

LAURA GONÇALVES RIBEIRO

# Target Tracking Using Optical Markers for Remote Handling in ITER



LAURA GONÇALVES RIBEIRO

Target Tracking Using Optical Markers  
for Remote Handling in ITER

ACADEMIC DISSERTATION

To be presented, with the permission of  
the Faculty of Information Technology and Communication Sciences  
of Tampere University,  
for public discussion in the TB109  
of the Tietotalo building, Korkeakoulunkatu 1, Tampere,  
on 2 June 2023, at 12 o'clock.

# ACADEMIC DISSERTATION

Tampere University,

Faculty of Information Technology and Communication Sciences,

Finland

*Responsible  
supervisor  
and Custos*

Professor Atanas Gotchev  
Tampere University  
Finland

*Supervisor*

University Lecturer, Dr. Sari Peltonen  
Tampere University  
Finland

*Pre-examiners*

Dr. Frederik Zilly  
Bosch  
Germany

Dr. Gregory Dubus  
Fusion for Energy  
Spain

*Opponents*

Professor Sven Lončarić  
University of Zagreb  
Croatia

Dr. Gregory Dubus  
Fusion for Energy  
Spain

The originality of this thesis has been checked using the Turnitin Originality Check service.

Copyright ©2023 Laura Gonçalves Ribeiro

Cover design: Roihu Inc.

ISBN 978-952-03-2881-8 (print)

ISBN 978-952-03-2882-5 (pdf)

ISSN 2489-9860 (print)

ISSN 2490-0028 (pdf)

<http://urn.fi/URN:ISBN:978-952-03-2882-5>



Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino  
Joensuu 2023



# PREFACE

The work described in this dissertation was developed during my employment at Tampere University, formerly Tampere University of Technology, between 2018 and 2022. The dissertation consists of a collection of four papers that have been published during this time. I would like to acknowledge the efforts of the reviewers of these publications and pre-examiners and opponents of the dissertation in revising the manuscripts and providing valuable feedback on the work described herein. I would also like to acknowledge the primary funding source of my studies: European Union's Horizon 2020 research and innovation program under the Marie Skłodowska Curie grant agreement No. 764951, Immersive Visual Technologies for Safety-Critical Applications as well as Fusion for Energy, under grant agreement F4E-GRT-0901, and the Eurofusion consortium.

The dissertation focuses on the development of a vision system for the maintenance of ITER's experimental fusion reactor. ITER is a joint global effort aiming to demonstrate the feasibility of nuclear fusion as a source of energy. I feel privileged to have contributed in a small way to this ambitious endeavor. One of my struggles throughout this process has been with the significance of one's time and efforts. I am grateful for the F4E/ITER projects that helped me find a sense of purpose.

When I am not occupying my mental efforts with the pursuit of purpose, I like to deconstruct and solve problems. I am grateful to have been able to make a living out of that during this time. I owe this foremost to my supervisor, Professor Atanas Gotchev, who hired me and brought together everything else that is mentioned here. Prof. Gotchev has provided me with a safe environment to exercise my interests and autonomy and self-develop. Throughout my employment at Tampere University, he has strived to provide me with as stable as possible work at the cost of his own personal time and efforts. I am not exactly sure how he manages such a large set of people and endeavors but it has certainly been a privilege to be around to see it happen. Professor Robert Bregović and MsC Olli Suominen have collaborated in

these efforts, and I would like to acknowledge their remarkable skills in managing projects, fundings, people and expectations.

I would like to specially acknowledge Olli's role in this whole venture. As demonstrated in the upcoming sections he has been my closest collaborator. Over the course of various quests, I have always felt that you were in my team. Thank you for taking, over these years, some of your time to guide and support me. I hope we keep finding ways to positively contribute to each other's lives and work.

I would like to acknowledge as well the roles of: Dr. Sari Peltonen for the prompt and pedagogically correct feedback; Ahmed, for all practical things; Sarianne for being a really good office mate; and all other peers at work that enriched my life with their different ways of being and doing.

I would like to acknowledge the dear people outside of work that supported me throughout this process in various ways. Thank you for sharing with me your own struggles and frustrations and accompanying me through my own troubles and joys. Particularly, but not exclusively, I am grateful for: the hosts and participants of Portuguese food events, the helpers of my 7 in 5 years apartment moves, second-hand furniture pursuits and automobile break downs, the late-night pick-me-ups at various airports and the companions of nature adventures, wisdom walks, sauna and winter swimmings. I am grateful for those that have humored me by allowing me to practice my Finnish language skills with them and that have fostered my hopeless integration within Finnish culture. A thanks to the BJJ gyms and training partners that contributed to my physical and mental well being throughout this process.

At the root of everything, I would like to thank my family, particularly the three resilient women that raised me, for striving to provide me an easier start in life than what they could have. And my dear cousin and aunt, that emotionally support me.

I have found that even when one is quite fortunate from the start, no life path comes without a certain degree of pain and I am endlessly grateful to those that facilitated the smaller and the bigger struggles and made the various ventures in this phase of life a little less scary.

Over the years I have rationalized and repeated countless times to myself and others that any venture that is progressing at a speed (even very slightly) higher than zero eventually comes to an outcome. This thesis will stand as, at least, proof of that!

Laura Gonçalves Ribeiro,  
November 2022.

# ABSTRACT

The thesis focuses on the development of a vision system to be used in the remote handling systems of the International Thermonuclear Experimental Reactor - ITER. It presents and discusses a realistic solution to estimate the pose of key operational targets, while taking into account the specific needs and restrictions of the application.

The contributions to the state of the art are in two main fronts: 1) the development of optical markers that can withstand the extreme conditions in the environment; 2) the development of a robust marker detection and identification framework that can be effectively applied to different use cases. The markers' locations and labels are used in computing the pose.

In the first part of the work, a retro reflective marker made up ITER compliant materials, particularly, fused silica and stainless steel, is designed. A methodology is proposed to optimize the markers' performance. Highly distinguishable markers are manufactured and tested.

In the second part, a hybrid pipeline is proposed that detects uncoded markers in low resolution images using classical methods and identifies them using a machine learning approach. It is demonstrated that the proposed methodology effectively generalizes to different marker constellations and can successfully detect both retro reflective markers and laser engravings.

Lastly, a methodology is developed to evaluate the end-to-end accuracy of the proposed solution using the feedback provided by an industrial robotic arm. Results are evaluated in a realistic test setup for two significantly different use cases.

Results show that marker based tracking is a viable solution for the problem at hand and can provide superior performance to the earlier stereo matching based approaches. The developed solutions could be applied to other use cases and applications.



# CONTENTS

1	Introduction . . . . .	17
1.1	Significance of the Research Topic . . . . .	17
1.2	Thesis Objectives and Research Questions . . . . .	20
1.3	Thesis Structure and Related Publications . . . . .	21
2	Visual Sensing Technologies in ITER's Remote Handling Systems . . . . .	23
2.1	The Divertor in the Context of the ITER Machine . . . . .	23
2.2	Visual Sensing Technologies at ITER . . . . .	26
3	Design of ITER Compliant Retro Reflective Markers . . . . .	31
3.1	Definitions and Background. . . . .	31
3.2	Requirements for a Retro Reflector at ITER . . . . .	33
3.3	ITER Compliant Retro Reflector. . . . .	34
3.4	Optimization of Retro Reflector Performance . . . . .	35
3.5	Performance Evaluation . . . . .	39
3.6	Summary . . . . .	41
4	A Unified Framework for Marker Based Pose Estimation at ITER . . . . .	43
4.1	Definitions and Background. . . . .	43
4.2	Marker Detection and Identification Pipeline. . . . .	47
4.3	Performance . . . . .	51
4.4	Pose Estimation . . . . .	53
4.5	Summary . . . . .	56
5	End-to-end Evaluation of Pose Estimation Accuracy . . . . .	57
5.1	Definitions. . . . .	57
5.2	Experimental Setup . . . . .	59
5.3	Reference Value Estimation . . . . .	59
5.4	Results . . . . .	60

5.5 Summary . . . . . 62

6 Discussion and Conclusions. . . . . 63

References . . . . . 67

## List of Figures

1.1	Schematic representation of the research problem - the need for a system that reliably extracts mission critical information from images of a safety critical environment to enhance the perception of a remote operator. . . . .	19
2.1	Cross section of model of the ITER machine. Highlighted are the blanket and divertor structures. Figure adapted from <i>iter.org</i> . . . . .	24
2.2	Model of a divertor cassette assembly (CA#4). Highlighted are the knuckle, pipe flanges and bridging link structures. Figure adapted from <i>iter.org</i> . . . . .	25
3.1	Illustration of the concept of diffuse (left), specular (middle) and retro (right) reflection. . . . .	31
3.2	Visual appearance of standard RRM and laser engravings on metallic pieces. . . . .	32
3.3	Plan view of design constraints. We aim at optimizing the number of rays arriving at a camera lens with diameter $D_{\text{lens}}$ , which is displaced from the light source by a fixed distance $\Delta h$ . The working distance is variable within a fixed range of values set by the application. . . . .	34
3.4	Retro reflected light distribution of a fused silica glass bead (left side), narrower distribution produced by a type of glass with higher refractive index (upper right corner) and by manipulating the distance between the lens and the reflective surface. . . . .	36
3.5	Concept of RR design alternative based on an array of ball lenses, with radius $R_l$ and a corresponding pattern of spherical reflective surfaces, with radius $R_m$ and separated by distance $d$ . . . . .	36
3.6	Light ray distributions for different RR marker configurations. The distance to the reflective surface ( $d$ ) is varied in relation to the focus distance ( $d_f$ ). The spheres have a radius of 0.2 mm. . . . .	38

3.7	Behaviour of RR marker configurations with varying distance to the reflective surface. The curvature radius and aperture of the primary lens are constant with $R_l=1.0$ and $a=2.0$ . The curvature radius of the reflective surface, $R_m$ , is changed accordingly, to fulfill the condition of retro reflection. The data is obtained using a commercial sequential raytracing software (OSLO by Lambdares [38]). . . . .	39
3.8	Images of manufactured ITER grade RRM. On the left hand side the developed marker, housed inside it's stainless steel housing. On the right hand side the marker is shown mounted in one of the operational pieces. . . . .	40
3.9	Experimental setup for RRM testing (not to scale). Samples are rotated around centre point O by angle $\theta$ . Camera and LED sources are stationary throughout the experiments. . . . .	40
3.10	Experimental evaluation of average intensity of all pixels belonging to the marker, in raw image values. . . . .	41
3.11	Diagram of full marker assembly. 1- Reflective Base. 2-Spacer Disk. 3- Spacer Ring. 4- Body. 5- Guiding Pins. 6-Spherical Glass Beads. 7-Glass Window. . . . .	42
4.1	Example target (bridging link in vessel connector): the functional structures in the central area restrict the placement of markers to a relatively small border area. . . . .	47
4.2	Proposed marker detection and recognition pipeline. . . . .	48
4.3	Example of successful detection and identification - knuckle use case . .	52
4.4	Example of successful detection and identification - pipe flange use case.	52
4.5	Example of successful detection and identification - bridging link (bottom) use case. . . . .	53
5.1	Relevant coordinate frames and rigid body transformations. . . . .	58
5.2	Experimental setup with manipulator, imager and light source on the left-hand side. . . . .	59
5.3	Alignment piece composed of a transparent polymer window engraved with an alignment axes. . . . .	60



## List of Tables

4.1	Appearance of relevant fiducial marker systems. . . . .	44
4.2	Example output matrix. Note that each matrix column refers to a separate network output. The matrix concatenates in its rows the outputs of different markers detected in the same image. Below the matrix are the chosen classes according to ‘maximum value’ strategy and our proposed approach. On the right hand side we show the joint probabilities of some example marker permutations. . . . .	51
4.3	Robustness measure of marker detection and identification at consecutive pipeline stages - knuckle (I), pipe flange (II) and bridging link (III) use cases. . . . .	54
4.4	Robustness measure of correspondence step with respect to occlusion of specific markers - knuckle (I) and pipe flange (II) and bridging link (III) use cases. . . . .	55
5.1	Average and Worst Case Accuracy and Precision of Estimated Translations. . . . .	61
5.2	Average and Worst Case Accuracy and Precision of Estimated Rotations. . . . .	62



