtained as a result of robotic system or WS state evolution step by step:

$$\tilde{X}_0 \to \tilde{X}_1 \to \cdots \to \tilde{X}_{n-1} \equiv \tilde{Y} \text{ or } \tilde{S}_0 \to \tilde{S}_1 \to \cdots \to \tilde{S}_{n-1} \equiv \tilde{S}^g.$$

Robotic system or WS state evolution occurs as a result of certain decision strategy used by robotic system:

$$\tilde{D}_0 \to \tilde{D}_1 \to \cdots \to \tilde{D}_{n-1} \equiv \tilde{D} \text{ or } \tilde{D} = \wedge_{i=0}^{n} \tilde{D}_i \tag{23}$$

and is expressed in actions or manipulations:

$$\tilde{A}_i = f\left(x_{i-1}, s_{i-1}^{'}, d_{i-1}\right), \tilde{A}_i = \tilde{M} v_i \vee \tilde{M} p_i.$$

$$\text{Limitations}: \left\| \tilde{A}_i - \tilde{A}_{i-1} \right\| = \mu_i\left(A_i\right) - \mu_{i-1}\left(A_{i-1}\right) \le \varepsilon_{ai},$$

$$\left\| \tilde{x}_i - \tilde{x}_{i-1} \right\| = \mu_i\left(x_i\right) - \mu_{i-1}\left(x_{i-1}\right) \le \varepsilon_{xi},$$

$$\left\| \tilde{s}_i - \tilde{s}_{i-1} \right\| = \mu_i\left(s_i\right) - \mu_{i-1}\left(s_{i-1}\right) \le \varepsilon_{si}.$$

For limitations' non-completion because of WS dynamic state, robotic system's insufficient properties will require initial decision plan $\overset{\to}{\tilde{D}}^{0}$ reformulation that has to compensate non-completion of limitations.

The final plan will include elements of previous plans:

$$\tilde{D}^{\text{final}} = \tilde{D}^0 \cup \tilde{D}^1 \cup \cdots \cup \tilde{D}^{k-1}.$$

Fuzzy set theory allows to describe robotic system state fuzziness and fuzziness in action selection at every step of discrete strategy planning process [22]. Decision-making process for the fuzzy system will consist of action execution sequence in conditions of fuzzy system characteristics:

$$\tilde{X}_0 * \tilde{D}_0 \Rightarrow \tilde{X}_1 * \tilde{D}_1 \Rightarrow \cdots \Rightarrow \tilde{X}_{n-1} * \tilde{D}_{n-1} \Rightarrow \tilde{Y}_n.$$

Herewith, each action is evaluated as a fuzzy variable.

In practice, fuzzy sets are proposed to use for mobile robot movements in WS description and to characterize moving robot in certain direction possibility, wherein belonging function constructing should be based on experience by accumulating IDMS FIS during functioning.

Maybe during the consideration of strategy planning system, everything should be oriented to the set that displays system purpose. Such a set will include one or more local purposes that as a result can be changed by the system.

In the case when purposes are set as a geometric position of the robot or objects, it is necessary to use sets depending on current working points or object state:

$$\tilde{P}_i = \left\langle \mu_0/\tilde{P}_i^0, \mu_1/\tilde{P}_i^1, \mu_2/\tilde{P}_i^2, \ldots, \mu_{n-1}/\tilde{P}_i^{n-1} \right\rangle, \tag{24}$$

$$\tilde{S}_i = \left\langle \mu_0/\tilde{S}_i^0, \mu_1/\tilde{S}_i^1, \mu_2/\tilde{S}_i^2, \ldots, \mu_{n-1}/\tilde{S}_i^{n-1} \right\rangle,$$

where $\tilde{P}$, $\tilde{S}$ are the appropriate point sets and system states. Of course, such sets must be interrelated, because the position determines the state and the state determines position. In the case when specified sets are realized by object-oriented programming methods (or other language tools are used) that allow to reload arguments fuzzy strategy planning system, the purpose will be set or as points set element $\tilde{P}$ or as states set element $\tilde{S}$. For example: $\tilde{P} = \langle 0.8/P_1, 0.6/P_2, 0.4/P_3, 0/P_4 \rangle$, where P_1 is the most desired position, P_2 is a desired position, P_3 is an allowable position, and P_4 is not an allowable position.

Let's consider a practical example of problem for a mobile robot in fixed WS displayed on Fig. 2.

Space characteristics are:

Dimension 100×110 standard units
Objects' presence:
Type object, type door, robot, obstacles
Robot possible actions:
Moving from point p_i to point p_j
Doors d_i opening (closure)
Loading (unloading) of the robot by object in point p_i
Object moving to point p_j

An example of a task can be moving *object_1* to point P_7. Task decision variants:

Variant 1	Variant 2
Move to the point P_1	Move to the point P_1
Take object_1	Take object _1
Move to the point P_2	Move to the point P_3
Open door_1	Open door_2
Move to the point P_5	Move to the point P_6
Move to the point P_7	Move to the point P_7

Thus, in each case, the decision is a consistent set of actions that must be done by the robot for the task purpose provision. Although the task purpose is clear, the choice of an action variant can be considered as a fuzziness based on fuzzy factors.

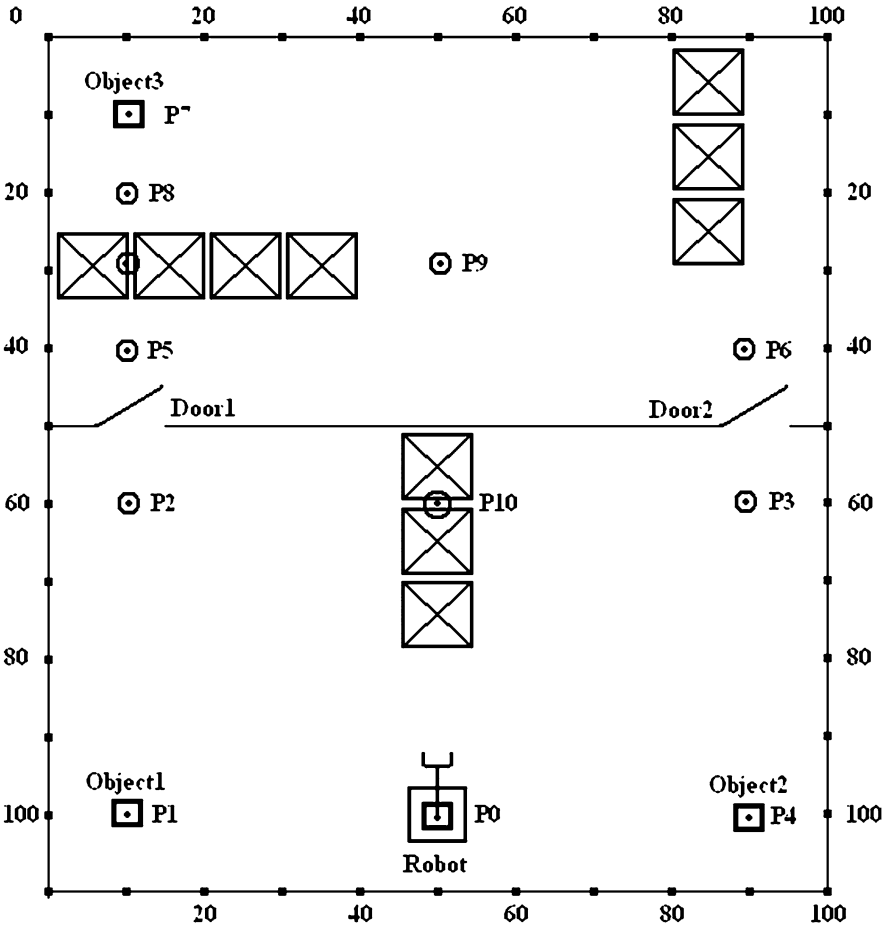


Fig. 2 Mobile robot world model

For such tasks, fuzziness sources are:

(a) State of the doors connecting the room
(b) Direction definition fuzziness in the presence of obstacles
(c) Obstacle fuzzy state which can be moved from one place to another after a certain period of time

These factors are considered as limitations affecting strategy planning.

Motion direction fuzziness depends on the availability of the alternative routes. Addressing the presentation path by Freeman code, we can write a fuzzy set that will characterize the transition from point P_0 to point P_1:

$$\tilde{P}_1 = \langle 0/P_0,\ 0.1/P_1,\ 0/P_2,\ 0.1/P_3,\ 0.8/P_4,\ 0/P_5,\ 0/P_6,\ 0/P_7 \rangle .$$

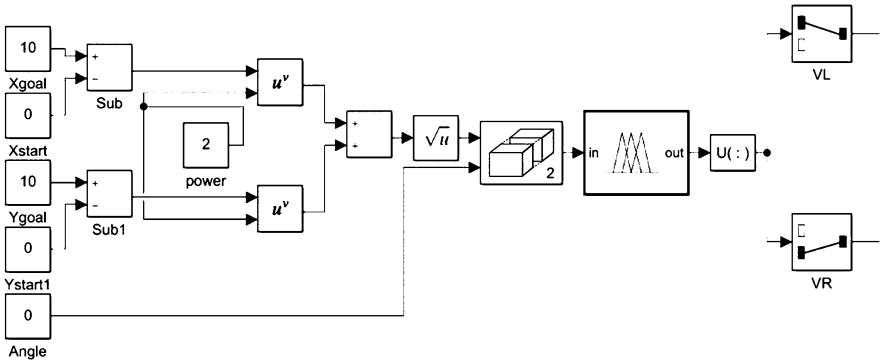


Fig. 3 Fuzzy adaptive robot control model

If the object_1 is already taken, another alternative may occur:

$$\tilde{P}_2 = \langle 0.4/P_0,\ \ 0/P_1,\ \ 0.7/P_2,\ \ 0/P_3,\ \ 0/P_4,\ \ 0/P_5,\ \ 0/P_6,\ \ 0/P_7\rangle\,.$$

Thus, the availability of the alternative routes will determine a fuzzy set that will characterize the robot movement. If we assume that, regardless of robot's actions in the working space, some obstacles are moving (e.g., truck devices using), we can set a fuzziness of obstacle presence at some point:

$$\tilde{P}_3 = \langle 0/P_0, 0/P_1, 0/P_2, 0/P_3, 0/P_4, 0/P_5, 0/P_6, 0/P_7, 0.8/P_8, 0.4/P_9, 1/P_{10}\rangle\,.$$

All these factors will affect the determination of the robot movement from point to point during the decision plan execution.

Consider in detail the example in Fig. 3 about the robot movement in an enclosed workspace. This system can be set in geometric form (specifying the desired points), e.g.:

$$\tilde{G} = \langle 0/P_0, 0/P_1, 0/P_2, 0/P_3, 0/P_4, 0/P_5, 0/P_6, 1/P_7, 0/P_8, 0/P_9, 0/P_{10}\rangle\,.$$

That is certain. In addition, in system there are limitations caused by doors and other obstacles (sets C^0 and C^1, respectively), more exactly the lack of obstacles caused by doors and other objects.

While solving the problem, the purpose and limit sets undergo modifications:

Step 1:

$$\tilde{G}_0 = \langle 0/P_0, 0.8/P_1, 0/P_2, 0/P_3, 0.6/P_4, 0/P_5, 0/P_6, 0/P_7, 0/P_8, 0/P_9, 0/P_{10}\rangle\,,$$

$$C_0^0 = \langle 1/P_0, 1/P_1, 1/P_2, 1/P_3, 1/P_4, 1/P_5, 1/P_6, 1/P_7, 1/P_8, 1/P_9, 1/P_{10}\rangle\,,$$

$$C_0^1 = \langle 1/P_0, 1/P_1, 1/P_2, 1/P_3, 1/P_4, 1/P_5, 0.2/P_6, 0.5/P_7, 0.4/P_8, 0.5/P_9, 0.6/P_{10} \rangle .$$

These sets describe fuzzy sets of targeted points, obstacles caused by doors, and obstacles, respectively.

Strategy planning will correspond to the fuzzy solution formulas that take into account the purposes and limitations. Here set element conjunction is used, which has the result:

$$\tilde{D}_0' = \langle 0/P_0, 0.8/P_1, 0/P_2, 0/P_3, 0.6/P_4, 0/P_5, 0/P_6, 0/P_7, 0/P_8, 0/P_9, 0/P_{10} \rangle .$$

For the final decision, selection expression conjunction will be used (maximizing); thus, it is elected element with a coefficient of 0.8:

$$\tilde{D}_0'' = \langle 0.8/P_1 \rangle .$$

Step 2:

$$\tilde{G}_1 = \langle 0.2/P_0, 0/P_1, 1/P_2, 0.6/P_3, 0/P_4, 0/P_5, 0/P_6, 0/P_7, 0/P_8, 0/P_9, 0/P_{10} \rangle ,$$

$$\tilde{C}_1^0 = \langle 1/P_0, 1/P_1, 1/P_2, 1/P_3, 1/P_4, 1/P_5, 1/P_6, 1/P_7, 1/P_8, 1/P_9, 1/P_{10} \rangle ,$$

$$\tilde{C}_1^1 = \langle 1/P_0, 1/P_1, 1/P_2, 0.2/P_3, 1/P_4, 1/P_5, 0.2/P_6, 0/P_7, 0/P_8, 0.5/P_9, 0/P_{10} \rangle ,$$

$$\tilde{D}_1' = \langle 0.2/P_0, 0/P_1, 1/P_2, 0.2/P_3, 0/P_4, 0/P_5, 0/P_6, 0/P_7, 0/P_8, 0/P_9, 0/P_{10} \rangle ,$$

$$\tilde{D}_1'' = \langle 1/P_2 \rangle .$$

Step 3:

$$\tilde{G}_2 = \langle 0/P_0, 0.2/P_1, 0/P_2, 0.5/P_3, 0/P_4, 0.9/P_5, 0/P_6, 0/P_7, 0/P_8, 0/P_9, 0/P_{10} \rangle$$

$$\tilde{C}_2^0 = \langle 1/P_0, 1/P_1, 1/P_2, 1/P_3, 1/P_4, 0.8/P_5, 0/P_6, 1/P_7, 1/P_8, 1/P_9, 1/P_{10} \rangle$$

$$\begin{aligned} \tilde{C}_2^1 &= \langle 1/P_0, 1/P_1, 0/P_2, 0.2/P_3, 0.5/P_4, 1/P_5, 0.8/P_6, 0.8/P_7, 0.8/P_8, \\ &\quad 0.2/P_9, 0/P_{10} \rangle \tilde{D}_2' \\ &= \langle 0/P_0, 0.2/P_1, 0/P_2, 0/P_3, 0/P_4, 0.8/P_5, 0/P_6, 0/P_7, 0/P_8, 0/P_9, 0/P_{10} \rangle , \end{aligned}$$

$$\tilde{D}_2'' = \langle 0.8/P_5 \rangle .$$

Step 4:

$$\tilde{G}_3 = \langle 0/P_0, 0.1/P_1, 0.2/P_2, 0/P_3, 0/P_4, 0.2/P_5, 0.4/P_6, 0.9/P_7, 0.2/P_8, 0.2/P_9, 0/P_{10}\rangle$$

$$\tilde{C}_3^0 = \langle 1/P_0, 1/P_1, 1/P_2, 1/P_3, 1/P_4, 1/P_5, 1/P_6, 1/P_7, 1/P_8, 1/P_9, 1/P_{10}\rangle ,$$

$$\tilde{C}_3^1 = \langle 0.1/P_0, 0.2/P_1, 0.2/P_2, 0.4/P_3, 0/P_4, 0/P_5, 1/P_6, 0.5/P_7, 0.5/P_8, 0.4/P_9, 0/P_{10}\rangle$$

$$\tilde{D}_3' = \langle 0.1/P_0, 0.1/P_1, 0.2/P_2, 0/P_3, 0/P_4, 0/P_5, 0/P_6, 0.5/P_7, 0.2/P_8, 0.2/P_9, 0/P_{10}\rangle \tilde{D}_3''$$
$$= \langle 0.5/P_7\rangle .$$

Thus, the decision in the system will be set:

$$\tilde{D}''' = \langle 0.8/P_1,\ \ 1/P_2,\ \ 0.8/P_3,\ \ 0.5/P_7\rangle .$$

As mentioned earlier, adaptive strategy planning system must be able to adapt (modify) the process of complex problem-solving according to changes in the system workspace. Typical examples of such systems are in [23, 24], and dedicated to other areas are in [25, 26].

In terms of fuzzy sets, the workspace state change affects the characteristics of individual purposes (sub purposes) and limitations on their achievements during the decision execution and will be displayed through changes in the coefficients of certainty.

By the way, the existence of limitations generally well characterize changes in the workspace. From this perspective, an approach based on consideration of the purposes and constraints is the best.

Thus, in moment t_0 in fuzzy system, next sets will exist:

Local purpose set $\tilde{G} = \{G_0, G_1, \ldots, G_{n-1}\}$
Limitation set $\tilde{C} = \{C_0, C_1, \ldots, C_{n-1}\}$, which will take into account that every moment exists according to the set C_i^j

System purposes (or single purpose) may be displayed by a set of facts, in particular, the robot presence (or other ODM) in a specific point of the workspace in a specific state, as seen from the example above.

From the very beginning (as for previous models) based on information about the workspace current state X^0 and the global goal Y, a purpose achievement plan must be developed. It takes into account both the discrete robot workspace (consisting of an object set) and discrete actions that can be executed by robot, discrete process of the plan execution (multi-stages).

Thus, system state will have an evolution:

$$\tilde{X}_0 \to \tilde{X}_1 \to \tilde{X}_2 \to \cdots \to \tilde{Y}. \tag{25}$$

Moreover, each state can be represented by a fuzzy set.

State fuzziness will determine action fuzziness, which has to be executed by robot strategy planning system, so:

$$\tilde{D}_0 \to \tilde{D}_1 \to \tilde{D}_2 \to \cdots \to \tilde{D}_{n-1}. \tag{26}$$

Action fuzziness means a possibility of decision selection at each step of strategy planning. So, every step will be characterized by a decision vector $\vec{\tilde{D}}_i = \left\{\tilde{D}_i^0,\ \tilde{D}_i^1,\ \tilde{D}_i^2, \ldots,\ \tilde{D}_i^{n-1}\right\}$, where $\tilde{D}_i^j$ is an element of a possible fuzzy decision set among which a variant with the highest certainty coefficient must be selected.

Thus, for concrete robotic system strategy planning, the result is a vector $\vec{D}$ that is a fuzzy set. Its particularity is a casual connection between elements. A certain variant selection is an unfuzzification of the strategy planning vector.

The selected strategy may be implemented at once or may be deferred at certain time. But deferment will affect the implementation quality negatively because workspace circumstances will change significantly.

Similar situation will appear if WS ODM is changed dynamically during decision execution. For decision vector $\vec{D}$, it will mean certainty coefficients of set elements in certain situations will change, and it's necessary to say about adaptability in strategy planning fuzzy system.

So, strategy planning fuzzy system adaptability will appear under decision-making fuzzy vector changes, which represent workspace and ODM (robot) dynamics.

Definition. Strategy planning system in which decision-making fuzzy vector changes is called adaptive.

For some cases at a certain step of task solution fuzzy by selection decision-making act $\tilde{D}_i = \mu_i / D_i$ will have a certainty coefficient $\mu_i < \varepsilon$, where $\varepsilon \geq 0$ is a selection threshold (for the simplest case) and will be considered unacceptable for the system. In this case, action generator must find a new decision for the step i, i.e., to generate $\tilde{D}_i'$ and all further decision-making acts sequence:

$$\tilde{X}_0 * \tilde{D}_0 \Rightarrow \cdots \Rightarrow \tilde{X}_i * \tilde{D}_i' \Rightarrow \tilde{X}_{i+1} * \tilde{D}_{i+1}' \Rightarrow \cdots \Rightarrow \tilde{X}_{n-1} * \tilde{D}_{n-1}' \Rightarrow \tilde{Y}_n, \tag{27}$$

In strategy planning system, description certainty coefficients for separate subgoals for such decision stages will be equal to 0, so the process can be represented in next way:

$$\vec{D} = \{\mu_0 \neq 0/D_0, \quad \mu_1 \neq 0/D_0, .., \mu_i = 0/D_i, .., \mu_{n-1} \neq 0/D_{n-1}\} \tag{28}$$

Such fact presence will mean there is a necessity of plan $\overrightarrow{D}$ reformulation beginning from D_i, so:

$$\overrightarrow{D}' = \left\{\mu_0 \neq 0/D_0, \quad \mu_1 \neq 0/D_0, \ldots, \mu_i' \neq 0/D_i', \mu_{i+1}' \neq 0/D_{i+1}', \ldots, \mu_{n-1}' \neq 0/D_{n-1}'\right\}. \tag{29}$$

Action list which will follow after μ_i'/D_i' will not necessarily change, so start plan will contain constant (c) and variable (t) parts:

$$\overrightarrow{D} = \overrightarrow{D}^c \cup \overrightarrow{D}^t. \tag{30}$$

For another part of decision acts with certainty coefficients $\mu_i \geq \varepsilon$, it's necessary to evaluate all the strategy planning sequences and to select vector $\overrightarrow{D}$, which has the highest total certainty. Calculations for such case are shown in the example for Fig. 2.

Thus, plan adaptation is a part $\overrightarrow{D}$ changing; moreover, this part analysis must be done constantly at each next step of strategy planning system work until purpose state achievement.

Fuzzy set methods using shows more flexible and accurate character of strategy planning process state description and possibility of detailed numeric evaluation of intellectual robotic system each step.

5 Fuzzy Adaptive Robot Control Modeling

For the case of a wheeled mobile robotic platform, fuzziness can be considered with the application of control system simulation methods.

With the help of MATLAB, a diagram of a system of fuzzy adaptive control of a robot was created, shown in Fig. 3.

At the input of the model – the coordinates of the endpoint of the movement (*Xgoal*, *Ygoal*), initial coordinates of the robot (*Xstart*, *Ystart*), and robot rotation angle (*Angle*), entering the system using a file from the Visual Studio or PyCharm, from the programs for identifying the robot, obtaining its coordinates (center of mass of the rectangle limiting the work), and from the program for obtaining the angle of rotation of the robot.

The route that the robot needs to pass is calculated (Fig. 4).

Further, the parameters of the path and angle of rotation of the work are transferred to the fuzzy controller block, which has the following form (Fig. 5).

The MotorVelo_2.File is loaded into this block as a base with rules.

Let's take a closer look at the creation of this file. It was created based on the Fuzzy Logic Designer package. The appearance of a fuzzy regulator, namely, a model with rules, is shown in Fig. 6.

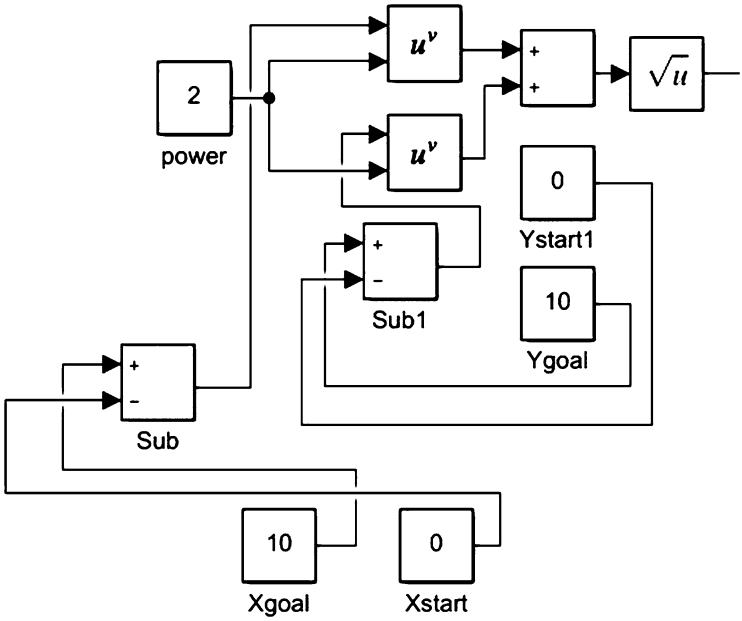


Fig. 4 Calculation of the robot route

There are two parameters at the input. The first is the path length and the second is the robot rotation angle. The outputs are the left and right speeds of the robot.

Membership functions for a parameter *Length* are presented in Fig. 7, membership functions for a parameter *Angle* in Fig. 8, and membership functions for parameters *VeloLeft* and *VeloRight* in Fig. 9. After that, the rules were entered into the fuzzy regulator block (Fig. 10).

We will carry out a revision of the model, according to the rules of the new indicators of the inputs of the regulator's errors.

The results are shown in Fig. 11 with two parameters [50; 40] and Fig. 12 with parameters [90; −120].

Dependence between path length –*Length*, the angle of rotation of the robot *Angle*, and left wheel speed *VeloLeft* are shown in Figs. 13, 14, and 15.

Consequently, the proposed fuzzy control system, the input of which is supplied with the required distance of movement and the angle of rotation of the robot, based on the input of linguistic variables and fuzzy rules, makes it possible to obtain the speed of rotation of the wheels of the robot.

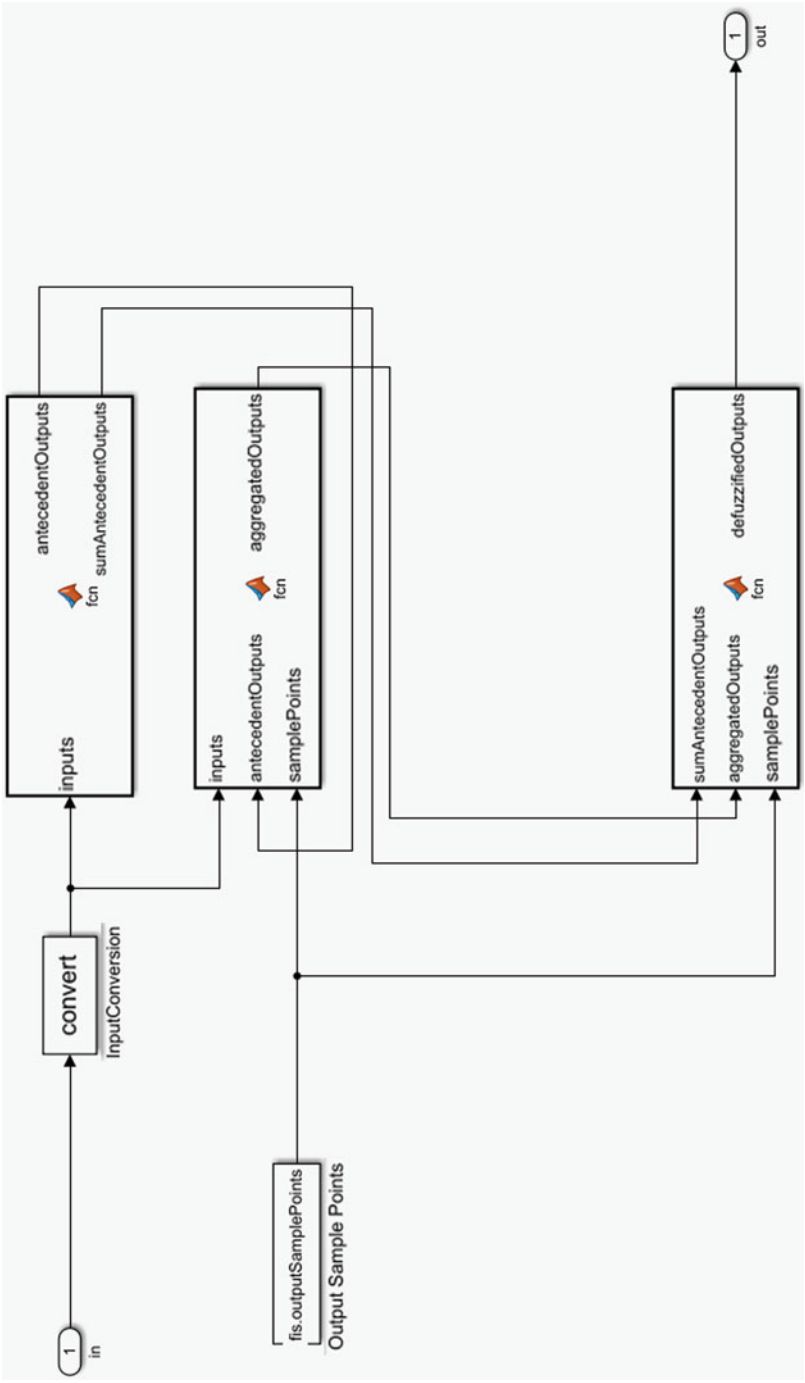


Fig. 5 Fuzzy control block

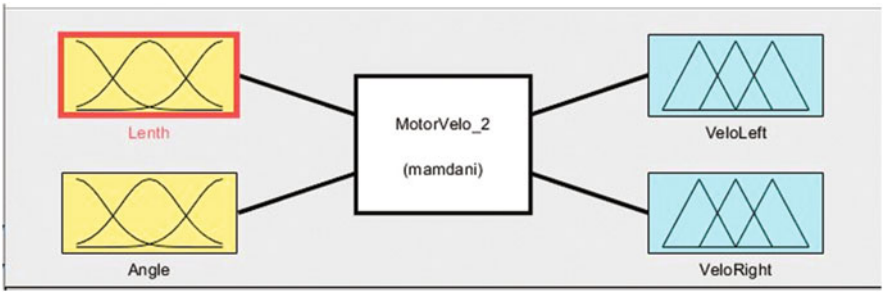


Fig. 6 Fuzzy control block model

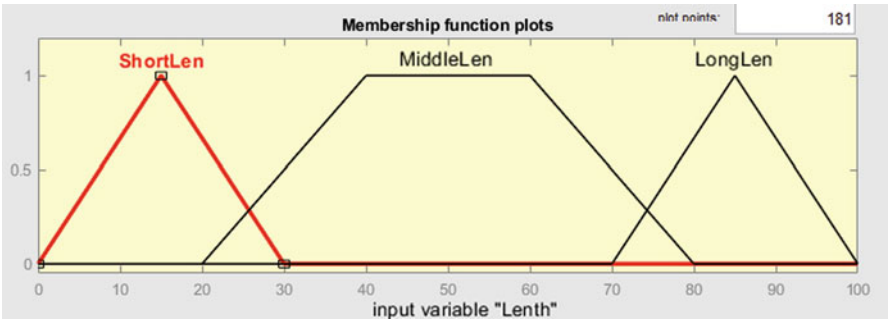


Fig. 7 Input parameter setting for *Length*

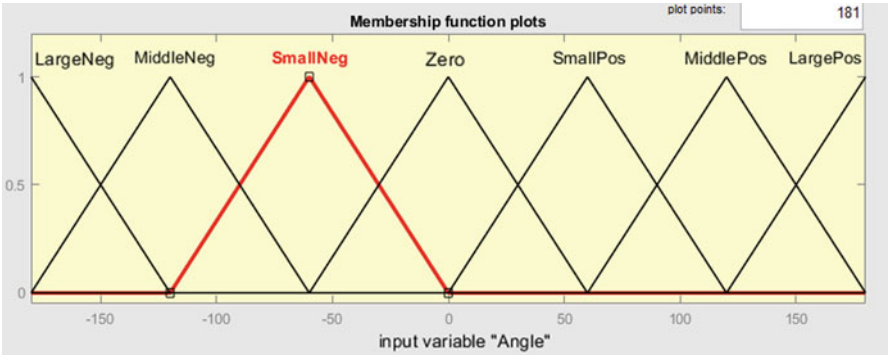


Fig. 8 Input parameter setting for *Angle*

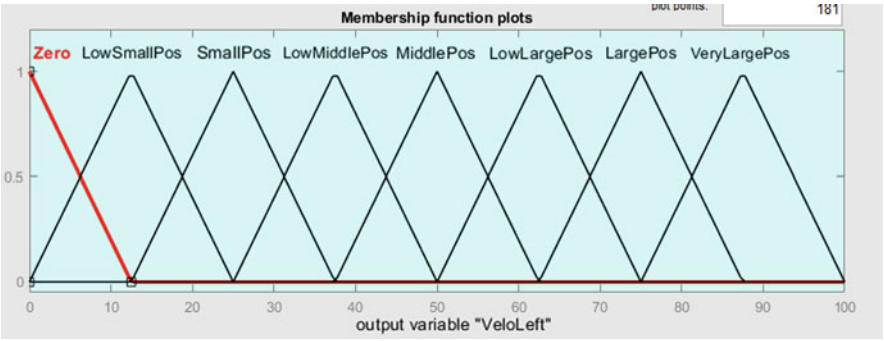


Fig. 9 Output parameter setting for *VeloLeft* and *VeloRight*

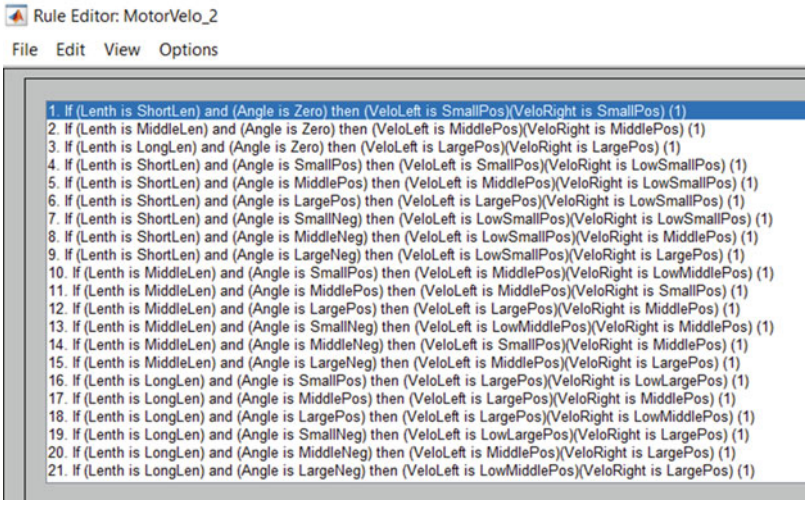


Fig. 10 Fuzzy regulator rules

6 Conclusions

The proposed chapter deals with theoretical results associated with strategy planning in conditions of uncertainty. As the basis of consideration, fuzzy sets were taken. Robot control intelligent system functioning is considered from the point of view of fuzziness of state changes both workspace and robot. It was shown from the point of view of robot control intelligent system multistage tasks that the most perspective is approach to their modeling using fuzzy sets.