# ABBREVIATIONS

CCC	Concentric Contrasting Circle
CS	Control System
DOF	Degree of Freedom
ICP	Iterative Closest Point
ITER	International Thermonuclear Experimental Reactor
JET	Joint European Torus
LED	Light Emitting Diod
LIDAR	Light Imaging Detection and Ranging
MSER	Maximally Stable Extremal Regions
RH	Remote Handling
RR	Retro Reflector
RRM	Retro Reflective Marker
TCP	Tool Centre Point
TOF	Time of Flight



# LIST OF PUBLICATIONS

- Publication I      L. G. Ribeiro, O. J. Suominen, S. Peltonen, E. R. Morales, and A. Gotchev, “Robust vision using retro reflective markers for remote handling in ITER,” *Fusion Engineering and Design*, vol. 161, p. 112 080, 2020. DOI: 10.1016/j.fusengdes.2020.112080.
- Publication II     L. G. Ribeiro, O. J. Suominen, A. Durmush, S. Peltonen, E. Ruiz Morales, and A. Gotchev, “Retro-reflective-marker-aided target pose estimation in a safety-critical environment,” *Applied Sciences*, vol. 11, no. 1, p. 3, 2020. DOI: 10.3390/app11010003.
- Publication III    L. G. Ribeiro, O. J. Suominen, P. Bates, S. Peltonen, E. R. Morales, and A. Gotchev, “Design and manufacturing of an optimized retro reflective marker for photogrammetric pose estimation in ITER,” *Fusion Engineering and Design*, vol. 191, p. 113 557, 2023. DOI: 10.1016/j.fusengdes.2023.113557.
- Publication IV    L. G. Ribeiro, O. J. Suominen, S. Peltonen, E. R. Morales, and A. Gotchev, “ITER-Tag: An adaptable framework for robust uncoded fiducial marker detection and identification,” in *2022 10th European Workshop on Visual Information Processing (EU-VIP)*, 2022. DOI: 10.1109/EUVIP53989.2022.9922671.

## *Author's contribution*

The author of the thesis is the first author of all the publications and was responsible for conceptualization, implementation, experimental evaluation and reporting of the main methods and writing of the manuscripts. In Publication I and Publication II, Olli J. Suominen and Ahmed Durmush worked together with the main author on the early RRM design and prototype. Ahmed Durmush was responsible for the physical

apparatus for the initial experimental setup. In Publication III, Olli J. Suominen worked together with the main author on the methodology for the manufacturing and testing of RRM's and in Publication IV, on the experimental setup and calibration procedures. In Publications II and IV, Olli J. Suominen generated the synthetic image data. Olli J. Suominen, Sari Peltonen, Emilio Ruiz Morales and Atanas Gotchev contributed with supervision, writing, reviewing and editing of manuscripts.

# 1 INTRODUCTION

## 1.1 Significance of the Research Topic

### ITER in the context of the global efforts towards fusion energy

In the quest to find a suitable primary world's source of energy, nuclear fusion remains a very attractive option towards which considerable global efforts have been applied [1].

One of the main challenges of fusion as a source of energy is the extremely high amount of energy required to start the reaction. Joint European Torus (JET)'s experiments have achieved world record fusion performance in a major technical breakthrough, having, however, consumed more energy than could be produced. Following up the efforts of JET, the International Thermonuclear Experimental Reactor (ITER) aims to demonstrate the viability of fusion as a source of energy by building, operating and dismantling a site where fusion is maintained for longer periods of time with a net energy gain. This energy will, however, not be harnessed. Beyond ITER lies the demonstration electricity-producing reactor, called DEMO, which is currently being designed.

ITER is, therefore, part of a series of ambitious global research undertakings. It poses a wide set of technical challenges and requires innovation and research in multiples areas. The technical solutions found for ITER might pave the way to address the technical challenges in DEMO and in the power plants of the future.

### Remote Handling needs and challenges

Inspection, maintenance and decommissioning are some of the key technical challenges in ITER. Nuclear fusion reactions produce byproducts that cause radiation damage and radioactivity of components in the environment. Throughout the life-

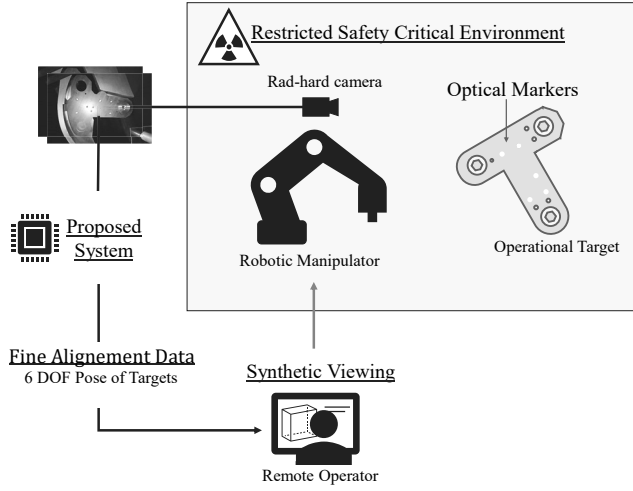
time of the reactor, damaged and activated components have to be inspected and changed or repaired as needed [2]. As operator access to activated areas is restricted due to the risk of ionizing radiation exposure, maintenance is required to be performed through remote handling (RH). ITER's RH follows a human-in-the-loop approach where the operator makes key decisions based on the available information of the remote environment. In order to take appropriate actions, the operator needs access to accurate and reliable information of the current status of the system. This is particularly important when navigating tight spaces and in operations requiring alignment of tools to targets with tight tolerances. ITER's infrastructure and RH systems are quite large and complex when compared to most applications. The existence of a large amount of moving parts, together with the fact that radiation and temperature might cause changes to the parts leads to the possibility that kinematic models do not depict with accuracy the relative poses of manipulators and targets.

To address this challenge, robotic manipulators will integrate sensors to gather updated information on the current status of the environment. However, ITER's hostile conditions limit the type of sensors that can be used and the quality of the visual data that can be gathered. For instance, radiation tolerant cameras usually provide low resolution images [3] and might suffer from radiation induced noise. Further, ITER's operational scenarios often provide poor visibility due to the narrow spaces, cluttered scenes and poor lighting and viewing angles [4].

Manual alignment of parts based on video streams of the remote environment poses significant challenges, both due to the poor viewing conditions and to how the human visual system perceives depth. Some of the natural depth cues cannot be provided by a single video stream. In this context, there is a need for *reliable, robust computational methods that can extract and refine mission-critical information from challenging images, in particular the 6 degree of freedom (DOF) pose of targets*. This information can be used to improve the operator's perception of the remote environment, as shown in Figure 1.1.

*Synthetic viewing* [5] will likely be the method of choice in presenting the acquired information to the operator. In this visualization strategy, live sensing data is used to update the existing models of the environment. Computer graphics tools are used to generate synthetic views of the environment. These views are often clearer than captured data in terms of contrast and image quality and allow the operator to explore advantageous viewpoints, where cameras could not be originally located.





**Figure 1.1** Schematic representation of the research problem - the need for a system that reliably extracts mission critical information from images of a safety critical environment to enhance the perception of a remote operator.

### 3D sensing and processing as an RH aid

In earlier literature, 3D Node has been proposed as a vision system to estimate the pose of a specific target at ITER [6], [7]. This system requires the use of two cameras that are rigidly attached to each other and calibrated. The output is used to estimate the 3D structure of the scene by establishing point correspondences across the two views. The recovered 3D point cloud is matched to the known 3D model of the target. This approach successfully recuperates the pose in advantageous viewpoints and light conditions. However, it is not as robust as required.

The main issue we identify in applying stereo matching approaches to the problem at hand is that most targets in ITER have large, flat, textureless surfaces, with large specular reflective areas. These surfaces break the assumption that image points are unique and look similar from different perspectives. For this reason, they pose a significant challenge in establishing the robust point correspondences needed for a good 3D reconstruction.

## Taking full advantage of the problem's constraints

The stereo matching-based approach does not exploit the knowledge of the geometry of targets when performing 3D reconstruction of the scene. Further, it does not exploit the application's human-made environment, where components can be, to a certain degree, designed taking into account their RH capabilities. Our hypothesis is that reliability can be improved by introducing *optical markers* in the targets' surfaces that can be detected more unambiguously and accurately and used to directly compute the pose. Retro reflectors (RRs) can be used as particularly robustly detectable marks that are easy to discriminate in considerably challenging visibility conditions. The main challenge we anticipate with this approach is achieving highly detectable marks that are compatible with the environmental and material constraints at ITER. Often RRs are made of polymers and adhesive materials that cannot withstand the temperature and radiation in most areas of the ITER environment. An adequate marker should be made of more resilient materials.

The marker based approach is highly desirable as it could be straightforwardly generalized to different operational scenarios due to the lower dependence on the visual characteristics of the targets. Further, marker based tracking would allow the simplification of the system to a single camera, which is one of the main objectives of further developments due to space constraints in most operational scenarios.

## 1.2 Thesis Objectives and Research Questions

In the aforementioned context, the main objective of the thesis is the development of a novel vision-based solution to estimate the 6 DOF pose of key pieces in ITER's divertor RH system. The solution should be accurate, robust and general enough to be feasibly applied to locating different operational targets. The work addresses, specifically, the following research questions:

**RQ1:** Can we devise a retro reflective marker (RRM) design that complies with the strict restrictions set by ITER divertor's environment and what methodology can be followed to optimize its performance?

**RQ2:** How can we build a framework that successfully detects and identifies both retro reflective and diffuse markers in different use cases in ITER's RH systems?

**RQ3:** How can we effectively evaluate the absolute end-to-end accuracy of the proposed solutions?

### 1.3 Thesis Structure and Related Publications

The structure of the thesis is as follows: Chapter 2 contextualizes the work within existing research. The thesis's contributions to the state of the art are presented in Chapters 3, 4 and 5. Chapter 3 focuses on the development of radiation tolerant RRM. An optimized RRM is proposed, manufactured and tested. Chapter 4 focuses on the development of computational methods to detect markers in images and estimate the pose. A detection pipeline is proposed and tested in relevant use cases. In Chapter 5, the experimental setup and methods to evaluate the overall accuracy of the proposed system are described. In Chapter 6, we discuss the results, draw conclusions and outline future development prospects.

In Publication I, we present the preliminary results of our proposed methodology using RRM to estimate the pose of the knuckle of the cassette locking system at ITER. In Publication II, we extend the previous work with a detailed description of the system from a computer vision perspective and an extensive study of the performance of the system. In Publication III, we propose a methodology for optimizing the performance of retro reflective optical markers while addressing ITER's environment constraints. In Publication IV, we propose a general framework for marker detection and identification that can be applied to different use cases. We evaluate the performance of the system from a robustness point of view and analyse three considerably different operational use cases.



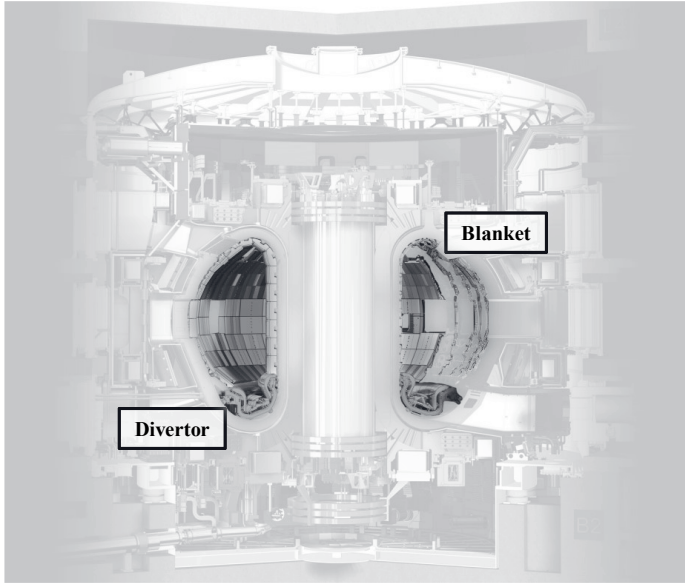
## 2 VISUAL SENSING TECHNOLOGIES IN ITER'S REMOTE HANDLING SYSTEMS

The thesis proposes a vision system to be used in three relevant use cases in ITER's divertor RH systems. In this chapter, we start by briefly contextualizing these use cases and the divertor within the ITER machine and its functionalities. More detailed overviews on ITER's RH systems can be found in the literature [8], [9]. In the second section, we discuss the usability of different sensing technologies in the divertor environment. We overview the vision solutions that have been developed for other RH systems at ITER. ITER's RH systems are characterized by different restrictions, requirements and vision tasks, as well as some commonalities. In analysing the overall picture we aim at proposing solutions that are generalizable to a wide set of use cases and systems.