Strategy planning model on the basis of fuzzy set theory provides strategy planning description in conditions of information about robotic system uncertainty action selection uncertainty at the strategy planning every step and on the basis of them provides task decision alternative way evaluation that robot must execute.

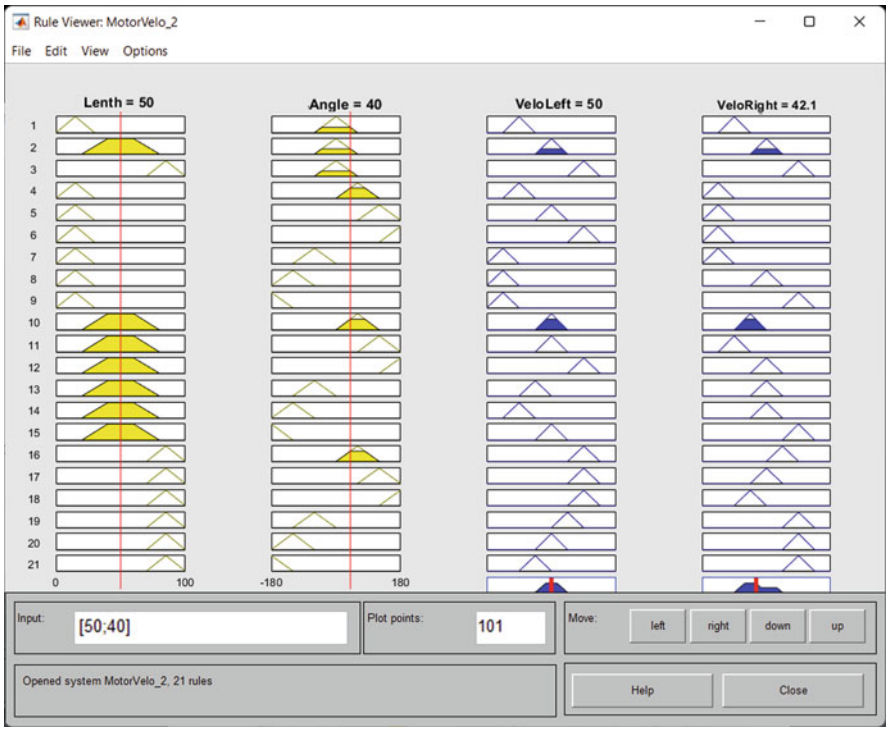


Fig. 11 Results with parameters [50; 40]

For the case of wheeled mobile robot, control fuzzy component of the model contains a detailed description of the procedures for the formation of linguistic variables associated with the dynamics of the movement of a mobile robotic platform and formalizes a set of production rules for a fuzzy regulator.

The practical significance of the obtained results is in the development of information and software for an adaptive visual control system for a mobile transport robot, which implements the proposed models and method, practically providing an analysis of the workspace using a computer/technical vision system and recognition and identification of objects in the workspace and robotic equipment, obtaining their spatial coordinates, and supporting for visual target designation of the route of the mobile robot and the functioning of an intelligent decision support system.

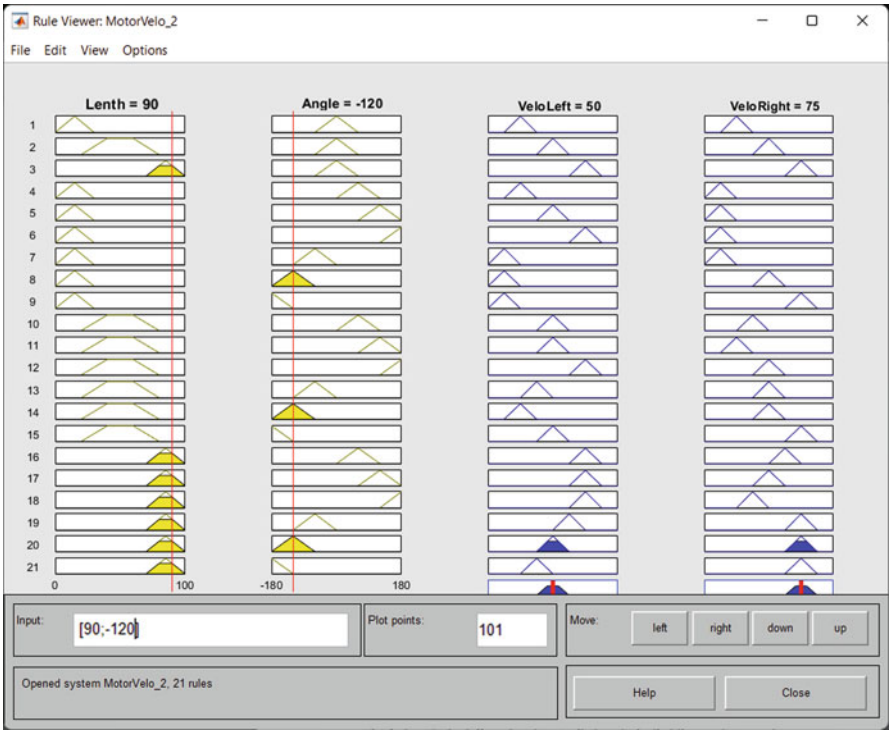


Fig. 12 Results with parameters [90; −120]

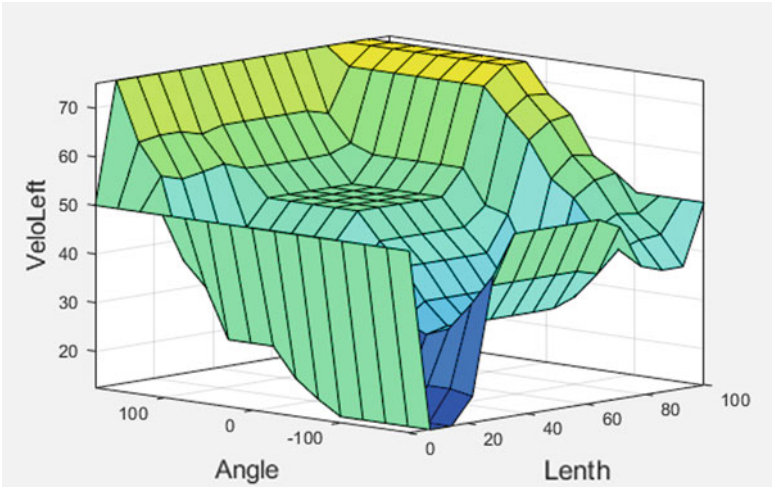


Fig. 13 Dependence between *Length*, *Angle*, and *VeloLeft*

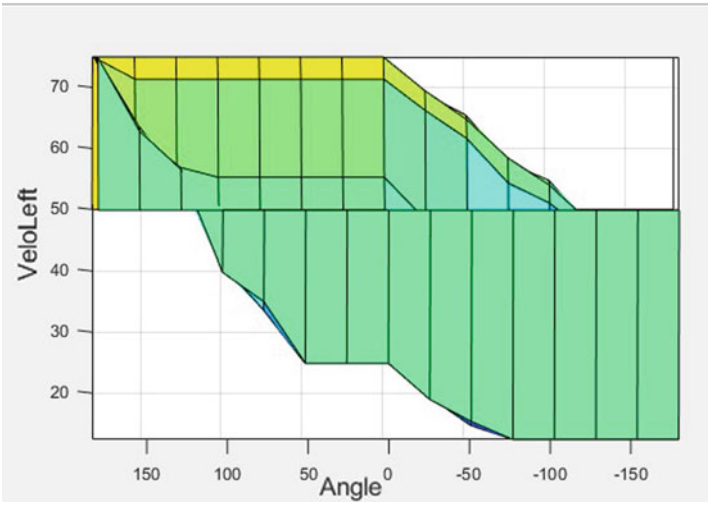


Fig. 14 Dependence between *Angle* and *VeloLeft*

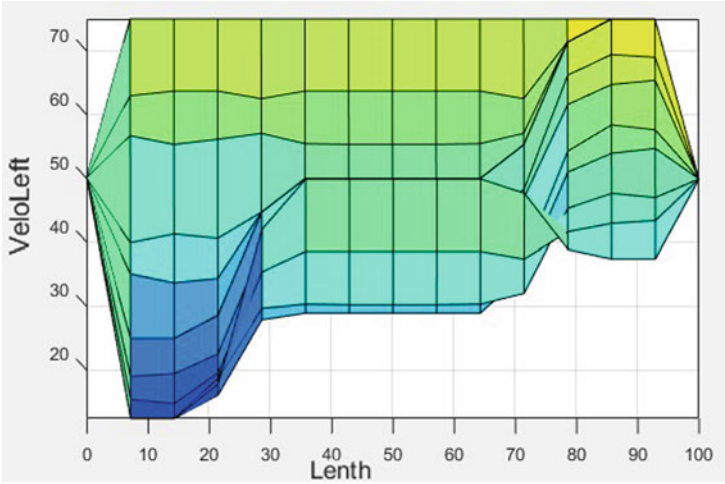


Fig. 15 Dependence between *Length* and *VeloLeft*

References

1. Kleene, Stethen. Mathematical Logic [Text] / S. Kleene. – New York: Wiley, 1967. – 398 p.
2. Shoenfield, Joseph. Mathematical Logic [Text] / J. Shoenfield. – A K Peters/CRC Press, 2001. – 356 p.
3. Stroustrup Bjarne. The C++ Programming Language [Text] / B. Stroustrup – 3d Edition, Addison Wesley Longman, Reading, MA. 1997.- 920 p.

4. Gerasimov, A.S. The development and implementation of search inference algorithm to extend unfinite-valued Lukasevitch predicate logic [Text]: Thesis for Candidate of Science (Physics and Mathematics) / A.S. Geramimov; [Sanct-Pereburg State University]. – SPb., 2007 – 194 p. (in Russian)
5. Karavayev, M.V. Fuzzy logic application for simulation of autonomous adaptive control [Text] / M.V. Karavayev // Works of Russian Academy of Science. – 2004, Vol.7. – P. 41-53. (in Russian)
6. Mironov, A.M. Fuzzy modal logics [Text] / A.M. Mironov // Fundamental and Applied Mathematics. – 2003, Vol. 9, № 1. – P. 201-230. (in Russian)
7. Jensen, R. Fuzzy-Rough Nearest Neighbour Classification and Prediction [Text] / R. Jensen, C. Cornelis // Theoretical Computer Science. – 2011, Vol. 412, No. 42. - P. 5871-5884.
8. Altrock, Constantine. Fuzzy Logic and Neuro-Fuzzy Applications in Business and Finance [Text] / Constantine von Altrock. – Prentice Hall Ptr, 1996. – 400 p.
9. Zadeh, Lotfi A. Toward extended fuzzy logic - A first step [Text] / Lotfi A. Zadeh // Fuzzy Sets and Systems. – 2001, Vol 160(21). – P. 3175-3181.
10. Russell, S. Artificial Intelligence: a modern approach [Text] / S. Russell, P. Norvig. – Prentice-Hall, 2-d Edition, 2002. – 1132 p.
11. Luger, G. Artificial Intelligence: Structures and Strategies for Complex Problem Solving [Text] / George Kuger. – Addison-Wessley, 5th Edition, 2004. – 928 p.
12. Surmann, Hartmut. Learning a fuzzy rule based knowledge representation [Text] / Hartmut Surmann // Proceedings of 2-nd. ICSC Symp. On neural computation, NC'2000, Berlin, 23-26 May, 2000. – P. 349 – 355.
13. Agarwal, Sudhir. Modeling Fuzzy Rules with Description Logics [E-Resorce] / Sudhir Agarwal, Pascal Hitzler // Proceedings of the Workshop OWL - Experiences and Directions, Galway, Ireland, November 2005. – Access: http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-188/sub5.pdf
14. Siler, William. Fuzzy expert systems and fuzzy reasoning [Text] / William Siler, James J. Buckley. – John Wiley & Sons. 2005. – 422 p.
15. Yuschenko, A.S. Reemote control of robots with application of fuzzy representations [Text] / A.S. Yuschenko // Artificial Intelligence. – 2002. – № 4. – P. 388 – 396. (in Russian)
16. Kiselyov, D.V. Learned system of mobile robot fuzzy control [Text] / D.V. Kiselyov, A.S. Yuschenko // Artificial Intelligence. – 2003. – № 3. – P. 181 – 189. (in Russian)
17. Berrabah, Sid Ahmed. Robot Navigation Based on Adaptive Fuzzy Controller [E-Resorce] / Sid Ahmed Berrabah, Cristina Gabriela Rozmarin. - Mechanical Department, Royal Military School, Belgium. 6 p. – Access: http://mecatron.rma.ac.be/pub/2009/RISE-BERRABAH-ROZMARIN.pdf
18. Cordon, O. Genetic Algorithms and Fuzzy Logic in Control Processes [E-Resorce] / O. Cordon, F. Herrera, E. Herrera-Viedma, M. Lozano. - Technical Report #DECSAI-95109, March, 1995. pp 1-28. – Access: http://sci2s.ugr.es/publications/ficheros/acs-5-135-168.pdf/
19. Liu, Zhi. A Probabilistic Fuzzy Logic System for Modeling and Control [Text] / Zhi Liu, Han-Xiong Li // IEEE Transactions on fuzzy systems. – 2005. – Vol. 13, No. 6. – P. 848-858.
20. Pugh, Alan. Robot Vision [Text] / Alan Pugh. – Springer-Verlag, 1986. – 320 p.
21. Wolfer, James. Fuzzy logic control for robot maze traversal: An undergraduate case study [Text] / James Wolfer, Chad A. George // World Congress on Computer Science, Engineering and Technology Education. – March 19 - 22, 2006, São Paulo, BRAZIL. – P. 46-50.
22. Klir, J. B. Fuzzy sets and fuzzy logics: theory and applications [Text] / J. Klir, B. Juan. – Prentice Hall PTR, 1995 – 574 p.
23. Weller, Joshua A. Neural Correlates of Adaptive Decision Making for Risky Gains and Losses [Text] / Joshua A. Weller, Irwin P. Levin, Baba Shiv, Antoine Bechara // Journal of Experimental Psychology. – 2007. – Vol. 18, Num 11. – P. 958-964.
24. Parasuraman, Raja. Adaptive Automation for Human-Robot Teaming in Future Command and Control Systems [Text] / Raja Parasuraman, Michael Barnes, Keryl Cosenzo // The International C2 Journal. – 2007. – Vol 1, No 2. – P. 43–68.

25. Thrun, S. Probabilistic Robotics [Text] / S. Thrun, W. Burgard, D. Fox. – The MIT Press, 2005. – 667 p.
26. Ross, S. Online Planning Algorithms for POMDPs [Text] / S. Ross, J. Pineau, S. Paquet, B. Chaib-draa // Journal of Artificial Intelligence Research. – 2008. – Vol. 32. – P. 663-704.

3D and 2D Visual Digital Technologies and Cultural Heritage Documentation for Conservation and Monitoring: A Critical Review and Assessment

Naif Haddad

Abbreviations

BIM	Building information modelling
CH	Cultural heritage
CS	Creative suite
GCI	Getty conservation institute
GIS	Geographical information system
H-BIM	Heritage BIM
ICOMOS	International council on monuments and sites
IT	Information technology
RG	Robotic group
TCA	Tourism and culture authority
TLS	Terrestrial laser scanner
TVS	Technical vision system
UAV	Unmanned aerial vehicle

1 Introduction and Scope

Cultural Heritage (after this referred to as "CH") "refers to monuments, groups of buildings and sites of heritage value, constituting the historic or built environment" [22]. It is appreciated through its tangible, intangible and spiritual character. Conservation is defined as the process of managing changes to the

N. Haddad (✉)
Queen Rania Faculty of Tourism and Heritage, Department of Conservation Science, Hashemite University, Zarqa, Jordan
e-mail: naifh@hu.edu.jo

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1_10

cultural significance of a historical monument and place based on internationally agreed international charters, conventions and principles. After any conservation intervention, CH documentation provides the base for monitoring, management and maintenance of a site, as well as a recording for posterity. Monitoring plays a central role in discovering and controlling when an irregular change arises and expands over time and is critical for proper maintenance, such as severe cracks, bulges, rising dampness in walls, warp roof beams and settlements in floors and foundations [35]. Before any intervention, CH documentation is an integral part of any conservation process. It is an essential prerequisite for a detailed understanding of its cultural significance and factors affecting its condition. Thus, it contributes to revealing pathological causes and providing practical guidelines on risk assessment, risk mitigation, conservation, rehabilitation, interventions, monitoring and proper maintenance.

Accordingly, assisting and identifying the causes of the risk and threats and its critical levels could be considered a preventive conservation measure if it started before deciding on a particular repair or recovery. Consequently, international charters, conventions and principles consider CH documentation before any intervention an integral part of the conservation process, even an essential prerequisite to forming an exhaustive understanding of a building's cultural significance and the factors affecting its condition [15].

According to Burra Charter [41], cultural significance means aesthetic, historical, scientific, social or spiritual value for past, present or future generations. It is embodied in the place itself, its fabric, setting, use, associations, meanings, records, related places and related objects (Article 1.2). Article 6.1 of the Burra Charter process recommends that "the cultural significance of a place and other issues affecting its future are best understood by a sequence of collecting and analysing information before making decisions". Understanding, however, cultural significance comes first, then developing policy and finally managing the place by the policy. Without being informed about what aspect of and why a CH structure or its significance for CH, the conservation work has no meaningful purpose for today or the future [23, p. 3]. International charters, conventions and principles illustrate wide-ranging guidelines, good practices and principles for the essential prerequisite of CH documentation within conservation and monitoring. They include, sometimes, technical and practical descriptions for their implementation. These charters request that CH documentation be the first and final step for any adequate conservation and intervention process. In addition to answering why, what, how, when, where and many other questions.

On the other hand, the world's CH faces many threats, vulnerabilities and risks; natural anthropogenic and technical factors cause disasters that drive its losses, even factors such as inadequate and inappropriate conservation [18]. In addition to neglect by urban planning of vast cities, climate change, structural deterioration, weathering, biodeterioration, tourism, armed conflicts worldwide, uncontrolled development and a host of other such factors are also among the reasons CH is threatened. Generally, many anthropogenic risks are also due to economic constraints and lack of funding observed in the maintenance, combined

with uncontrolled tourism growth, limited management plans and lack of compatible criteria for modern reuse [10, 12]. As a result, today, the world is losing its CH more rapidly than documented in many cases [27, p. 5]. Moreover, CH is being neglected and not sufficiently protected; even the heritage community cannot do much. These serious reasons make why its documentation is a critical matter. This is even more critical with the recently increased humanitarian crises and armed conflicts worldwide, especially in the Middle East.

The importance of CH documentation is evident in research, protection, conservation, reconstruction, stabilisation, identification, interpretation, management, awareness and serving in educating the public regarding the CH values data recovery when mitigating losses resulting from construction for future generations. Also, in retrieving of stolen cultural objects according to the convention on the international return of stolen or illegally exported cultural objects, vandalism, and destruction of CH in a crisis time. According to Resolution on Information as an Instrument for Protection against War Damages to the CH [38], "the destruction of historical records, monuments and memories serves the purpose of suppressing all that furthermore bears witness that the threatened people were ever living in the area". CH after conflict can be destroyed, ruined, damaged or faster than accessed and documented. Therefore, any feasible and conceivable way to preserve its reminiscence is crucial and must be well-thought-out.

According to ICOMOS [50], understanding, documenting and interpreting the setting is "essential to defining and appreciating the heritage significance of CH". Meanwhile, inclusively understanding the setting "requires a multi-disciplinary approach and the use of diverse information sources". As well known, documentation provides a record of the existing condition and setting of a CH at a particular point in space/time. Documentation work also enabled archaeologists to date monuments accurately. It serves as a diagnostic tool in analysing, monitoring and interpreting the conservation problem. Thorough documentation is critical for future monitoring and evaluation as a baseline record. The most helpful documentation articulates changes due to degradation, modifications and conservation intervention over time. All the above serious issues mark why CH documentation is a must, especially in critical emergencies of a humanitarian crisis and armed conflicts [19]. However, during the last few decades, CH has undergone a noticeable shift in why it is at risk and what can be prepared and done to safeguard it and mitigate risks [21].[1] Therefore, it is vital to develop an approach to monitor CH monuments and sites'

[1] According to Heritage and Resilience; Issues and Opportunities for Reducing Disaster and [27, p. 6] the Getty Conservation Institute two publications on recording: *Recording, Documentation, and Information Management for Conservation of Heritage Places*, Guiding Principles (2007) and *Recording, Documentation, and Information Management for Conservation of Heritage Places, Illustrated Examples (2007)*, CH documentation should deal with and respond to the following:

- We must assess the values and significance of the heritage in question because we care about the creative history of mankind.
- If we care about our common shared history, there is only one thing that can be done: preserve that history, by guiding the process for progress of conservation.

behaviour methodically and their interventions to mitigate any risk to their valuable and fragile material and fabric. The existing condition of any CH is a prerequisite to any conservation action or plan to provide the needed data for archaeologists, architects and conservators to support them in developing conservation interventions and site presentation schemes.

Several digital visual documentation /recording and surveying tools and techniques are available, such as laser scanning technology, photogrammetry, thermal imagery and GIS, which act as a catalyst for documentation, interpretation, promotion and CH management and monitoring. However, it is vital to detect and identify how documentation can answer questions or issues connected with the CH and update an appropriate conservation strategy, taking into consideration numerous factors such as accessibility, site particularity, typology, scale, the fragility of the existing materials and the agreed inventories of different categories of CH assets, etc. Of these main questions when dealing with CH documentation/recording or surveying is why we do it and what it should answer? Who does or which team does this documentation/recording or surveying? To summarise, according to Schmid [6], the best typical response to the matter of why CH documentation is "to create a permanent record, a sort of warranty against loss". In fact, "we cannot stop the loss of cultural heritage. However, we can better document heritage" [27, p. 9]. So, CH documentation can produce a lasting record if it is lost. Planning, however, to create a schedule and establish a strategy to collect data consistently and systematically, with no gaps, is critical in CH documentation. Planning is an essential principle to address the lack caused during an emergency—time, money and skills. Figure 1 shows generally the sequence of the recommended considerations for 3D CH documentation and economical means of obtaining the needed information; non-destructive techniques should be used whenever appropriate; care should be taken to assure that documentation efforts do not duplicate previous efforts; and report of the main results of any recording should be disseminated and published.

This chapter reviews the state-of-the-art 3D and 2D visual digital computer-based technologies for CH documentation techniques. It aims to present a critical review, assessment and investigation of these technologies within conservation and monitoring, from data collection to data sharing for conservation. It examines and discusses the following questions. Does CH documentation indeed contribute to the conservation process, and do we need to increase its use and practice? How does one define CH documentation in the context of CH conservation and monitoring? Who is the CH documentation project team? Is there a need for sharing and documentation standards or guidelines? Will 3D and 2D visual digital technology completely replace the traditional and more labour-intensive methods for CH documentation? Is there a need to integrate digital visual documentation such as photogrammetry and

- Meanwhile, this can be achieved in two means: by conserving the physical fabric and/or by creating a record to preserve the knowledge if that physical fabric cannot be conserved.
- Only documentation offers a tool for monitoring and managing CH while creating an essential record, and communicating the character and preserving the cultural significance.

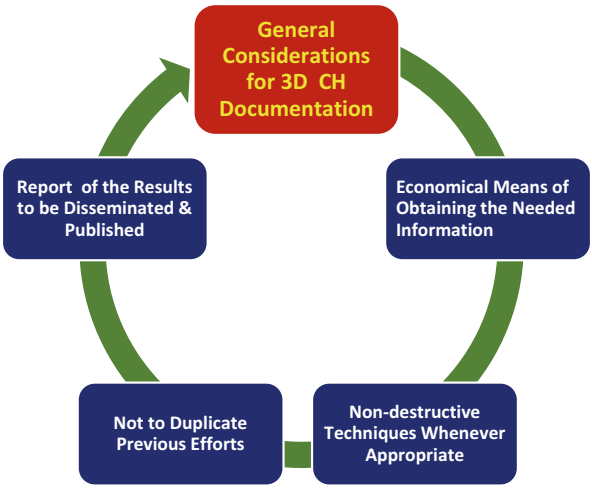


Fig. 1 General sequence of the recommended considerations for 3D CH documentation

3D scanning within the traditional hand survey and measuring techniques? How should we evaluate, deal with and choose 3D and 2D visual digital technologies for CH documentation within conservation and monitoring?

2 A Critical Review and Assessment of the Difference Between CH Terms Within Conservation and Monitoring: Surveying, Recording and Documentation

There is a different understanding and use of what specifies the main terms and definitions of CH documentation. This section presents a brief critical review of three main terms extensively used by the CH community. These terms are surveying, recording and scientific record, documentation and CH documentation methodology. Surveying is just one part of CH documentation. It is "part of cultural heritage documentation includes all methods available to record the geometry of objects and/or topography" [5, p. 3, 15]. More analytically, measured survey for CH documentation produces measured drawings on-site by hand measurement and/or using various heritage recording tools (total stations, photogrammetry and 3D laser scanners). Sometimes, it applies a kind of mathematical formalism in order to improve the resolution of 3D scans, for example, neural networks [39].

By reviewing the related international charters and principles about using the term, we can ensure that it is limited compared to the other two terms: recording and documentation [15]. For example, in the Charter for the Protection and Management of the Archaeological Heritage [23], we have a special section for the first time under the section "Survey"; in Article 4, it states that "General survey of archaeological

resources is, therefore, an essential working tool in developing strategies for the protection of the archaeological heritage". Subsequently, the archaeological survey "should be a basic obligation in protecting and managing the archaeological heritage". Another example for the term survey has only one reference in Burra Charter [41] regarding the condition of fabric; "Survey the fabric sufficiently to establish how its physical state will affect options for the treatment of the fabric".

However, since surveying has become a specialised field, this does not mean that it is sufficient to leave the final documentation decision methodology and needed conservation information to a land surveyor or a photogrammetrist. As far as the surveying part is concerned, a board of specialists should decide on the most effective method of surveying. However, in uncomplicated cases, it may be decided that simple surveying methods are sufficient and no further professional help from the surveying side is necessary. On the other hand, several methods must be applied in complex cases. Therefore, the survey provided a base to record conditions and coordination required at this level. Thus, any cooperation between surveyors and recorders is crucial for CH documentation.

On the other hand, heritage recording is "the capturing of graphic and photographic information" [31, p. 92]. Record is defined as collecting related information items (as in a database) treated as a unit. Recording means "the capture of information which describes the physical configuration, condition and use of monuments, groups of buildings and sites, at points in time, and it is an essential part of conservation process" [22, p. 49].

CH records include "tangible as well as intangible evidence and constitute a part of the documentation that can contribute to an understanding of the heritage and its related values". According to ICOMOS [24], "evidence provided by the fabric of a place should be identified and understood through systematic research, recording, and analysis. The recording is an essential part of the physical investigation of a place. It informs and guides the conservation process and its planning. Systematic recording should occur prior to, during, and following any intervention. It should include the recording of new evidence revealed, and any fabric obscured or removed. Recording of the changes to a place should continue throughout its life".

Thus, cultural significance and its value led to an understanding of the importance of recording and informing the decisions made to conserve CH. Five main principles for a recording of monuments, groups of historic buildings and sites [22], put together in Sofia, define aspects of recording as (1) reasons for a record, (2) responsibility for recording, (3) planning for recording, (4) content of records, and (5) management, dissemination and sharing of records. Meanwhile, the term "scientific record" for documentation, according to Letellier et al. [28], is the output of study, investigation and conservation activities consisting of research/investigation records provided by different conservation professionals. After going through the process of interpretation and selection, the correlation of all relevant research/investigation and sharing records, such as the collection of information prior from the archive, Internet, newspapers, secondary sources, etc., as also satellite imagery and other secondary data sources including interviews from recent first-hand accounts that could be useful, mainly social media provide a complete picture of the current sci-

entific knowledge about a cultural heritage place. "Heritage record" plus "scientific record" is equal to "complete record" [15].

Regarding the differences or similarities between CH documentation of "ancient" or the archaeological monuments and sites and the novel architecture sites or the built heritage of the modern era, we can make some general notes. Generally, CH documentation is based on the values offered by the archaeological and architectural heritage resources for diverse people, groups and societies and governments. Since there is no severe and complete research on the semantic values, the complexity of documentation of these two CH categories depends on the number of identified values typified by Tolou Del et al. [46] by 40 values. They are developed and evaluated through these values. The question is, "which values should be preferred in presenting the heritage and which values can be ignored?" [34, p. 113].

Unlike the novel architectural sites, ancient ruins' documentation presents many challenges due to intense looting that demolished the site while being unique and non-renewable places with cultural and natural components and associated values. Moreover, they often have a long history of research with a corpus of incompatible and underutilised documentation, and, as a result, their documentation is not cumulative. Today, they cannot be just assessed as scientifically researched fields due to the growing public interest and tourism that has widened the scope of archaeological sites [34, p. 174]. Therefore, their documentation for reconstructions of the building remains hypothetical, if not somewhat speculative [49]. Reconstruction means building "again as closely as possible to a documented earlier form, using new materials" [24].

Nevertheless, in the case of archaeological sites, it represents a broader scope of values. It is recognised that they consist of essential resources through scholarly studies, cultural and natural conservation and amenity factors, including tourism [34, p. 69]. Presentation and reconstruction at archaeological sites and monuments is a complex matter. It focuses on many issues that had not been previously anticipated or fully understood. This is also reflected in their documentation. In this respect, documentation of CH is not the domain of archaeological activities alone; thus, CH documentation studies cannot be seen from the sole perspective of archaeology. Overall, there is still no standard method of documentation and evaluation shared by scholars. Meanwhile, in the case of the built heritage of the modern era, because properties and sites under this category were under threat, they are increasingly subject to severe alteration or destruction, but without a proper debate and assessment of the values embedded in them. For example, the whole picture of what constitutes the South-East Asian modern architectural heritage remains mainly unclear [51, p. 118].

From the evolutionary point of view, the documentation of these two categories, we can assume that they depend on four main factors: complexity and dynamics and physical archaeological setting, values and symbols of the monument and site, particularity of integrity and authenticity and the required information of interpretation, presentation and management and site design. To conclude, the documentation of the traditional architectural heritage category includes the archaeological sites and monuments. Also, modern properties and sites must be considered worthy of

preservation and transmission to future generations for cultural identity aspects about continuity and change [51].

2.1 How Does One Define the Term CH Documentation in the Context of Conservation and Monitoring?

Generally, CH documentation includes two main actions: the gain of information regarding monuments, buildings and sites and their physical characteristics, history and problems and the process of organising, interpreting and managing that information [27, p. 5]. It means "collecting, recording, keeping, and managing information about a place and its cultural heritage value, including information about its history, fabric, and meaning; information about decisions taken; and information about physical changes and interventions made to the place" [24].

According to ICOMOS [24], Article 11, in the document and archiving section, regarding the CH value and CH significance, and all issues conservation, "should be fully documented to ensure that this information is available to present and future generations. Documentation includes information about all changes to the place and any decisions made during the conservation process. Documentation should be carried out to archival standards to maximise the record's longevity and should be placed in an appropriate archival repository. Documentation should be made available to connected people and other interested parties".

Several tangible and intangible techniques, such as oral, written, graphic and photographic, can produce CH documentation [1, pp. 11–27]. Agreeing with Addison [6], it is also the basis of the whole thing that moves into dissemination, presentations, education, television documentaries and games that help children learn. It can offer material for educational purposes and promotion. However, it is the medium through which this knowledge is recorded, collected and stored. It involves providing or formulating descriptive, technical, photographic and other relevant documents. It is a planned process that requires an assessment of existing records and a rationale to prepare needed records. It is, though, both the product and action of meeting the information needs of heritage management, as also is the compilation of records and information on CH collected to note its condition and monitor any changes.

Thus, it is related to the systematic collection and archiving of records of CH assets to preserve them for future reference. It also refers to the stock of existing information, while recording is the active process of creating and collecting new records [15, 28, p. 117]. It is a graceful balance; it is a continuous balance between essential and longer-term needs. However, it is not only describing "the context of the materials and their relationship in space and time to geological deposits, and significant architectural features, but also the answer for questions of a risk assessment and monitoring of the remains of past human activities" [17]. Generally, it refers to damage to assets, with a detailed presentation, i.e. pinpointing the

location, type and extent of the damage, identification of the assets and the materials from which they are made and their composition, features and other characteristics; measurements, soundings, tests, analyses and research work; and any research results obtained.

To conclude, CH documentation has many dimensions, such as technical, social, economic and environmental, that contribute to social and cultural well-being [15]. This has to be reflected in assessing and monitoring CH risk. It also illustrates how knowledge of the practice(s) was transmitted between generations and any vital data concerning the control or possession of that knowledge to raise public awareness about the increasing risks facing CH resources. Thus, it describes the CH significance of the heritage resource for both the resident population and broader local, national and global specialised involved and related team [17, 19].

For all the above, the term documentation is adopted in this research. CH documentation, certainty, is a highly scientific discipline that requires knowing the latest digital visual technology, guided by principles, methodology, standards, organised structure and the international charters, conventions and principles [15]. Meanwhile, CH documentation methodology offers a referential framework to guarantee that the recording process can deal with the variety of CH, managed by one entity and be replicated consistently. The methodology is dependent on various parameters, such as the objectives of recording. In addition, it is dependent on the step in the conservation process when the record is compiled; site typology and significance, which often dictates the level of precision and technique to be considered; and present conditions (environment, location, condition of the material fabric, accessibility), which can also indicate the technique to be considered [35].

3 Critical Review, Assessment, and Investigation of 3D and 2D Visual Digital Technologies for CH Documentation and Project Team(s) Within Conservation and Monitoring

While CH documentation may consist of direct examination or survey/record of the resource; published or archival material; Website information; written, graphic or photographic resources; iconographic sources; or oral sources [19], it can be categorised and classified within conservation and monitoring by three types (Fig. 2) as follows:

1. Textual and oral descriptions/written CH documentation and material testing methods.
2. Technical graphic/geometric CH documentation (non-photographic graphic) such as hand-measuring and tachometry.
3. Visual and photographic inspection techniques and tools for CH documentation contain the following: old photos and records of the current state in colour, black-white or digital technique, including satellite and aerial survey records—panoramas, perspective, silhouettes, dominants, exterior and character-

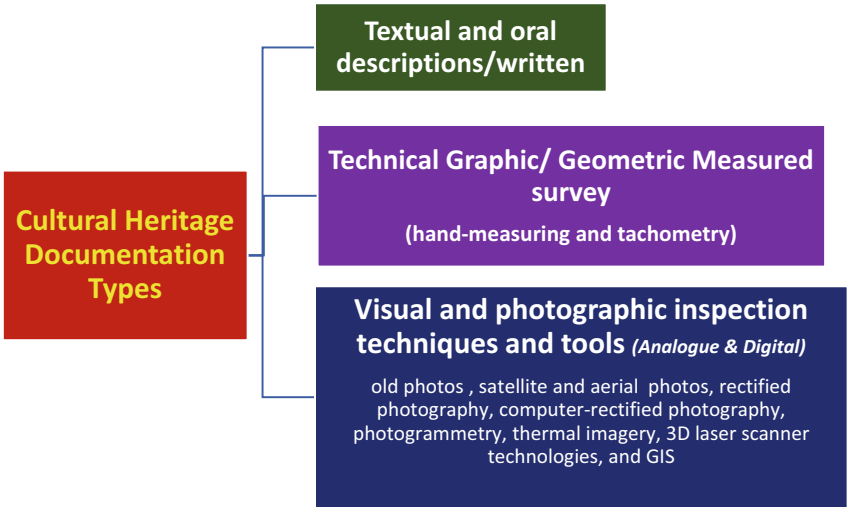


Fig. 2 The three main types/categories classified within conservation and monitoring

istic details of a structure(s)—extracted from, such as photography, rectified photography, computer-rectified photography, photogrammetry, thermal imagery, 3D laser scanner technologies and GIS, which are essential tools for urban and CH planning. These visual digital technologies' positive constructive and crucial role in documentation, conservation, identifying, interpretation and monitoring is discussed in the following section.

3.1 Visual and Photographic Inspection Techniques and Tools for CH Documentation

Visual and photographic inspection of CH documentation has gained a recognised position in CH studies, research and applications. In the last two decades, smart technological, scientific and cultural development has allowed the diffusion of different digital applications oriented towards past communication using three-dimensional content. Today, we must acknowledge that digital and virtual (computer-based visualisation) has become an established way of rebuilding and dealing with the past. It has become a specialised discipline, practised by trained professionals, requiring theoretical and technical expertise concepts and knowledge, such as 3D scanning, photogrammetry, thermal imagery and geographic information systems (GIS). GIS is well-known and controlled mainly by the geomatics people. Geomatics is a term that refers to the integration of geodesy within information technology science for surveying, analysis and management of earth-based data.

More analytically, GIS computer-based systems collect, archive, manage, retrieve, analyse and produce geographic and other related attribute data [13]. It can play a significant role in identifying risks and uncovering preparedness needs. The GIS includes four main functional subsystems for the data input, storage, retrieval, manipulation, analysis of data and output and display of data [40]. However, data integration or linking of information that is difficult to associate through any other means in different forms, for example, GIS, is well-known for combinations of mapped variables to build and analyse new variables; even it has been acknowledged as a vital component in the development of many virtual CH collections and archives.

However, building information modelling (BIM) can be applied as a complementary tool for GIS for the management of CH for providing an integrated representation of both physical and functional characteristics of CH, based on 3D documentation digital and traditional techniques (photogrammetry, TLS, classical surveying and their integration). The integration of GIS and BIM makes the spectrum of applications even broader [47, pp. 25, 30]. BIM recently is more and more effectively used as H-BIM (Heritage BIM) for automatic generation of conservatory CH documentation, even in risk assessment, monitoring the acoustic, thermal, moisture, structural analysis and material quantity and visualisations for CH preservation. However, H-BIM implementation still has many challenges as an emerging technology that requires research at visual and spatial sensing and sensor systems, computer vision and image processing technologies [7, p. 90]. When the parametric objects are modelled using CH documentation and the laser scanning data, libraries of the modelled elements are generated, thus encapsulating H-BIM. The significance of these new H-BIM libraries is permitting the diagnosis, management, rehabilitation, reconstruction and maintenance processes of CH to become more straightforward, precise and quicker during the rest of its life cycle. However, the lack of international H-BIM libraries and insufficient international collaboration has limited the potential use of H-BIM [33, p. 2]. According to López et al. [33], there is a need to create a "universal and free-access H-BIM library containing all the valuable information for architects, designers, archaeologists, historians, engineers, and conservators of architectural heritage".

These digital hi-tech tools and techniques are now available to precisely document and record the spatial configuration of CH, mainly used for producing as-built building models. Each method ranges inaccuracy, impacting the fabric, time and expertise required to acquire and process data [14, 35]. 3D models allow users to deliver comprehensive risk analysis not only in 2D but also in 3D or even in four dimensions if the acquisition is carried out at different times [10–12, 29]. The accuracy, however, of the results of the risk assessment needs to be monitored and evaluated regularly [48]. It can be characterised and considered based on factors that influence the perfect tool and instrument to be chosen accuracy needed in combination within authorisation to use the method (e.g. aerial photogrammetry or thermal imagery may not be permitted), availability of instruments and power supply, accessibility of object into the observation stations, permission to touch the object and even the type of intervention and conservation projects to be undertaken.

For example, it is not the same for projects and works for adaptation, restoration, conservation, asset structure, preventive and protective measures or reconstruction works.

The photographic documentation can quickly provide information on the critical condition, i.e. before, during and after restoration (PC = previous condition, IC = intermediate condition, FC = final condition). Of meaning here is that old photographs taken previously can be compared with new photographs taken from the same position and lighting to monitor and gauge the extent of lost CH fabric and features. Of course, observation of the light and its effect on the subject and time of day is a significant concern and challenge. However, determining the best time to capture the pictures is critical regarding solving the shade and shadow issue and getting excellent lighting. However, there are many limitations regarding the large objects or large, complicated façades. This issue can be solved via UAV cameras; however, it is, in many cases, challenging to use for security reasons [3, 52]. Nevertheless, it is advised to have at least 60% overlapping between the images to be registered.

Using a handheld digital camera and rectified photography, on the other side, has several advantages in architecture, urban planning and CH conservation. It is more than adequate for simple, quick and inexpensive [11]. This method is the most proper when the CH monument surface is geometrically flat. It also needs minimum training and demands no high-tech equipment. Image rectification can be taken out with or without measurement control points on the object but with slight variations in accuracy and reliability. CH monuments with multiple flat surfaces in different planes can be rectified; each can be separately rectified and then brought into one reference plane. Rectified photography allowed the streetscapes to be viewed in their entirety in a measurable, organised way to improve conservation planning and design [8, pp. 47, 48]. Selected images can be rectified in a repetitive, structured procedure using Adobe Photoshop CS2 (Creative Suite). Adobe Photoshop CS2 software has a specific Lens Correction filter that performs the same correction function. Panorama Tools, a free plug-in added to Adobe Photoshop CS2, can also be used to correct lens distortion.

Nowadays, photogrammetry and 3D laser scanner technologies are the two most effective techniques mainly relevant for accelerating spatial data collection from existing buildings and for rapid intervention in risky conditions [14, 30, 33, p. 2]. Photogrammetry is an affordable measurement technique that is a very accurate tool for capturing CH structures on a much bigger scale, and it is less expensive than scanning. The main idea of a photogrammetric image-based approach is to create correspondences between primitives extracted from a minimum of two but frequently more images, which are converted into a 3D point cloud using a mathematical model [47, p. 12]. The data is an excellent record that can monitor the CH recording condition over time. Meanwhile, PhotoModeler uses several photos from different angles to create 3D models and export the measured data as a dxf file. It can extract precise measurements and details even by using Google Earth Images in PhotoModeler Scanner.

Metigo software application produces rectified imagery plans; true ortho-projections on planar, cylindrical and conic surfaces; 3D photorealistic models; and measurements from photography and mappings on a photographic base. Metigo MAP 3.0 is a software for digital mapping in the rectified image/photomap developed with conservators and restorers. Agisoft Metashape is a software that generates 3D spatial data to be used in GIS applications, CH documentation and visual effects production and indirect measurements of objects of various scales, while Agisoft PhotoScan is an advanced image-based 3D modelling solution to create professional-quality 3D content from still images. Meanwhile, an object in Microsoft Photosynth needs between 20 and 300 photos for 3D reconstruction. After photo selection, no more interaction is required. The result can be uploaded automatically to the Web (http://labs.live.com/photosynth).

Finally, a combination of non-destructive recording digital or photographic imaging techniques and tools, such as integration of 3D terrestrial laser scanning point cloud, digital photogrammetry, thermal imagery, multispectral sensors and infrared reflectography for CH documentation [30], has a wide variety of applications which are recently commonly used in geodesy, architecture, construction, landscape, archaeology, history and virtual and augmented reality. It gives the ability to realise invisible and hidden threats from view. The synergy of these data permits comprehensive information on the actual state and increases the amount, quality and accuracy of the geospatial information about CH. The synergy is, sometimes, the only approach to achieve the complete documentation of CH characterised by complicated geometry or located in an area that proves difficult for traditional measurements, as it provides the opportunity to produce more accurate results of CH documentation, cartographic studies, architectural and construction documentation, 3D models and visualisations [37, p. 7].

On the other hand, laser-based technologies provide 3D support in architectural, urban and cultural landscape surveying and documentation with different levels of detail in the form of millions of points, with geometric coordinates (X, Y and Z). Recent terrestrial laser scanner (TLS) devices contain three main components: (1) a rotary mirror, which gathers data at different heights and from different directions, (2) a laser source that sends and receives the signal and (3) the data storage module [47, p. 7]. 3D scanning uses laser light to scan the object directly, without contact with it. Having evolved from developments in measurement devices for mechanical engineering and manufacturing, 3D TLS has seen growing acceptance for CH documentation in the past few years. 3D models derived from TLS point clouds are used mainly in many virtual representations of CH. The processed point clouds can be incorporated into building information modelling (BIM) platforms as a new paradigm evolved significantly in CH documentation and management. In a virtual environment, it can represent the actual conservation condition of the analysed CH. BIM platforms are based on components based on stored parametric elements in object libraries [33, pp. 2, 10].

As laser-based scanners are extensively used in the different domains to obtain physic dimensions and become practical tools used on 3D coordinates measuring systems [42, 43], it is essential to understand the technical requirements of laser

scanners and have some ways to define statistical data that supports its reliability [42]. Regarding the technical requirements with the potential of 3D scanning application in CH documentation, we should consider the following factors [2, 11]:

1. *Acquisition and post-processing data:* One of the main reasons this tool becomes broadly used is its data acquisition process. Due to its high-speed data collection and processing, high density of measured points, identifying the shape of CH objects and its anomalies as distortions, asymmetry, and the advantage of the time of analysis. Therefore, mages are mapped onto the model to get a virtual copy of the actual object. However, we face difficulties on some material surface overlap to complete all scans. High-resolution TLS with powerful point cloud processing software is an innovative tool for many applications [36, p. 81]. Concerning the number of viewpoints, recording points, scanning time needed in the field, processing time in the lab and reducing the amount of processing data, increasing computer performances offers the management of colossal point clouds and helps realise exciting new perspectives utilising 3D models. However, laser scanners and post-processing software's costs are still expensive and still limited in conservation and restoration projects, and limited expert academic staff.
2. *Kind, type, scale, material and texture of the cultural heritage object and structure:* Asymmetric objects, free-forms, irregular surfaces such as caves, large and high-detail structure and complex, the rough and dark of the object, the surrounding light intensity, the colour and surface of scanned objects, scanning surface angle, capturing and detecting sharp edges and lines of intersection and morphometric data are essential for a correct representation of CH object that should be considered. However, the more complex structure is, the more challenging it is to get adequate overlap for data registration. For example, utilising TLS in a cave requires multiple setups to cover up areas of significance, so overlapping must be organised and well planned before data acquisition [36, p. 79].
3. *Accuracy:* It affects the protection of CH and the decision-making process. Scan resolution and accuracy, coverage and range have significantly improved. It depends on the length of the scanner base and the object distance between the object and the scanner for a manufacturing characteristic of artistic works. It also depends on the suitability of 3D scanning equipment. However, sometimes, the specifications stated by the producers are not comparable. However, as done in photogrammetry, laser scanning accuracy cannot reach accuracy through larger image scales.
4. *Benefit ratio and cost:* Several modes of 3D scanners have been developed for CH assessment applications, and the use of laser sensors in the 3D part measurement process has presented a significant improvement in the data acquisition process regarding time and cost [11]. Despite the advantages of the 3D scanner, it has still not been used broadly to produce CH projects. Laser scanning has the main advantage of directly being used for 3D visualisation, for point-to-point measurements, like if the user were physically present and

stored for subsequent use. Thus, offering the acquisition of dense data sampling with high accuracy is considered one of the most current efficient instruments for 3D monitoring. Meanwhile, it is relatively expensive, and sometimes its capabilities are overrated; it has a remarkable impact on the public seeing 3D products and sharing 3D models via the Internet to provide data management and dissemination of information about CH. However, the shortage of incorporating this technology into the curricula, especially architecture schools, should be urgently re-examined and solved.

On the other hand, according to Sepulveda-Valdez et al. [43], there is a need to focus on the principal sources of uncertainties on implementing a laser scanner based on dynamic triangulation, primarily on its passive element, mainly by examining the processed signal on different environmental conditions. Of interest to mention that the optimised communication within a robotic swarm or group (RG) in a tight obstacle ambient is crucial to optimise group navigation for efficient sector trespass and monitoring. However, the optical laser scanning sensor, integrated with intelligent data management, can offer a more effective dead reckoning of the RG [44]. Sergiyenko et al. [45] described the data transfer in the group of robots while moving in an area with a high-level density of obstacles to increase their movement speed by creating and synchronising an area map made by each robot separately. Ivanov et al. [25] offer autonomous robotic group behaviour optimisation during a distributed area in a cluttered hazardous terrain. The navigation scheme can use the original real-time technical vision system (TVS) based on a dynamic triangulation principle.

Meanwhile, thermal imagery can identify problems with CH structures. It allowed increasing the knowledge of alterations in the building and exterior walls (e.g. invisible moisture shapes due to water infiltration and plaster damages as initial detachment processes). Moreover, its services identify inner structural details such as pilasters lying at winded façades, beams and lintels at the top of the windows. Infrared reflectography is a powerful tool to examine the state of CH objects; it uses a specialised digital detector or heat-sensitive film to capture absorption and emission characteristics of reflected infrared radiation between 750 and 2000 nanometres [26, p. 129]. It is simple, quick and effective in investigating surface conditions by detecting original faded or hidden drawings and penetrating the upper layers of over-painted surfaces. It has been successfully applied in CH and monitoring daily temperature variations. Near-infrared reflectography has also proved its usefulness in pigment identification and distinguishing replicas, as also infrared radiation is involved in thermographic analyses [47, p. 19].

Spectral imagery is commonly associated with 2D CH documentation, mainly relating to flat 2D objects such as historical wall paintings. However, the range of information that can be obtained increases when the spectral information is combined with the reconstruction of the 3D shape. Therefore, of importance is to make sure that there is an integration of multispectral photography with digital cameras and thermal imagery to provide relevant and accurate information to assess the state of risk and preservation of these monuments [11, 14, 29, 30].

Nevertheless, the integration of spectral and geometric data was a concern in much research; the issue remains a challenge shortly, according to Tobiasz et al. [47]. On the other hand, damage and threats detection are one of the most critical applications of spectral imaging; meanwhile, integration of spectral data with other CH documentation techniques increases the potential of such analyses in many CH fields, such as architecture and art history, conservation, restoration, interpretation and maintenance.

To conclude, there are many software tools to perform the modelling and complete 3D and 2D documentation of CH. 3D and 2D visual digital information technology is revolutionising how we are documenting and preserving CH, including the databases of all possible threats, damages, deterioration and safeguards to consider in a system. Thus, depending on one or a group of the above techniques that may influence the selected method, complexity of the documentation project's measures is needed for the nature of interventions to be undertaken. However, visual and photographic tools and non-photographic (graphics) documentation techniques are merging into one process in which the digital visual image of the 3D model technology is the main base [17, 18]. 3D digital models can illustrate the current physical condition of CH combined with the GIS database collecting information about its "historic evolution, material composition, state of conservation, etc., enhance the potential of CH modelling and consequently gains popularity in this field" [47, p. 26]. Meanwhile, assessing CH risks requires a conceptual framework to identify and quantify their various components [14, 18, p. 7].

Finally, it is crucial for the direction and control of the digital visual data collection and data interpretation through a two-stage process: the risk analysis process (scope definition, information resources/assets identification and risk prioritisation) and risk management process (decisions to control, ignore or reduce risks) [32]. It, thus, supports the risk assessment process and risk mitigation if the following key factors are taken into consideration: (1) understanding the object by identifying and reviewing the purpose of the documentation, (2) putting all the digital visual collected information together and creating a model and (3) making digital visual what an interpretation (virtual and digital modelling techniques) is.

3.2 Who Is the CH Documentation Project Team(s?)

Regarding the question, who does, or which team does, CH documentation/recording or surveying? The answer may be more difficult and complex. CH documentation consists of contributions of many scientific branches. In many cases, it cannot be accomplished without cooperation between experts of different disciplines. Therefore, it included a team of professionals working together, including historians, archaeologists, architects, engineers, geomatics professionals and surveyors, academics, computer scientists, art historians, conservators, analysts, CRM people, filmmakers, virtual and augmented reality specialists and even the public and local people and collecting documents and information from the public

and local people, such as old photos, videos or photographs and documents, as also information about how and when did the loss occur. What was the monument like before? Are all essential parts of the CH documentation projects team(s) process? Thus, all the above should be involved in the CH documentation to preserve our cultural history and diversity.

However, the combination of all the above professionals and local people can be turned out to be a severe challenge. Though defining responsibilities, each person on the project team is critical, as efforts could not be duplicated or, worse, areas or sites missed. So, outlining clear responsibilities is a key in the CH documentation project team(s) guiding principle. This can answer who was responsible for a clear reporting structure. Of meaning is to highlight who can be blamed if assessment results are incorrect.

Furthermore, how can each team member get what he wants from the assessment process? A unique research team should investigate the applied documentation methods in virtual digital images of information content [19]. The different professions involved in a project and are needed to produce the virtual digital images might have very different understandings of certain words and terms. Therefore, the precise and accurate terminology to describe the CH documentation needs must be consistent and clear for all professionals. During the design of the CH project, one or more teams should be set up for any given protected asset. The project team must formulate and organise the documentation research data and project design programme to ensure the timely initiation of the reports for professional supervision of the project and act on the notes resulting from such management.

4 Discussion of Visual Digital CH Documentation Tools and Techniques, Sharing and Standards and Design Issues: Critical Evaluation

As mentioned, there are various visual digital tools, techniques and methods for documenting and recording the CH environments, either separately or in an integrated manner. However, combining and integrating different risk identification tools and techniques often lead to the best results and less time. In addition, risk identification data can be tied to real-world coordinates, which makes it possible to be used for the technology of GIS and H-BIM or spatial statistical analysis to map and monitor signs of CH risks and deterioration. But what about the integration of digital and traditional in situ information? Does it allow for more accurate visual models for the presentation and interpretation of risk management?

However, due to rapid development in complex and software much complete 3D digital visual documentation, there is a severe discussion about replacing, or even disappearance of, some traditional CH surveying and recording methods. These scurrying tools and techniques provide ground-truth data and reference other image-based and range-based measurements. They are also applied to measure an

object's shape and determine its location on the Earth's surface, which is crucial for monitoring the CH structure's deflection over time [11, 47, p. 30]. Currently, the best practices for CH documentation are not widely exchanged inside the conservation field. There are not enough international periodicals or free Websites that permit experts and the community to share their knowledge. Also, "less than satisfactory levels of human and financial resources are dedicated to documentation activities" [27, p. 8]. Any adopted digital visual documentation method should serve various end-users with different agendas and offer access rights management and privileges for different groups according to the responsibilities and interests [48].

On the other hand, in a dialogue between three well-known specialists in CH documentation within conservation and from the Getty Conservation Institute (GCI), under the title "People and Technology: A Discussion about Heritage Documentation", Schmid [6, p. 14] believes that "multi-disciplinary cooperation in conservation projects is an old idea, but it rarely really works". For Schmid, what happens is that everybody documents their proper part of the project. However, there is "rarely an integration or correlation of data, a real interdisciplinary exchange and evaluation, which is, from a conservation point of view, an absolute requirement for a successful project". As IT-based technologies strongly influence CH documentation within conservation works, it must be fast, efficient, low cost and easy to use, train others, be easily shared and manage and importantly, have the capability to process and extract further information later. Addison [6, p. 13] raises the issue of a timeline, saying that "stone lasts thousands of years, wood and paper hundreds of years, and digital media—cd, magnetic tape—tens of years. Even worse than the medium is the format: how it is encoded on the software. This lasts in the single digits at most". That looks to Addison the "central problem". So, the concern is that there is more work to be done, as we are at the beginning to bridge the gap between the geomatics professional providers and creators of CH information (e.g. photographers, heritage recorders, photogrammetrist and surveyors), and the users for the conservation of our shared CH. Addison [6, p. 15] assumes that it would be valuable "to communicate to the makers of digital tools—be they digital camera or laser scanner manufacturers—our needs as a community". Unfortunately, they either are not aware of what is needed to make their data usage in the future, do not have time or are not easy.

The gap between them arose from the starting point that most of the geomatics professional providers are not qualified or trained to define and deal with the needed data for the specific users for the conservation process that must be collected, processed and analysed before any action can be taken. Therefore, the primary goal component of CH documentation that should be developed according to the user's specific requirements is how this process should be carried out. Addison [6, p. 12] consequently thinks that the issue is just about training, education and guidance. In this concern, it is necessary to explain the relations between the different teams/parts of the structure/site users and other decisions. In addition, it is necessary to document all their discussion findings and results during the training education and guidance progress.

Finally, how should we evaluate, deal with and choose digital visual technologies for CH documentation within conservation and monitoring? Harrison [20] examines the questions of precision and accuracy in historic surveying buildings, considering the different requirements of both scholarship and preservation; modern digital technology has changed matters significantly and promises to continue to bring change. Whereas the dilemma was once measured as precisely as possible or as precisely as a scaled drawing could display, now we should ask, how should we evaluate and deal with digital visual technologies for CH documentation? The particularity of the documented object, structure, monuments and sites comes first, and then comes the needed information, the needed accuracy and the expected results giving:

1. Cost and time: the cost can inevitably be one of the deciding factors between different recording methods. However, it should not be employed to decide the level of documentation/survey.
2. Technical considerations or logistics based on the potential of applying digital methods.
3. Whether the documentation can fulfil the technical and academic requirements and be agreed upon and applied by the end-user.
4. International indexes.
5. Risk assessment and mitigation issues.
6. Monitoring past human activities remains [11, 17].

Figure 3 shows a chart of the suggested sequence of how should we evaluate and deal with 3D CH documentation based on the potential of the application of digital visual methods for conservation and monitoring.

4.1 Is There a Need for Sharing and Documentation Standards or Guidelines?

With the rapid improvement of digital technology, specifically in the last two decades, many international heritage organisations have been working in close collaboration towards defining and standardising CH documentation and the needs of the built environment. However, few international standards exist for recording and documenting cultural heritage [27, p. 7]. Furthermore, a systematic documentation methodology and procedures framework is needed to structure the collection of these records to be standardised and utilised as references, according to Muhammad and Chabbi [35]. However, one of the vast potentials of the new information technologies and information management tools is that it offers us "a better way to communicate and share our results" [6, p. 14]. However, according to LeBlanc and Eppich [27], communication is more complicated without guidelines.

Indeed, organising and classifying data is needed to make informed decision management, dissemination and sharing of records because sharing information is

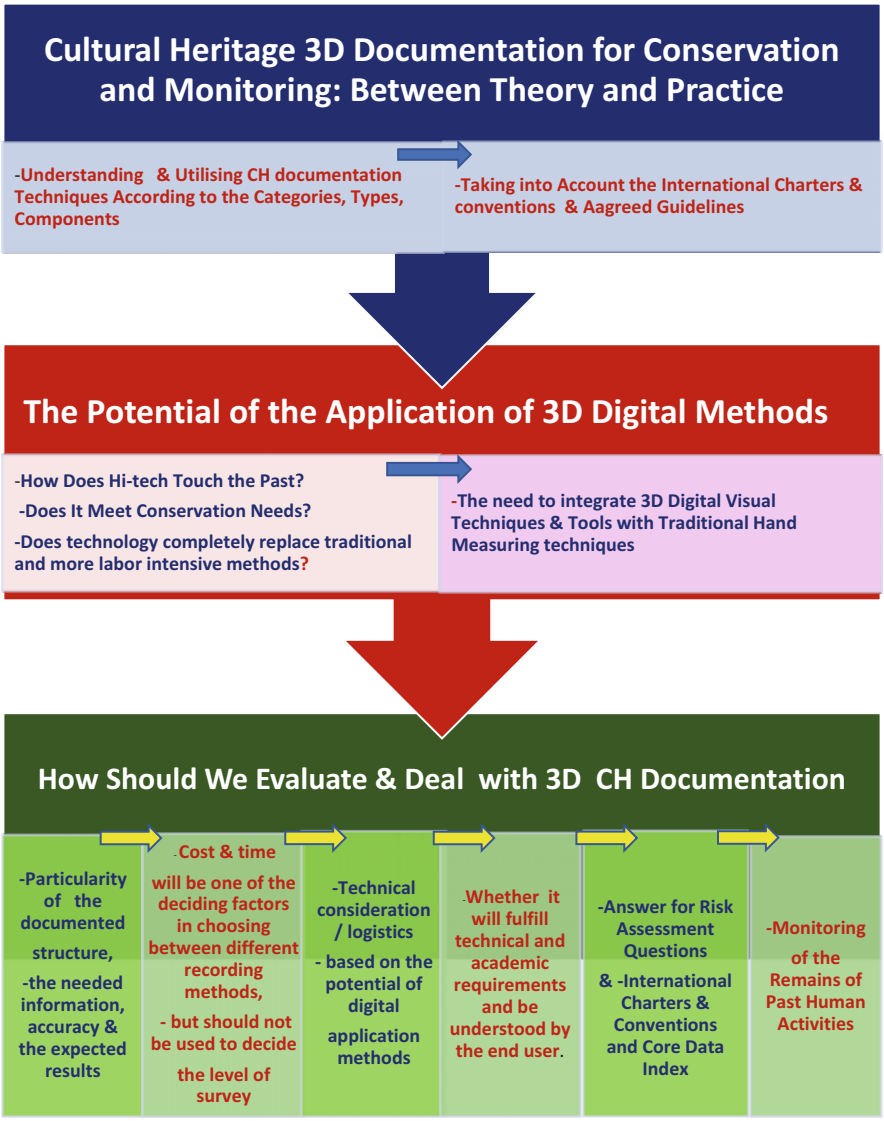


Fig. 3 The suggested sequence of how should we evaluate and deal with 3D CH documentation based on digital visual methods for conservation and monitoring

vital for an effective response. Information sharing is a prerequisite for the better understanding and effective management of CH. It is also essential for other reasons, including promoting and interpreting the heritage for vital economic purposes, such as cultural tourism and regional development. Addison looks ahead to when there is "a global database of all projects that can be cross-referenced and cross-indexed so

that multiple people can share and communicate their results together in a global archive that can have longevity beyond individual project lives" [6, p. 14]. The question, however, is if the quality of information is sufficient to provide good information for the required conservation decisions' needs standards. We cannot also ignore the strengthening of cultural and social identity at regional, national and international levels and the ability to fight the theft of and illicit traffic in cultural property on a global scale by sharing information and maintaining contact with fellow professionals in their own and associated fields. The challenge, however, for practitioners is to develop a rigorous methodology and procedures that adopt and adapt these basic principles to local requirements [35].

An excellent example in the Middle East of how international standards have been adapted to the local and practical needs in the last couple of years is the case of the "TCA Abu Dhabi's Department of Conservation". It has derived from adopting best international practices and guidelines and establishing a system for documenting and monitoring Abu Dhabi's CH. The documentation methodology developed by TCA Abu Dhabi choose the proper tools and techniques to collect, analyse and manage records as a function of the various building and site types, and building on the cumulative experience from projects. The TCA Abu Dhabi case also illustrates how this methodology is used for specific surveys, assessments, investigation, emergency conservation programs, monitoring systems and maintenance regimes [35]. In addition, the projects showed how documentation of buildings requires a well-thought-out protocol that considers several criteria. As a result, it was agreed to create a documentation laboratory to improve recording, documentation and information management methods continually. The laboratory can also be developing cutting-edge tools best suited to Abu Dhabi's heritage, which can be used as a model for the region.

On the other hand, however, recently, international standards are not accredited by much CH documentation among international conservation experts. According to Addison [6], "the problem is changing projects and changing needs". Thus, Addison prefers the idea of guidelines to the idea of standards. As an information technology user, Addison argues that documentation and conservation are "case-by-case issues" and that it is very challenging to define detailed standards for the field. He thus considers that what is missing are guidelines. He believes that the main issue is with the introduction of digital technology in CH documentation "is that we have on one side, a segment of rather computer-illiterate conservation professionals" and information technology specialists trying to sell their products on the other side.

More analytically, Addison argues that guidelines are needed to provide technology users "some information on the basic functioning of these tools and explain the pros and cons and the costs". Addison concludes that the users must have sufficient data to select the method that best matches up to their needs and to be able to communicate with information technology specialists in a more productive approach [6, p. 13]. Thus, for Addison, "standards can backfire since people are highly resistant and because they take so long to get everyone's agreement that they are obsolete by the time they are agreed upon". Bryan also agreed that this is an area where he sees the "Getty being well placed—not to create standards but to

offer guidelines on how standards ought to be developed and maintained across the heritage sector" [6, p. 13].

Finally, Levin points that "what matters is the human quality and guidelines, which are not tool-based. Communication, integration, the multi-disciplinary approach—these seem to be the themes of this conversation" [6, p. 15]. He argues that whether operating in a region with access to everything or working in one that only has the essential tools, "one can fail in both places or succeed in both places". Therefore, Levin concludes that in both instances, what is needed requires guidelines recognised and followed by outstanding communication from the beginning of a project among all team members.

4.2 Will Visual Digital Technology Completely Replace Traditional and More Labour-Intensive Methods for CH Documentation?

One serious query, though, is how to bridge the gap between the two groups: the non-technical users who are not qualified or trained to maintain the rapid changes in digital knowledge; they need, thus, to be continually updated so that when new visual digital technology comes along, it cannot be effectively integrated into the process. On the other side, there are many concerns about how hi-tech touches the past. Does it meet conservation needs? Eppich and Chabbi [9] designed some exciting questions and revealed unexpected results from a literature review of CH documentation within the conservation process to evaluate this matter.

GCI evaluated the literature review more analytically, reading over 700 articles, books, conference proceedings and white papers based on the following questions. First, is a conservation issue clearly stated? If so, is the scope of the issue addressed by documentation? Second, is there a correlation between the documentation objectives and the conservation process? Third, are the documentation tools proper to address the conservation issue regarding costs, details, precision, time and availability? Finally, are the tools effective in informing the conservation process? Eppich and Chabbi's [9] analysis revealed unexpected, exciting results. The results can be summarised in four concerns:

1. Only 1/6 of the reviewed literature is powerfully relevant to conservation, and hi-tech tools often do not deal with the current needs of conservation.
2. Meanwhile, there are increasingly articles that feature hi-tech tools and complicated procedures; unfortunately, conservators often do not use these tools.
3. These published documentation works often target the already up-to-date and expert professionals mainly published extensively in Europe. Thus, restricting the audience and decreasing the influence of essential ideas, techniques and methods.
4. Of significance for the research is that "simple and easy-to-use tools are not published or widely disseminated".

While CH documentation projects, based on the visual digital photographic current tools and techniques, can reinforce the vital goal of producing a correct diagnosis and monitoring, to extend the CH sites' and monuments' life cycle for the benefit of present and future generations, many heritage users are in a dilemma about the traditional tools' and techniques' future. More specifically, will digital technology completely replace traditional and more labour-intensive methods for documentation? Or should we fight to keep using the advanced 3D and 2D visual digital tools and techniques combined with the traditional manual hand survey?

In the author's view, however, we should realise that by the hand survey/manual CH documentation, the importance and the necessity of *close* in situ observation in the site/field support the parallel use of the two methods, the traditional one and visual digital documentation method. The manual hand survey as an in situ practice has the advantage not only of touching, smelling, feeling, discovering and living the moment with the monument, but also it is giving the ability for understanding the spiritual, mental, feeling the sense of place and even the hidden significance of humanity and humankind. The use of the five senses at the exact moment cannot be fully achieved by a 3D visual digital. They are behind the scenes, even with simulation associated with feeling the sense of place using audiovisual effects because the perception of cultural communication is no longer created based on a genuine sense.

Many can consider this is a romantic, old-fashioned approach. Yes, it is, but if we completely lose it, CH might suffer from many misunderstandings of anamneses and memories needed to keep in touch with. Yes, the author believes that visual digital has been widely used to document the process before, during and after conservation as the best tool for monitoring. However, what about if the cultural community loses such people full of love? Even their lives have changed because they were part of the daily history, talking, whispering and grasping their legacy across these monuments. Biryukova and Nikonova [4] concluded that creating virtual cultural storages based on their analysis of virtual forms of CH preservations context of interaction between contemporary society and cultural tradition support the author's vision. They conclude that virtual forms "do not allow to preserve the real reflection of memory, history and tradition the same way a real museum does and consequently, the axiological meaning of the term heritage is lost".

Retaining to the dialogue mentioned above, "People and Technology", to further clarify the author's point of view, Bryan explained that of importance is not ignoring the sequence of the cycle concept that "starts with understanding the historic environment". Bryan states that "once people understand it, they start to value it—and if they value it, they'll start to care for it". This caring makes people enjoy the heritage—and once they enjoy it, it can be a desire for more understanding. Thus, the documentation results are required to preserve, record, represent and disseminate and present [6, p. 12].