### 2.1 The Divertor in the Context of the ITER Machine

One of the core challenges in implementing nuclear fusion is the confinement of the ultra hot reacting plasma. ITER aims at demonstrating the resolution of this challenge using a strong magnetic field that manipulates the plasma. This type of reactor design is referred to as tokamak. ITER's tokamak has, at its core, a vacuum vessel where the fusion reaction occurs. Figure 2.1 shows a model of the ITER machine. The vacuum vessel is the torus with "D" shaped cross-section at the centre. Covering the inner walls of the vacuum vessel and shielding the remaining structures from the high-energy neutrons and high thermal loads is the *blanket*. Below the vacuum vessel is the *divertor* that extracts unwanted particles generated during the reaction that would, otherwise, dilute its fuel.

The *in vessel viewing system* [10] is the vision system that is closest to the core of ITER and the most challenging in terms of environment conditions. It surveys the status of the blanket and divertor plasma facing components using dedicated sensors



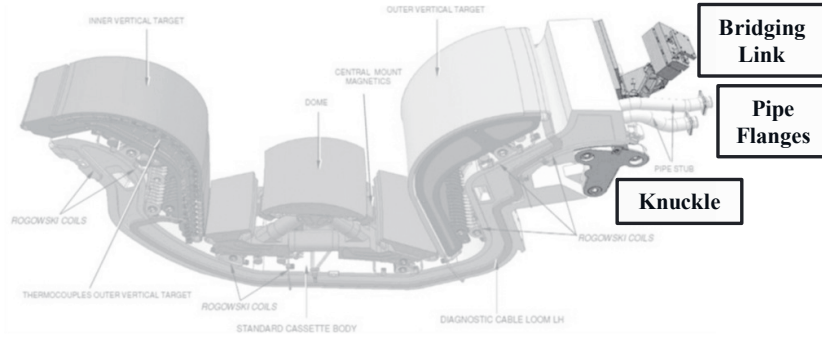
**Figure 2.1** Cross section of model of the ITER machine. Highlighted are the blanket and divertor structures. Figure adapted from *iter.org*.

that have been developed to address specifically this challenging case.

Due to the expected erosion of the *blanket* and the *divertor*, their maintenance is foreseen to be repeatedly needed in ITER's lifetime. The replacement of these large scale components is done in parts. The blanket is comprised of 440 modules each weighting around 5 tons and the knuckle is constituted of 54 cassette assemblies each weighting around 10 tons. In both cases installation and removal of the heavy constituting parts requires that a remotely controlled manipulator performs certain actions with millimetric tolerances. These actions include locking and unlocking, connecting and disconnecting of specific mechanisms, as well as, pipe cutting and welding. The need for operations on pipes comes from the need for circulating water to cool the blanket and the divertor. Circulating water would be, in future projects, an essential part of the energy production strategy.

Throughout the thesis we will address specifically the estimation of the pose of three specific targets in the divertor RH system:

- The *knuckle*, which is part of the cassette locking mechanism [11] and needs to be operated using a wrench tool, a pin tool, and a jack to fixate or release the cassettes [12].



**Figure 2.2** Model of a divertor cassette assembly (CA#4). Highlighted are the knuckle, pipe flanges and bridging link structures. Figure adapted from *iter.org*.

- The *pipe flange*, a structure that needs to be clamped by dedicated tools in order to cut and re-weld the divertor cassettes cooling pipes.
- The *connector between the bridging link and in-vessel wall*. Some of the divertor cassettes host diagnostics systems requiring cables to route electrical and optical signals from the divertor to the diagnostics hall. In replacing the diagnostic cassettes there is a need to operate the *connector between the bridging link and in-vessel wall* [13].

Figure 2.2 shows a model of one of the divertor cassette assemblies, where the abovementioned structures have been highlighted.

The manipulator that conducts these tasks is attached to and brought in during maintenance periods by a dedicated transporter that carries also the replaced parts. This transporter exits the vessel through dedicated ports where it interacts with a transfer cask system.

The *transfer cask system* transports the heavy activated loads between the tokamak and the hot cell building, where they can be refurbished or disposed as radioactive waste material. The transfer cask should follow suitable trajectories that maximize distance to other objects and should dock to the dedicated port plugs. Navigation and docking constitute two different handling and vision tasks with specific requirements. Navigation has less strict positioning tolerances than docking due to larger margins.

While the primary targeted method for navigation is path following of predefined trajectories [14], [15], there is a need for verification and backup systems. Earlier work has proposed a secondary cask localization system based on a well established fiducial marker library and detection strategy [16], [17]. This approach requires at-

taching markers to the casks and would make use of the supervision cameras that may be positioned around the main building's galleries. The authors consider, specifically, glued or painted markers. This strategy is not directly applicable to use cases with stricter constraints, where the use of polymer based materials is restricted. Laser engraving of markers could be considered. However, the markers in this library are made up of relatively fine details and their detection relies on sharp corners and edges of highly contrasting areas. High contrasts might be difficult to achieve with engraving techniques. The docking task essentially consist on locating a target, the port plug, with respect to the cask.

Besides the RH systems we have discussed thus far, ITER is planned to incorporate the *neutral beam cell RH system* and the *multi-purpose deployment system*.

The neutral beam cell functionality is related with the heating of the plasma. Routine maintenance is not foreseen. However, remote replacement of components might be needed due to failure. Expected vision tasks include the localization of specific targets with respect to a lifting crane and of specific regions where payloads should be dropped. The environment restrictions are less strict than in the blanket or divertor.

The *multi purpose system* [18] consists of a mobile platform equipped with various sensors and manipulators. It is designed for support in unexpected scenarios where RH fails. For the sake of redundancy, it will likely integrate systems that differ from the primary vision solutions. As it is built to operate in abnormal scenarios and handle unstructured scenes it is furthest from the subject that we address in the thesis.

## 2.2 Visual Sensing Technologies at ITER

In accurately determining the 6 DOF pose of operational targets, understanding the 3D structure of the environment is of utmost importance. While video cameras do not readily provide the depth of imaged targets, advanced methods have been developed either to directly or indirectly sense it or compute it. Novel technologies keep arising in this area and pushing forward performance limitations. However, ITER's nuclear environment poses a set of unique challenges that are not often addressed in commercial development. High radiation levels are the main limiting factor and considerations of how sensing technologies survive and perform in high temperature



and radiation levels at ITER are essential.

Visual sensing technologies can be classified as active or passive, depending on whether their working principle requires dedicated rather than ambient lighting to image a scene. Active sensors based on the time of flight (TOF) principle resolve distance based on the time the emitted light signal travels between the detector and the scene [19]. A Light Imaging Detection and Ranging (LIDAR) sensor uses a laser to illuminate the scene. Generally sensors that measure the TOF are relatively complex, requiring, for instance, active light sources, scanning mechanisms, precise timing circuitry, etc. They are technically more challenging to achieve as radiation tolerant devices. We are not aware of existing radiation tolerant sensing technologies based on the TOF principle, even though there are ongoing works on the development of radiation resistant LIDAR parts [20].

A dedicated LIDAR [21] has been developed to address the challenging case of the in-vessel viewing system. The design achieves the targeted radiation resistance by locating its active components, such as the laser and the photo-detectors outside of the bio shield. The probe which effectively accesses the in-vessel environment contains only passive optical components and the scanning head. It is made from materials that have been tested to withstand the environment, namely, stainless-steel and fused silica. The acquired signals are routed through radiation tolerant optical fibre. As it is built for static inspections due to the practical constraints that routing the signal through the optical fibre requires, it is not applicable to our targeted use cases.

An earlier work [22] has analysed the use of several commercial active sensing technologies in ITER: a structured light based sensor, a scanning LIDAR and a radar. Structured light is an active sensing method that computes depth based on the distortion of a projected light pattern, that should differ from environmental light and other elements in the scene. The scenes at ITER are considerably different from natural scenes and specific challenges in using structured light might arise. Particularly, in how the pattern is reflected by the large specular surfaces. The authors highlight the fact that none of the tested sensors would survive a large radiation dose. As radiation tolerant versions of these sensors are currently not available, these solutions are not applicable to our targeted use cases.

The most mature and inexpensive radiation resistant sensing technology is based on passive video cameras. Recent work has focused on developing and testing radia-

tion hardened digital CMOS cameras that can withstand a total ionizing dose up to 10 MGy [22], [23]. Cameras are at the moment the most viable option, particularly in use cases with more strict environmental restrictions, such as the divertor.

In using a video camera, depth can be computed by establishing point correspondences across multiple views. Multiple views can be captured using multiple cameras or a moving camera. However, establishing robust point correspondences across multiple views might be challenging and computationally expensive. This is because the visual appearance of image points might not be unique. Advanced techniques and complicated operations might be required to figure out that similarly looking points might not be a correct match. In ITER, the characteristics of the scenes, such as the large textureless surfaces and specular surfaces pose additional challenges to finding strong point correspondences.

Marker based pose estimation circumvents establishing point correspondences in unstructured scenes by introducing markers on specific targets that can be tracked unambiguously. Markers might be passive elements (such as diffuse or retro-reflective marks) or active emitters, such as Light Emitting Diodes (LEDs) [24]. Given a marker location in a set of 2D images, it is possible to determine the world position of the mark by triangulation. Alternatively, if the relative positions of a set of marks are known and their correspondence to 2D image points is established, the 6 DOF pose of the target can be computed from a single image. This type of approach has been used in wide range of applications [24]–[27] and can be particularly advantageous when targets would otherwise be uniform or low in contrast and not provide enough identifiable features [24], [28].

For the reasons presented above, the thesis focuses on marker based pose estimation. The first use case that the thesis considers is the *knuckle of the cassette locking system*. The localization of this specific target has been addressed in earlier literature [6], [7]. These earlier works used a stereo matching approach to estimate the geometry of the remote environment as acquired by the cameras. The approach attempted to circumvent the challenges in establishing robust point correspondences in textureless surfaces by extracting edge information from images as a prior step to stereo matching. The transformation between the estimated point cloud and reference model was calculated using an iterative closest point (ICP) algorithm. The results were overall not robust enough. For some poses, the algorithm was unable to provide a reliable estimate. After outlier removal, the repeatability of the system

was reported to be in the  $\leq 3$  mm error range for operating distances of 500 to 1000 mm. The authors do not report an estimate of the absolute accuracy of the system.

The aforementioned approach does not exploit all the constraints provided by ITER's human-made environment and predictable tasks. We hypothesize that more reliable results can be achieved by detecting the known geometry, or, even better, through the introduction of marker objects in the cassettes' surfaces that can be tracked by cameras unambiguously and accurately. Further, the marker based approach would allow a straightforward simplification of the system to a single camera, which is highly desirable due to the space constraints.

In the last part of our work we apply the similar principle to other use cases in ITER's divertor: the localization of the pipe flange and the bridging link. We are not aware of any earlier work that has addressed specifically these use cases.

The marker based approach has applicability to other use cases and remote handling systems in ITER. Particularly in the neutral beam cell and hot cell RH systems, where the environmental restrictions are less strict. It could be straight forwardly applied to the transfer cask docking task. Marker based tracking might be particularly interesting in the blanket RH systems, where the environment is so harsh that it might cause changes to the appearance of surfaces. Finding suitable marks that could survive that environment is an ongoing challenge.

Overall, marker based pose estimation is a very attractive option for pose estimation in the ITER application, particularly for predicted tasks, where known targets need to be recurrently located. There is a definite benefit in the development of a unified framework for marker placement, detection and pose estimation for the RH systems of ITER. This thesis aims to constitute a consistent contribution towards that objective.