The author assumes that the logical sequence presented by Bryan, to start to care for heritage, is not the same as when using digital tools and techniques compared with traditional in situ close observation documentation. In the author's view, losing the traditional and manual hand survey in situ know-how for the future digital

generations is analogous to an actor playing without standing on an actual theatre stage. Yes, he might know the inexplicable enigmas of playing his role in the cinema with the latest virtual and augmented reality, but he might lose his talent to deal with the crowd at the end. The author cannot imagine, till now, that an actor without marching on the theatre stage is genuine. Addison [6, p. 15] clarified that it is possible to get further photorealistic reconstruction, but clarifying the essence of the CH interpretation is the most important. What seems typically missing from these interpretations is "documentation of the documentation", explaining how they reached this conclusion. So, it is not as much the technology—"it's down to us, the people, informing, marking, and explaining everything we do so that someone can understand it at a later stage".

Bryan [6, p. 13] cited an excellent example to clarify his position of using simple digital technology; he explained that "we're applying close-range laser scanning on some of our projects—but not on every single one. We have even got non-specialists who want to generate data on some projects. At this point, imaging is probably a more effective tool that could be used by most people, where all they need is a hundred-dollar digital camera". What is needed now is to recognise that this in situ archetypal documentation model invites others interested in CH documentation to question this initial position. Not all specialists can agree with the appraisal provided above, and more fundamentally, some researchers, especially the geomatics, can be opposed to the author's consideration of the in situ archetypal documentary value of CH.

Such keen considerations should be brought to bear on CH documentation considering the intention demonstrated here, but such an investigation is beyond the scope of this chapter and is better left to someone with the inclination. However, as one of the critical focuses of this study, the interplay of close in situ observation, new digital visual technology and connoisseurship exists at the intersection of old traditional and new digital methods which is within the formal study of CH documentation. However, analytical tools available in our time should be ethically utilised in the best interest of bridging the gap between past and present. This is because supplementing traditional in site documentation reports with intensive photographic evidence mapped onto the virtual environment can be practised creatively and increasingly as digital visual technology becomes more affordable for conservation edutainment.

Schmid [6, p. 13] agrees that while 3D laser scanning is perhaps the most detailed measured survey tool, there are alternative methods in most cases. He queried about the need in a particular recording project and the options to respond to that need. Schmid concludes that it may perhaps not just be a laser scanner. In territories where the workforce is less expensive than equipment, they could do it with hand measurements even with the same helpful result. Based on his extensive experience, Schmid states that getting to actual, considerable results in investigation projects requires the correlation of various parameters—partly scientific, partly technical, graphical or whatever. His vision is that a computer cannot carry out the same work "as a human mind in these cases. A computer might help", but in the end, to make a reasonable interpretation means to integrate and consider different parameters

from different areas of expertise [6, p. 15]. Now, the computer is incapable of doing that. Biryukova and Nikonova's position [4, p. 171] offered a clear answer for the creators of digital reconstructions; they refuse to accept the fact of the fundamental "disadvantage of digital technologies—the death of authenticity. They create a special computer database software in which 'collect' electronic copies of objects with various details about styles, techniques or materials to create a simulated authenticity".

A vast amount of communication for CH documentation and conservation needs to happen in situ. For Addison, communication is another crucial factor and has three phases; "First, communication before you start so that the needs are understood. Second, communication during. This is a ripe area for the technology developers—for example, finding a way for the surveyor to communicate his data to the photographer while capturing it, or for the photographer to link what he is capturing to the archaeologist so that in the field or during the recording process, there is communication among all players. Finally, there needs to create an information management system after the recording. Nascent information management systems" [6, p. 14]. Thus, Bryan [6, p. 15] rightly supposes that even in the coming 10 or 20twenty years, where the tools that the documenter uses can increase in speed use 3D more, and some may even become automated, the process will still be based on "human involvement in data gathering and input". In addition, limited access to tools is also another concern, especially in the third developing world; for Levin [6, p. 13], tools that are accessible in one section of the world may be much less obtainable in another part, "yet the need for documentation is no less". More important, however, is Addison's view [6, p. 13] that "one hard-coated sheet that states, this is the metadata that you should attach, whether you do it by hand or otherwise". This moves back to putting together some guidelines on what different technologies can do [6, p. 13].

To conclude, meanwhile documentation process can be very fast so that a more significant number of the users for conservation, such as many architects and archaeologists, especially in many countries with rich CH legacy cannot join the action on the ground, coordinated to collect documentation that can be analysed and provide additional information later. For the moment, laser scanners are among the best CH for monitoring; this issue should be more emphasised in CH monitoring studies and development research. As mentioned, the traditional CH documentation prepared and conducted by hand survey is likely in a dilemma in countries where even human work is cheaper than technological solutions, in which the field materials consisted of pencil, paper and a tape measure. There are many increased cases where the traditional one is not an option. An example is an underwater archaeology, where many monuments and sites are underwater. Of course, it is interesting to have information about such a CH, but sending people to do it is not always the best option, or it is not possible. In this sense, technical solutions are essential, and novel digital technologies are obligatory.

However, the transition from the proper circles of interaction with cultural objects to the virtual sphere "weakens interpersonal communication in the context of connection to tradition" [4, p. 172]. Properly leveraging digital multimedia

can help raise awareness and educate [16]. Their control, in fact, of the digital CH documentation tools already available and their skills that can work together is a critical point to be stressed and discussed by the different groups of the CH community to overcome this severe issue. On the other hand, time and planning between the two groups can also be solved and stressed as essential in CH documentation efforts. Bridging the gap efforts between them must be now seriously discussed in many joined meetings, not only through social media, such as crowdsourced mapping initiatives, but also by conducting unique intelligent and creative curricula, workshops and training providing beneficial product information.

Finally, while an increase in IT-based visual imaging supports CH documentation, recording, interpretation and specification within conservation in the future, "if the documentation itself is to be better, it can be because communication and guidelines are better", as Levin and Bryan conclude [6, p. 15]. However, Biryukova and Nikonova [4] believe that CH's role in the upbringing and education of modern-day man "disappears and fades into the background, under the pressure of the flow of information". The critical issue can be related to the two main groups, the geomatics technical professionals' providers and creators of CH digital visual and virtual information from one side and the non-technical users of the CH community. The needs to friendly communicate to define how and what we need from what the author calls her, the "semi-technical specialist", the architects, planners and some archaeologists in this case, after reclaiming or more rightly sharing and discussing their active previously role from and with mainly the geomatics people.

5 Summary and Concluding Remarks

This research offers a comprehensive review of relevant aspects of digital visual documentation. CH documentation has many dimensions, such as technical, social, economic and environmental, that contribute to social and cultural well-being. However, CH documentation and risk assessment is a fundamental decision-making process in developing any conservation and recovery CH plan, thus enabling the definition of the conservation recovery policy. Meanwhile, CH documentation comprises a wide-ranging field of items, matters, disciplines, skills, tools and technologies, and because of those enormous varieties, every CH documentation project has its particularity, needed precision and specific team. Therefore, the 3D and 2D non-destructive digital visual technologies for CH documentation and the 3D photo models show novel products in planning, presentation and interpretation by documentation. Therefore, it should practice the fitting techniques and range to produce precise visual information based on the existing condition, location, type, extent of damage and tests (Fig. 3). A collaboration between the different team members of scientific and professional disciplines is a must. This might make the mission challenging. However, this challenging mission makes it very inspiring and rewarding to share and work together in this field.

Therefore, special attention must be given to precise and accurate words and terms to describe the documentation for conservation and monitoring of CH. This needs to be consistent and clear for all professionals, especially the people of geomatics. Several reasons make this a vital issue, including misunderstanding or intimidation by the conservation language or technology between archaeologists, architects, conservators and geomatics. 3D virtual digital imaging in CH conservation can face and monitor risk issues and maintenance and management. However, digital data's long-term preservation and accessibility must also be considered. Digital archives and heritage agencies should be worried to be ready.

As one of this study's key focuses, the interplay of close in situ observation, new digital visual technology and connoisseurship at the intersection of old traditional and new digital methods must be considered within the formal study of CH documentation. In the best interest of bridging the gap between past and present, all tools available are to be utilised. Bridging the gap between the information user and the information provider means that institutions and universities should train experts and students to know many methods and specialists for specific digital visual and virtual branches. Unifying the fragmented efforts between the different parties and adopting a unified model from documentation to virtual presentation and interpretation from all sides to facilitate the exchange and sharing of information should be prioritised. Supplementing traditional in site documentation reports with intensive digital photographic evidence mapped onto the virtual environment can be practised increasingly as digital visual technology becomes more affordable for conservation edutainment.

Creating a national centre for digital information on CH in every country that includes all studies in this field is necessary. In addition, there is an urgent need to organise more smart workshops and conferences to bring together academics, planners and architects, CRM people, geomatics, IT, virtual and artificial intelligence people, conservators and risk analysts to discuss how digital visual and virtual technology can enhance learning CH legacy, risk mitigation and management best practices. CH documentation for conservation projects must be organised to make the relevant information easy to use and archived in hard and soft copies. In addition, there is a pressing need for guidelines to be a shared global archive to clarify if the work is carried out in teams; who does what, where and when? Ultimately, sharing knowledge, promoting the new digital visual and virtual technology, and promoting the CH documented data should respect the following aspects/ principles:

1. Meanwhile, good CH documentation for conservation projects is based on informed decisions; information to make these decisions is, in significant part, obtained through digital and traditional documentation and recording tools following general principles. CH documentation and recording tools and techniques of the traditional building techniques are needed to survive now more than ever if we accept that the context of tradition, legacy and heritage is a need in these fields of human existence. The old black and white photographic archives even have more longevity today than most digital data. According to the Burra Charter, traditional techniques and materials are preferred for

conserving significant fabric [41, p. 4.2]. However, conservation programmes, activities and management need to enhance the CH documentation of traditional construction techniques and materials. In addition, there is a parallel need to prepare intelligent, friendly edutainment and serious game materials on and from CH documentation starting from hand surveys to 3D laser scanning to cover even public information needs and awareness.

2. There is a crucial need for awareness, education and training among the related students, considering the above-mentioned challenges. To address the best practice documentation approaches and techniques of the various related kinds and types of CH risks raising awareness, especially for the students of architecture and heritage studies, based on the growing body of adequate online 3D digital documentation and virtual software to look, we should explore the wonders of their rapid development. Web-based digital visual CH documentation and virtual and augmented reality training also could be provided for professionals, students and the local community.
3. Digital visual technology for CH documentation is not the only possibility for its perception. There is also a need to organise and conduct other tools, even interviews, to document the history of use, construction and interventions of historic buildings and sites from other several sources. However, witnesses and testimonies should be analysed to identify disagreements as memory can sometimes play tricks. This collected information can be used, managed and disseminated as a tool to deal with at least first aid preservation strategies. The critical and debatable concern remains if digital virtual image technology based on digital visual CH documentation may prevail over their real/traditional images in our consciousness in the future.
4. By implementing different documentation and recording tools and techniques, we can produce condition reports, conservation management plans, monitoring reports and various reports related to the fieldwork before and during emergency conservation. As a result, different methods often lead to the best results. Combining more than one method gives accuracy and makes it easier to know the measurement and less time. Integrating photogrammetry and terrestrial laser scanning, spectral imagery and thermal imagery into GIS even with H-BIM for the documentation for conservation, monitoring and presenting CH information in user-defined periods should be now widely implemented. These can play a significant role in monitoring, considering the long-term storage of digital information. However, despite the advantages of the 3D scanner, it has still not been used broadly to produce CH projects.
5. The actual items of digital visual CH documentation for conservation works, which starts with the process of planning how to manage the project and researching the structure's history, cultural significance, the international charters and conventions and heritage principles and ethics, should be part as also for the geomatics and virtual and augmented reality people education and seminars. According to the Burra Charter, understanding cultural significance and values led to understanding the meaning and required documentation and presentation approach. Several tools and techniques can be wisely nominated

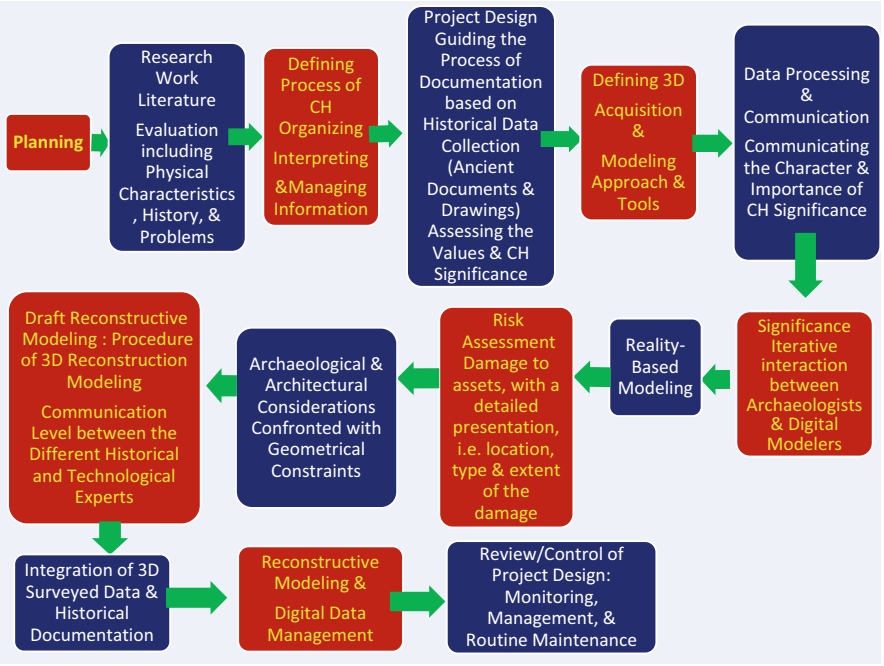


Fig. 4 The suggested sequence process of the 3D visual documentation from the planning to project design and monitoring

and combined to reach the appropriate needed information and the needed level of 3D documentation by evaluating the condition and cultural significance of any CH site and conservation and management objectives. Figure 4 illustrates the suggested sequence process of the 3D visual documentation from the planning phase, to project design to data processing and communication, to monitoring and management.

Finally, the approach process for progress, which is required to safeguard CH in an organised and structured manner, is now compulsory to be explained and shared through innovative seminars, guidelines and workshops concerning the role of each one of the team: the geomatics person, architect, archaeologist, the conservator, the virtual reconstruction people and the decision-maker. Knowledge of these principles, tools and their uses are readily available. However, many of the geomatics, virtual reconstruction people and decision-makers might be unaware, uninformed or unconvinced because they see only the result of the work.

References

1. Andrews, D. D., Jon Bedford, and Heather Papworth. 2009. Measured and drawn: techniques and practice for the metric survey of historic buildings. Swindon: English Heritage. 11-27.
2. Armağan Güleç S. Korumaz & Grazia Tucci & Mustafa Korumaz & Valentina Bonora, Opportunities of Geometric Documentation of Historic Buildings with Terrestrial Laser Scanner, Examples from Aksaray/Turkey, International Journal of Architecture & Planning, Volume 6, Issue 2, pp: 246-273, 2018 DOI: 10.15320/ICONARP.2018.54–E-ISSN: 2147-9380
3. Baik, Ahmad Alitany, A., From Architectural Photogrammetry Toward Digital Architectural Heritage Education, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-2, 2018 ISPRS TC II Mid-term Symposium "Towards Photogrammetry 2020", 4–7 June 2018, Riva del Garda, Italy.
4. Biryukova, Marina V., Nikonova, Antonina A. 2017, The Role of Digital Technologies in the Preservation of Cultural Heritage, *Muzeológia a kultúrne dedičstvo*, 2017, 5:1:169-173.
5. Böhler, W., Heinz, G., 1999: Documentation, Surveying, Photogrammetry. XVII CIPA Symposium, Recife, Brazil, Proceedings. (www.i3mainz.fh-mainz.de.)
6. Conservation, The Getty Conservation Institute Newsletter, 2005, People and Technology: A Discussion about Heritage Documentation, By Alonzo Addison, Berkeley, Paul Bryan, and Werner Schmid, talk with Rand Eppich and JeVrey Levin Volume 20, Number 3.
7. Cheng, Hung-Ming, Yang, Wun-Bin, Yen, Ya-Ning, 2015, BIM applied in historical building documentation and refurbishing, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-5/W7, 2015 25th International CIPA Symposium 2015, 31 August – 04 September 2015, Taipei, Taiwan.
8. Elwazani, Salim and Lerma, José Luis, 2007, Recording Streetscapes, In Eppich, R. and Chabbi, A, *Recording, Documentation, and Information Management for the Conservation of Heritage Places:* Illustrated Examples. Los Angeles, CA: Getty Conservation Institute.
9. Eppich, R. and Chabbi, A. 2006, How Does Hi-tech Touch the Past? Does It Meet Conservation Needs? Results from a Literature Review of Documentation for Cultural Heritage, The 7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage VAST (2006), M. Ioannides, D. Arnold, F. Niccolucci, K. Mania (Editors).
10. Haddad, Naif, Criteria for the Assessment of Modern use of Ancient Theatres and Odea, *International Journal of Heritage Studies, 2007a, Vol. 13.No. 3, pp. 265-280.*
11. Haddad N. From ground surveying to 3D laser scanner: A review of techniques used for spatial documentation of historic sites. Journal of King Saud University—Engineering Sciences. 2011;23:109-118
12. Haddad, Naif, Towards creating a dialogue between the specialised technician and non-technician users of the 3D laser scanner. In: XXI International CIPA Symposium; Athens, Greece. 2007b. pp. 350-355
13. Haddad N A, Al-Khader Sh A, Fakhoury L A. Al Mujib natural reserve in Jordan: Towards an assessment for sustainable ecotourism management plan utilising spatial documentation. Natural Resources and Conservation. 2013;1(3):65-76.
14. Haddad NA, Akasheh TS, Lerma JL, Khresat BR (2015) Towards a Risk Management and Conservation Plan for the Djin Blocks at the World Heritage Site of Petra, Jordan: The Case of Djin Block No. 9, Conservation and Management of Archaeological Sites, 17:2, 175-191.
15. Haddad NA, Fakhoury LA, Sakr YM. (2021), A Critical Anthology of International Charters, Conventions & Principles on Documentation of Cultural Heritage for Conservation, Monitoring & Management, Mediterranean Archaeology and Archaeometry, Vol. 21, No 1, (2021), pp. 291-310
16. Haddad, N.A. (2016), "Multimedia and cultural heritage: a discussion for the community involved in children's edutainment and serious games in the 21st century", Virtual Archaeology Review, Vol. 7 No. 14, pp. 61-73.
17. Haddad, Naif, (2019), "Digital heritage documentation and risks assessment: towards bridging the gap between theory and practice", Symposium, 15-16.11.2018, Responsibility for Cultural Heritage through Geomatics, Karlsruhe, Germany. 15-29.

18. Haddad, Naif Adel Leen Adeeb Fakhoury, Talal S. Akasheh, (2018) "Notes on anthropogenic risks mitigation management and recovery of ancient theatres' heritage: Qualitative assessment and recommendations", Journal of Cultural Heritage Management and Sustainable Development, Vol. 8 Issue: 3, 222-256, https://doi.org/10.1108/JCHMSD-11-2016-0062
19. Haddad, Naif A. and Fakhoury, Leen A. (2019), "The Role of Digital Heritage Documentation in Risks Assessment and Managing the Reconstruction Process in Syria", The first international scientific conference for reconstruction, Syria homeland construction: Between reality &ambition, 11-13-7-2019, Damascus University, Damascus, Syria.
20. Harrison, E. 2002. How should we measure "an ancient structure?" Nexus Network Journal 4, 18–24.
21. Heritage and Resilience; Issues and Opportunities for Reducing Disaster, 2013, UNISDR Global Platform (Geneva, 19-23 May 2013).
22. ICOMOS (International Council for Monument and Sites) (1996), Principles for the Recording of Monuments, Groups of Building and Sites. ICOMOS 11th General Assembly, Sofia, Bulgaria, 59 October 1996.
23. ICOMOS Charter for the Protection and Management of the Archaeological Heritage (1990). Prepared by the International Committee for the Management of Archaeological Heritage (ICAHM) and approved by the 9th General Assembly in Lausanne in 1990.
24. ICOMOS New Zealand Charter for the Conservation of Places of Cultural Heritage Value (Revised 2010)
25. Ivanov, M. Sergyienko, O. Tyrsa, V. Lindner, L. Flores-Fuentes, W. Cesar Rodríguez-Quiñonez, J,. Hernandez,W,. and Mercorelli,P,. Influence of Data Clouds Fusion From 3D RealTime Vision System on Robotic Group Dead Reckoning in Unknown Terrain, IEEE/CAA Journal of Automatica Sinica, Vol. 7, No. 2, PP. 368-385, March 2020, DOI: https://doi.org/10.1109/JAS.2020.1003027.
26. Kim, by Soon-Kwan, Reading Interventions, 2007, In In Eppich, R. and Chabbi, A, *Recording, Documentation, and Information Management for the Conservation of Heritage Places:* Illustrated Examples. Los Angeles, CA: Getty Conservation Institute.
27. LeBlanc, François and Eppich, Rand, 2005, Documenting Our Past for the Future, The Getty Conservation Institute Newsletter, Volume 20, Number 3, Jeffrey Levin (Editor).
28. Letellier, Robin, Werner Schmid, and François LeBlanc. 2007. *Recording, Documentation, and Information Management for the Conservation of Heritage Places: Guiding Principles*. Los Angeles, CA: Getty Conservation Institute. http://hdl.handle.net/10020/gci_pubs/recordim
29. Lerma JL, Akasheh T, Haddad N, Cabrelles M. 2011, Multispectral sensors in combination with recording tools for cultural heritage documentation. Change Over Time. 2011; 1(2):236-250
30. Lerma JL, Cabrelles M, Akasheh TS, Haddad NA 2012, Documentation of Weathered Architectural Heritage with Visible, Near Infrared, Thermal and Laser Scanning Data, International Journal of Heritage in the Digital Era, volume 1m number 2, 2012,252-275.
31. Lianos, N. Stamnas, A. (2016,) A Recording and Documentation System of Building Stock: The Case of Pentalofos Settlement in Kozani (Greece), Proceedings of the 8th International Congress on Archaeology, Computer Graphics, Cultural Heritage and Innovation *'Arqueológica* 2.0' in Valencia, Spain), Sept. 5 – 7. pp. 89-94
32. Lichtenstein S. Factors in the selection of a risk assessment method. Information Management & Computer Security. 1996;4(4):20-25
33. López, Facundo José, Lerones, Pedro M., Llamas, José, Gómez-García-Bermejo, Jaime. and Zalama, Eduardo,2018, A Review of Heritage Building Information Modeling (H-BIM), *Multimodal Technologies and Interact. 2018, 2, 21*
34. Mosler, Aslı Saruhan, 2006, Landscape Architecture on Archaeological Sites: Establishing landscape design principles for archaeological sites by means of examples from West Anatolia, Turkey, dissertation, Technical University of Munich.
35. Muhammad, S., and Chabbi. A. (2012), "Documentation and recording for the conservation of built heritage in Abu Dhabi Emirate." The 3rd International Architectural Conservation Conference & Exhibition, IACC 2012, Dubai, UAE, 17-19 December.

36. Oludare Idrees M. and Pradhan B., 2016. A decade of modern cave surveying with terrestrial laser scanning: A review of sensors, method and application development. International Journal of Speleology, 45 (1), 71-88. Tampa, FL (USA) ISSN 0392-6672 https://doi.org/10.5038/1827-806X.45.1.1923
37. Przemyslaw Klapa, Bartosz Mitka, Mariusz Zygmunt, Application of Integrated Photogrammetric and Terrestrial Laser Scanning Data to Cultural Heritage Surveying, World Multidisciplinary Earth Sciences Symposium (WMESS 2017), Earth and Environmental Science 95 (2017), 1-8.
38. Resolution on Information as an Instrument for Protection against War Damages to the Cultural Heritage, 1994, the Swedish National Commission for UNESCO and ICOMOS, Sweden on 10 June 1994
39. Rodríguez-Quiñonez, J. C., Sergiyenko, O., Hernandez-Balbuena, D., Rivas-Lopez, M., Flores-Fuentes, W., & Basaca-Preciado, L. C. (2014). Improve 3D laser scanner measurements accuracy using a FFBP neural network with Widrow-Hoff weight/bias learning function. Opto-Electronics Review, 22(4), 224-235.
40. Saleh, B., Sadoun, B. Design and implementation of a GIS system for planning. Int. J. Digit. Libr, 2006, 6, 210–218.
41. The Australia ICOMOS Charter for the Conservation of Places of Cultural Significance (The Burra Charter) 1999 (Australia 427 ICOMOS).
42. Sepulveda-Valdez, C., Sergiyenko, O., Tyrsa, V., Flores-Fuentes, W., Rodríguez-Quiñonez, J. C., Murrienta-Rico, F. N., & Kolendovska, M. (2020, June). Geometric analysis of a laser scanner functioning based on dynamic triangulation. In 2020 IEEE 29th International Symposium on Industrial Electronics (ISIE) (pp. 1398-1403). IEEE. DOI: https://doi.org/10.1109/ISIE45063.2020.9152268.
43. Sepúlveda-Valdez, C. A., Sergiyenko, O., Hernandez-Balbuena, D., Tyrsa, V., Mercorelli, P., Flores-Fuentes, W., ... & Melnik, V. (2019). Circular scanning resolution improvement by its velocity close loop control. In 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), pp. 244-249, IEEE. DOI: https://doi.org/10.1109/ISIE.2019.8781135.
44. Sergiyenko O. Y. and Tyrsa, V. V. "3D Optical Machine Vision Sensors With Intelligent Data Management for Robotic Swarm Navigation Improvement," in *IEEE Sensors Journal*, Vol. 21, no. 10, pp. 11262-11274, 15 May 15, 2021, DOI: 10.1109/JSEN.2020.3007856.
45. Sergiyenko OY, Ivanov MV, Tyrsa VV, Kartashov VM, Rivas-López M, Hernández-Balbuena D, Flores-Fuentes W, Rodríguez-Quiñonez JC, Nieto-Hipólito JI, Hernandez W, Tchernykh A, Data transferring model determination in robotic group, Robotics and Autonomous Systems, Volume 83, 2016, Pages 251-260.
46. Taher Tolou Del MS, Saleh Sedghpour B, Kamali Tabrizi S. The semantic conservation of architectural heritage: the missing values. *Heritage Science***8,** 70 (2020).
47. Tobiasz A, Markiewicz J, Łapiński S, Nikel J, Kot P, Muradov M. 2019, Review of Methods for Documentation, Management, and Sustainability of Cultural Heritage. Case Study: Museum of King Jan III's Palace at Wilanów, *Sustainability* 2019, *11*(24), 7046; https://doi.org/10.3390/su11247046
48. Vileikis O, Cesaro G, Santana Quintero M, van Balen K, Paolini A, Vafadari A. Documentation in World Heritage conservation: Towards managing and mitigating change—the case studies of Petra and the silk roads. Journal of Cultural Heritage Management and Sustainable Development. 2012;2(2):130-152.
49. Vranich, A. Reconstructing ancient architecture at Tiwanaku, Bolivia: the potential and promise of 3D printing. Heritage Science 6, 65 (2018). https://doi.org/10.1186/s40494-018-0231-0
50. Xi'an Declaration on the Conservation of the Setting of Heritage Structures, Sites and Areas, Adopted in Xi'an, China by the 15th General Assembly of ICOMOS on 21 October 2005 Final version - 22.10.2005.
51. World Heritage papers 5, Identification and Documentation of Modern Heritage, Published in 2003 by the UNESCO World Heritage Centre with financial contribution from the Netherlands Funds-in-Trust, Compiled and edited by R. van Oers and S. Haraguchi.

Optoelectronic Navigation Systems of Autonomous Mobile Ground Robots in Non-deterministic Environment

Oleg Sergiyenko

Abbreviations

ACO	Ant colony optimization
APF	Artificial potential field
BFO	Bacterial foraging optimization
CCD	Charged-coupled device
CMOS	Complementary metal-oxide-semiconductor
CS	Cuckoo search algorithm
DCD	Digital count distortion
DDJ	Data-driven jitter
EKF	Extended Kalman filter
FA	Firefly algorithm
FL	Fuzzy logic
FOV	Field of view
GA	Genetic algorithm
LMS	Least mean square
MADS	Mesh adaptive direct search
MR	Mobile robot
NN	Neural network
OMA	Other miscellaneous algorithms
PJ	Periodic jitter
PL	Positioning laser (system)
PSO	Particle swarm optimization
RJ	Random jitter

O. Sergiyenko (✉)
Department of Applied Physics, Instituto de Ingeniería, Universidad Autónoma de Baja California, Mexicali, BC, Mexico
e-mail: sergiyenko@uabc.edu.mx

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1_11

RM	Roadmap method
SA	Scanning aperture
SFLA	Shuffled frog leaping algorithm
SLAM	Simultaneous localization and mapping
TVS	Technical vision systems
UAV	unmanned aerial vehicle

1 Introduction

1.1 Machine Vision Systems in Robotics

Nowadays, different robots are widely used in almost every area of human activity. They are able to work in hard and extreme conditions; perform operations dangerous for the human person, as well as routine operations; and successfully work in areas where the price of human mistakes is high.

Optoelectronic devices in robotic systems play one of the most important roles. Such sensory devices are greatly on demand in robotics due to their naturally remote character, and their ability to detect and quantify many physical phenomena and parameters from significant distance, mostly several times bigger than own dimensions of robot. It helps to increase robot's safety in non-deterministic environment. This circumstance converts optoelectronics practically in inseparable matter of robotic technology. In this chapter, we will analyze typical cases of optoelectronic applications in robotics, mainly on the example of autonomous navigation of mobile ground robots.

For this purpose and better clarity of the further material, it is required to refresh some basic concepts of the mobile robotics. Robots can be classified according to the following indicators: by the area of its use, industrial, military (battle, supporting), exploratory, and medical; by the working environment, ground, underground, overwater, underwater, aerial, and space; and by the level of mobility, stationary, mobile, etc.

The conditions of robot functioning, which are determined by the type of area of its usage and pattern of the working process, can be classified into two categories: deterministic and non-deterministic. The deterministic category includes engineered and human-made surroundings. In the surroundings of the second category, the structural organization is almost missing; therefore, they are called completely unorganized (non-deterministic). These surroundings include particularly natural surroundings and surroundings resulting from natural emergencies, or under the destruction of the engineered and human-made surroundings, i.e., when buildings and facilities are destroyed.

In the modern world, the development of ground-based mobile robots performing actions in a natural environment is an urgent task, for example, scouting on a terrain, patrol and demining, military operations (including in cases of radioactive, chemical, and bacteriological contamination of the terrain), and performance of tasks on the surfaces of other planets. Under the destruction of human-made environments, mobile ground robots are capable of performing rescue operations

in destroyed facilities, conducting clearing of blockages, and scouting and fighting in urban conditions.

Mobile robot (MR) can be represented as a set of the following main systems: transport, special technological, navigation, and control. Data such as control of movement and operation of special technological equipment, perception of information about the environment, as well as adaptive control of the power device and chassis, taking into account the interaction of the transport system with the environment, is provided using the mobile robot control and navigation system. The target of automatic control of a ground mobile robot, which is moving on a rough surface, regardless of the presence of a pavement in non-deterministic environment, is hard enough. Because its spatial orientation most often is exposed to non-deterministic surface effect, by which it's moving, the probability of meeting with chaotically distributed obstacles also is the highest in comparison with other classes of robots. Such a class of machines includes work for movement on the surface of other planets and elimination of the consequences of technogenic (accident at enterprises and nuclear power plants) or natural cataclysms (earthquake, tsunami), so an additional complication in the designing of these devices is the high probability of presence of impacts such as radiation or other types of pollution.

That is the task of automatic control of the movement of such a robot which is a hard and important theoretical and practical task. It requires simultaneous solutions of several significant aspects: capability of viewing the surrounding area with the help of a machine vision system, necessity to determine its own spatial orientation, estimation of a frequency instability of the onboard time reference for a possibility of greater coherence of all actions of the controlled object, and formation of the trajectory of its movement in a given area.

The vision of the machine provides machine's ability (with assistance of technical measuring equipment and subsequent computer-based mathematical processing) to remotely probe and get information about surrounding area for further analytical processing. There are several significantly different approaches for constructing such systems. For example, such approaches include the use of direct contact of a sensory element with a preassigned direction of the course (discrete magnetic, permanent inductive, or simply visual), adaptation of CCD or CMOS cameras for imitation of binocular vision of a human, multidirectional vision at the base of cameras with lens of a fish-eye type, and things like that.

Technical vision systems (TVS) of mobile robots, operating in non-deterministic environments, now usually use several lasers or ultrasonic range finders. Their capabilities to search for obstacles and detailing their profile definition are insufficient to solve the tasks which robots facing.

For systems for determination of the objects' spatial orientation, the most recognized among them are gyroscopic and accelerometric platforms, but they also have disadvantages that prevented them from becoming widely used. Gyroscopes are too energy-consuming sensor systems for autonomous navigation, because for functioning they consume a significant portion of the energy of an onboard power source, and in addition, are complex and expensive systems for massive use. Accelerometer systems, by their characteristics, are much more useful for

massive use. However, in a dynamic mode (especially in the presence of mechanical vibrations), they suffer from interference in the informational channel; therefore, they also did not receive the massive implementation.

1.2 Timing of Self-Positioning in Robotics

In the methods of the robot's onboard clock validation, or even the methods of measuring stability of the frequency standard in general, there is no definite strategy at all. The use of expensive onboard time standards is not economically efficient for most categories of robots. The use of corrections by referring with expensive satellite standards of time via a special communication channel is not possible for all categories of robots; in addition, it leads to a complex of additional theoretical problems, associated with time delays in a channel, and a stochastic nature of the obstacles in it. The method recognized in a laboratory conditions for accurate measurement of the frequency of a reference relies on observations of the behavior of Allan's curve for a long time, usually 24 hours. This method is reliable enough, but due to its long-term duration, it is not useful for navigation tasks. In addition, it is not able to estimate adequately the short-term frequency instabilities that occur during the observation period. Specifically because of that, the goal of robot's autonomous navigation still is looking for a reliable and effective method for quick registration of changes in frequency, which can provide high accuracy. In the list of the most important disadvantages of mobile robotics, there are still significant discrepancies between the desired and real trajectory of MR due to inaccurate determination of the current spatial orientation because of slippages in MR movers, and significant and partially accumulated errors in determining the time it spent on the trajectory.

A significant contribution to the development of autonomous navigation of the MR, both at the system-theoretical and at the hardware-software levels, was made by researchers [1–27].

Navigation of robots is the core problem in robotic science. Systematically, the algorithms of robot navigation are global or local, with respect to environmental conditions in robot's operating zone. In global navigation, the robot operates in the known environment, and its path trajectory to avoid obstacles can be preliminarily planned. Because the robot's environment is unknown in local navigation, it needs to use various kinds of environmental sensors to detect obstacles and prevent collisions. Previously, a number of researchers for robot navigation task have designed many algorithms [1, 3, 5, 6, 8, 11, 14, 16, 17, 20, 22]. This chapter provides a detailed analysis of the problems encountered by a mobile robot avoiding obstacles in a static environment during navigation, claiming together several aspects of this complex and multi-parametric problem. Traditionally, specialists in robotic navigation are solving different parts of this problem independently. However, it is of paramount importance to solve them taking into account the cross talk effects of these parts during their simultaneous performance. The menu-driven, in which the start and end points of the mobile robot's trajectory are defined, and obstacles of

various standard sizes and shapes are modeled, controls the environment [28]. After that, the navigation of a mobile robot avoiding collisions with obstacles is carried out, followed by reaching a certain point with the specified parameters with minimal energy loss.

None of the above methods has received widespread international use due to the disadvantages and limitations that impede the automatic navigation of the MR.

1.3 Subtasks in Mobile Robotic Navigation

Thus, following the analysis, the relevant research-applied problem is a generalized and agreed-upon solution simultaneously of the following theoretical problems, inextricably related with a complex solution of automatic navigation of the MR. To be mentioned are the next problems:

- Developing a theory and methods of machine vision to be able to see the sector in front the MR.
- Development of a method for checking the frequency stability of the MR onboard clock.
- Method of determining the self-orientation of the MR for the effective functioning of the robot in non-deterministic environments in the presence of polluting factors.

As mentioned in [5], there are certain problems in the MR reactive navigation control method, such as inefficient definitions of contributing factors and a large number of rules and their causal interactions, such as a robot longitudinal and lateral wheel slippage, self-coordinate determination uncertainty, etc.

It can be concluded that a hybrid approach to autonomous navigation of robots is a modern trend, which is confirmed by the analysis of sources [8]. To solve the problem of autonomous navigation of a mobile robot performing an emergency task in a partially known environment, a specially adapted hybrid approach is used.

In the context of limited robot capabilities and known/unknown obstacles, such navigation should be fast and flexible enough to correct errors in known parts of the environment (unforeseen obstacles), requiring a task-independent autonomous preprocessing phase, to build a roadmap in known sectors, with nearly time-optimal reference trajectories [8]. Such a task must be performed in an online process that provides smooth transitions between these navigation modes, combining reactive navigation with trajectory fixation. Completing navigational tasks in relatively challenging environments and being able to handle unexpected obstacles is the goal of this. In the case of emergency tasks, the lead time can be reduced compared to other waypoint navigation methods aimed at optimizing the path.

In recent decades, in the developed methodologies for controlling mobile robots and navigation, as shown in [17], MR have a significant positive impact in various aspects of the life of modern society, for example, in medicine, in industry, and in the service sector and socialization. A review of various works on the topic of

autonomous mobile robots, including books, covering a wide range of decision-making areas for navigation and control in [17], helps researchers in this area to better understand and structure the whole complex of problems related with autonomous navigation. The following aspects of wheeled MR are strongly related with successful dead-reckoning: path and motion planning, conventional control, affine model-based control, invariant manifold-based control, vision-based control, model reference adaptive control, sliding-mode control, fuzzy and neural control, dynamic modeling, kinematic modeling, localization and mapping, and control and software architectures.

An important area of research has long been the navigation of mobile robots [12], in which there are still several unsolved problems. Various attempts have been made to solve problems related to the navigation of mobile robots in both mapping and non-mapping environments. Simultaneous localization and mapping (SLAM) methods have also been extensively researched, and several possible solutions have been proposed to address the problem [12]. In [18], the operation of the algorithm, which is an evolutionary technique with the ability to self-adaptation, self-planning, and self-organization, is shown. To adapt to the undesirable and changing conditions of a complex system, characteristics such as high speed of convergence and reliability are used. Minimizing the number of iterations and making it easier to get the optimal path in the shortest time is due to the possibility of obtaining a better position compared to its initial position. It is widely used nowadays due to its effectiveness in solving many problems, including MR fault-finding problems, reliability-redundancy optimization problems, combinatorial optimization and learning problems, dynamic environment perception problems, cooperative network problems, mixed variables structural optimization problems, problems of dispatching economic emissions, and training on a demonstration problem [18].

2 Mobile Robot Navigation Approaches/Techniques

In the literature, we can find various methodologies on navigational methods. Different methods applied for MR navigation are broadly classified [4] as two big groups: classical and reactive methods. Most of them use optoelectronic devices as TVS in robotic systems.

MRs is widely used in all areas, including industry, education, agriculture, mining, rescue, military and space industries, and so on. For navigation, the robot uses a variety of intelligent devices to localize the position and to simulate the environment.

It is necessary to control MR motion, to detect and recognize the obstacles, and to escape clash with hindrances by navigational techniques application. For any navigational technique, planning a safe path from the start (initial) to the final (target) position is one of the most important functions. A decisive step in planning the route of a mobile robot performing its tasks in a simple or complex environment is the optimized choice of navigation method. At the moment, the area of navigation

of mobile robots is one of the most studied topics. It can be defined that mobile robot navigation consists of three categories: global navigation, personal navigation, and local navigation.

Global navigation consists in the ability to determine the position of an object in the environment relative to the original coordinate system and move it to a predetermined goal [4].

Local navigation is related to the recognition of the environmental dynamical conditions and the definition of positional relationships between various objects. The personal navigation is the control of the various elements within the sector of interest, relative to each other, by considering their position. It is important to mention that these three kinds of navigations are not separated or independent; each MR during its operation must perform all these three kinds of navigations, mostly simultaneously. The only difference is the percentage of these three parallel tasks, which the MR is performing in the present instance. This is the most high-end topic of the modern robotic research. To find the optimal algorithm, which in the most proper and natural way will combine all kinds of information, obtained by TVS, and most efficiently use it for MR collaboration with environment, this is the goal of most researches in area of mobile robotics. The main concepts engaged in the functioning of the robot are systematized in Fig. 1. Let's try to analyze the elements of Fig. 1 with the aim to generate the example of the efficient approach to MR navigation further in the text, taking into account the careful congruent combination of the multiple subtasks related to the global task of safe path planning of MR in non-deterministic environment.

In [4] the navigation planning has been classified based on the preliminary data about the surrounding environment required for path devising. In global navigation, the MR strongly needs the preceding information about the surrounding, i.e., obstacle and goal positions within the considered sector of surroundings, while in the local navigation, MR does not require such a prior information of the environment. Herein below, we will try to propose the original approach able to combine carefully the properties of both variegations.

2.1 Popular Methods

The popular methods, according to their name, were the earliest solutions to the navigation problem. These are quite different approaches, but as a common point, we can mention that in popular methods the main idea is to detect the presence of obstacle (result) or its absence (no result), and further localization of the detected obstacle within the structured environment. The main disadvantage of such approach is significant computational cost and low prediction of the uncertainty in the changing environment. It is broadly considered that nowadays classic methods are not preferable in real-time tasks, with comparison to modern methods of artificial intelligence. Meanwhile, for me it is still a big question if computational cost of classic methods cannot be less than computational cost of some advanced

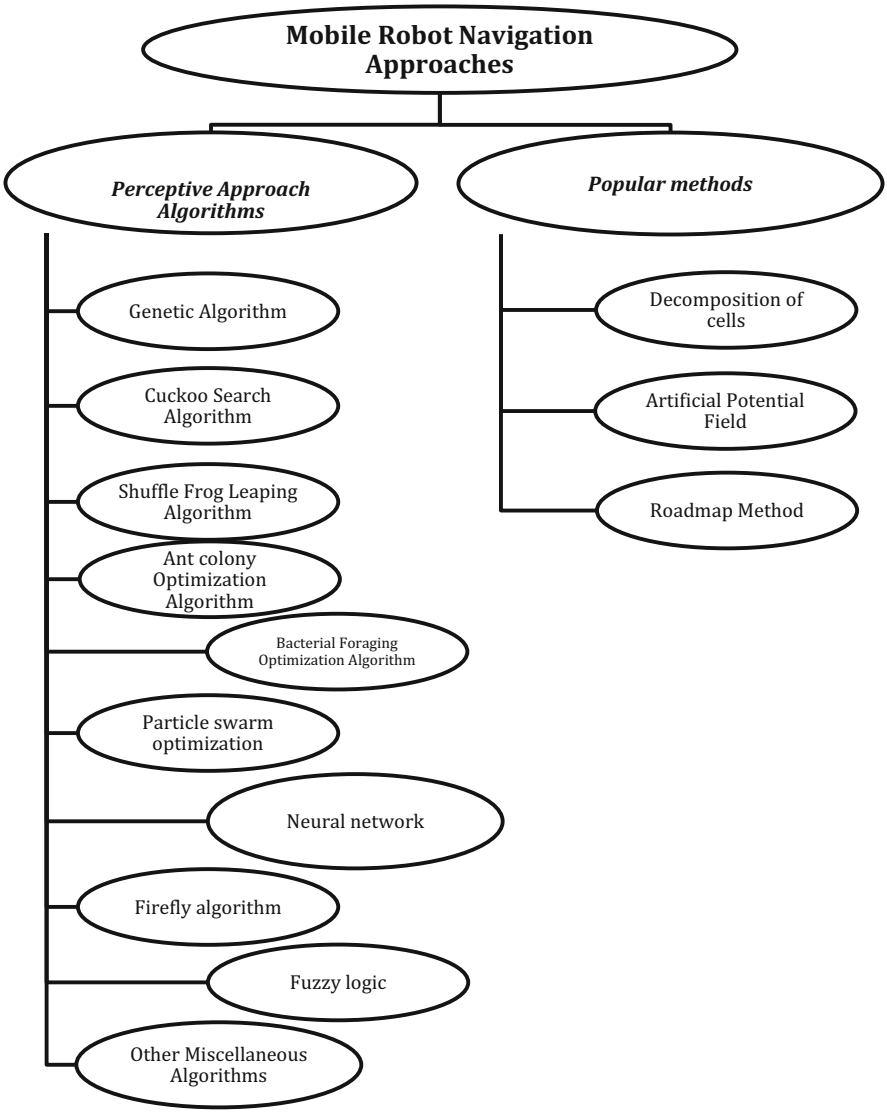


Fig. 1 Classification of mobile robot navigation

artificial intelligence algorithms. The most professional way is to calculate carefully in each particular application the time of computation for each proposed solution. Because of this, further in the text will be described some (not all) of popular methods, such as CD, RM, and APF.

2.1.1 Method of Decomposition of Cells

In the method of decomposition of cells, the region of interest is preliminarily divided into nonoverlapping grids, called cells. Connectivity graphs are used to track the transition of MR through cells in order to achieve the final goal. Throughout the path planning, the cells without obstacles are considered as save from the path from the start coordinates to the destination (target position). Cells containing obstacles are sequentially separated into two new cells to form a new empty cell. An empty cell is automatically added to the sequence of available cells to form a safe optimal path from the start coordinates to the target position. In the decomposition cell approach, the start position is represented by the start cell, and the target position is represented by the target (end) cell. The sequence of validated empty cells, joining its positions within the sector structure, represents the required path [4]. The decomposition cell approach can be classified as exact and self-adaptive. Usually, during cell decomposition, cells do not have a certain size and shape.

2.1.2 Artificial Potential Field (APF) Method

In [4] is described the APF method for MR navigation. The target point and obstacles are considered as positive/negative charged surfaces. On the MR also is created the virtual total potential with corresponding sign and its proportional imaginary force in such way the robot is attracted to the target and repelled from the obstacle with the help of an imaginary force. In order to avoid obstacles and to achieve the final goal of the trajectory, the mobile robot must follow a negative gradient. In this classical approach, a harmonic function can be applied to avoid the problem and conditions of a local minimum, as well as the dynamic properties of robot navigation and versatile variations of the laws of electrostatics.

2.1.3 Roadmap Method (RM)

This method is known in literature as roadmap method or the highway approach. It supposes the method to move from one point to another one passing through the free spaces. The trajectory of movement in this case is represented by a set of one-dimensional waves [4]. Using a set of such curves, the roadmap is built. The computer will then search for the ideal location from a set of homogeneous paths to obtain an optimized path for the MR using various game nodes: usually the RM tries to form the shortest path from its initial position to its final position. The basis of the roadmap is based on Voronoi diagrams and visibility graphs: to form an extended trajectory, map nodes are connected with the initial and final (target) positions.

2.2 Perceptive Approach Algorithms

The perceptive approach algorithms, as it clear from their generalized name, are rising from the researcher's desire to invent a tool which has a high ability to solve the MR navigation problem with a possibility to react to dynamically changing environment and compensate the negative effects with adequate and proportional reaction in control system, obtained directly from input data processing.

In this group can be mentioned (but in no way limited only with these algorithms) such approaches as cuckoo search algorithm, shuffled frog leaping algorithm, ant colony optimization, bacterial foraging optimization, particle swarm optimization, neural network, firefly algorithm, fuzzy logic, and other miscellaneous algorithms. They possess a great property to handle the environmental uncertainty presented in the TVS. That is the reason why various reactive approaches are compared and explained below.

2.2.1 Genetic Algorithm (GA)

GA is a known search-based development tool rising from the principle of natural selection in genetics. The parts of genome in real biological systems are crossed quasi-arbitrary from two independent sets of parent genomes; the same idea is practically used in engineering tasks: the long sets of parameters influent on the technical process are divided also quasi-arbitrary, and then are simulated new sets of system states searching for the combination when the objective function becomes globally better. The application of GA to the field of computer science and automatics is fruitfully explored for the past several decades. The use of GA is specific in that meaning that the promising results happen in each practical task almost arbitrary; it is not easy to predict if the results will be excellent for a whole class of applications. Nowadays, it has wide range of applications in different fields of science and technology. One of them is robot navigation.

2.2.2 Cuckoo Search (CS) Algorithm

Same as the PSO, the CS algorithm is also a metaheuristic algorithm. It is a well-known fact that cuckoos lay their own eggs in the nests of other birds. Analysis of this behavior lead to creation of the CS algorithm based on the next three rules for the optimization problem:

1. The host nest is chosen arbitrarily and one cuckoo means only one laid egg (this corresponds to the implementation of any of the heuristic algorithms, and the decision-maker randomly selects one of the elements from the package and places it in one of the given blocks),
2. The eggs of the host may be present in the nests. The cuckoo can discard any previously placed egg of the previous owner and place its own (this is consistent

with the analysis of the preliminary packaging result and leaving the best part of the alternative solution).
3. The owner of the egg can find someone else's cuckoo egg and discard it or build a new nest. This corresponds to iterative procedures in heuristic algorithms.

The advantage of the CS algorithm is an increased convergence rate and efficiency, for example, in application of mobile robot navigation to be optimized. Also, there exists the improved CS algorithm (ICS), when the iterations are not simple, but defined by predetermined complex function.

The CS algorithm usance for navigation of a wheeled robot in a static environment is described in [4]. The environment is partially unknown, and CS-based algorithm performs for receiving better results of navigation in erratic surroundings by combining the different heuristics.

2.2.3 Shuffled Frog Leaping Algorithm (SFLA)

This is a memetic metaheuristic path to optimization developed for solving combinatorial optimization problems. It is inspired by the behavior of frogs during the food search; it is a population-based cooperative search metaphor. This algorithm combines consecutively the local search elements with further global information exchange. This algorithm researches the virtual population of interacting frogs. This population is divided into several separated groups called in SFLA as memeplexes. Frogs in virtual populations also are virtual units. The virtual frogs supposed as carriers/bearers of informational units are called in SFLA as memes. In every memeplex, the SFLA simultaneously executes the independent local search, performed using a method appropriated for discrete problems. The virtual frogs are periodically replaced and shuffled into new groups to guarantee the global character of exploration. Finally, for ameliorated information random generation, random virtual frogs are simulated and replaced within the population.

Nowadays, the SFLA is widely applied in different branches of engineering optimization, thanks to such advantages as improved solution speed, low parameter quantity, easy implementation, enhanced success rate, and higher search ability under increased uncertainty. Among such engineering optimization applications, SFLA is commonly used for navigation of a mobile robot, for a direction planning strategy avoiding/skipping a narrowly most favorable solution problem using a adjusted fitness function for optimal direction generation. SFLA is frequently cooperating with particle swarm optimization-like techniques. The search of the globally best frog, and its position, permits to control the moving of the MR between the static and dynamic handicaps. SFLA is also useful in a multi-objective approach [4]. This guarantees optimum track length, track safety, and a smooth ride during navigation. In comparison with the GA in a static environment, the SFLA shows better path smoothness. It is probed as a navigation strategy for 3D underwater as well as for solving UAV 3D trajectory planning for unmanned aerial vehicles.

2.2.4 Ant Colony Optimization (ACO)

ACO is a swarm intelligence algorithm using a population-based approach applied to solve a combinatorial optimization problem. This algorithm comes from the analysis of the ants' behavior. Ants possess the capability to detect the right way from their home to a fount of food. The ACO algorithm is also used to solve the mobile robot navigation productively both avoiding obstacles and path planning. In order to improve speed, computational efficiency, solution variability, and dynamic convergence behavior compared to other algorithms such as GA, ACO can be used when planning the direction of a mobile robot. Especially, ACO has a great potential for swarm navigation, applying collision avoidance strategy for several MR within a static environment using a specialized function for better selectiveness. If the ant reaches a dead corner, after this a penalty function is applied to prevent the path deadlock of the MR by the trail intensity. Modifications of ACO-fuzzy-based hybrid approach are more efficient in static cluttered environment, both for multiple humanoid and wheeled robots in real time. Frequently, ACO can be combined with Petri net to obtain a better commitment.

2.2.5 Bacterial Foraging Optimization (BFO) Algorithm

BFO was created based on an algorithm that simulates the behavior in nature of *Escherichia coli* and *Myxococcus xanthus*. "Social" bacteria cooperate to hunt for food and survive bacteria by creating the optimization of energy received per unit time. Bacteria communicate to each other by specific chemical signals, and the BFO algorithm represents a chemotaxis processing the chemical gradients of this collective communication, by means of chemotaxis, swarming, reproduction, elimination, and dispersal. Bacteria always move searching the more nutrient regions within considered sector. Bacteria with enough food live a longer life divided into two equivalent parts. Bacteria with lack of food in its location are going to disperse and die. Both these bacteria reproduce the chemical signals to other members of the group: (1) in the more nutrient region, bacteria by these signals are attracted to others and (2) bacteria in the lesser food area send a warning signal to other bacteria advising to don't lose the energy to travel to this region. Bacteria always try to move with the goal to reach a higher nutrient region within considered sector. Bacteria are motivated to explore within considered sector for a new area of food.

The BFO algorithm for mobile robot navigation has a preferred application in a static environment. At variable velocities, it is characterized by uniform Gauss and Cauchy distributions [4]. To improve the performance of wheeled MR, using the technique of attraction force for the target and the repulsion force of obstacles, and when there are many obstacles for real-time navigation, this algorithm is most effective. To select the optimal direction vectors that direct the search to a promising area with the best local search, the negative feedback of the algorithm is checked. For many robots and drones using BFO, the same approach is available. In this case,

BFO tries to avoid complex factors and tries to get optimal search parameters in 3D space.

2.2.6 Particle Swarm Optimization (PSO)

PSO represents a metaheuristic nature-based algorithm modeling the social behavior of biologic beings' big groups, to be mentioned fish schooling, ant colonies, bird crowds, locusts, etc. [29]. Currently, in many areas of research, it finds its universal application as an optimization tool for solving various problems of science and technology. The PSO simulates the social animals'/insects' behavior. In a difference to real animals'/insects' groups, PSO does not require the presence of the leader within the group to reach the target. For example, if the group of animals/insects is moving with the aim to find food, the group does not require any leaders. Instead of this, members of group are following one of the members, which has most close position to the food. It is possible to say that the group reaches the required solution only by right communication within the group between its n members. The main idea behind the PSO algorithm is that it requires a group of particles to be used, each of which in this case will be a potential solution to the existing problem. One of the typical examples of PSO practical application is a complex of problems appearing in the field of MR navigation, such as the mapping and localization issues in the unknown environment applying a multi-agent particle filter in order to minimize the computation costs and get global uncertainty decrease with characteristics of more stable convergence and obtain smoother and precise trajectory avoiding trapping in local optima [4]. Frequently, PSO algorithm is combined with MADS algorithm (mesh adaptive direct search) which gives a better result compared to the GA and EKF (extended Kalman filter).

2.2.7 Neural Network (NN)

The neural network also called the artificial neural network represents itself the smart system. It is inspired by structure of nervous system of any higher biological organisms, and consists of many densely interconnected simple processing elements designed to make decisions based on information about the environment and its changes. These elements are interconnected according to certain predefined rules. They can transfer the information, thanks to its capability to respond to external inputs in dynamic state. The NN is represented by structured and organized layers of interconnected nodes as shown in Fig. 2 of [30]. These nodes are called neurons, and they are forming layers. Layers can be open and hidden. The function of neurons in a hidden layer is mostly defined by type of NN. The numbers of hidden neurons usually are found empirically (from practical experiments), under criterion of error convergence to a threshold error minimum, and the numbers of hidden layers are also found empirically. Usually, they are multiple, because just one hidden layer can't give the possibility to reach the necessary network learning rate under a

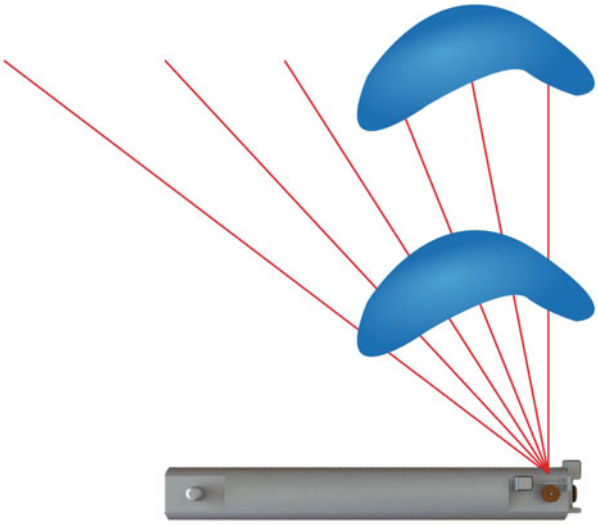


Fig. 2 Visualization of 3D laser scanning for mobile robot navigation

specified error limit. The main function of nodes is that of activation. The reduction of uncertainty in environment recognition realized by TVS means is achieved through the recognition of patterns by the NN mechanism, specifically its input layer [31, 32]. Then these patterns are transferred forward to hidden layers. In hidden layers, patterns are processed by a weighted connections system, and then connect to output layer in order to generate the required decision. Many characteristics of the NN, such as the ability to learn and generalize, distributed structure, fault tolerance [33], and high parallelism, make it useful for navigating mobile robots [4, 26]. The employment of an NN for a wheeled mobile robot navigation in a totally or partially unknown environment nowadays is a popular research trend; it is used to reduce the probability of collision in the path by finding a trajectory which avoids the nearest obstacle at the moment. In the most recent research, this leads to the creation of automated learning strategies, in which the NN adds or excludes new hidden layers during its training, accomplishing the navigation task.

2.2.8 Firefly Algorithm (FA)

The FA gets its name from the fireflies flashing way of acting; it is also known in scientific literature as the metaheuristic algorithm. It is auspicious to processes naturally stochastic, as random states and general identification of fireflies' movements. The firefly (winged beetle of the Lampyridae family) produces light very quickly. It is known as bioluminescence. Fireflies use this property for various purposes, for example, to send messages, to scare away animals that try to eat them, or to choose

a partner. FA has been applied as an optimization tool, and one of its applications could be mobile robot navigation.

The generalized idea of FA is the next. Each firefly is simply substituted in algorithm with MR or other objects in the sector. Properties of these participants of the process are adequate to firefly brightness. For detailed understanding, it is necessary to add some concerns about fireflies' behavior. The rhythm of flashlight totally defines the attraction between fireflies of different genders. It also depends on the time window when the flash of light occurred and on the speed of the light flashing. In this case, the fireflies' flashing light is the main function to be optimized. It helps to create the algorithm of optimization. One firefly is attracted to another at a rate that is directly proportional to the higher light intensity. This is precisely the basic principle of the functioning of the firefly. It is also based on three additional FA rules: the degree of attractiveness is proportional to the brightness of the firefly, so for any two flashing fireflies, the less bright one tends to the brighter one; regardless of gender, all fireflies are attracted to each other equally; the greater the brightness, the smaller the distance between the two fireflies. In the case of equal brightness of the flashing fireflies, the random move was observed.

In mathematical aspect, the objective function is stated as evaluation of the brightness of fireflies. The equation, as the rule, combines the next parameters: attractiveness of firefly on the different distances, light absorption coefficient, individual position of first and second firefly (considered in pairs), and current values of mutual positioning ij, at kth dimension, obviously considering the randomization parameter.

Movement of one firefly toward the brighter one in this case is an equivalent to the inverted behavior in pair robot – obstacle for the case of MR navigation.

2.2.9 Fuzzy Logic (FL)

The FL approach was proposed more than five decades ago; now it is used in majority of research and development fields. The main idea of fuzzy control is supposing three consecutive steps: the fuzzification, inference, and defuzzification. At the initial stage of fuzzification, the input and output data of each real value will be converted into membership classes in accordance with the established fuzzy conditions using the real environment. The second step is when the inference engine combines the facts obtained from the first step, structures them according to rules, and further starts the process of fuzzy reasoning. The fuzzy inference stage uses several forms of the membership function. The fuzzy rule for analysis is applied after the correct selection of the membership function and the definition of input and output variables. In the third step, a defuzzification block is provided to convert subsets of the output computed by the inference engine into a control function. In the navigation tasks, fuzzy logic controller mostly operates as inputs of the angle and the distance of MR body orientation; the speeds of the left and right wheels act as the output of the controller. In MR navigation, FL is used when there is an enhanced degree of uncertainty, complexity, and nonlinearity in environment to analyze, for

better data classification, pattern recognition, and autonomous MR control, and to make an optimal decision. It is possible to say that the FL concept is inspired by the notable human ability to manage the perception-based information using the logical consequence "If-Then," also called the human-supplied rules, formalizing this logic construction into mathematical formalism. It helps the system designer and executive device (computer) to obtain more complete and rigorous information about the systems' performance in the real world, and it is naturally appropriated for MR path planning.

2.3 Other Miscellaneous Algorithms [OMA]

In no way the abovementioned algorithms are not the complete list of the computational theorized approaches to the task of MR navigation improvement. The variety of mathematical tools for similar task solutions practically limits the possibility to provide such a complete systematization of all possible and substantially differed methods. The abovementioned algorithms are listed as the most applied in engineering practice by the moment, and which are mostly recognized by researchers due to their helpful results. However, every month several new mathematical tools are newly invented, and potentially show a promising perspective. The only point of delay is a quality demonstration of their advantages by many-years efficient application in MR navigation. To accomplish the task of navigating mobile robots in various environmental conditions, the following intelligent methods can be mentioned: harmony search algorithm, differential evolution algorithm, invasive weed optimization, bat algorithm, reactive obstacle avoidance based on following the boundary using sliding mode, collision avoidance based on geometric approach, control, and so on [4].

2.4 Methods to Solve Subtasks in Mobile Robotic Navigation

As shown in Fig. 1 and mentioned above, the navigational strategies cited in the literature can be classified into classical and reactive approaches. Classical methods are attributed with several lacks such as enhanced computational cost, trapping in local minima, maximum uncertainty nonuniformity and independence on controllable factors, narrowed information about the surrounding environment, the constraints by the sensing mechanism precision for real-time navigation, and so on. Consequently, for classical approach performance, there always exists the ambiguity if a solution can be obtained, or for the applied mathematical formalism, such a solution cannot be reachable. It is difficult to use popular methods in real time due to their low prediction ability and dependence on many physical factors. However, many advantages of the classical approaches cause the recent endeavors to develop several upgraded strategies, which can be classified as hybrid

algorithms. Such strategies are not performing better in real time in comparison to reactive approaches, but are preferably winning for navigation in known scenes, mostly due to their deterministic formalism in involvement of the detailed and certain initial information about the working environment. Due to the ability of reactive approaches to cope with a high level of environmental uncertainty during operation, they are used to navigate mobile robots in an unknown environment [4]. Reactive approaches are easy to realize in software/hardware, and they are self-adjustable in the meaning of the intelligence and efficiency; therefore, for real-time navigation reactive approaches, application is more frequent in comparison to classical approaches. At the same time, reactive approaches are not always better and show several disadvantages with respect to classical approaches. It is expedient to note such unpleasant factors as longer time of required calculations; sophisticated topology and design; indispensable existence of the learning phase, which practically means post-processing; and critical requirements to memory capacity. All this complicates the reactive approaches' application for low-cost robots, which are the most significant part of the potential marketplace.

All mentioned above lead us to the indispensable conclusion: at the moment, there is no evidently winning, unequivocally uncontested strategy to realize the automatic and autonomous navigation of the mobile robot.

That is why in the present work is presented the own original and progressively reasoned approach to this global problem solution.

The approach to MR navigation must be focused on its feasibility to low-cost mobile robot platforms with low-cost and reliable performance processors. The abovementioned problems must be solved simultaneously in the same processor, with the possibility of easy access to all these three data sets in order to complement one channel with data obtained in another one with the aim to improve MR behavior at whole.

The objective and tasks of the present study, as well as its main purpose, are to develop the combined theory and improve autonomous navigation systems of mobile ground robots in non-deterministic environments.

To achieve the objective, it is necessary to solve the following theoretical problems:

- Develop a theory of autonomous navigation of mobile ground robots [22, 26] in non-deterministic environments, based on the integrated use of information about the robot's own spatial orientation and time scale and about the surrounding terrain in a single matrix of the robot's state.
- Develop a method of laser sounding of the space surrounding the robot, based on the dynamic triangulation method [22–28, 31, 32], which provides ability of scanning in an extended spatial sector with the ability to mapping surface of obstacle discretized by a coordinate grid in real time and in a unified coordinate system.
- Develop a measuring method frequency and time intervals of signals using the rational approximation by the mediant, which ensures the coincidence of some integer number of measured intervals with an integer amount of reference intervals without a residue [34–36].

- Propose a new criterion of grading coincidences without a residue of a certain integer number of reference intervals with an integer number of measured time intervals, which will contribute to its simple and reliable hardware implementation, based on usage of mediant [35] with numerator in the format "an integer with a certain number of zeros" [37].
- Investigate in practice the dependence of the proposed method of measuring time intervals and the frequency of signals on the natural factors that disturb the measurement, primarily jitter [36] and the instability of the pulse width (duration) [38] of pulse sequences.
- In the robots TVS to investigate the possibility of improving the estimate of the temporal position of a pulse with a complex shape of the envelope, in order to determine the spatial position of the center of the illuminated laser spot on the obstacle surface [39, 40] within the limited time, will provide significant advantages in the distinctive ability of the three-dimensional coordinates on the edge of the sector considering by TVS.
- Explore the possibility of functioning of the TVS laser scanner with a variable scanning step [41], which will optimize the search for an obstacle in the field of view of robot and speed up the scanning process without significant loss of information content.
- Identify opportunities and propose a method to form a reliable area of the robot's field of view by filtering large errors in the scan results based on the use of neural network training algorithms [31, 32] and their mutual comparison, which will increase the distinctive ability of the TVS.
- Propose a method for determining the optimal MR trajectory of movement [23, 26, 28] in a given sector of space by a criterion of the maximum smoothness and minimum path of a trajectory (minimizing the influence of the steering mechanism), based on the use of data obtained from the proposed TVS, as well as data from the accelerometric definition system of the MR spatial orientation and onboard clock of the robot.
- Develop for non-deterministic environments a combined theory of autonomous navigation of mobile ground robots. It is based on the complex use of the robot in a single matrix state, its own spatial orientation and time scale, as well as information about the surrounding space of the robot.
- Propose a method of laser sensing of the space surrounding a robot based on a method of dynamic triangulation [22, 42–44], which differs from those known for the possibility of mapping the surface of an obstacle and the possibility of scanning in a volumetric sector, facilitates the adoption of control decisions in one coordinate system and in real time.
- A new method for measuring time intervals and frequency of signals [34, 37], the method of rational approximation by the mediants' approximation, which unlike the known ones does not describe the irrational remainder from dividing a measured integer number of indivisible standard intervals, and ensures coincidence of some integer number of measurement intervals with an integer number of reference intervals without residue.

- It has been proved that the proposed new method for measuring time intervals and frequency of signals is invariant to jitter and instability of the pulse width (duration) of pulse sequences [36, 38].
- For the first time, a method of saturation and differentiation [40, 45] was proposed for positioning increase, which differs from the known ones in that it allows on the surface of the obstacle, near the illuminated laser spot, determines the spatial position of the energy center under conditions in limited time with sufficient accuracy, and provides significant advantages at the edge of the TVS review sector.
- Improved method of functioning of the TVS laser scanner with a variable scanning step, which differs from the known ones in that it allows in the field of view of the robot, improving the search for obstacles according to the criteria of maximum information content and minimum time spent on scanning.
- Improved method of forming a durable zone of visual fields of a robot by filtering inaccurate scan results with significant errors, based on the use of the Levenberg-Marquardt algorithm, which allows for a real-time increase in the distinctive ability of the TVS.

The developed scientific provisions and methodological principles for the implementation of autonomous navigation of a mobile ground robot in a non-deterministic environment provide significant advantages. The results obtained are of great practical importance.

The use of the method of dynamic triangulation in a laser TVS for the first time provides an opportunity to increase the area of scanning in front of a robot to an acceptable level, in which the determination and localization of objects become deterministic [22, 45]. The scan zone allows to create a digital map of obstacles, which is not only simplifying the plot of trajectory of the robot's movement, but also can be beneficial for navigation of a group of robots, and for describing and mapping unknown spaces, first passed by the robot.

Dynamic triangulation in a non-deterministic environment allows in comfortable for fast transformations of the Cartesian coordinate system define dangerous boundaries of obstacles, allowing them simultaneously to rank their level of hazard based on their knowledge of distances to obstacles. Its practical advantage is also the fact that without additional transformations, the control system places all objects related to navigation in a single global coordinate system, which significantly reduces the processing time and eliminates the possibility of conversion errors during the transformation.