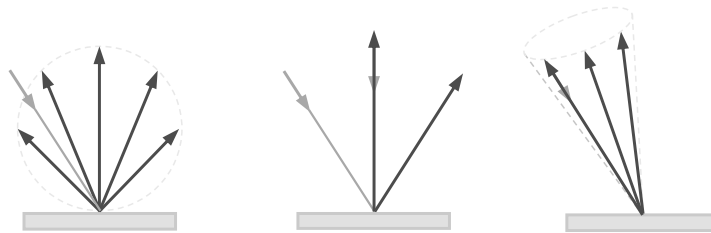


### 3 DESIGN OF ITER COMPLIANT RETRO REFLECTIVE MARKERS

In this chapter, we address RQ1: Can we devise an RRM design that complies with the strict restrictions set by ITER divertor's environment and what methodology can be followed to optimize its performance? We define the functional requirements and restrictions for a retro reflector to act as a marker in the targeted application. We propose a methodology to design a suitable RRM that is easily detectable and fulfills the environmental restrictions. Lastly we evaluate the performance of the proposed markers.

#### 3.1 Definitions and Background

A *retro reflector* is an optical element that reflects incident light in a cone around its source, as depicted in Figure 3.1 (right side), for a wide range of incident directions. This behaviour differs from diffuse reflection, where incident light is reflected in a wide range of directions, as shown in Figure 3.1 (left side). Retro reflection makes it possible to achieve a considerably higher percentage of incident light returned at or near the light source than a diffuse reflector. The behaviour also differs from specular reflection, as shown in Figure 3.1 (middle), where light is only returned near the source for an incidence angle of zero.



**Figure 3.1** Illustration of the concept of diffuse (left), specular (middle) and retro (right) reflection.

RRs are used in a wide range of applications, from traffic safety and road markings to distance measurements in space [29], industrial [30], and robotics applications [31].

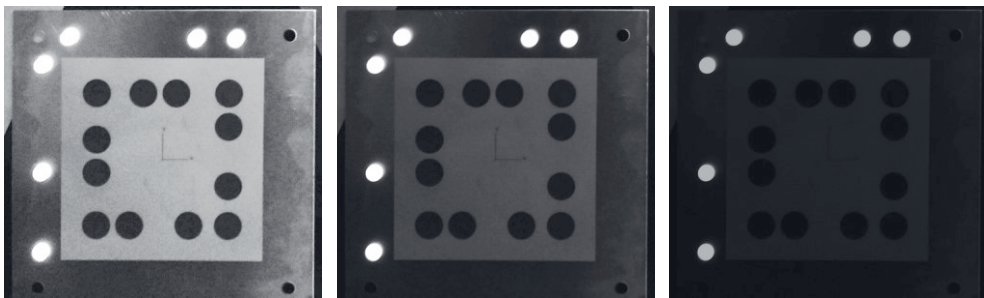
RRMs are used in photogrammetry to provide easily trackable, high contrasting features [24], [28]. They provide considerably higher contrast than naturally occurring structures and diffuse marks [32]. Segmentation of RRM in images can be largely facilitated by controlling image exposure and under-exposing and eliminating other structures in the environment [33]. This method is effective in image noise suppression [32]. Figure 3.2 shows the appearance of standard RRM and engravings on metal pieces. It is visible how RRM can be used to facilitate segmentation by underexposing the background.

These characteristics deem RRM quite advantageous for the ITER application, where poor imaging conditions, including poor signal to noise ratios of the cameras are to be expected.

RRs are characterized by the following main properties [34]:

- *Brilliancy*: reflectance for a specific observation angle,
- *Divergence*: maximum angular distribution of retro reflected light,
- *Angularity*: brilliancy as a function of the incident angle.

The ideal RR allows large entrance angles (high angularity) and returns as much light as possible in as small a cone of light as possible (high brilliancy and low divergence). However, the ideal retro reflective behaviour is not desirable in all applications. Particularly, in applications such as photogrammetry where the light source and detector cannot usually be co-located, the optimal retro reflective pattern would cover the useful retro reflecting angles, without overly exceeding them.



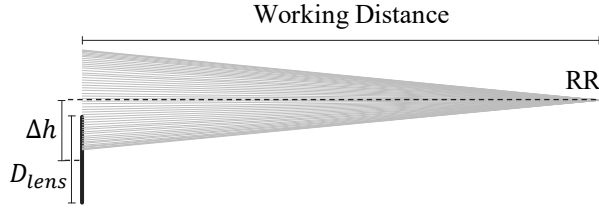
**Figure 3.2** Visual appearance of standard RRM and laser engravings on metallic pieces.

The most common ways of implementing retro reflection is either through a corner cube or a lens and mirror design. Corner-cube RRs rely on the reflection of light on three mutually orthogonal surfaces. They can be manufactured either as a truncated cube corner of a transparent material (i.e., a prism) or as a hollow structure. Cataphotes or cat's eye retro-reflectors are made-up of a primary lens with a secondary mirror located at its focus [34]. These are very commonly implemented using a sphere of a transparent material, embedded in or coated with a reflective material. Spherical retro-reflectors generally have higher divergence due to spherical aberration and lower retro-reflectivity. On the other hand, corner-cube RRs usually have higher retro-reflectivity, while their fabrication is potentially more challenging [34], [35]. Due to their characteristics, availability, and manufacturing considerations, glass bead retro reflectors are the most commonly used type in photogrammetry [32].

### 3.2 Requirements for a Retro Reflector at ITER

The retro reflector for use as a marker for pose estimation at ITER has the following functional requirements:

- *Divergence*: the retro reflected light pattern should cover without overly exceeding the viewing angles as to optimize the amount of light arriving at the surface of the camera lens. The viewing angles can be easily quantified for a specific setup as a function of the working distance, the displacement between the camera and the light source, denoted by  $\Delta h$ , and the diameter of the camera lens, denoted by  $D_{\text{lens}}$ . A visual representation of these measures is shown in Figure 3.3. In ITER divertor's remote handling use cases the working space is severely limited. The imaging distance is constrained by the available space and is expected to be in the range of 300 to 500 mm. The maximum available space for the camera and lighting envelope is 50×50 mm.
- *Acceptance*: We aim at maximizing the uniformity of light return for entrance angles in the range of  $-20^\circ$  to  $20^\circ$ . We expect an operator to be able to make an alignment with minimal effort well in this range.
- *Brilliance*: There is no minimum specification for the percentage of light returned in useful directions. We aim to optimize its value.



**Figure 3.3** Plan view of design constraints. We aim at optimizing the number of rays arriving at a camera lens with diameter  $D_{lens}$ , which is displaced from the light source by a fixed distance  $\Delta h$ . The working distance is variable within a fixed range of values set by the application.

- *Size:* The size of markers should be sufficiently small and have a relatively short embedment depth, such that they can be mounted to targets without compromising their structural integrity.

The developed marker should tolerate and keep its intended functionality in the environmental conditions of ITER’s divertor. The conditions can be summarized as follows:

- High total integrated gamma radiation doses (around 10 MGy).
- Operating temperature up to 200 degrees Celsius.

In practice, the environmental restrictions deem unfeasible the use of most common materials in commercial retro reflectors, such as polymers and adhesives. The best candidate materials to make up the markers are: stainless steel, which makes up most of the other structures in the divertor, and fused silica glass, which is planned to be used in diagnostic windows and fiber optics. These materials have already been tested and cleared for use in ITER’s divertor.

### 3.3 ITER Compliant Retro Reflector

To the best of our knowledge, there is no off-the-shelf retro reflector that satisfies the application’s requirements and has a suitable performance. This assessment was made in a preliminary study where candidates’ performance was evaluated in a practical test setup simulating the intended use case.

For this reason, we developed, in Publication I, a custom-made marker object, consisting of an array of spherical glass beads. Each glass bead acts as a cat’s eye retro reflector, where the primary lens focuses light into a reflective back metal plate. Due



to the restriction on the use of adhesives and paints, the beads are held in place by a stainless steel housing and a sieve-shaped cover. This design provides attachment and protects the glass beads while simultaneously providing a means of attaching the marker to the targets. This design is further described in Publication II.

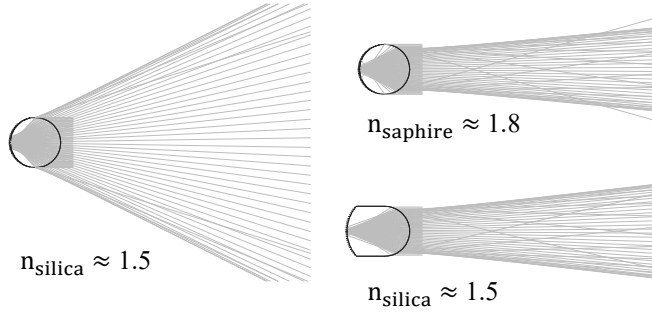
This design serves the purposes of the application and complies with the restrictions. Stainless steel is the main material component in ITER and satisfies all requirements and fused silica is approved as a material for all vacuum classes. Moreover, according to the study in [36], gamma radiation should not induce significant changes to the refractive index of fused silica glass in the relevant wavelengths (380-700 nm). The reported changes in absorption do not influence our application since our measurements are not affected by minor intensity loss.

We investigated as well the alternative of manufacturing a stainless steel hollow corner-cube array. It could technically be implemented and would have the advantage of being made from a single piece and a single material, potentially making up a sturdier piece. However, we were unable to identify and implement a suitable manufacturing strategy for this configuration.

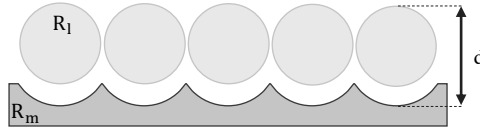
### 3.4 Optimization of Retro Reflector Performance

The early RR design for use at ITER had two degrees of freedom: the radius of the sphere ( $r$ ) and the refractive index of the sphere's material ( $n$ ). Given the refractive index of fused silica ( $n_{\text{SiO}_2} = 1.46$ ) and a sphere radius of 1 mm, the marker has a divergence of approximately 25 degrees, as shown in Figure 3.4, left side. The retro reflected light distribution significantly exceeds the useful angles. The use case's observation angles can be led to be a few degrees, even at shorter working distances. For a maximum camera-light source distance of 30 mm and a minimum working distance of 300 mm, the useful retro reflecting angle is about 6 degrees. A lower divergence value could be achieved by using a type of glass with a higher refractive index, as shown in Figure 3.4, upper right corner. However, not all forms of optical glasses are suitable for ITER's environment. Fused silica is, in overall, the best option due to availability, cost, radiation resistance and has already been tested and cleared for use. While using fused silica, it is not possible to optimize the retro reflector behaviour within the existing design.

In Publication III, we proposed an alternative design with an additional degree of



**Figure 3.4** Retro reflected light distribution of a fused silica glass bead (left side), narrower distribution produced by a type of glass with higher refractive index (upper right corner) and by manipulating the distance between the lens and the reflective surface.



**Figure 3.5** Concept of RR design alternative based on an array of ball lenses, with radius  $R_l$  and a corresponding pattern of spherical reflective surfaces, with radius  $R_m$  and separated by distance  $d$ .

freedom that makes it possible to achieve a lower divergence value. The additional freedom is provided by varying the distance between the focusing element and the reflective backing, as shown in Figure 3.4, lower right corner. The radius of the reflective backing should be changed accordingly to ensure that the conditions for retro reflection are met.

The proposed design is implemented using an array of spherical glass beads that are carefully aligned and kept at a fixed distance to a reflective plate. The reflective plate has a corresponding pattern of concave indentations, acting as spherical mirrors. Each lens and corresponding mirror form one retro reflecting element. The beads and the mirrors have coinciding optical axes and a common aperture diameter, as shown in Figure 3.5.

A similar design could be achieved using an array of plano convex or biconvex lenses as the focusing element. These options can technically be implemented and have the advantages of facilitating the alignment and mounting of the optical elements. However, custom manufacturing of these kind of arrays is rather resource consuming. Further, it is difficult, with most manufacturing methods we are aware

of, to achieve the high values of aperture that would be required to achieve high RR acceptance values. The main advantage of the bead array option is the ease of manufacturing and consequent low cost of spherical glass beads and their intrinsically high aperture.

We have developed a methodology to calculate the RR parameters that optimize brilliancy at the observation angles. It is composed of the following sequence of steps:

- **Step 1:** Establishing the size of each individual retro reflecting unit, i.e. size of the glass beads.