The possibility of using a variable scanning step [41] makes the proposed TVS an adaptive information-gathering tool, which allows to optimize the scanner's work by simply changing the command in the guide program, adapting it to the more detailed inspection, or to determine the presence of obstacles as quickly as possible.

The proposed method of measuring the frequency [34–38, 46] allows to record quickly all changes in frequency that have occurred in the onboard standard of time. This can be useful for identifying external causes that, by their appearance, caused the corresponding instability. In practical application, it provides the following

advantages: the results of the scanner measurements are refined (improves the accuracy of calculating the angles of dynamic triangulation), and the accumulation of errors in the determination of all MR parameters decreases, which are summarized during the MR stay on the motion path, and allows you to clarify significantly the readouts of all onboard sensors in the frequency domain.

The method of suppressing measuring noise in accelerometers on the base of the LMS-filtration algorithm [47] makes it possible to improve significantly the determination of the parameters of the current self-spatial orientation of the MR.

The method to determine the optimal MR trajectory in a given sector of space by the criterion of the minimum distance traveled and the maximum smoothness of the trajectory [24, 28] minimizes the number of impacts by the steering mechanism, resulting in reduced energy costs of the MR and deterioration of its mechanical parts, as well as improved directional stability and reduced possibility of drifts and slippage.

The proposed methods of probing the surrounding space by the TVS, as well as determining the frequency and time intervals possible to use in other areas, for example, in static monitoring of important engineering structures [48, 49] and building onboard standards of other objects.

Herein below for the first time, the concept of increasing robot survivability in non-deterministic non-contrast media is proposed. It is considered in the presence of radiation and other types of pollution, based on the use of technical electronic, electrical, mechanical, and computing tools, resistance to external factors, and the rational simplification of data reception systems. It is described how processing optical signals and computational algorithms are designed to process information on board the robot.

3 Fundamentals and Problems in Mobile Robotic Navigation

3.1 Practical Specialties of the 3D Laser Scanner Functioning

For better understanding of the general concepts of the spatial coordinates' acquisition, it is expedient to consider in deep detail all physical processes passing both simultaneously and consecutively during 3D laser scanner functioning. The 3D laser scanning itself was explained detailed several times in [22, 25, 27], so it is not useful to repeat it. We will just mark the most crucial points of this process visualized in Fig. 2.

3D laser scanner (or TVS) has laser ray emitter and reflected laser ray receiver. Laser ray can move in the horizontal plane with the help of stepper motor. It covers the necessities of 2D (plane) scanning. To add the third coordinate, the system is equipped with one more stepper motor rotating all the TVS in the vertical plane, or in other words moving the horizontal scanning plane along the vertical axis. This model of laser scanning is very simple and reliable and provides us with not-

abundant set of 3D coordinates. These coordinates have deterministic nature and well-described uncertainty distribution.

The main practical problem of such scanning model is well described in [28], and supposes the discretized field of view (FOV). From step to step of stepper motor, the laser ray forms unavailable sectors ("dead zones") within FOV, where the presence of an object cannot be detected. It becomes a more significant problem when the scanned object is located at longer distance from TVS. The efficient solution of this problem is given in the previous publications [41, 50]; it implies substitution of the stepper motor with device able to diminish the step size tending to cero. It can be the simple electromechanic reductor as in [41] or system based on DC motor emulating by control function the stepper motor behavior, but with variable size of step [50]. This is not so principal for optimized TVS functioning; more important here is to observe the next natural phenomena, which are doing the TVS functioning not so easy to implement (and which we will try below to eliminate or, at least, decrease).

The most challenge in this case is that (as shown in Fig. 2) the closer point has more accuracy. For the same size of scanning steps, the same object (blue shape in Fig. 2) has a different number of the detected points with measured coordinates when it is located on a different distance from TVS (known in literature as striking distance). This generates two main problems to solve:

- It undermines always the existence of the minimal obstacle size, which can be detected by TVS on the specified striking distance (so-called problem of minimal distinguishable grade).
- The nonuniform TVS uncertainty distribution as a function of striking distance.

This last problem can be illustrated as shown in Fig. 3.

The detailed experimental research of this natural phenomenon was presented in the past publication (Fig. 14 of [23]). The most precise measurement is in dark-green zone, less accuracy possesses the light color, and within blue zone, the system measurement results can be only considered as reference estimation. This

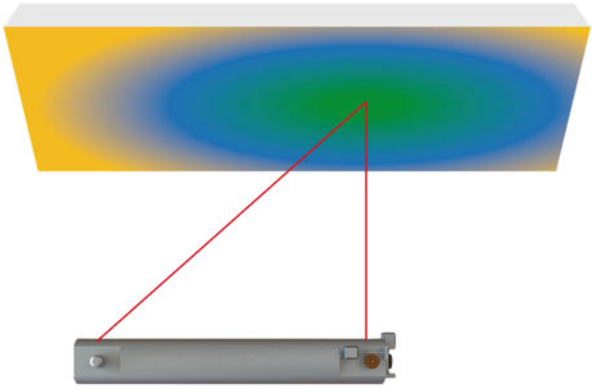


Fig. 3 Visualization of 3D laser scanning uncertainty/accuracy distribution

circumstance must be undermined on the stage of the MR navigation algorithm design.

It is important to note that the nonuniform uncertainty distribution does not depend only on striking distance, but also caused by symmetric uncertainty increase from the center to the edges of TVS, both left and right. Therefore, the techniques to decrease this disadvantage are not so simple as focus distance adjustment as per cameras, but become a nontrivial task, which obviously must combine both geometric and electronic approaches.

Let us analyze with more attention what are the main physical reasons for such TVS functioning. The main base of this angular geometrical uncertainty is the theoretical assumption that laser ray is a perfectly straight line with zero width, while the real ray is a cone, opening as more as longer it is. It led to a situation that a laser ray on the scanned surface forms not a geometric point, but certain sector of geometric figure, whose size and shape depend on many factors. This circumstance will be analyzed below.

3.2 Problems of Laser Spot Shape Imperfections

This analysis will provide with the help of visualization of Fig. 4. It is possible to say that angle of acquisition acts as a measure of distortion. Certainly, as shown in Fig. 4, the laser ray projects on the surface as a small circle spot (similar to geometric point) only in one position, when it is a local perpendicular to this surface.

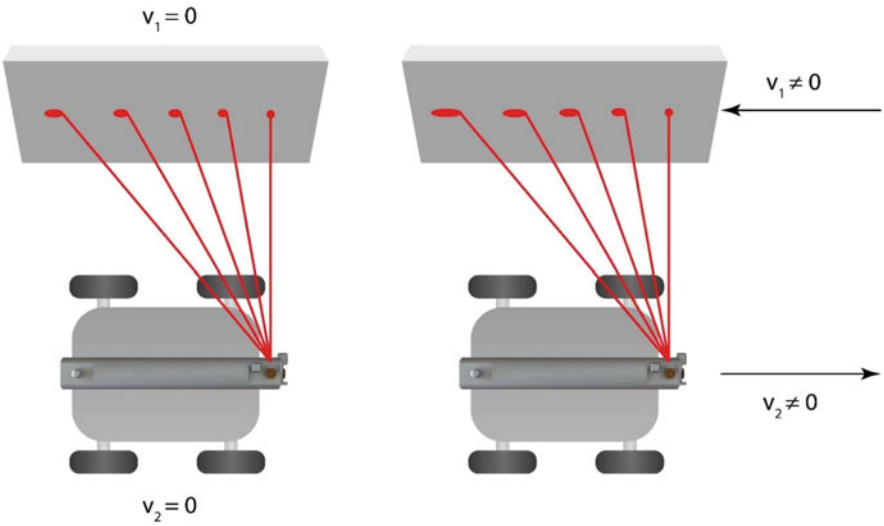


Fig. 4 Visualization of laser projection on the scanning surface

In other positions, this spot is geometrically deformed. As we can see from Fig. 4, to the left from perpendicular position, both for static (left part of Fig. 4) and for dynamic (right part of Fig. 4) MR positions, the projection of laser ray becomes more extended and changes its geometric shape.

The next important issue is that the laser ray projection even in perfect perpendicular position to the surface also never looks as a circle. It is small spot/blur with no smooth envelope line. Its shape imperfections become more visible and detailed also when the distance between laser source and projected blur is longer. These imperfections of the blur shape have different reasons, to be mentioned nonhomogeneous medium, light source fluctuations due to non-stable voltage, thermal effects, mechanical vibrations both of TVS and scanned surface, etc. The stochastic character of these fluctuations from the smooth line causes situation, when the simple geometric analysis of the obtained shape becomes inefficient: first of all due to unreasonably long time of such operation.

The reasonable approach to this problem solution would be given below.

The next important factor to analyze is clearly observed in Fig. 5. Mutual displacement of the observer's point and object of interest also causes the significant changes of geometric situation with projected laser spot shape.

The case of Fig. 5a represents a classic static monitoring, same as Fig. 4 (left). We can observe that the ray in perpendicular position forms a quasi-circular spot, but any other angle different from 90° on the right or on the left forms an oval spot. The distortion of spot raises from the center to the edges of FOV. Fig. 5b–e represents different kinds of dynamic cases. From Fig. 5b it follows that when MR start to move, even when the obstacle is still static, we can observe a kind of effect similar to Doppler effect for acoustic waves: the distortion of spot raises on the passed part and decreases on the part of scanned surface which MR still not have reached. This distortion of spot on behind and in front becomes as more evident, as more significant is vector sum of $V1$ and $V2$ (velocities of MR and obstacle correspondingly). It can be observed in comparison to Fig. 5b, c. The case of Fig. 5c shows that if $V1$ and $V2$ have opposite directions, the distortion of laser spot on behind and in front becomes less significant. In a particular case when $V1 = V2$, it converts into a static case as shown in Fig. 5a due to vector sum of $V1$ and $V2$ becoming zero.

It is expedient to note next important factor influent in whole uncertainty formation. Not only MR velocities cause the geometric distortions in ray tracing and triangulation scheme. As presented TVS can be classified as rotational laser scanner [27]; the rotational scanning velocity (angular velocity of electric motor, which moves the ray of laser) is also very influential on 3D coordinate measurement results. During many experiments was established the rule about rotational and linear displacement of laser ray: as faster is a displacement, as more uniform becomes general TVS accuracy. It can give a simple but useful conclusion. One of the most simple ways to decrease the TVS uncertainty is just to increase the scanning velocity, if application permits this.

Finally, one more important conclusion can be derived from observation of all cases in Fig. 5. As the accuracy of laser scanning depends on mutual positioning

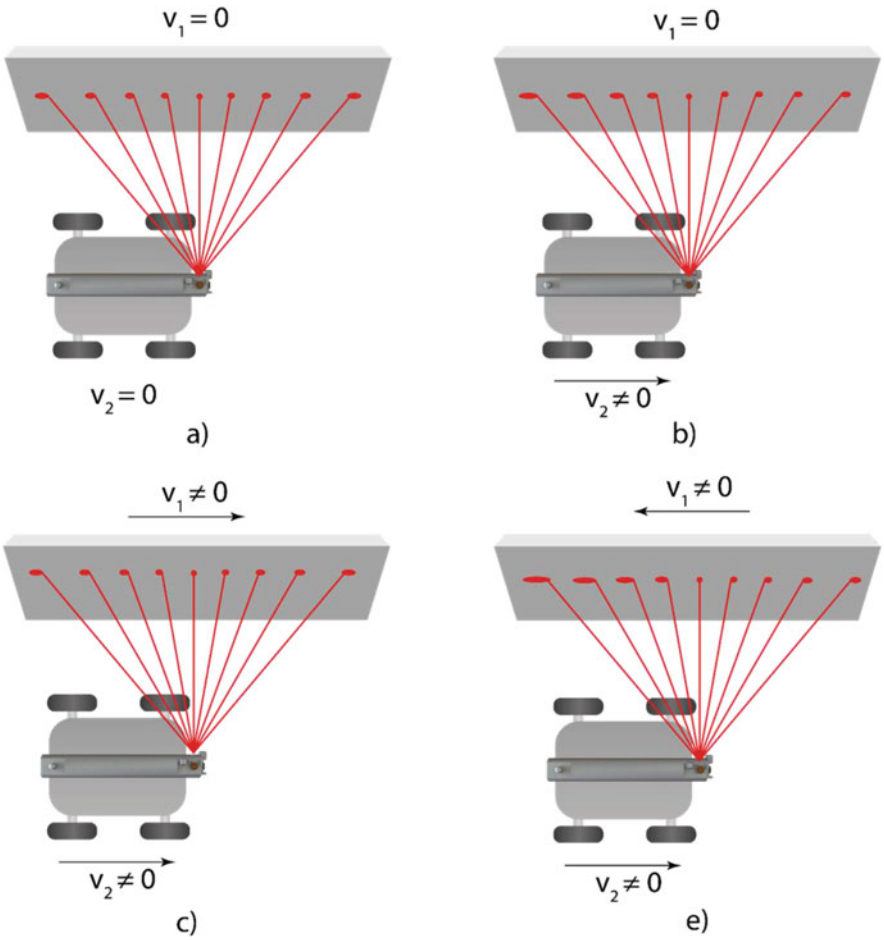


Fig. 5 Variations of laser projection spot on the scanning surface due to dynamic cases (**a**) static surface and static robot (**b**) static surface and robot in parallel displacement (**c**) surface and robot in parallel displacement with the same direction and (**d**) surface and robot in parallel displacement with opposite direction

of MR and obstacle, as well as other geometric aspects, so for more adequate information about same point is required data acquisition from different points, at least two.

This is the main reason why recently the monitoring tasks, especially of the enhanced important structure or landscapes, are preferably performed not by single MR, but by robotic swarms [26, 51–53]. That is why below we will consider some aspects of robotic group vision and navigation.

3.3 Problems of MR Group

The task performance by robotic group (frequently also called in literature as robotic swarm) [24, 26, 51–53] raises up the price of the task as many times as many robots involved. However, in addition to this single drawback, many advantages appear. To mention them, let us visualize the real scene where several robots are operating, and observe in detail aspects of the collective work. Before this, let us choose the example of any commercial robot, which can be used for such task performance. For this purpose, it is expedient to select the Pioneer 3-AT research model, because this research robotic platform was reported in the most literature sources for the last decade worldwide, and its further simulation here can be easily compared by many researchers from different laboratories without extensive recalculations of the MR working characteristics. In addition, even though the Pioneer 3-AT is not the most recent MR example, it naturally possesses the most typical electromechanical parameters established in the class of wheeled mobile robots for the last decades. The general view of Pioneer 3-AT is presented in Fig. 6.

As widely known, the minimal requirement for precise information is observation of the same point of interest at least from two different positions (better n), because according to metrology axiom, the increase of measurement n times is decreasing the uncertainty $\sqrt{n}$ times. Of course, the observation it is not the same as measurement, due to impossibility of repeatability and reproducibility conditions. However, the general behavior would be the same. This is the reason why below we will consider which practical specialties of laser scanning by group of robots what role plays.

Fig. 6 Pioneer 3-AT research robot model

3.4 Scanning from Various Positions of the 3D Laser Scanner

It is extremely important first of all to cover the topics and crucial points related with collaborative character of collective work within the robotic group. It is broadly known that the information can be characterized by its both quantity and quality. During the process of entropy decrease, both parameters are changing simultaneously. But only in the case of robotic swarm, designer according to task preferences can adjust the second parameter of information quality. In the case of scanning by single MR, this parameter is only stochastic and non-variable, defined by nature and system constants. The next three figures (Figs. 7, 8, and 9) will help us to explain the details mentioned above. Figure 7 shows the simplest geometric case when three mobile robots are moving in the same course parallel to each other and perpendicularly to obstacle. Even in this case, the obtained information will not be the same. Due to the natural unevenness of the surface, the data from the points located far will come later. If the whole information will be collected in the central computer (on the bottom of Fig. 7), the data transmission must be intelligent, and take into account the natural delays of 3D point data acquisition by TVS of each of n robots.

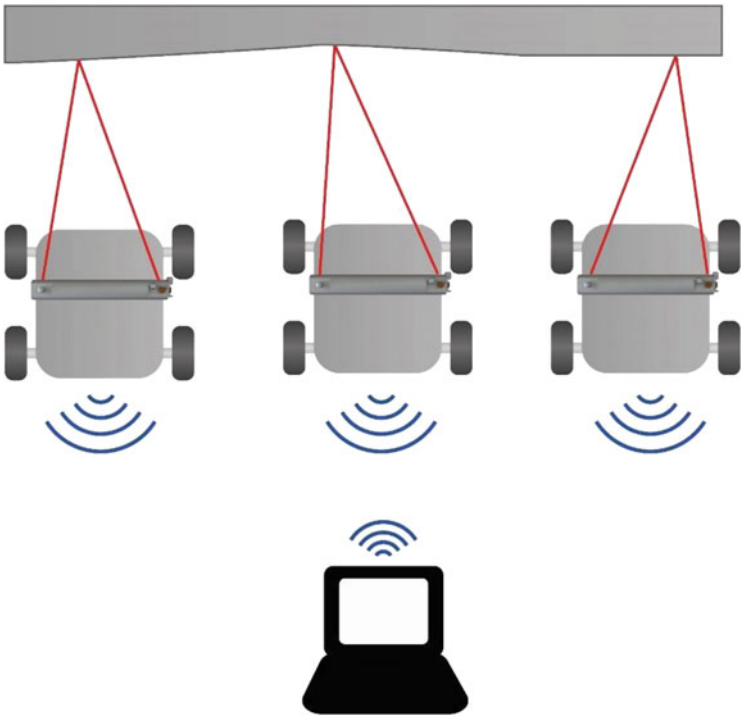


Fig. 7 Robotic swarm approach to the wall surface

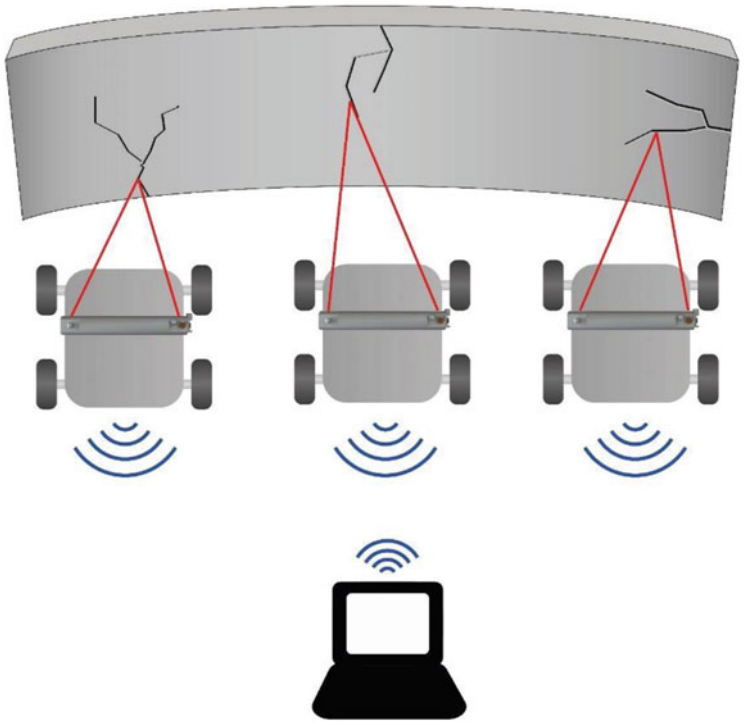


Fig. 8 Robotic swarm approach to the wall surface with cracks

The next problem is better illustrated in Fig. 8. When robots are moving in the same order, their arbitrary location with respect to defects (in considered example, cracks on the wall) on the scanned surface causes that mentioned defects will be detected in different moments of time. This is defined mostly by the individual distance "defect – MR number *n*-th." However, it also depends on the angular position of the MR with respect to the local perpendicular to the wall. This position, also called the angle of attack, initially is arbitrary.

Nevertheless, it is obvious that for optimal informational entropy decrease, it is preferable to adjust for each of *n* robots the individual angle of attack with respect to abovementioned facts, as it is shown in Fig. 9.

It is clear that a quantity of information, obtained simultaneously (parallel at the same time) from different points, and especially from the different angles of view from the beginning will give more enriched information, with additional details which, probably, can be inaccessible and out of reach in a case of single MR.

In this case, one of the most serious additional problems becomes the issues of properly designed algorithm for data transference between each robot and central computer, the data fusion of each *n*-th scan in the common 3D view of the scanned object, and the properly organized access for each of n robots to this commonly obtained 3D view.

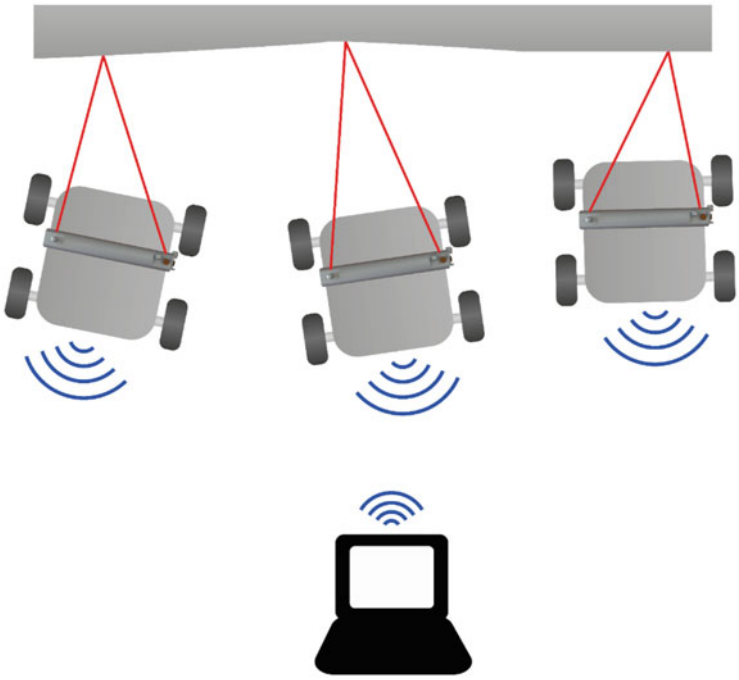


Fig. 9 Robotic swarm approach with different angle of attack to the wall surface

Assuming mentioned above, we can conclude that the parallel scanning from various points requires well-organized simultaneous data fusion.

Finally, we can add that on this stage the important questions are also exact quantity n of robots are required for obtaining of explicit 3D image, and where exactly is the smart spatial location for the mentioned central computer, with simplified bidirectional access for each *n*-th MR.

All these issues convert this task into nontrivial mathematical problem, deserving the detailed analysis separately.

3.5 Problems of Simultaneous Data Fusion of 3D Laser Scans

The detailed solution of the mentioned above problem requires a long-time analysis, and previously was explained in the past publications [24, 28]. It is not expedient to repeat it completely; it is enough just to refresh the most significant conclusions. They are the next: (1) it is possible, (2) the richness of information both depends on distance between MRs and angle of attack, and (3) the optimal quantity *n* of MRs does not exists as is: usually, it maintains the rule "more robots – more rich information," but for every practical application, it is very easy to detect the quantity

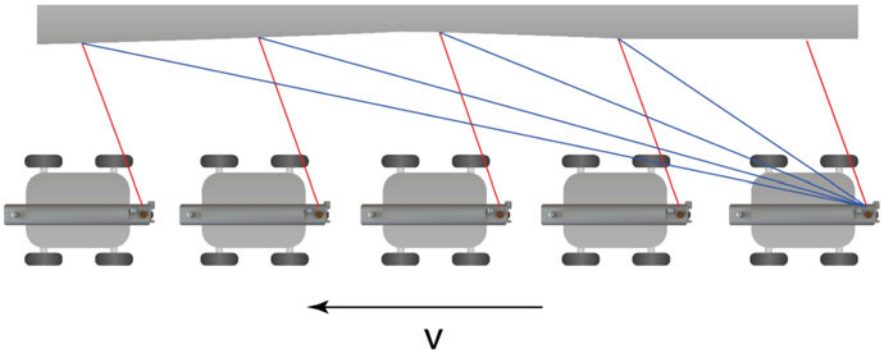


Fig. 10 Robotic swarm overlapping of TVS FOVs

n after what the informational quality increase becomes not critical or important for further performance.

It is only expedient to note that sometimes the quantity *n* is also limited from the other side: too much MRs within the group, especially on the limited size operational zone, leads to negative effect when TVS of each robot covers the sector of responsibility of neighbor robot, and causes unpleasant cross-talk effects, as it is shown in Fig. 10.

Therefore, the quantity of MRs in swarm must be not too small, but not redundant also.

Some geometrical details about informational quality and cloud density of points within the obtained 3D scan will be explained and visualized with the help of further figures.

Fig. 11 presented the basic case when three MRs are scanning the simple geometrical three-dimensional figure of more or less same size as robots themselves from the different sides and from relatively short distance. For easy explanation of facilities, the FOV of each MR is painted in different colors (red, green, blue), and in Fig. 10b the obtained points of 3D digitized shape are of the same color as the responsibility zone of scanning of *n*-th MR. Certainly, such a model can be considered as almost perfect order for unknown objects and 3D mapping. It provides a sharp image of real object with uniform coverage by digitized points, and each MR contributes more or less same part for final 3D image.

It is obvious that if robots are operating in unknown surroundings and approaching to obstacles arbitrarily, it is low probably that they will get a positioning as shown in Fig. 11. More probable is situation as shown in Fig. 12. When robots are approaching from far away to zone of obstacles, it is highly probable that they will meet primarily the different obstacles, and start to scan different objects. It will cause additional difficulties on the stage of data fusion. Therefore, some additional tools must be applied in order to classify the situation. MRs must estimate the relative distance between them, compare this distance with the size of already

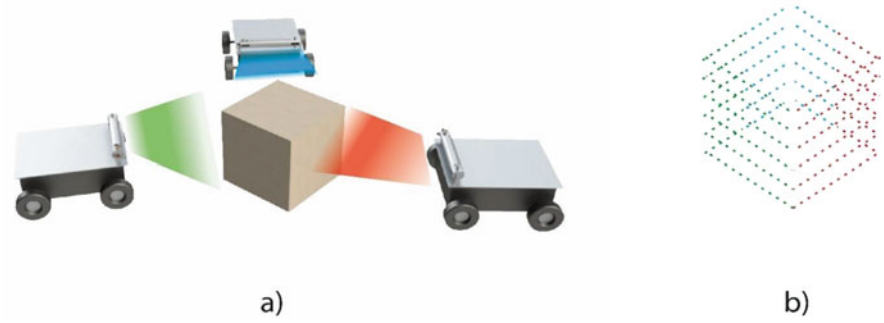


Fig. 11 Three robots scanning one object with the TVS (**a**) representation of three robots scanning one object (**b**) point cloud obtained by three robots scanning one object

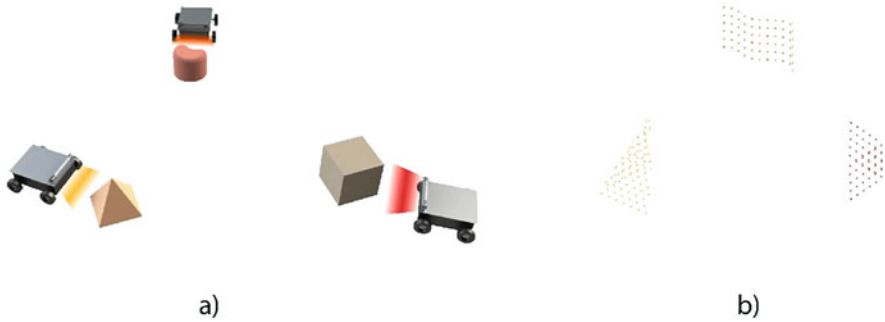


Fig. 12 Three robots scanning three objects with the TVS (**a**) representation of three robots scanning three objects (**b**) point cloud obtained by three robots scanning three objects

scanned part of surface, and analyze if the shape of scanned fragments and distance between them consider that they can be parts of the same object.

Figure 13 illustrates the next important issue in collective/collaborative scanning. In a comparison to Fig. 11, the geometric situation is the same, but the number of robots changes from three to four. At the first glance, it looks redundant. Five MRs are not necessary to scan such a simple object, and mutual overlapping of scanning sectors is clearly visible in Fig. 13b. Nevertheless, this can be helpful for the detection of the irregularities on the surfaces of the regular geometric shapes. In addition, the difference of the measurements of the same place can serve as indicator of noise, it can help to detect the instant rate of uncertainty between two adjacent robots in order to correct it, and it serves similar to photogrammetry of flat images in 3D reconstructions of scenes.

Figure 14 shows that same five robots on the scene of Fig. 11 will detect the shapes within the collective FOV more firmly. However, the distribution of which obstacle will maintain with less information about is still keep arbitrary. As it is not productive to maintain the quantity of MRs above the number of possible obstacles within the collective FOV, we will try further to find solution for this.

Fig. 13 Five robots scanning one object with the TVS (**a**) representation of five robots scanning one object (**b**) point cloud obtained by five robots scanning one object

Fig. 14 Five robots scanning three objects with the TVS (**a**) representation of five robots scanning three objects (**b**) point cloud obtained by five robots scanning three objects

Figure 15 represents the same scene of Fig. 13, but has two robots more. In this case, the redundancy is obvious; moreover, it is another form to illustrate the case of Fig. 10: the overlapping of responsibility sectors between several adjacent robots leads to situation, when the endeavor of uncertainty correction will cause more uncertainty.

Figure 16 is trying to show less significant, but still interesting, detail. When the quantity of MRs is naturally growth, it is possible to get again almost perfect situation. On the scene similar to Figs. 12 and 14, we can obtain much more realistic 3D obstacle mapping. It is only meaningful to suggest the algorithm when additional robots will join with that subgroup where the scanning surface from the first scanning steps looks as more complex. However, such swarm configuration is expedient to apply only when the quantity in swarm from the beginning is big, and robots have flexible task schedule. It is impractical to add robots to the group on purpose.

Fig. 15 Seven robots scanning one object with the TVS (**a**) representation of seven robots scanning one object (**b**) point cloud obtained by seven robots scanning one object

Fig. 16 Seven robots scanning three objects with the TVS (**a**) representation of seven robots scanning three objects (**b**) point cloud obtained by seven robots scanning three objects

Assuming the above analysis, it is expedient to say some next things. In the practice, it is unproductive to neglect the fact that robots in a real spatial configuration on cross-terrain can have quite significant differences in the three-dimensional orientation of the central axis of their field of view [23] (as shown in Fig. 17), as well as the effects of shading in some sectors, where, as we know, the signal amplitude in the photodetector can vary very significantly [31, 50]. Due to all mentioned above, herein below is presented the original proposition how to solve this inextricable triune problem: how to navigate MR with enough 3D data about environment, to sense the instant spatial 3D orientation of itself, and to synchronize the time axis of all these simultaneous processes and validate the instability of own MR's onboard time standard.

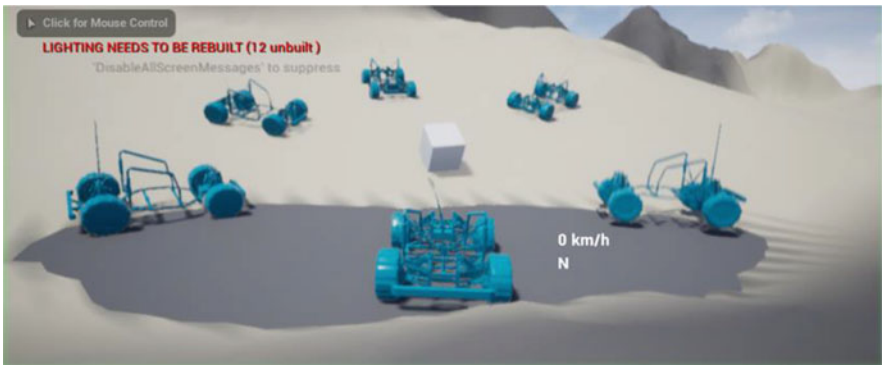


Fig. 17 Robotic swarm in real conditions of TVS functioning

4 Strategy of Mobile Robot Navigation

The state of the theory and methods of autonomous navigation and automatic motion control of mobile ground wheeled robots in non-deterministic environments are reviewed above. The research framework has been defined with the following conditions: a mobile ground wheeled robot operating offline is considered; the quasistationary nature of non-deterministic media; with irregularities that can be overcome, with a height difference of 0.25 the radius of the wheel of the robot; with the present possibility of contaminating factors (radiation, intoxication, etc.).

The importance of solving the strategic and tactical tasks that confront the robot, the possibility of using TVS data and systems for determining its own spatial orientation, and measuring the frequency of the onboard clock is indicated.

An analysis of various principles and methods for constructing the said MR systems is given, their defects are analyzed, and based on the results of the analysis, technical requirements are formulated for the parameters of the integrated autonomous navigation system of mobile ground robot. It was determined that for the implementation of autonomous navigation of the MR, it is necessary to obtain detailed information on all objects in a given wide spatial sector in the form of a coordinate grid of their surfaces, obtain reliable information about the current spatial orientation of the robot, determine the stability of the frequency of onboard clock of a robot, and be able to simultaneously reduce all this information in the matrixes of direct and inverse kinematics of the MP, which makes it possible to plot the optimal trajectory of the robot in non-deterministic environments.

Analysis of the technical aspects of this complex issue showed that the scientific and methodological apparatus used in the analysis and synthesis of algorithms for the functioning of mobile ground robots is imperfect and does not allow solving actual scientific and applied problems at a sufficiently high level. This gives rise to contradictions between the level of development of the scientific and

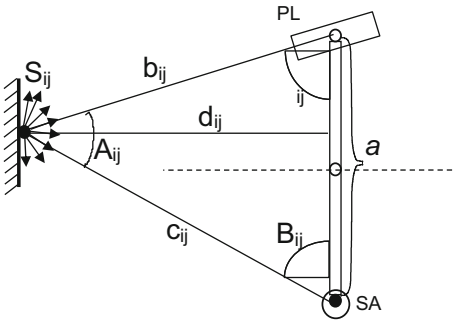


Fig. 18 The scheme of measuring coordinates by the triangulation method

methodological apparatus of this subject area and the requirements imposed in practice.