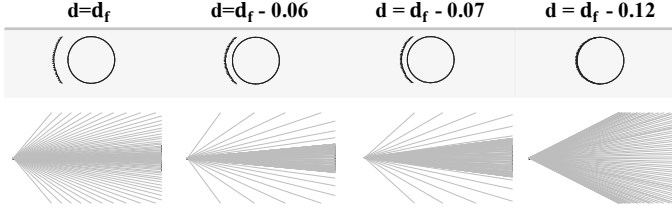
For smaller sizes of the units, the marker would be perceived as a uniform object from the point of view of the camera. This might facilitate processing at later stages. However, smaller sizes of the features also correspond to tighter requirements on the manufacturing tolerances for the reflective array, and holding and aligning structures. Holding and aligning several independent glass beads of small dimensions at a fixed distance to the reflective plate has some technical difficulty.

- **Step 2:** Establishing the distance between the lens and the reflective surface ( $d$ ) with respect to the distance at which the paraxial rays converge ( $d_f$ ). The paraxial focal length of the lens can be calculated as [37]:

$$d_f = \frac{R_l(2 - n)}{2(n - 1)} + 2R_l. \quad (3.1)$$

Different values of the distance  $d$  produce different retro reflective patterns, as shown in Figure 3.6. We can see that some configurations produce a narrower cone of retro reflection. Further, some configurations produce a denser light ray distribution in the periphery of the retro reflection cone, which is advantageous in optimizing the number of rays arriving at the lens.

By performing a ray tracing analysis, of different configurations it is possible to find the most advantageous one. Figure 3.7 illustrates such an analysis. It quantifies the percentage of light rays incident in the RR that are returned at a circular surface representing the camera lens. This analysis considers a single light source that is vertically displaced in relation to the circular surface of the lens.



**Figure 3.6** Light ray distributions for different RR marker configurations. The distance to the reflective surface ( $d$ ) is varied in relation to the focus distance ( $d_f$ ). The spheres have a radius of 0.2 mm.

The most advantageous option ensures that the percentage of light returned in useful directions is as high as possible over the widest range of incident angles. It is also advantageous to obtain as uniform behaviour as possible over the range of expected working distances. In the example in Figure 3.7,  $d = d_f - 0.33$  is the most advantageous option. We see that it has a relatively uniform behaviour over the range of working distances: 300 to 500 mm.

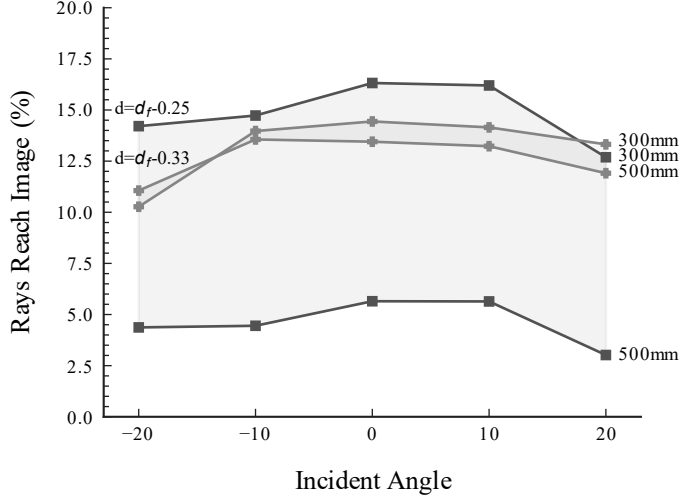
- **Step 3:** Fulfilling the retro reflection condition. In the proposed RRM configuration, the retro reflection condition is:

$$R_m = R_l + d. \quad (3.2)$$

This condition ensures that the reflective surface is kept at the same relative distance to the focus regardless of the incident angle. In this manner, we ensure a uniform RR behaviour over a wide range of incident directions.

- **Step 4:** Establishing the aperture value,  $a$ . In the proposed design, the aperture of the focusing elements is equal to their diameter. This provides the widest range of accepted angles. Figure 3.7 gives a good indication of what would be the range of accepted angles. The aperture of the reflective surfaces might be slightly lower due to manufacturing constraints.

A similar analysis can be done for the alternative plano convex or biconvex lens options. Please refer to [P3] for the derivation of the conditions presented above, as well as for an in depth analysis of the benefits and trade offs of the various marker types.



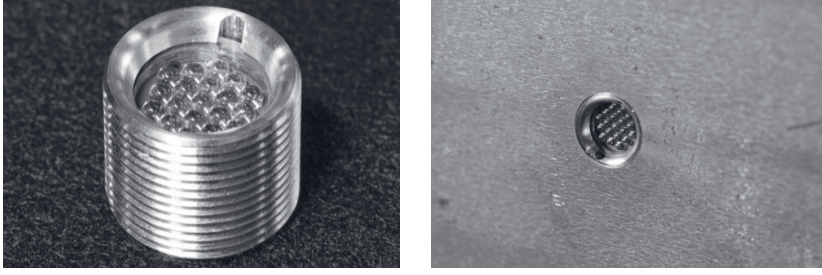
**Figure 3.7** Behaviour of RR marker configurations with varying distance to the reflective surface. The curvature radius and aperture of the primary lens are constant with  $R_1 = 1.0$  and  $a = 2.0$ . The curvature radius of the reflective surface,  $R_m$ , is changed accordingly, to fulfill the condition of retro reflection. The data is obtained using a commercial sequential raytracing software (OSLO by Lambdares [38]).

### 3.5 Performance Evaluation

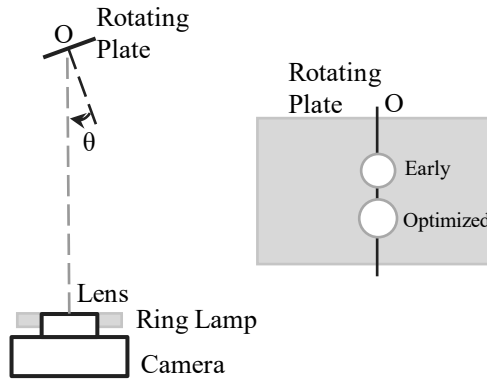
A marker configuration with a bead radius of 1 mm and a distance to the reflective surfaces of 0.39 mm was implemented and tested. These RR parameters are obtained following the proposed optimization methodology for the viewing angles in the experimental setup. Small compromises were made to accommodate manufacturing considerations. The final assembly, with metallic aligning and holding structures is shown in Figure 3.8 and Figure 3.11. Details on the manufacturing process and manufacturing compromises are further discussed in Publication III.

#### Experimental Setup

The performance of the developed RRM is measured in a photogrammetric setup, similar to what would be used in the operational scenario. With this method, we are able to obtain performance curves that directly represent the performance in the intended use case. The experimental setup is composed of a digital camera and a ring light, as shown in Figure 3.9.



**Figure 3.8** Images of manufactured ITER grade RRM. On the left hand side the developed marker, housed inside it's stainless steel housing. On the right hand side the marker is shown mounted in one of the operational pieces.

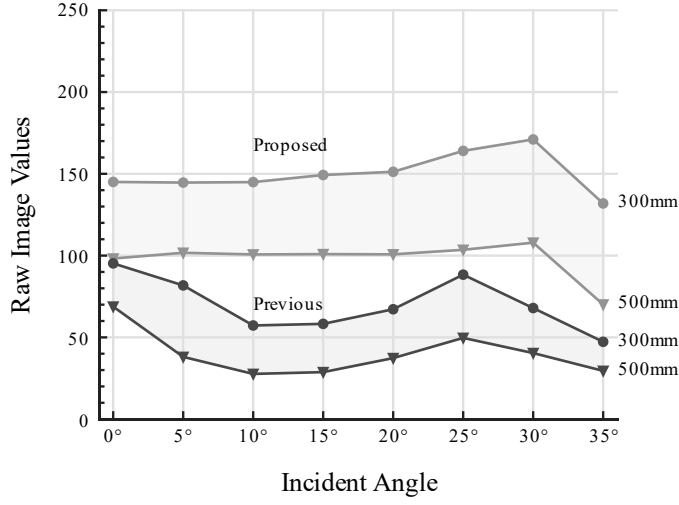


**Figure 3.9** Experimental setup for RRM testing (not to scale). Samples are rotated around centre point O by angle  $\theta$ . Camera and LED sources are stationary throughout the experiments.

The camera has resolution of  $1920 \times 1200$  pixels and the lens has a horizontal field of view of 71.2 degrees. The camera resolution corresponds to approximately 0.4 mm per pixel at the sampling distance of 300 mm and 0.5 mm per pixel at the sampling distance of 500 mm.

The light source is composed of four diametrically opposite LEDs. The distance from the centre of each LED to the centre of the camera lens is 22 mm. The observation angle (between the centre of the camera and the light source) is, therefore, fixed for a given working distance and is approximately 4 degrees at 300 mm and 3 degrees at 500 mm.

Angular performance is evaluated by rotating the test samples around their centre points and varying the entrance and viewing angles by rotating. The response data is acquired at two discrete values (300 mm and 500 mm), corresponding to the limits of the targeted working range. Response curves are symmetrical around the 0 degree



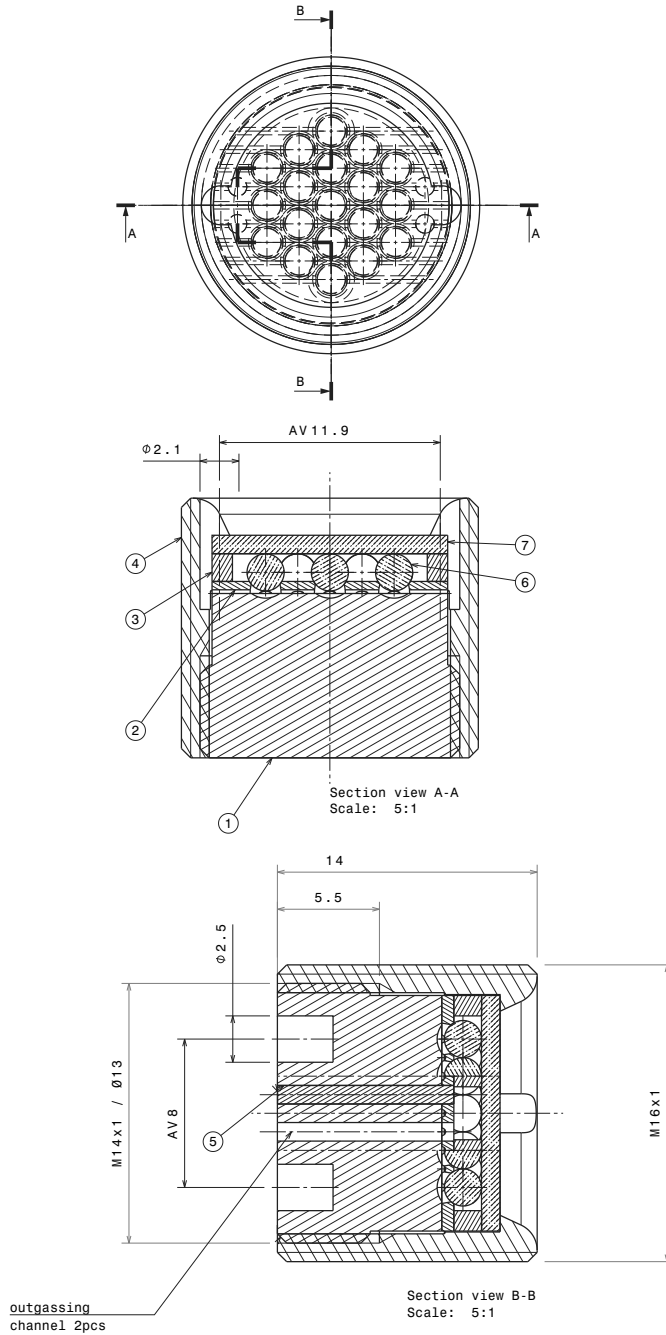
**Figure 3.10** Experimental evaluation of average intensity of all pixels belonging to the marker, in raw image values.

incident angle, due to the added contributions of opposing light sources.

Figure 3.10 shows the acquired response curves for the optimized and the early RR design. We observe that there is considerable gain in performance with respect to the non optimized version in the relevant range of working distances.

### 3.6 Summary

In Chapter 3, we have analysed how retro reflection can be implemented in practise and proposed (in Publications I and II) an early RRM prototype, proving that it is possible to develop a marker that complies with the application's restrictions. In Publication III, a methodology was devised to optimize RR's behaviour, specifically the amount of light returned to the camera in a programmatic setup. The methodology has been employed for a specific experiment setup and the performance was evaluated, showing that the optimized version has considerable gains over the early version.