4.1 Practical Specialties of the 3D Laser Scanner Functioning

The proposed methods and scanning vision system for autonomous navigation of a ground robot using the principle of dynamic triangulation are considered in this paragraph.

The laser beam, with the help of two stepping drives, moves along a certain path in horizontal and vertical directions. At each step it stops, forming a triangle during a very short time (Fig. 18). Using the scanning aperture (SA), located at a fixed distance from the laser transmitter, the temporary position of the apex of this triangle is fixed, and the instantaneous coordinates of the reflection point x, y, and z on each i-th horizontal and each j vertical scanning steps are determined using the (2–5) formulas.

Coordinate calculations become possible due to the fact that during the existence of a triangle, three of its elements become known: the angle of positioning of the laser beam C_{ij}, the return angle of the diffusely scattered laser beam B_{ij}, and the base distance a between the optical centers of the SA and laser positioning system (PL).

The slope distance d_{ij} from the base to the points illuminated by a laser is determined by the formula.

$$d_{ij} = a \cdot \frac{\sin B_{ij} \sin C_{ij}}{\sin\left[180^\circ - \left(B_{ij} + C_{ij}\right)\right]}, \tag{1}$$

where a is the distance between the axis of rotation of PL and SA is the scan base. If $a = 1$ m, calculations are significantly simplified and scanning time is shortened. Using angle values, $B_{ij}, C_{ij}, \sum_{j=1}^{j} \beta_j$, and the basics of a, you can calculate the

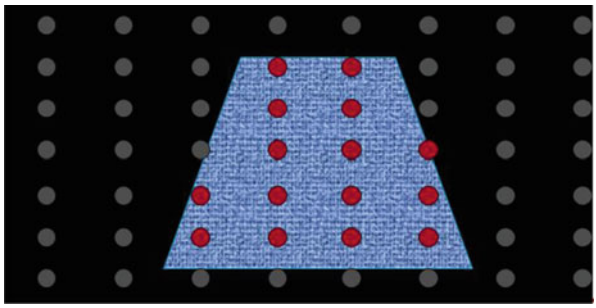


Fig. 19 The scheme of step-by-step scanning of the sector of the field of view of the robot and finding the obstacle by the laser beam

coordinates of each point highlighted by a laser on the surface of the obstacle (Figs. 20 and 18) in the rectangular Cartesian coordinate system OXYZ of TVS:

$$x_{ij} = a \cdot \frac{\sin B_{ij} \sin C_{ij} \cos \sum_{j=1}^{j} \beta_j}{\sin\left[180^\circ - \left(B_{ij} + C_{ij}\right)\right]} \tag{2}$$

$$y_{ij} = a \times \left(\frac{1}{2} - \frac{\sin B_{ij} \cos C_{ij}}{\sin\left[180^0 - \left(B_{ij} + C_{ij}\right)\right]}\right), \tag{3}$$

if $B_{ij} \leq 90^\circ$, (rotation to the right),

$$y_{ij} = -a \cdot \left(\frac{1}{2} + \frac{\sin B_{ij} \cos C_{ij}}{\sin\left[180^0 - \left(B_{ij} + C_{ij}\right)\right]}\right), \tag{4}$$

if $B_{ij} > 90^\circ$, (rotation to the left),

$$Z_{ij} = a \cdot \frac{\sin B_{ij} \cdot \sin C_{ij} \sin \sum_{j=1}^{j} \beta_j}{\sin\left[180^0 - \left(B_{ij} + C_{ij}\right)\right]}. \tag{5}$$

Thereby, the vertex of such a dynamic triangle slides on the surface of all objects in a given sector, covering them with a grid of discrete points (Fig. 19). A methodology has been developed for transforming a continuous surface of an obstacle into a grid of discrete points covering it, with specific detected 3Dcoordinates of each *ij*-th point.

Figure 20 shows the main elements of TVS and its placement on a MR. The TVS is located in the upper front part of the MR and contains a powerful laser with a collimator, installed directly in the laser positioning system (PL), a photo-receiving scanning aperture (SA), and a control system for its elements.

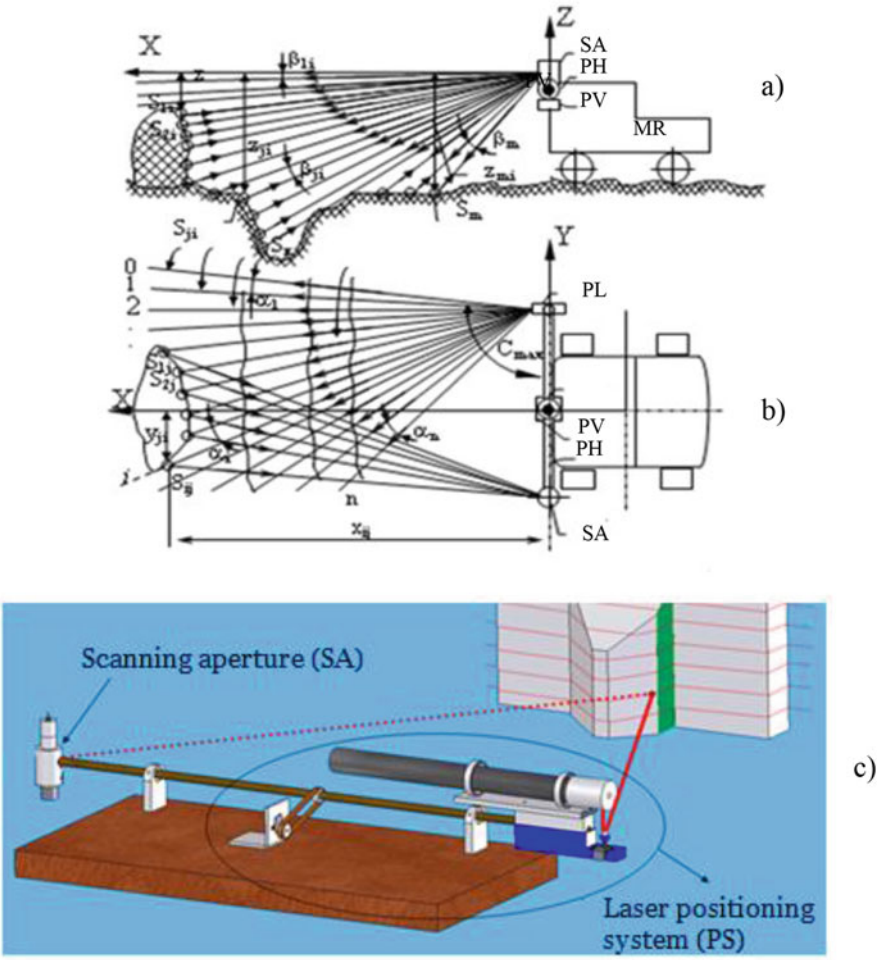


Fig. 20 The principle of operation of the TVS based on dynamic triangulation. (**a**, view from the side; **b**, a top view; **c**, a prototype of a laser scanner)

The main characteristics of the proposed triangulation method were studied by theoretical and experimental methods using the developed experimental sample TVS, which Fig. 21 shows. The studies were conducted on obstacles, different in material, size, and shape, with various absorbing and reflecting features of surfaces placed in different points of the field of view of the TVS.

Experimental studies of the accuracy characteristics of the system were performed on a special optical Table. A variety of reflecting and absorbing objects was placed at the nodes of the high-precision grid. An analysis of the results of measurements of the coordinates of objects allowed us to identify the following

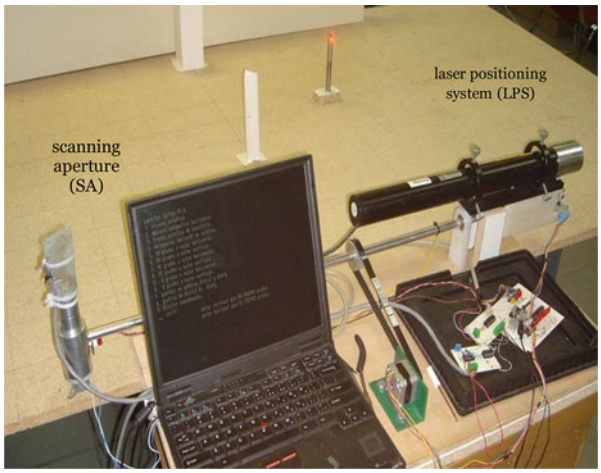


Fig. 21 The experimental setup used in laboratory studies

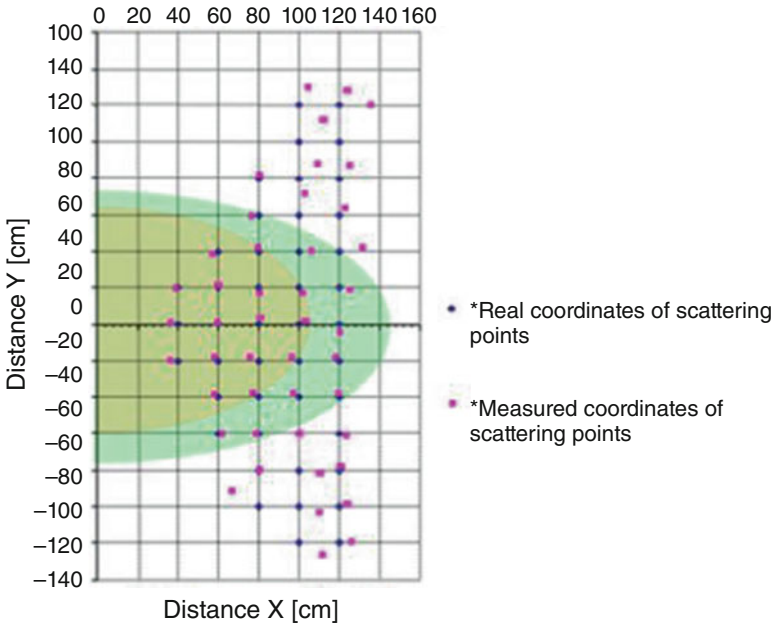


Fig. 22 Three zones of TVS accuracy identified during laboratory studies

zones in the TVS field of view: reliable (relative measurement error of coordinates $\delta \leq 1\%$), acceptable ($\delta \leq 5\%$), and peripheral ($\delta > 5\%$) (Fig. 22).

Such a laser scanning scheme is preferable for mobile robot TVS because it allows to decrease the most of mentioned above physical constraints.

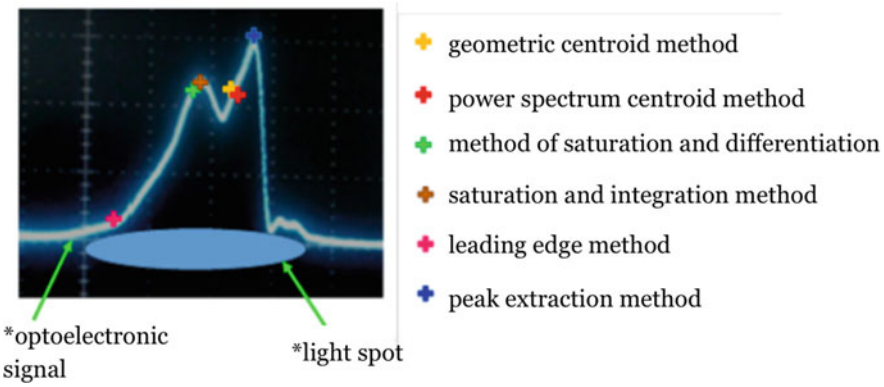


Fig. 23 The comparison of methods for estimating the localization of the energy center of a laser spot.

4.2 Problems of Laser Spot Shape Imperfections

As have been mentioned above, in 3.1 by Figs. 4 and 5, one of the most significant sources of uncertainty is the irregular shape of the laser reflection spot. To process this spot geometrically will generate a huge raise of computational cost. Nevertheless, we can avoid this by simple physical analysis. In theory, we are considering laser ray as straight line. For the distorted spatial body (laser cone), this line represents its rotational axis. Consequently, on the projected laser spot, the projection of this axis is the geometric center of the obtained irregular figure, independently of its real shape or grade of irregularity.

Additionally, we will find this center not on real light spot, but on the result of its conversion into proportional electric signal. These methods exist, and have been explicitly analyzed in our previous publications [39, 40, 45].

The method of mathematical computer simulation was used to analyze the known methods for estimating the energy center of a laser signal spot scattered on interference: the geometric centroid method, the power spectrum centroid method, the saturation and integration method, the leading edge method, and the peak extraction method (Fig. 23).

The analysis showed that considered methods do not provide sufficient accuracy in determining the spatial position of the energy center of the illuminated laser spot on the surface of the obstacle under conditions of a limited time.

A new method of saturation and differentiation is proposed for estimating the temporal position of the center of the pulse with a complex shape of the envelope that satisfies the set conditions. This contributes to the improvement of the distinctiveness proposed by the TVS by determining the energy center of the laser spot. The technical parameters of the TVS were determined: coordinate measurement accuracy, speed (obtaining coordinates of a single point takes 0.2 ms),

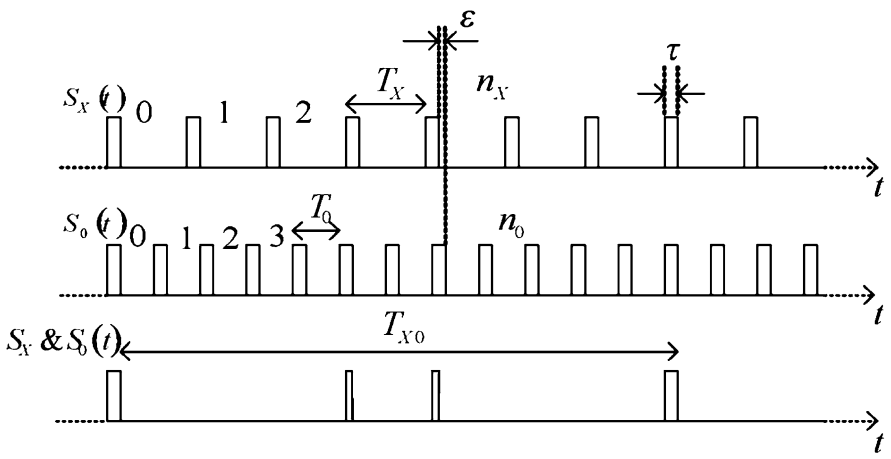


Fig. 24 The pulse coincidence principle

and range (for a range of 20 m, the value of the required laser power is 1.26 W; for a range of 100 m, the value of the required laser power is 31.5 W).

4.3 Onboard Robot Reference Clock Validation

For perfect navigation, it is extremely important not only to know 3D coordinates of obstacles but also to know precisely how much time MR moves along the calculated trajectory. Any errors in time readings are immediately converted into errors in self-positioning on this trajectory. Time readings' uncertainty mostly is caused by not stable frequency of the onboard standard of time (MR clock). Therefore, the onboard clock frequency must be controlled at any moment when MR is moving along its trajectory. The next section describes a new method for quickly determining the frequency of a signal (onboard robot reference clock). It is a method based on the theory of the coincidence of impulse sequences by packets, with the determination of the "best" matches using the mathematical apparatus of Farey fractions or rational approximations of an unknown value by mediants.

The essence of the method is illustrated (Fig. 24) when measuring the frequency: $S_x(t)$ is a sequence of pulses to be measured; $S_0(t)$ is the pulse sequence of the reference frequency. The final measurement interval T_{x0} corresponds to the complete coincidence of the sequences of pulses. As can be seen from the figure, during the measurement, incomplete partial coincidences of the pulse sequences occur, which do not lead to the end of the measurement process.

Pulse coincidences occur when the inequality holds:

$$|n_X T_X - n_0 T_0| \leq \varepsilon \tag{6}$$

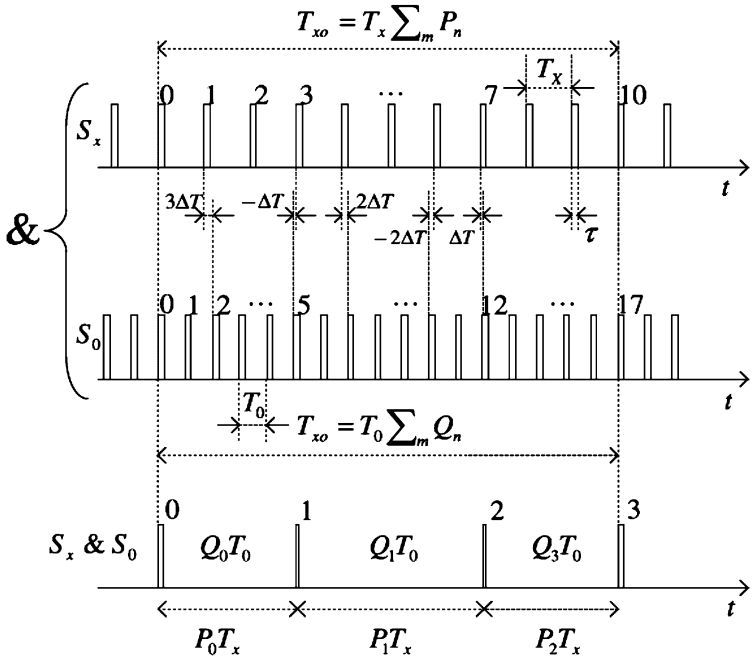


Fig. 25 Geometrical scheme of pulse coincidence

where ε is the permissible error (difference value between the time intervals n_0T_0 and n_XT_X (Figs. 24 and 25). With the worst partial coincidence, when the pulses overlap is zero, the maximum measurement error is 2τ.

The left side of Eq. (6) shows that the T_X/T_0 approximation uses rational numbers $f_X = 1/T_X$ which allows writing.

$$\left| f_X - \frac{n_0}{n_X} f_0 \right| \leq \frac{\varepsilon f_X f_0}{n_0} \tag{7}$$

In (Eq. 7), f_x is a hypothetical effective estimate of the unknown frequency, and f_0n_0/n_x is the frequency obtained by measurement. Relative measurement error is.

$$\beta = \left| f_X - \frac{n_0}{n_X} \right| / f_X \leq \frac{\varepsilon}{n_0 T_0} \tag{8}$$

With the full coincidence of the pulses (when $\varepsilon = 0$, Fig. 25), at the moment T_{X0}

$$|n_X T_X - n_0 T_0| = 0 \tag{9}$$

and each time interval n_0T_0 and n_XT_X is equal to the measurement time T_{X0}

$$n_0T_0 = P_nQ_n\Delta t_{n-1}, \tag{10}$$

$$n_XT_X = P_nQ_n\Delta t_{n-1} \tag{11}$$

where Q_n and P_n are the indicators of counters of the reference and measured frequencies, counting n_0 and n_x for n-th coincidence, and Δt_{n-1} is the difference indicator of two integer intervals n_0T_0 and n_xT_x; it has the same order of magnitude as the expected relative measurement error β.

It is appropriate that the permissible error in (Eq. 6) be $\varepsilon = \Delta t_{n-1}$. In number theory (in the Euclidean algorithm used to search for the greatest common factor), this parameter Δt_{n-1} is known as the greatest common factor. Under the condition of decimal calculation for both periods T_X and T_0, $T_X < 1$ and $T_0 < 1$, the expression for the greatest common divisor Δt_{n-1} is

$$\Delta t_{n-1} = (T_X, T_0) = \frac{1}{10^r}\left(A, 10^{r-s}\right) \tag{12}$$

where A, r, and s are integers and $r > s$, r is an exponent for the expected order of magnitude β and r-s is the difference between the relative error β and the order of magnitude of the standard pulse sequence.

If numbers A and 10^{r-s} in Eq. (12) are mutually simple, then $\Delta t_{n-1} = 10^{r-s}$ ($a = 1$), and if they are not mutually simple, then $\Delta t_{n-1} = a/10^r$, where a is an integer and

$$Q_n = 10^{r-s}/a \tag{13}$$

Since the counting codes by pulse counters of unknown and known frequencies are equal to the whole number of pulses counted on the n-th coincidence (Fig. 25), the number of periods of the measured pulse sequence is

$$n_X = A \cdot 10^{r-s}. \tag{14}$$

Expression (14) is a numerical condition that is proposed to stop the measurement process. The criterion for the end of measurement is formulated as follows [37]: the best approximation of the measured value for a given period of measurement takes place when registering the coincidence of the pulses of the measured and reference pulse sequence; the counter of the number of pulses of the measured sequence shows "an integer with zeros."

The search for the best approximation of the measured frequency is illustrated in Fig. 26.

As can be seen from Fig. 26, each circle corresponds to the approximant frequency found when the pulses coincide. Among them, the group is clearly defined group convergent (circled in the figure), which are much faster than others

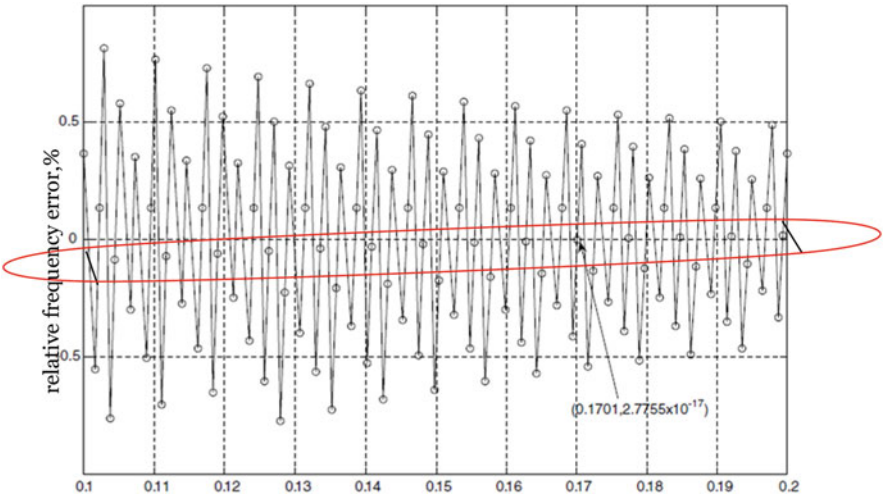


Fig. 26 The search for the best approximation of the measured frequency with zero error

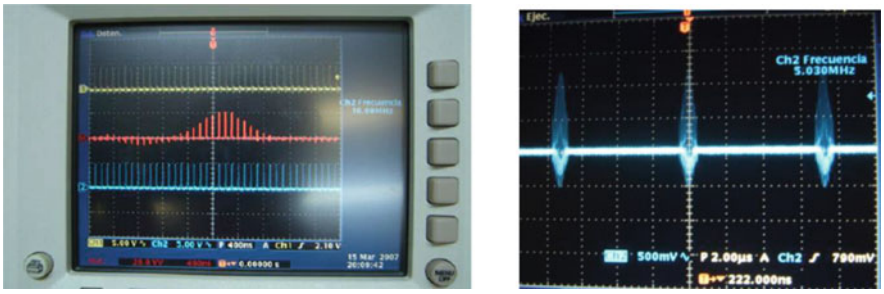


Fig. 27 The type of coincidence of pulses in the experiment (left). A jitter of coincidences packets (right)

approaching zero error. The one which the criterion mentioned above is fulfilled in 0.1701 s; the error is equal to zero.

The study of the proposed method of measuring the frequency was performed by the mean of mathematical computer simulation and experimentally (Fig. 27).

The research results showed that this method quickly enough allows you to register changes of the current frequency value.

The analysis of the effect on the accuracy of this method of random phase shifts of the signal pulses (jitter) is given [36]. It can be seen from Fig. 28 the influence on the coincidence process of various combinations of four typical types of jitter. These types are random oscillations, also known as random jitter (RJ); periodic oscillations, or periodic jitter (PJ); distortion of the fill factor, also known as digital count distortion (DCD); and initial dependent oscillations, or data-driven jitter (DDJ). This combined jitter deforms the form of convergent groups of the

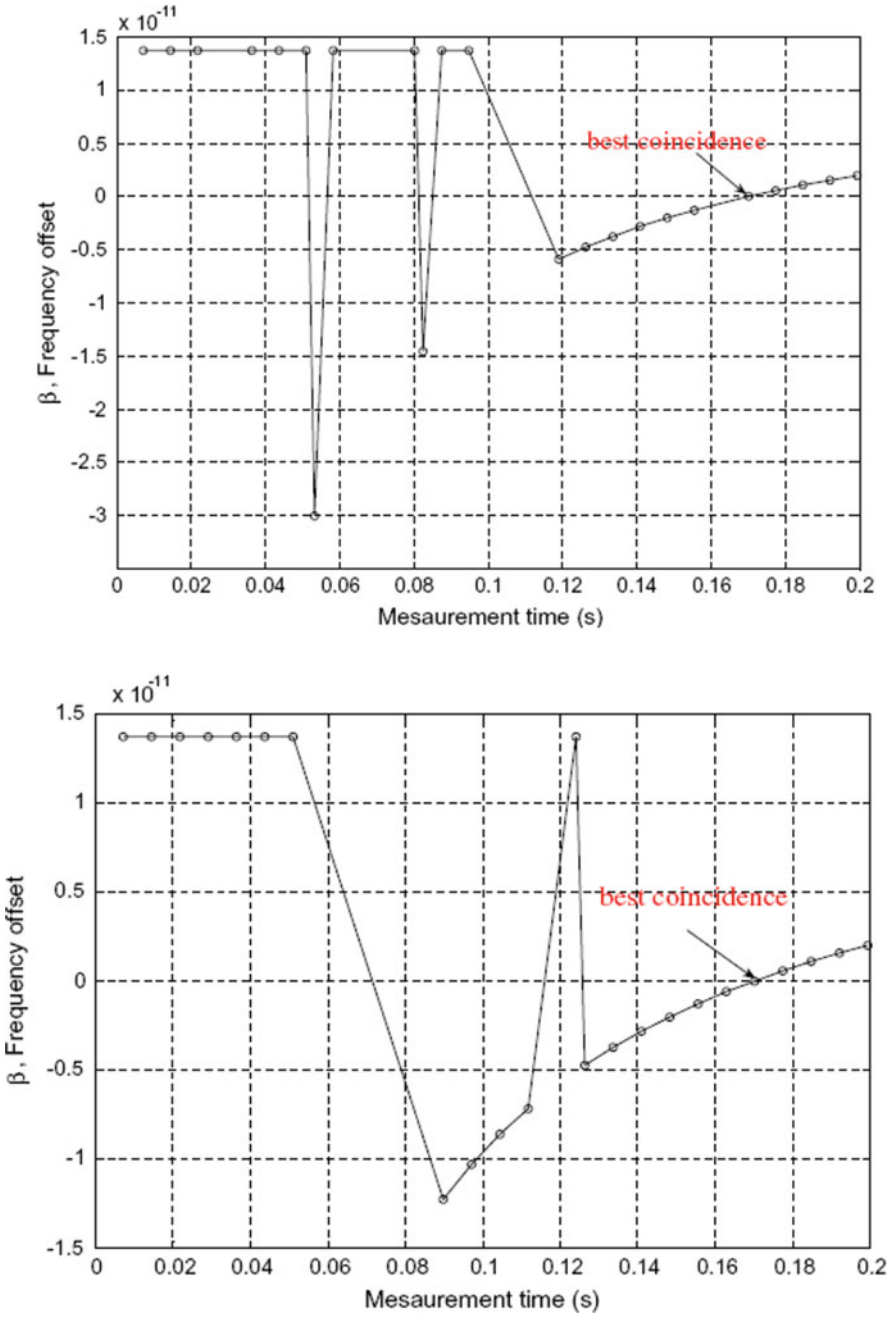


Fig. 28 Search for the best approximation of the measured frequency with zero error in the presence of jitter

ideal case, but a place of best match remains the same for all cases. This indicates the invariance of the proposed method for measuring jitter. Fig. 28 shows just two different screenshots, selected from several experiments described in [36].

4.4 Neural Networks' Application on 3D Measurement Error Decrease

This section describes the methods of information processing in the TVS, allowing to improve the performance of the system as a whole in terms of distinctive ability, performance, and noise immunity.

A method has been proposed for expanding the reliable zone of the field of view by filtering gross errors of scanning results and creating a digital map of the visible surface of the obstacles identified using the Levenberg-Marquardt algorithm. This algorithm is implemented in a digital filter for processing measurement results, which allows to improve the accuracy of the initial scan results and rejects anomalous errors.

Its comparative analysis [31] was carried out with other methods most recognized in the world literature, such as the Polak-Ribier and quasi-Newton methods, which are used alternately in a neural network with two steps of learning. The feedback step in the neural network calculates an error in the gradient descent and propagates this result back to each neuron in the input layer and then in the hidden layer.

In the second step, the weights' coefficient and deviations are calculated repeatedly, and the output from the activated neurons is transmitted forward from the hidden layer to the original layer. The network was calibrated with random weights and deviations, and then trained using all three algorithms mentioned above. The training package consisted of 60 data samples selected from 80 actual practical measurements. Each sample is taken in the form of mutual ratification.

In that way, the network was trained to provide an absolute measurement error of the x, y, z coordinates for all conditions at once. The length of the training data was 60 points. The network contained five neurons and three layers and was trained until an acceptable percentage of error was reached. The test data consisted of a residue of 20 samples (results of measuring x, y, z coordinates) from each test (Fig. 22).

The training (one by one for the same package $60 + 20 = 80$ data) was performed in accordance with the methods of Polak-Ribier, quasi-Newton, and Levenberg-Marquardt, in order to compare their effectiveness for a given task.

The comparison results are illustrated in Figs. 29 and 30.

It is possible to assume that the idea of neural networks' application, due to the finite and nonredundant quantity of data within typical FOV, has a good contribution and allows increasing the general efficiency of the system at whole.

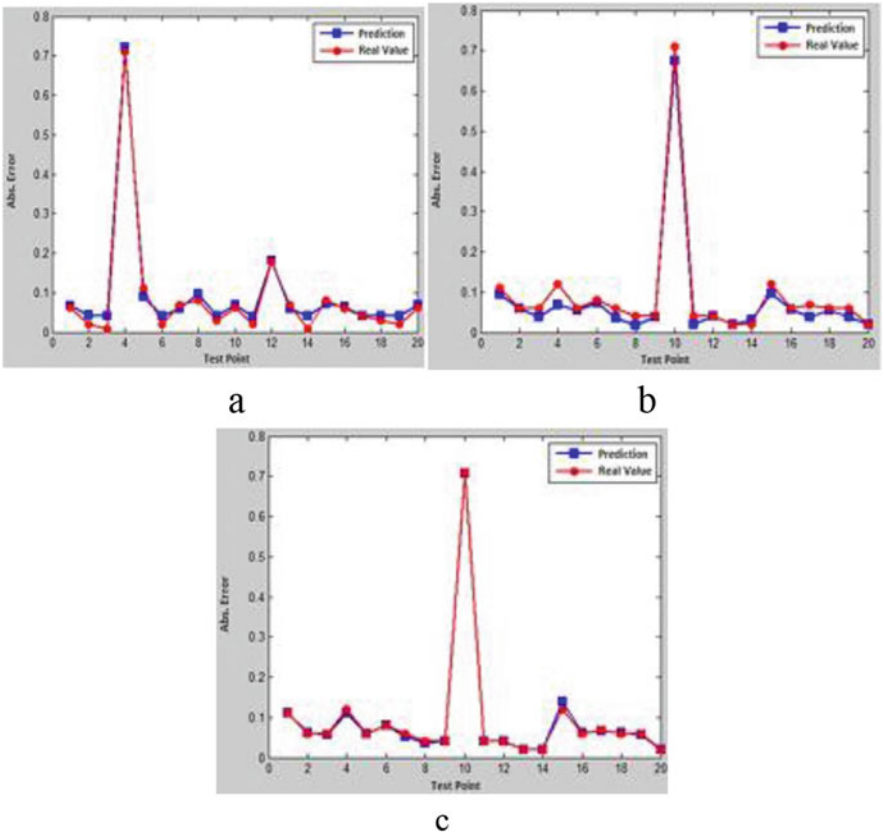


Fig. 29 The predictions of errors by a neural network trained in the following methods: (**a**) Polak-Ribier, (**b**) quasi-Newton, and (**c**) Levenberg-Marquardt, for each test point of the scanner compared to its real error

4.5 TVS Functioning with a Variable Scanning Step for Faster Search

A method for the operation of the TVS laser scanner with a variable scanning step, which allows reducing the time to search for an obstacle and a digital description of its dangerous edge, is proposed. To determine the optimal angles of the scan step, a mathematical model of the field of view of the work has been developed, which includes 1067 scenarios: obstacles with different shapes, sizes, and numbers, as well as their diverse location in the field of view sectors and grouping options are used. Evaluation of the scanning efficiency was performed by the parameter $Z = P/t$, where P is the number of points identified on the way and t is the scan time.

For all scenarios of the model, informative graphs are obtained, shown in Fig. 31. By pair comparison and elimination of the worst, the three most informative angles

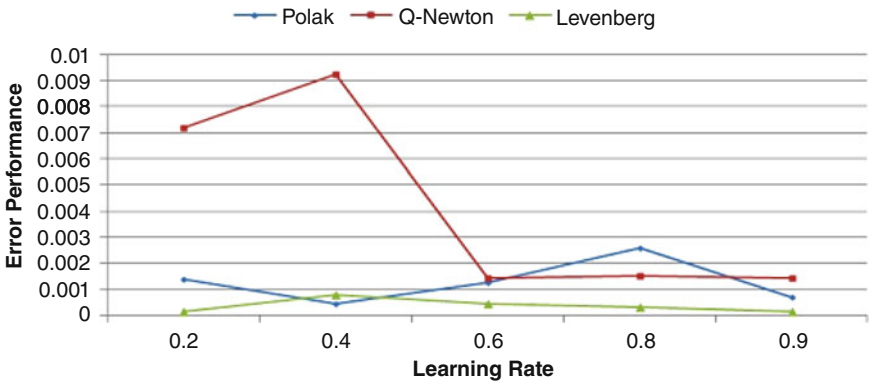


Fig. 30 Comparison of the prediction error of the three methods for different indicators of the coefficient of learning

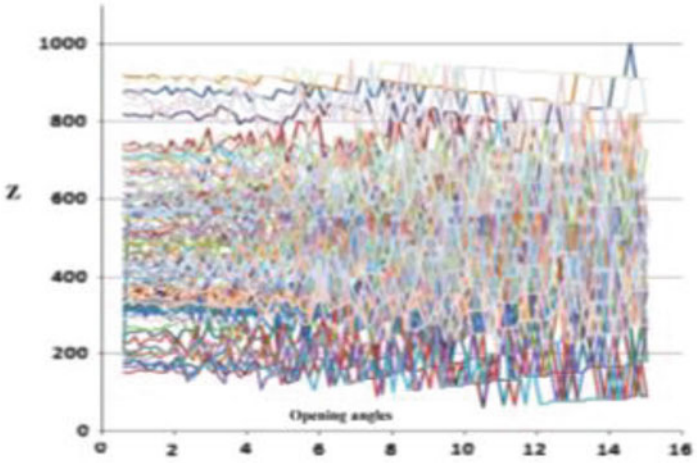


Fig. 31 Comparison of the informativity of the TVS for different angles of the scan step

of the scanning step are defined: small, medium, and large (indicated on the graph in Fig. 32). As the simulation results indicate, combining the angles of the scanning steps makes it possible to significantly speed up the search for obstacles in the field of view of the robot.