**Figure 3.11** Diagram of full marker assembly. 1- Reflective Base. 2-Spacer Disk. 3- Spacer Ring. 4- Body. 5- Guiding Pins. 6-Spherical Glass Beads. 7-Glass Window.



## 4 A UNIFIED FRAMEWORK FOR MARKER BASED POSE ESTIMATION AT ITER

In this chapter, we address RQ2: How can we devise a framework that successfully detects and identifies both retro reflective and diffuse markers in different use cases in ITER's RH systems? We overview state of the art methods for detecting and identifying constellations of markers in images and discuss their applicability to the targeted application. We propose and describe a novel framework and evaluate its performance in relevant use cases.





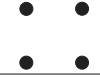

### 4.1 Definitions and Background

Fiducial markers are elements placed in the field of view of an imaging system that act as references in performing image based measurements. Among other applications, they can be used to compute the 6 DOF pose of a camera with respect to the marks.

A wide set of fiducial marker systems have been proposed for pose estimation [39], [40]. However, the vast majority of existing approaches cannot be directly applied to the ITER use case. The most widely used frameworks, such as AprilTag [41], [42], Aruco [43] or ARTag [17], [44] make use of fine details and highly contrasting corners. The expected imaging resolution and working distances at ITER make it impossible to discern such fine details. Some recently proposed systems make use of greyscale gradients [45], [46] or even colours [47], [48] to achieve high distinguishability. From the point of view of manufacturability this type of marks are not feasible. It would be advantageous to implement a *unified fiducial marker framework at ITER*, that can be used both with laser engravings and retro reflective marks. A unified framework would have advantages from the convenience and cost perspectives, taking into account the need for adequate testing and maintaining the detection software. ITER's RRM's look like dots, as described in Chapter 3. For the sake of achieving a unified system, laser engraving shall be dots as well. Therefore, we focus

our overview on dot-based or concentric circle-based fiducial marker systems using a relatively small number of marks. Systems relying on a large number of dots, like RUNE-Tag [49] or thin concentric rings like Concentric Circles [50] are not considered. In building constellations for operational pieces at ITER, the number of marks should be kept at a reasonable level due to practical considerations. This is particularly relevant when using RRM, as their mounting requires drilling into operational pieces. Table 4.1 shows the visual appearance of selected approaches.

**Table 4.1** Appearance of relevant fiducial marker systems.

System	Appearance	Approach
Pattern of five CCC [27], [51]		Classical
Circular ring marker [52]		Classical
Pi-Tag [53], [54]		Classical
WhyCon [55], [56]		Classical
4 Dots [57], [58]		Learning-based
MarkePose [59]		Learning-based

Dots are a frequently used shape as they can be easily detected in blurry or low resolution images. Concentric contrasting circles (CCCs) are an alternative to dots that can improve the distinguishability of the marks from other contrasting image regions. Their uniqueness is based on the unlikelihood that arbitrary image regions would have contrasting circular regions with nearly coincident centres. The estimated centres of the regions can be compared as a check on detected regions [60]. Similarly, the ratio of the area of inner and outer circles can be compared to an expected ratio as an additional check to eliminate false marker candidates [60]. These checks are, however, not projection invariant. This means that the assumptions that circle centroids are coincident within a certain threshold or that areas have a certain ratio may not hold for all viewpoints. The expected viewpoints and corresponding thresholds need to be carefully checked.

Projective invariance is an important consideration in computer vision and in

detecting fiducial marks. As points are projected in an imaging system, their appearance changes. For example, circles show up as ellipses if the imaging plane is not exactly parallel to the circle's plane. However, certain properties, often referred to as projective invariants, are kept under perspective projection, such as [53], [61]:

- Invariance of the class ellipse, meaning that ellipses are always ellipses under perspective projection;
- Collinearity of points, deriving from the fact that straight lines remain straight under perspective projection;
- Angular ordering of coplanar points;
- Cross ratio of four collinear points;

Usually, approaches will make use of those to be able to identify targets at different viewpoints. The invariance property of ellipses is fundamental in detecting circular marks.

In most fiducial frameworks using circular marks [27], [51], [53], [58], [59] each circular mark provides one control point. The control point corresponds, for example, to an estimate of the centre of the detected shape. Even though the centres of projected ellipses are not exactly coincidental with the centre of the original circle the difference is negligible for most practical purposes when the size of circle is small enough [62].

Some methods compute the pose of targets using a single camera by establishing a correspondence between image and world coordinates of  $n > 4$  coplanar control points [61]. To establish the correspondence, they organize marks in identifiable marker constellations. Constellations should be built taking into account the risk of potential pose ambiguities. For example, a constellation of four marks placed at the corners of a square has a roll ambiguity, for multiples of a 90 degree rotation. The individual markers are identified by applying a set of rules based on the geometry of the constellation and on projective invariants. For example, in [27], [51] marks are identified by trying to find the three collinear marks. The remaining points are identified based on their location with respect to the line. A similar approach with more complex heuristics and a slightly different marker distribution is followed in [52]. In these approaches, the identification strategy is very likely to fail if any of the marks are occluded or missing. Pi-Tag [53], [54], is an approach using 12 dot marks distributed along the sides of a square (4 per side, with shared corners). It uses

collinearity and cross ratio based rules in identifying the marks. The higher amount of marks creates some redundancy and provides a moderate robustness to occlusion. However, it needs at least 7 markers to function without any occlusion resilience.

WhyCon [55], [56] follows a significantly different approach. It computes the pose from a single CCC, based on the parameters of a fitted ellipse. In this case, the accuracy of pose estimation is highly dependant on the quality of the ellipse fitting step. Ellipse fitting is particularly challenging and might fail when resolution is low. For this reason, this kind of approach is not applicable to the targeted application.

Other works detect and identify marks with learning based approaches. In [57], [58] the authors extract and exhaustively classify image patches. The approach is trained using patches of real data that are annotated using a simple detection algorithm that is reset manually upon failure. The samples are rotated and lightened or darkened to artificially increase the variation in the training set.

MarkerPose [59] follows a deep convolutional neural network approach similar to that of Deep Charuco [63]. Both share the network structure of Superpoint [64], an approach that estimates key point coordinates and their descriptors on natural images. This approach takes full images as an input. The image coordinates of control points and corresponding IDs are estimated in a black box manner, where it is not possible to fully discern the inner workings of the identification process. Deep-Tag [65] formulates the problem differently and implements a two stage pipeline of object detection followed by keypoint detection in the warped image patch. Deep, end-to-end approaches have been successfully applied to a wide variety of computer vision problems in recent times [66], including object detection [67], [68], feature detection [64] and feature matching [69], [70]. These solutions can often achieve very good results, even in very challenging conditions. However, their interpretability and explainability are a concern when working on safety critical applications. In general, it is desirable to construct solutions that can achieve adequate results with as simple models as possible. The task of finding highly distinguishable marks (such as RRM)s in a controlled environment has a much lower degree of complexity than finding descriptors in natural images. It should be possible to solve that task with acceptable performance with a model of lower complexity.

Further, deep learning approaches require large, realistic, labelled datasets and the determination of a potentially high number of hyper-parameters. Acquiring and labeling real data is a labour intensive process that scales up with the amount

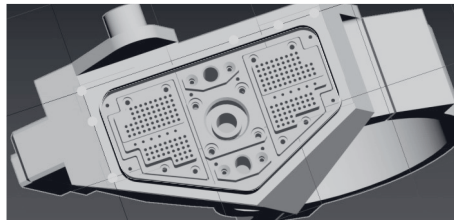
of needed data. The efforts in generating synthetic datasets do not scale up with the amount of data. Generating realistic synthetic dataset might, however, require considerable efforts as a start up step.

## 4.2 Marker Detection and Identification Pipeline

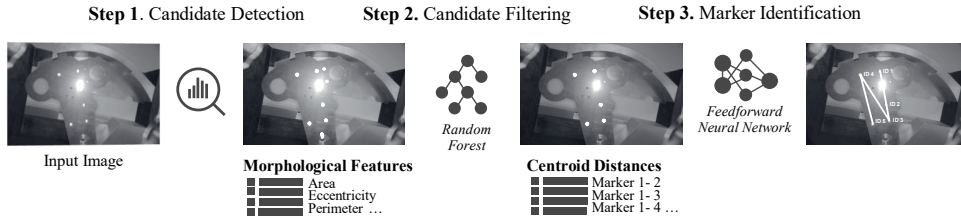
The early strategy for marker identification presented in Publication I and described in detail in Publication II follows a set of rules based on projective invariants, as some of the techniques we have described thus far. It was developed to address one specific use case: the knuckle of the cassette locking system. It is able to correctly identify the position of marks in the expected constellation in most cases where none of the marks are missing. Handling missing marks is generally a challenge with rule based approaches using a relatively small number of fiducials, as seen in earlier literature [51], [53]. It is not feasible to come up with a set of rules that can handle all possible occlusion cases.

Another challenge with this type of approach is that it is not easily adaptable to different constellation geometries. In creating marker constellations for different operational pieces at ITER, there is a need to take into account their shape and the available area where marks can be placed. As an example, Figure 4.1 shows a model of the connector of the bridging link in vessel socket, which is one of the operational targets at ITER. It is evident how the connecting structures invalidate the placement of marks in a large portion of the target's area.

Applying the same constellation geometry to all operational pieces is not feasible and devising a new set of heuristics for each target is not a viable approach. Therefore there is a need to develop an identification strategy that can be easily adapted to different constellation geometries.



**Figure 4.1** Example target (bridging link in vessel connector): the functional structures in the central area restrict the placement of markers to a relatively small border area.



**Figure 4.2** Proposed marker detection and recognition pipeline.

The proposed approach in Publication IV attempts to reach a balance in the considerations discussed thus far. It consists of a hybrid pipeline that extracts marker candidates from images using classical methods and establishes correspondences using a shallow learning approach. The pipeline is illustrated in Figure 4.2. It aims to resolve the problem at hand with the required accuracy employing as simple model as possible for the sake of interpretability of the overall solution. Further, it aims at minimizing the amount of manual labelling tasks and being straightforwardly applicable to different constellation geometries. The pipeline is composed of three consecutive steps:

### Step 1: Candidate Detection

Candidate markers are extracted from images using the maximally stable extremal regions (MSER) approach [71]. It iterates over segmentation thresholds and identifies areas that are most stable throughout threshold variation. The algorithm has two input parameters: the step for threshold evaluation and the maximum area variation between extremal regions. We have empirically found that the algorithm reliably generates suitable sets of candidates for a wide range of values of the input parameters.

### Step 2: Candidate Filtering

Once marker candidates are established, they are further discriminated between true and false candidates, based on their morphological characteristics and image intensity values. The approach in Publication II applies a set of preliminary thresholds on the size and shape of candidate blobs. The candidates that pass the first test are clustered based on their characteristics using a hierarchical clustering strategy. The cluster

containing the markers is identified based on the number of elements in each group. The assumption behind this approach is that the cluster with the expected number of markers likely groups the correct candidates. This assumption does not hold in case any of the markers are occluded or missing. Further, outlier candidates might have a very variable set of characteristics and assumptions should not be made on how they might be organized by the clustering strategy.

In Publication IV, the filtering task is re-formulated as a classification rather than a clustering one. Subsequently, it makes no use of manually set thresholds and solves the problem with occluded or missing markers.

The input features for each candidate are its area, circularity, solidity, extent, mean intensity and centroid. Circularity is calculated as  $4 * \pi * area / perimeter^2$  and evaluates the "roundness" of blobs. It has a value of 1 for a circular blob and lower values for other elliptical shapes. Solidity is the ratio of the area to convex area and extent is the ratio of area to the bounding box area.

The classification problem is solved using a random forest algorithm [72]. We have empirically found that this approach provides fairly good results using only a few tens of training images. Each labelled image provides a varying number of training samples, depending on the number of candidates that are generated at the previous pipeline stage. Labeling is done through a semi-automatic procedure, where the candidates are presented as highlighted spots in the image and the user selects the ones corresponding to actual markers.

### Step 3: Marker Identification

The correspondence between detected markers and the model of the target is solved using a learning based approach. The coordinates of marker centroid are used as input and normalized as follows [61]:  $x' = K^{-1}x$ , where  $K^{-1}$  represents the inverse of the camera's intrinsic matrix, estimated in a prior camera calibration step, and  $x$  and  $x'$  are the original and the normalized values respectively.

One advantage of using the candidates centroids as an input is that a realistic synthetic dataset can be easily generated. Synthetic data can be created by picking random camera poses and projecting the known world point to the image using the rules of perspective projection. This is a rather straightforward process versus the alternative of generating realistic synthetic images.

We formulate the identification task as a classification problem, where each marker

constitutes one input, to be classified into categories: “1”, “2”, “3”, etc. One-hot encoding is used to convert the categorical input into a numerical form. The features used to identify each marker are the distances in x and y to each of the other  $N-1$  markers detected in the same image. Pairs of  $(x,y)$  distances are sorted based on their ascending x-coordinates to reduce the permutations, while keeping the correspondence between the two  $(x,y)$  dimensions.

We make the classification using a shallow, single hidden layer feed-forward neural network with a sigmoidal activation function. This type of network should, according to the universal approximator theorem, be able to approximate any continuous function, given that it has enough neurons in the hidden layer. The error is expected to be reduced by the increase of the number of neurons in the hidden layer up to point after which the overall performance is decreased due to the loss of generalization. We have empirically found that 30 neurons is a suitable number of neurons after which performance is not improved further.

The output of the network consists of an array of values between 0 and 1 for each class. These values sum up to one and can be inferred as a probability distribution. For a set of  $N=5$  markers detected in one image, we run the classification 5 times and obtain 5 output arrays. These can be concatenated in an output matrix similar to that of Table 4.2 (left side). In this representation, each matrix column represents the network output of one marker.

In choosing the most suitable class for each marker, we want to find a method that enforces the following conditions:

1. The same marker cannot belong to more than one class, meaning that only one match is correct in each column.
2. There cannot be more than one marker of the same class in an image, meaning that only one match is correct in each row.

The most common approach, corresponding to choosing the maximum value in each column enforces condition 1. However, it does not enforce condition 2. In order to enforce both conditions, we evaluate the joint probability of each possible permutation of markers. Permutations of 5 markers are for example: ‘54321’, ‘54312’, ‘54231’. The correct set should have the highest joint probability value. In Table 4.2 we present a concrete example that illustrates this concept. We see that choosing the maximum probability values leads to two markers being classified as “2”. The result



is that one marker in the image is wrongly identified. By estimating the joint probability of marker permutations and enforcing both conditions, we arrive at a fully correct answer.

The joint probability shall be obtained by multiplying the individual probability values. In Publication IV, the probabilities had been summed and consequently the results are overall slightly worse than those presented here.

**Table 4.2** Example output matrix. Note that each matrix column refers to a separate network output. The matrix concatenates in its rows the outputs of different markers detected in the same image. Below the matrix are the chosen classes according to ‘maximum value’ strategy and our proposed approach. On the right hand side we show the joint probabilities of some example marker permutations.

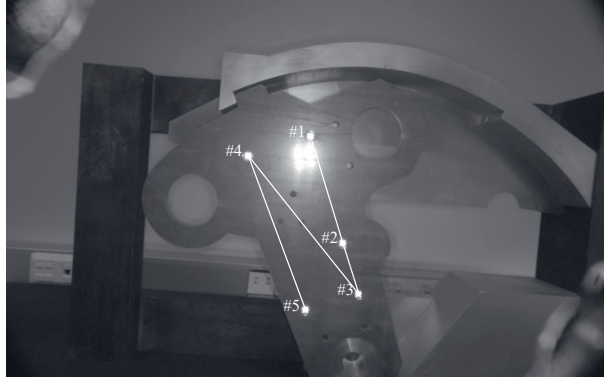
Output Matrix					Permutations of Markers	Joint Probability
2.8e-12	3.2e-13	<b>0.0646</b>	0.4366	1.7e-7	× "54321"	3.1e-19
6.3e-5	3.2e-7	0.9336	1.6e-4	<b>0.9965</b>	× "54312"	4.9e-19
<b>0.9999</b>	2.0e-8	6.4e-9	5.7e-15	0.0035	× "54213"	1.7e-31
5.3e-12	5.8e-6	1.8e-3	<b>0.5632</b>	8.7e-8	× "54132"	2.5e-13
3.1e-5	1.000	1.2e-8	1.2e-10	2.1e-6	✓ "35142"	<b>0.04</b>
Maximum Value	✓ "3"	✓ "5"	× "2"	✓ "4"	✓ "2"	
Proposed	✓ "3"	✓ "5"	✓ "1"	✓ "4"	✓ "2"	
					× ...	...

### 4.3 Performance

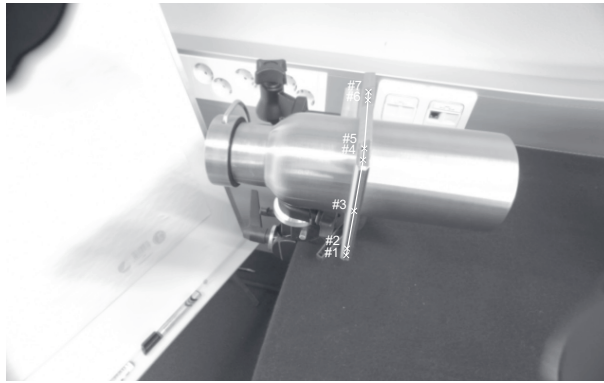
The methodology for marker detection and identification described in previous sections is applied to three different use cases in ITER’s divertor: the knuckle of the cassette locking system, the pipe flange and the bridging link in-vessel connector. The knuckle, whose structure can be inferred from the image in Figure 4.3, has a large flat textureless area. It does not pose a significant challenge to fitting a marker constellation. It is fitted with five RR markers. The characteristics of the markers are as described in Chapter 3.

The pipe flange, whose structure can be inferred from the image in Figure 4.4, is a more challenging target. In this use case the available working area is particularly limited and the expected viewpoints are similar to what is seen in the figure. For this reason, markers had to be placed along the flat area in the edge of the flange. Due to the very small available area, retro reflectors cannot be used. The piece is marked with laser engravings of smaller size (3 mm diameter).

The performance of marker detection in the two use cases described thus far



**Figure 4.3** Example of successful detection and identification - knuckle use case

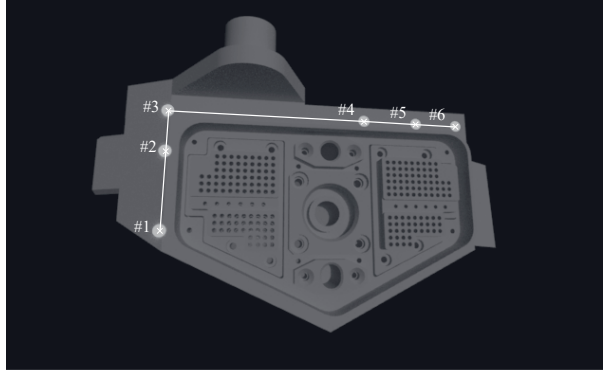


**Figure 4.4** Example of successful detection and identification - pipe flange use case

is evaluated on real image data of true scale realistic prototypes. The camera and lighting setup for obtaining the images is as described in Chapter 3. To emulate the use of radiation tolerant cameras, we convert the images to grayscale in-camera and downsample the resolution to 640x400 pixels. This produces a fairly realistic approximation of the image quality of the application.

The bridging link use case is analysed based on synthetically generated images. The bridging link structure can be inferred from Figure 4.5. This use case permits the use of RR markers. However, there are significant limitations on how the constellation can be constructed. The connector has a large inner area covered in functional structures where markers cannot be placed. Markers are placed along the outer rim of the piece, as seen in the figure.

For each of the use cases, the performance of the marker detection and identification methodology is evaluated in a pass or fail manner at each step of the pipeline



**Figure 4.5** Example of successful detection and identification - bridging link (bottom) use case.

stages. Results are shown in Table 4.3.

As seen in Table 4.3, the percentage of detected markers is high for all use cases. However, the amount of false candidates is considerably higher in the case using diffuse marks. This is to be expected as small dark diffuse marks might have a similar appearance to background structures. The proposed approach succeeds at identifying markers at a very high rate, provided that markers have been correctly detected and filtered. It struggles if marks are missing or wrongly detected, particularly in the pipe flange use case, leading to a lower full pipeline success rate.

Table 4.4 analyses missing marker cases in more detail. It quantifies the performance of the identification step in situations where one specific marker is missing or not detected. In use case I, the detection strategy recuperates relatively well from one missing mark. However, when edge marks (for example, markers 6 or 7) are not detected in use case II, identification rates are considerably lower. In these cases missing marks might lead the constellation geometry to appear similar at different perspectives. Therefore the identification becomes ambiguous. Constellations should be built, whenever possible, taking into account these cases.

## 4.4 Pose Estimation

Once image and world point correspondences have been established, and provided that camera calibration parameters are known a priori, the direct linear transform homography estimation method [61], [73] can be used to estimate the pose. The computation requires a minimum of four co-planar point correspondences. This

**Table 4.3** Robustness measure of marker detection and identification at consecutive pipeline stages - knuckle (I), pipe flange (II) and bridging link (III) use cases.

		I	II	III
<b>Candidate Generation</b> <sup>1</sup>	Detection Rate	99%	95%	100%
	$N_{\text{detected markers}}/N_{\text{markers}}$			
	False Candidate Rate	13%	74%	5%
	$N_{\text{false candidates}}/N_{\text{candidates}}$			
<b>Filtering</b> <sup>2</sup>	True Positive Rate	100%	98%	100%
	$N_{\text{true positives}}/N_{\text{positives}}$			
	False Positive Rate	2%	3%	0%
	$N_{\text{false positives}}/N_{\text{positives}}$			
<b>Identification</b> <sup>3</sup>	True Identification Rate	100%	100%	100%
	$N_{\text{true identifications}}/N_{\text{markers}}$			
<b>Full Pipeline</b> <sup>2</sup>	Detection Rate	95%	70%	100%
	$N_{\text{successful images}}/N_{\text{images}}$			

<sup>1</sup>Over the whole image set. <sup>2</sup>Over the test set

<sup>3</sup>Over the set of images where all markers were detected

estimate is used as an initial estimate to start an iterative approach that minimizes the re-projection error as a non-linear least squares problem.

The minimization of the re-projection error is solved using the Levenberg–Marquardt algorithm [74], [75], and can be written as follows:

$$\arg \min_{r,t} \sum_{i=1}^N [m_i - P(M_i, K, r, t)]^2, \quad (4.1)$$

where  $P$  describes the projection of world points  $(X, Y, Z)$  into image points  $(x, y)$ :

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \lambda K \begin{bmatrix} R & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (4.2)$$

**Table 4.4** Robustness measure of correspondence step with respect to occlusion of specific markers - knuckle (I) and pipe flange (II) and bridging link (III) use cases.

True Identification Rate	I	II	III
#1 Missing	72%	79%	78%
#2 Missing	100%	75%	80%
#3 Missing	75%	92%	100%
#4 Missing	100%	80%	77%
#5 Missing	82%	97%	57%
#6 Missing	-	70%	35%
#7 Missing	-	58%	-

$M_i$  and  $m_i$  represent world and image point coordinates, respectively and  $\lambda$  represents a scaling factor. The three-dimensional vector  $r$  represents the axis-angle representation of the rotation matrix,  $R$ , obtained using the Rodriguez formula [76].

The intrinsic matrix,  $K$ , describing the internal parameters of the camera is written as function of the camera focal length ( $f_x, f_y$ ) and principle point ( $c_x, c_y$ ) as follows:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.3)$$

The internal parameters of the camera are estimated using the well established method of [73], based on a set of images of a checkerboard pattern. Please refer to Publication II for a more elaborated discussion on the calibration process.

In the case where two cameras are used to provide a marker based estimate, each of the cameras provides an independent measure of the pose, provided that the camera's extrinsic parameters are known. In this case, the average of the two estimates is taken as the initial estimate and the minimization is as follows:

$$\arg \min_{r, t} \sum_{i=1}^N [m_i^l - P(M_i^l, K^l, r, t)]^2 + [m_i^r - P(M_i^r, K^r, r, t)]^2. \quad (4.4)$$

Superscripts l and r represent each camera.

## 4.5 Summary

In Chapter 4, the requirements of a suitable marker detection and identification framework for ITER's RH systems were established. In Publication IV, a general framework was proposed, applied and tested in three relevant use cases, using either RRM's or laser engravings. Results show that the approach successfully generalizes to different use cases and provides a moderate level of occlusion resilience. Lastly, we detail how the 6 DOF pose of targets can be computed from a set of correctly identified marker coordinates.