Algorithms for filtering the useful signal in the subsystem of the accelerometric determination of the robot's own spatial orientation and the useful signal of the photodetector, which is influenced by noise due to mechanical vibrations in the scanning aperture, are proposed. By comparative analysis of various types of filters known from the literature, it is recommended to use an LMS filter, which provides the best performance. The results of physical modeling showed that the use of the

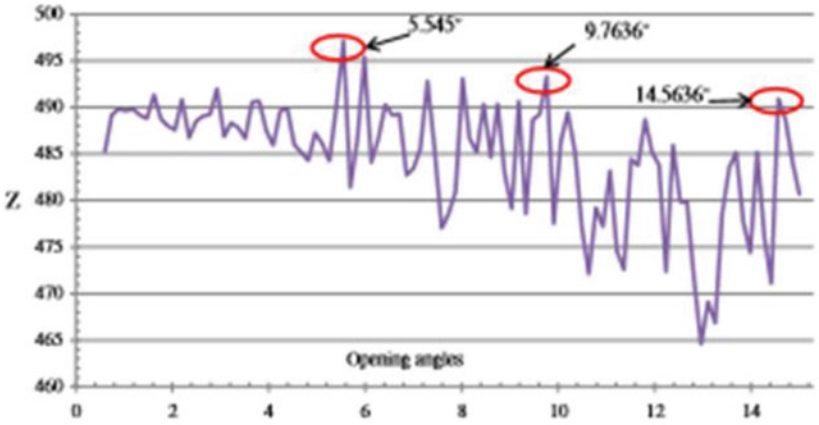


Fig. 32 Election of the most informative angles of the TVS scanning steps

filter significantly improves the quality of determining the own spatial orientation of the work (up to 0.2° for each of the coordinates).

4.6 Path Planning for MR Navigation

The next section shows the use of the results obtained in Sects. 4.2, 4.3, 4.4, and 4.5, to determine the trajectory of the MR for its autonomous navigation.

The state of the robot is described by a matrix mathematical model that uses data from the TVS, positioning accelerometers, and the onboard clock.

The following equations are defined to control the state and orientation of the mobile robot in three-dimensional space (Fig. 33) by a fixed base coordinate system:

$$^{0}T_r = \text{Trans}\,(x, y, z) \times \text{RPY}\,(\varphi, \theta, \psi) \tag{15}$$

$$^{0}T_r = \text{Trans}\,(x, y, z) \times \text{Rot}\left(\overrightarrow{a}, \varphi\right) \times \text{Rot}\left(\overrightarrow{o}, \theta\right) \times \text{Rot}\left(\overrightarrow{n}, \psi\right) \tag{16}$$

$$^{0}T_r = \begin{bmatrix} 1 & 0 & 0 & x \\ 0 & 1 & 0 & y \\ 0 & 0 & 1 & z \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} c\varphi & -s\varphi & 0 & 0 \\ s\varphi & c\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} c\theta & 0 & s\theta & 0 \\ 0 & 1 & 0 & 0 \\ -s\theta & 0 & c\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & c\psi & -s\psi & 0 \\ 0 & s\psi & c\psi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{17}$$

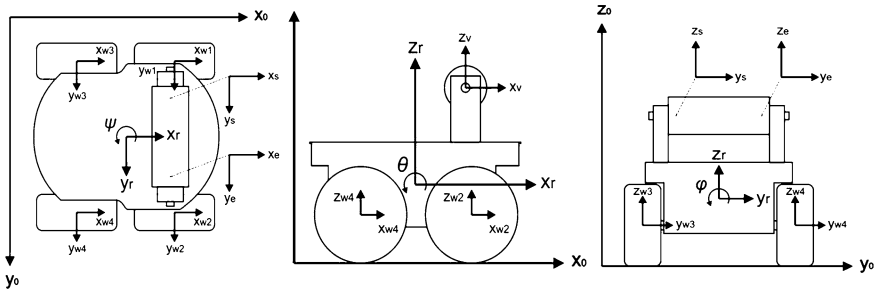


Fig. 33 The binding scheme of the coordinate systems of the mobile robot

$$ {}^{0}T_r = \begin{bmatrix} c\phi c\theta & c\phi s\theta s\psi - s\phi c\psi & c\phi s\theta c\psi + s\phi s\psi & x \\ s\phi c\theta & s\phi s\theta s\psi + c\phi c\psi & s\phi s\theta c\psi - c\phi s\psi & y \\ -s\theta & c\theta s\psi & c\theta c\psi & z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{18} $$

where the first three columns in Eq. (18) are unit vectors for the orientation of the mobile axes of a robot $\left(\left(\vec{a},\ \vec{o},\ \vec{n}\right)\right)$, the last column is the position vector of the center of mass of the robot in Cartesian coordinates (*x, y, z*), ${}^{0}T_r$ is the transformation of the matrix from the coordinate system r (associated with the robot) to the coordinate system *0* (base coordinate system), and $c\varphi$ and $s\varphi$ are the designation of the functions cosφ and sinφ (similar designations are used for the functions of the angles θ and ψ).

The sequence of *RPY* angles determines the robot's own spatial orientation: φ, rotation around the axis of the robot (roll R); θ, rotation around the axis of the robot; pitch P, rotation around the axis (yaw Y).

Assuming the desired positions and orientations in Cartesian coordinates and RPY angles are known, from the direct kinematic model:

$$ {}^{0}T_r = \begin{bmatrix} r_{11} & r_{12} & r_{13} & px \\ r_{21} & r_{22} & r_{23} & py \\ r_{31} & r_{32} & r_{33} & pz \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} c\phi c\theta & c\phi s\theta s\psi - s\phi c\psi & c\phi s\theta c\psi + s\phi s\psi & a \\ s\phi c\theta & s\phi s\theta s\psi + c\phi c\psi & s\phi s\theta c\psi - c\phi s\psi & b \\ -s\theta & c\theta s\psi & c\theta c\psi & c \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{19} $$

The inverse kinematic model is found when the coordinates *px, py,* and *pz* are incremented as the robot moves:

$$ \varphi = \mathrm{ATAN2}\,(r_{21}, r_{11}) \tag{20} $$

$$\theta = \text{ATAN2}\left(-r_{31}, r_{11}c\phi + r_{21}s\phi\right) \tag{21}$$

$$\psi = \text{ATAN2}\left(-r_{23}c\phi + r_{13}s\phi, r_{22}c\phi - r_{12}s\phi\right) \tag{22}$$

where φ, θ, and ψ are unknown RPY angles.

The inverse kinematic model of a mobile robot is used to control the position of the MR and allows determining the value of each angle in order to give the mobile robot the desired position and orientation.

A method has been developed for determining the MR movement trajectory in a given sector of space by the criterion of the minimum path and maximum smoothness of the trajectory (minimizing the influence of the steering mechanism). The first approach was to use the minimum number of points needed to form a curve. Figure 17 shows the curve obtained by three points; it successfully avoids obstacles and reaches the goal, but the distance covered by the arc is of excessive length (Fig. 34a).

Another approach uses a larger number of points to construct a trajectory and form a curve. The simulation results, in which 10 points were used to build the trajectory, are shown in Fig. 34b. As you can see, the curve not only bypasses the obstacles and reaches the goal but also reduces the path length by 12–15%, despite the fact that the time spent on the calculations increased by only 1.2 ms. This compromise allows you to significantly save resources of a robot, and the operating time of the control system increases slightly.

The research results showed that to safely pass through obstacles along the shortest path with maximum smoothness (which ensures the minimum power consumption of the steering and minimal wear of its mechanism in any implementation scheme), an increase in the number of points to build the desired trajectory is impractical: further increase in time for trajectory calculations does not lead to energy benefits.

In conducting physical experiments, the technical and geometric characteristics of the Pioneer 3-AT research model were used (Fig. 6).

5 Conclusions and Outlook

The work contains theoretical generalizations and the solution of the actual scientific and applied problem of parallel coordination of the following theoretical tasks, which are inseparably connected with the complex solution of automatic navigation of the MR: development of the theory and methods of machine vision for being able to see the sector before the MR, developing a method for checking the frequency stability of the onboard clock, and further developing the theory of determining the MR's own orientation, by creating new and improving known methods and mathematical models for probing the surrounding space, determining the coordinates and motion parameters of a mobile robot, and forming the optimal

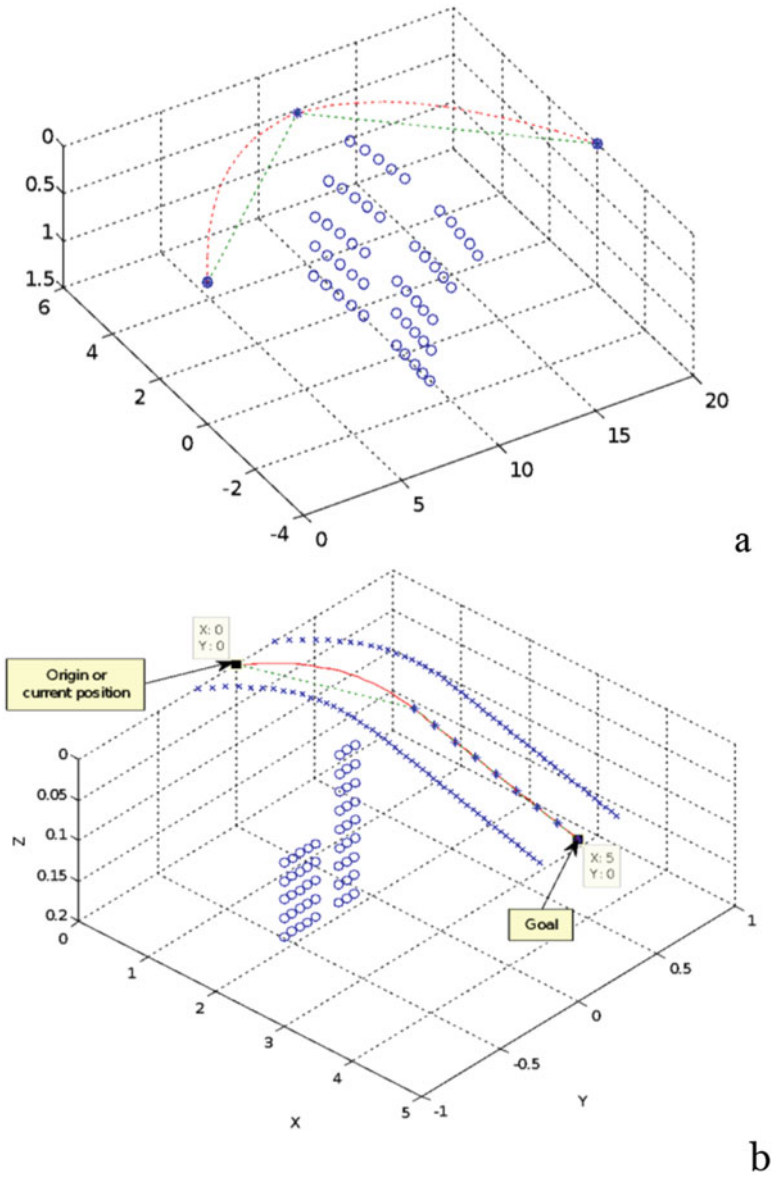


Fig. 34 The scheme of simulation of a mobile robot trajectory in MATLAB using (**a**) 3 points (scale in meters) and (**b**) 10 points (scale in meters)

trajectory of its movement non-deterministic environments subject to the presence of polluting factors. At the same time, practical advantages have been obtained in implementing the navigation of a mobile ground-based wheeled robot offline, which are listed below.

The main results of the work are as follows:

- The theory of autonomous navigation of mobile ground-based robots in non-deterministic environments is developed based on the integrated use of information about the environment of the robot and its own spatial orientation and time scale in a single matrix of the state of the robot.
- A new method of laser sensing of space around a robot is proposed – a method of dynamic triangulation – which makes it possible to increase the scanning area of a space in front of a robot to an acceptable level, in which the definition and localization of objects become deterministic. The method allows you to create a digital map of obstacles that simplifies the construction of its trajectory. It allows you to identify dangerous edges of obstacles in the Cartesian coordinate system convenient for fast transformations; rank their danger by distance to them, without additional transformations; and place all objects in a single global coordinate system, which significantly reduces control time and eliminates the possibility of conversion errors. The created digital map can be useful for navigating a group of robots and for describing and mapping unknown spaces first passed by the robot.
- It is indicated that the field of view triangulation TVS has three zones, each of which is characterized by certain indicators of the accuracy of measuring the coordinates of objects: reliable (relative measurement error $\delta \leq 1\%$), acceptable ($\delta \leq 5\%$), and peripheral ($\delta > 5\%$).
- A new method for measuring time intervals and frequency of signals has been developed – a method of rational approximation by median approximation – which is based on the coincidence of some integer number of measurement intervals with an integer number of reference intervals without a remainder. The method allows you to quickly register all frequency changes that occurred in the robot's onboard ideal for a certain time, and provides the following benefits: refinement of the scanner measurement results (accuracy of calculating the angles of dynamic triangulation), reducing the accumulation of errors in determining all the MR parameters that are summed up during the MR stay on the trajectory of movement, and clarification of indications of all onboard sensors of frequency type.
- A new matching criterion without residue certain integer interval with a reference number of the measured time intervals based on the use mediant with the numerator in the format "integer with a certain number of zeros." It allows a fairly simple way to implement frequency measurements at the hardware or software level.
- It is proved that the proposed new method of measuring time intervals and frequency of signals is invariant to jitter and instability of the pulse duration of pulse sequences.
- A method for estimating the energy center of a pulsed signal is proposed – the method of saturation and differentiation – which allows determining the spatial position of the energy center of the illuminated laser spot on the obstacle

surface under conditions of limited time and provides significant advantages of distinctiveness at the edge of the STZ review sector.

- The method of functioning of the STZ laser scanner with a variable scanning step has been improved, which makes it possible to optimize the search for an obstacle in the field of view of the robot by the criteria of the minimum time spent on scanning and the maximum information content. The possibility of using a variable scanning step makes the proposed FCZ an adaptive information gathering tool that, by simply changing commands in the control program, optimizes the scanner's performance, adapting it to the more detailed inspection, or to the most rapid definition of obstacles, according to current conditions.
- A method is proposed for expanding a reliable zone of the field of view of a robot by applying neural network training methods to filter scanning results with large errors based on the use of the Levenberg-Marquardt algorithm, which allows for a real-time increase in the distinctive ability of an TVS. The method of suppressing measurement noise in accelerometers based on the LMS filtering algorithm makes it possible to significantly improve the determination of the parameters of the current self-spatial orientation of the MP.
- A method is proposed for determining the optimal MR trajectory in a given sector of space by the criterion of the minimum path and maximum smoothness of the trajectory, which is based on the use of data obtained from the proposed FCZ, as well as data from the accelerometric definition system of the MR spatial orientation and onboard clock of a robot. Its use minimizes the number of impacts by the steering mechanism, as a result of which the energy costs of the MR and the wear of its mechanical parts are reduced, and the directional stability is improved and the possibility of skids and slippings is reduced.
- Introduced the concept of increasing the robustness of the robot in non-deterministic non-contrast media in the presence of radiation and other types of pollution through the use of technical electronic, electrical, mechanical and computational tools that are resistant to external factors and rational simplification of systems for receiving and processing optical signals and computational algorithms.

References

1. Hoy, M., Matveev, A., & Savkin, A. (2015). Algorithms for collision-free navigation of mobile robots in complex cluttered environments: A survey. Robotica, 33(3), 463-497. doi:https://doi.org/10.1017/S0263574714000289
2. Bonin-Font, F., Ortiz, A. & Oliver, G. Visual Navigation for Mobile Robots: A Survey. Journal of Intelligent Robotic Systems, vol. 53, 263 (2008). https://doi.org/10.1007/s10846-008-9235-4
3. R. Abiyev, D. Ibrahim, B. Erin, Navigation of mobile robots in the presence of obstacles, Advances in Engineering Software, Elsevier, Volume 41, Issues 10–11, 2010, Pages 1179-1186, ISSN 0965-9978, https://doi.org/10.1016/j.advengsoft.2010.08.001

4. B.K. Patle, Ganesh Babu L, Anish Pandey, D.R.K. Parhi, A. Jagadeesh, A review: On path planning strategies for navigation of mobile robot, Defence Technology, Elsevier, Volume 15, Issue 4, 2019, Pages 582-606, ISSN 2214-9147, https://doi.org/10.1016/j.dt.2019.04.011
5. O. Motlagh, S.H. Tang, N. Ismail, A.R. Ramli, An expert fuzzy cognitive map for reactive navigation of mobile robots, Fuzzy Sets and Systems, , Elsevier, Volume 201, 2012, Pages 105-121, ISSN 0165-0114, https://doi.org/10.1016/j.fss.2011.12.013
6. Bouraine, S., Fraichard, T. & Salhi, H. Provably safe navigation for mobile robots with limited field-of-views in dynamic environments. Autonomous Robot, Springer, 32, 267 283 (2012). https://doi.org/10.1007/s10514-011-9258-8
7. M. S. Miah, J. Knoll and K. Hevrdejs, "Intelligent Range-Only Mapping and Navigation for Mobile Robots," in IEEE Transactions on Industrial Informatics, vol. 14, no. 3, pp. 1164-1174, March 2018, doi: https://doi.org/10.1109/TII.2017.2780247
8. M. Hank, M. Haddad, A hybrid approach for autonomous navigation of mobile robots in partially-known environments, Robotics and Autonomous Systems, Volume 86, 2016, Pages 113-127, ISSN 0921-8890, https://doi.org/10.1016/j.robot.2016.09.009
9. Sungjoon Choi, Eunwoo Kim, Kyungjae Lee, Songhwai Oh, Real-time nonparametric reactive navigation of mobile robots in dynamic environments, Robotics and Autonomous Systems, Volume 91, 2017, Pages 11-24, ISSN 0921-8890, https://doi.org/10.1016/j.robot.2016.12.003
10. Dayal R. Parhi, Mukesh Kumar Singh. Navigational strategies of mobile robots: a review. International Journal of Automation and Control, Inderscience, Vol. 3, No. 2/3, 2009, Pages 114–134. https://doi.org/10.1504/IJAAC.2009.025237
11. F. Cuesta, A. Ollero. Intelligent mobile robot navigation. Springer. – 2005, 204 pages. Intelligent Mobile Robot Navigation | Federico Cuesta | Springer
12. Das Sharma K., Chatterjee A., Rakshit A. (2018) Experimental Study II: Vision-Based Navigation of Mobile Robots. In: Intelligent Control. Cognitive Intelligence and Robotics. Springer, Singapore. https://doi.org/10.1007/978-981-13-1298-4_9
13. Chatterjee A., Rakshit A., Singh N.N. (2013) Mobile Robot Navigation. In: Vision Based Autonomous Robot Navigation. Studies in Computational Intelligence, vol 455. Springer, Berlin, Heidelberg. – 222 pages. https://doi.org/10.1007/978-3-642-33965-3_1
14. Mohammed Algabri, Hassan Mathkour, Hedjar Ramdane, Mansour Alsulaiman, Comparative study of soft computing techniques for mobile robot navigation in an unknown environment, Computers in Human Behavior 50, 2015, 42-56, ISSN 0747-5632, https://doi.org/10.1016/j.chb.2015.03.062
15. Gerald Cook, Feitian Zhang. Mobile Robots. Navigation, Control and Sensing, Surface Robots and AUVs. Second Edition, 2020, John Wiley & Sons, Hoboken, New Jersey, 327 pages.
16. Singh Gill, J., Tomaszewski, M., Jia, Y., Pisu, P. et al., "Evaluation of Navigation in Mobile Robots for Long-Term Autonomy in Automotive Manufacturing Environments," SAE Technical Paper 2019-01-0505, 2019, https://doi.org/10.4271/2019-01-0505
17. Tzafestas, S.G. Mobile Robot Control and Navigation: A Global Overview. Springer, Journal of Intelligent Robotic Systems, vol. 91, pages 35–58 (2018). https://doi.org/10.1007/s10846-018-0805-9
18. B. K. Patle, Dayal Parhi, A. Jagadeesh, O. P. Sahu. Real Time Navigation Approach for Mobile Robot. Journal of Computers, Volume 12, Number 2, March 2017, Pages 135 – 142, jcp1202-05.pdf (jcomputers.us)
19. Zelinsky A., Yuta S. (1994) A unified approach to planning, sensing and navigation for mobile robots. In: Yoshikawa T., Miyazaki F. (eds) Experimental Robotics III. Lecture Notes in Control and Information Sciences, vol 200. Springer, Berlin, Heidelberg. https://doi.org/10.1007/BFb0027613
20. Giralt G., Chatila R., Vaisset M. (1990) An Integrated Navigation and Motion Control System for Autonomous Multisensory Mobile Robots. In: Cox I.J., Wilfong G.T. (eds) Autonomous Robot Vehicles. Springer, New York, NY. https://doi.org/10.1007/978-1-4613-8997-2_31
21. Filimonov, A.B., Filimonov, N.B. & Barashkov, A.A. (2019) Construction of Potential Fields for the Local Navigation of Mobile Robots. Optoelectronics, Instrumentation and Data Processing, Springer. Vol. 55, Pages 371–375. https://doi.org/10.3103/S8756699019040071

22. O. Yu. Sergiyenko. Optoelectronic System for Mobile Robot Navigation. Springer/Allerton Press, Inc., Optoelectronics, Instrumentation and Data Processing, Vol. 46, No. 5, October, 2010, pp.414-428
23. L.C. Básaca-Preciado, O. Yu. Sergiyenko, J.C. Rodríguez-Quinonez, X. García, V. Tyrsa, M. Rivas-Lopez, D. Hernandez-Balbuena, P. Mercorelli, M. Podrygalo, A. Gurko, I. Tabakova, O. Starostenko. Optical 3D Laser Measurement System for Navigation of Autonomous Mobile Robot. Optics and Lasers in Engineering by Elsevier, Vol. 54, 2014, pp. 159–169
24. O.Yu. Sergiyenko, M.V. Ivanov, V.V. Tyrsa, V.M. Kartashov, M. Rivas-López, D. Hernández-Balbuena, W. Flores-Fuentes, J.C. Rodríguez-Quiñonez, J.I. Nieto- Hipólito, W. Hernandez, A. Tchernykh. Data transferring model determination in robotic group, Robotics and Autonomous Systems by Elsevier, vol. 83, September 2016, pp. 251-260
25. O. Sergiyenko, W. Flores-Fuentes, J. C. Rodríguez-Quiñónez, M. Rivas-López, L. Lindner. Control theory and signal processing in machine vision for navigation. International Journal of Advanced Robotic Systems, SAGE, 17(4), August 2020: pp. 1–2. Editorial https://doi.org/10.1177/1729881420926470
26. O. Sergiyenko; and V. Tyrsa. 3D Optical Machine Vision Sensors with Intelligent Data Management for robotic swarm navigation improvement. IEEE Sensors Journal, 2020. Print ISSN: 1530-437X, Online ISSN: 1558-1748, https://doi.org/10.1109/JSEN.2020.3007856
27. "Machine Vision and Navigation". Edited by Oleg Sergiyenko, Wendy Flores-Fuentes, and Paolo Mercorelli. Editorial: Springer, November 12, 2019, 851p. ISBN 978-3-030-22586-5
28. M. Ivanov, O. Sergiyenko, V. Tyrsa, L. Lindner, W. Flores-Fuentes, J. C. Rodríguez-Quiñonez, W. Hernandez, and P. Mercorelli. Influence of data clouds fusion from 3D real-time vision system on robotic group dead reckoning in unknown terrain. IEEE/CAA Journal of Automatica Sinica, 7 2 2020, 368-385.
29. Chapter 13: Data Exchange and Task of Navigation for Robotic Group. M. Ivanov, O. Sergiyenko, V. Tyrsa, L. Lindner, M. Reyes García, J. C. Rodríguez-Quiñonez, W. Flores-Fuentes, J. E. Miranda-Vega, M. Rivas-López, D. Hernández-Balbuena/ in Book "Machine Vision and Navigation". Edited by O. Sergiyenko, W. Flores-Fuentes, and P. Mercorelli. Editorial: Springer, November 12, 2019, 851p. ISBN 978-3-030-22586-5
30. G. Capi, S. Kaneko, B. Hua, Neural Network based Guide Robot Navigation: An Evolutionary Approach, Procedia Computer Science by Elsevier, Volume 76, 2015, pp. 74-79, ISSN 1877-0509
31. J.C. Rodriguez-Quiñones, O. Sergiyenko, F. González-Navarro, L. Basaca-Preciado, V. Tyrsa. Surface recognition improvement in 3D medical laser scanner using Levenberg–Marquardt method, Signal Processing by Elsevier 93, 2, 2013, 378–386.
32. J.C. Rodriguez-Quiñonez, O. Sergiyenko, D. Hernandez-Balbuena, M. Rivas-Lopez, W. Flores-Fuentes, L.C. Basaca Preciado. Improve 3D Laser scanner measurements accuracy using a FFBP neural network with Widrow-Hoff weigt/bias learning function. Opto-Electronic Review by Springer, 22, 4, 2014, 224-235, https://doi.org/10.2478/s11772-014-0203-1
33. O. Sergiyenko, V. Tyrsa, A. Zhirabok and A. Zuev, "Sensor Fault Identification in Linear and Nonlinear Dynamic Systems via Sliding Mode Observers," in IEEE Sensors Journal, 2021. https://doi.org/10.1109/JSEN.2021.3080118
34. Sergiyenko, O.Y., Hernandez Balbuena, D.; Tyrsa, V.V.; Rosas Mendez, P.L.A.; Hernandez, W.; Nieto Hipolito, J.I.; Starostenko, O.; Rivas Lopez, M. Automotive FDS Resolution Improvement by Using the Principle of Rational Approximation. IEEE Sensors Journal, Volume: 12, Issue: 5, May, 2012, pp. 1112 - 1121. ISSN 1530-437X
35. O. Sergiyenko. The mediant-method for fast mass/concentration detection for nanotechnologies. Inderscience, International Journal of Nanotechnology, 13, 1-3, 2016, 236-246. ISSN: 1741-8151
36. O. Sergiyenko, D. Hernandez Balbuena, V. Tyrsa, P.L.A. Rosas Méndez, M. Rivas Lopez, W. Hernandez, M. Podrygalo, A. Gurko. Analysis of jitter influence in fast frequency measurements. Elsevier, "Measurement", Volume 44, Issue 7, August 2011, Pages 1229-1242. ISSN: 0263-2241. https://doi.org/10.1016/j.measurement.2011.04.001

37. D. Hernández Balbuena, O. Sergiyenko, V. Tyrsa, L. Burtseva, M. Rivas López. Signal frequency measurement by rational approximations. Elsevier, "Measurement", Volume 42, Issue 1, January 2009, Pages 136-144. ISSN: 0263-2241
38. F.N. Murrieta-Rico,; O. Sergiyenko; Petranovskii, V.; Hernandez-Balbuena, D.; Lars Lindner; Tyrsa, V; Moises Rivas-Lopez; Juan I. Nieto-Hipolito; and V.M.Karthashov. Pulse width influence in fast frequency measurements using rational approximations. Elsevier, "Measurement", Volume 86, May 2016, Pages 67-78. ISSN: 0263-2241. https://doi.org/10.1016/j.measurement.2016.02.032
39. W. Flores-Fuentes, Rivas-Lopez, M.; Sergiyenko, O. ; Rodriguez-Quiñonez, J. ; Hernandez-Balbuena, D. ; Rivera-Castillo, J. Energy Centre Detection in Light Scanning Sensors for Structural Health Monitoring Accuracy Enhancement. IEEE Sensors Journal, 14, 7, 2014, 2355-2361. ISSN: 1530-437X
40. W. Flores-Fuentes, O. Sergiyenko, J.C. Rodriguez-Quiñonez, M. Rivas-López, D. Hernández-Balbuena, L.C. Básaca-Preciado, L. Lindner, F.F. González-Navarro. Optoelectronic scanning system upgrade by energy center localization methods. Springer/Allerton Press, Optoelectronics, Instrumentation and Data Processing, Volume 52, Issue 6, 1 November 2016, pp.592-600. – ISSN 8756-6990
41. X.M. Garcia-Cruz, O. Yu. Sergiyenko, V. Tyrsa, M. Rivas-Lopez, D. Hernandez-Balbuena, J.C. Rodriguez-Quiñonez, L.C. Basaca-Preciado, P. Mercorelli. Optimization of 3D laser scanning speed by use of combined variable step. Optics and Lasers in Engineering by Elsevier, Vol. 54, 2014, pp. 141–151, ISSN 0143-8166 https://doi.org/10.1016/j.optlaseng.2013.08.011
42. L. C. Básaca, J. C. Rodríguez, O. Sergiyenko, V. V. Tyrsa, W. Hernández, J. I. Nieto Hipólito, O. Starostenko. Resolution improvement of Dynamic Triangulation method for 3D Vision System in Robot Navigation task, Proceedings of IEEE-36th Annual Conference of IEEE Industrial Electronics (IECON-2010), Glendale-Phoenix, Arizona, USA, 7-10 November, 2010, pp. 2880-2885. ISBN: 978-1-4244-5226-2/ISSN: 1553-572X
43. O. Real-Moreno, J. C. Rodriguez-Quiñonez, O. Sergiyenko, L. C. Basaca-Preciado, D. Hernandez-Balbuena, M. Rivas-Lopez, W. Flores-Fuentes. Accuracy Improvement in 3D Laser Scanner Based on Dynamic Triangulation For Autonomous Navigation System. Proceedings of IEEE 26th International Symposium on Industrial Electronics (ISIE 2017), Edinburgh, Scotland, UK, June, 19 - 21, 2017. Pages: 1602 – 1608. ISBN: Electronic ISBN: 978-1-5090-1412-5
44. C. Sepulveda-Valdez; O. Sergiyenko; V. Tyrsa; W. Flores-Fuentes ; J. Rodríguez-Quiñonez ; F. N. Murrieta-Rico; J. E. Miranda-Vega ; P. Mercorelli ; M. Kolendovska. Geometric analysis of a laser scanner functioning based on dynamic triangulation, 2020 IEEE 29th International Symposium on Industrial Electronics (ISIE), Delft, Netherlands, 17-19 of June 2020, pp. 1398-1403. https://doi.org/10.1109/ISIE45063.2020.9152268
45. M. Rivas López, O. Sergiyenko, V. Tyrsa, W. Hernandez Perdomo, D. Hernández Balbuena, L. Devia Cruz, L. Burtseva, J. I. Nieto Hipólito. Optoelectronic method for structural health monitoring. SAGE Publications, "Structural Health Monitoring: an International Journal", Vol.9, No.2, March, 2010, pp.105-120. ISSN 1475-9217 /. https://doi.org/10.1177/1475921709340975
46. D. Avalos-Gonzalez, O. Yu. Sergiyenko, D. Hernandez-Balbuena, V. Tyrsa, V. Kartashov, M. Rivas-López, J. C. Rodríguez-Quiñonez, W. Flores-Fuentes, F. N. Murrieta-Rico. Constraints definition and application optimization based on geometric analysis of the frequency measurement method by pulse coincidence. Measurement by Elsevier, 126, 2018, 184-193. ISSN: 0263-2241
47. W. Hernandez, J. de Vicente, O. Sergiyenko, and E. Fernández. Improving the Response of Accelerometers for Automotive Applications by Using LMS Adaptive Filters. MDPI, Sensors, 2010, 10(1), Basel, Switzerland, pp. 313-329; ISSN 1424-8220 /. https://doi.org/10.3390/s100100313
48. M. Rivas López, O. Sergiyenko, V. Tyrsa, W. Hernandez Perdomo, D.Hernández Balbuena, L. Devia Cruz, L. Burtseva, J. I. Nieto Hipólito. Optoelectronic method for structural health monitoring. SAGE Publications, "Structural Health Monitoring: an International Journal", 9, 2, , 2010, 105-120.

49. J. Rivera-Castillo, W. Flores-Fuentes, M. Rivas-López, O. Sergiyenko, F. Gonzalez-Navarro, J. C. Rodríguez-Quiñonez, D. Hernández-Balbuena, L. Lindner, L. C. Básaca-Preciado. Experimental image and range scanner datasets fusion in SHM for displacement detection, Structural Control and Health Monitoring, Wiley, 24, 10, 2016, 1-17. ISSN: 1545-2263.
50. L. Lindner, O. Sergiyenko, J. Rodríguez-Quiñonez, M. Rivas-López, D. Hernández-Balbuena, W. Flores-Fuentes, F. N. Murrieta-Rico, and V. Tyrsa. Mobile robot vision system using continuous laser scanning for industrial application, Industrial Robot by Emerald, 43, 4, 2016 360-369. ISSN 0143-991X
51. Şahin E., Girgin S., Bayindir L., Turgut A.E. (2008) Swarm Robotics. In: Blum C., Merkle D. (eds) Swarm Intelligence. Natural Computing Series. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-74089-6_3
52. Şahin E. (2005) Swarm Robotics: From Sources of Inspiration to Domains of Application. In: Şahin E., Spears W.M. (eds) Swarm Robotics. SR 2004. Lecture Notes in Computer Science, vol 3342. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-30552-1_2
53. Brambilla, M., Ferrante, E., Birattari, M. et al. Swarm robotics: a review from the swarm engineering perspective. Swarm Intelligence by Springer, 7, 2013, 1–41. https://doi.org/10.1007/s11721-012-0075-2

Index

O. Sergiyenko (ed.), *Optoelectronic Devices in Robotic Systems*,
https://doi.org/10.1007/978-3-031-09791-1