## 5 END-TO-END EVALUATION OF POSE ESTIMATION ACCURACY

In this chapter, we address RQ3: How can we effectively evaluate the absolute end-to-end accuracy of the proposed solutions? We propose a methodology to evaluate the end-to-end performance of the marker based pose estimation strategy, conduct experiments and report the results.

### 5.1 Definitions

Remote handling requires considering the pose of rigid bodies, such as robotic manipulators and their end effectors, targets and cameras, with respect to each other. The pose of a rigid body can be represented by an arbitrarily placed coordinate system. The mapping between two object coordinate systems is described by a rigid body transform. A rigid body transform has 6 DOF in 3D space. There are 3 DOF related to the translation and 3 DOF related to rotation.

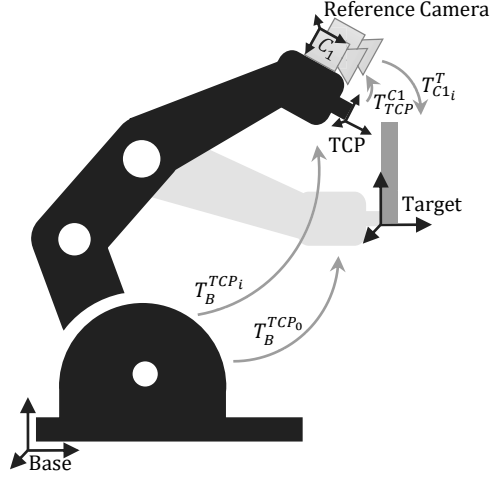
The composition of several transforms is not straightforward to compute and can be largely facilitated using an algebraic formalism. In this formalism, a homogeneous rigid body transformation,  $T$ , is represented in a 4x4 matrix as follows:

$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}, \quad (5.1)$$

where  $R$  represents the  $3 \times 3$  rotation matrix and  $t$  represents the  $3 \times 1$  translation vector.

The relevant coordinate systems and rigid body transformations that will be referenced throughout the following sections are depicted in Figure 5.1.

The task of pose estimation refers to the calculation of the rigid body transformation between a camera and a target ( $T_{Ci}^T$  at each position  $i$ ) based on a set corre-



**Figure 5.1** Relevant coordinate frames and rigid body transformations.

spondences between detected 2D image points and 3D world point coordinates.

For this estimate to be meaningful and convertible into actionable instructions for a manipulator's control system (CS), there is a need to estimate the transformation between the camera and the tool centre point (TCP) of the manipulator ( $T_{TCP}^{C1}$ ). This problem is often referred to as *hand-eye calibration*. In this work, hand-eye calibration is estimated using the well established method of [77].

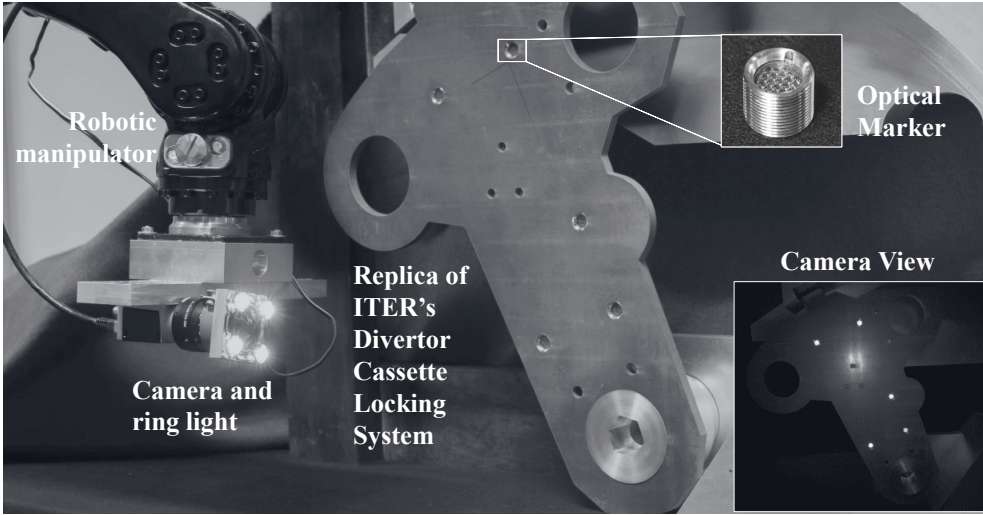
The definition of the TCP position is referred to as TCP calibration. TCP calibration estimates the transformation between the robot flange and the end effector of a specific tool. In this work, TCP calibration is resolved using the well established XYZ Four-Point TCP Calibration Method. This is often a built-in function of manipulators. The method consists of bringing the desired tool center to a known position and recording several points as the position is approached from different orientations.

Robot's CS can usually report the pose of the TCP with respect to its base's coordinate frame ( $T_B^{TCP_i}$  at each position  $i$ ) with varying degrees of precision. For the manipulators that can be used in the ITER environment, the quality of that estimate is not yet known and is not to be relied upon for location purposes. The industrial robot used in this work's laboratory experiments is able to report that transformation with a high degree of precision. The reported pose will be used in the establishment of reference values to evaluate the results of the vision based system.



## 5.2 Experimental Setup

The KUKA KR 16 L6-2 industrial robot was used as the manipulator in our experiments. The vision system is composed of a Basler acA1920-50gc industrial machine vision camera equipped with a close fitting ring lights that create a uniform light pattern that covers the entire target for the expected range of camera poses. The lenses are high quality, low distortion wide angle lenses with a nominal 6 mm focal length. With these particular cameras, the lenses produce an opening angle of roughly 82 degrees. The output of the cameras is converted to grey scale and down sampled to emulate the use of radiation tolerant cameras. The experimental setup is shown in Figure 5.2.

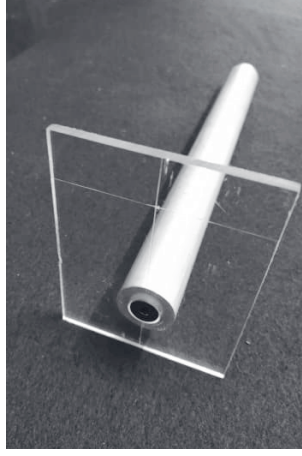


**Figure 5.2** Experimental setup with manipulator, imager and light source on the left-hand side.

## 5.3 Reference Value Estimation

The reference values for assessing the accuracy of the system were obtained by using an alignment piece, shown in Figure 5.3, made of a transparent polymer that had been engraved with alignment axes. Reference values were obtained by manually positioning the alignment piece in the aligned position and acquiring the transformation provided by the manipulator's CS,  $T_B^{TCP_0}$ .

In the aligned position, and provided that the TCP has been assigned to the cross



**Figure 5.3** Alignment piece composed of a transparent polymer window engraved with an alignment axes.

hair of the alignment piece, the TCP and target coordinate systems coincide. In this setting, for a given test position,  $i$ , the rigid body transformation between the camera and target coordinate systems,  $T_{C1_i}^T$ , is given as follows:

$$T_{C1_i}^T = (T_{TCP}^{C1})^{-1} (T_B^{TCP_i})^{-1} T_B^{TCP_0} T_{TCP_0}^T, \quad (5.2)$$

where  $T_B^{TCP_i}$  is provided by the robot's CS at position  $i$  and  $T_{TCP}^{C1}$  is the hand-eye transformation. For each test image,  $i$ , the pose estimated by the vision system is compared to the reference value ( $T_{C1_i}^T$ ).

## 5.4 Results

Given a rigid body transformation,  $T_{est}$ , and the corresponding reference value,  $T_{ref}$ , the absolute translation error ( $e_t$ ) in millimetres is calculated as follows:

$$e_t = \sqrt{(t_{est_x} - t_{ref_x})^2 + (t_{est_y} - t_{ref_y})^2}. \quad (5.3)$$

To calculate a total rotational error, the residual rotation matrix is converted to axis-angle form. The axis angle representation parameterizes a rotation by a 3D vector that describes its direction and an angle describing the magnitude. The magnitude value is extracted as the total angular error.

Results are presented in Table 5.1 and Table 5.2. The performance of marker

based pose estimation is evaluated using a single camera and two cameras so that a comparison can be drawn. A comparison is also established with the earlier stereo matching based approach [7] and the multi view approaches of [78].

**Table 5.1** Average and Worst Case Accuracy and Precision of Estimated Translations.

$e_t$ (mm)	Accuracy		Precision	
	Average	Worst Case	Average	Worst Case
Monoscopic	<b>2.1944</b>	3.0295	0.5597	0.9691
Stereoscopic	2.2924	<b>2.9813</b>	0.5033	0.9707
Stereoscopic ICP [7]	3.6017	5.7915	1.2287	2.9162
Multi view 1	3.4751	4.2541	0.6178	1.1204
Multi view 2	2.5606	3.2318	<b>0.4797</b>	<b>0.9304</b>

To compare our proposed approach with the stereo matching based approach, the same scene and sequence of poses were captured with increased lighting. The ground truth pose with some added noise was used as an initial estimate to the algorithm to make sure it converges to the correct solution, so we can reliably measure its accuracy without concern about parameter tuning. We observe that with small initial offsets ( $|\Delta\alpha| = 5$  deg,  $|\Delta t| = 25$  mm), it converges closely to the expected values. The statistics in the table refer to this offset range. With larger amounts of offset, the convergence deteriorates quickly and no longer ends up with correct estimates. To be able to establish a comparison with the marker-based multi view estimation methods of [78] (Multi view 1 and Multi view 2), for each position, we recorded an extra set of five random poses.

The marker-based method outperforms the earlier version of the system based on stereo depth estimation. The marker based option using two cameras performs slightly better on the metrics of orientation estimation, but does not perform in overall significantly better than monocular and it actually loses to monocular in average translation accuracy. There is no definite improvement in marker based pose estimation from the introduction of a second camera.

Regarding the approaches using multiple shots taken with the same camera, both approaches outperform monoscopic estimation in maximum angular error. Method "multi view 1" estimates rotation consistently better than method "multi view 2". Overall, the results of multi view estimation do not significantly outperform the

estimate from a single camera.

**Table 5.2** Average and Worst Case Accuracy and Precision of Estimated Rotations.

$e_R$ ( $^\circ$ )	Accuracy		Precision	
	Average	Maximum	Average	Maximum
Monoscopic	0.4105	1.6886	0.2421	1.2781
Stereoscopic	<b>0.3291</b>	<b>0.7950</b>	<b>0.1563</b>	<b>0.4659</b>
Stereoscopic ICP [7]	0.5367	1.5510	0.3098	1.0143
Multi view 1 [78]	0.4735	1.0074	0.2292	0.5339
Multi view 2 [78]	0.5381	1.2164	0.2590	0.6783

## 5.5 Summary

In Chapter 5, the end-to-end performance of the proposed system was evaluated in realistic test setups of two significantly different use cases: the pipe flange and the knuckle of the cassette locking system. A methodology was devised to estimate a reference value of the pose. Results were compared with those of other relevant approaches. The marker-based approach outperformed the previous stereo matching based approach.

## 6 DISCUSSION AND CONCLUSIONS

The thesis has focused on the development of methods to estimate the 6 DOF pose of operational targets in ITER divertor's environment. The main characteristics and challenges of the targeted application were identified. The context of existing 3D sensing technologies was identified, taking into account, particularly, the environmental restrictions imposed in the nuclear environment. A unified marker based pose estimation system was proposed. The system uses retro reflective markers, whenever possible, or alternatively laser engravings placed on the available areas of the targets. The contribution of the thesis has been in two main fronts: the development of suitable markers (RQ1) and the development of a methodology to estimate their location in images and estimate the pose (RQ2). It also presents a methodology to evaluate the end-to-end pose estimation accuracy of the overall system (RQ3).

In applying marker based tracking to the ITER application it was necessary to consider how markers survive and perform in the nuclear environment. It was demonstrated that it is possible to obtain an RRM that is made up of materials that withstand the divertor's environmental conditions, particularly stainless steel and fused silica. A methodology was developed to optimize the RR behaviour, i.e. the amount of incident light that is returned to the camera. The proposed RR design fulfills the environment restrictions and optimizes the amount of light returned to the cameras, thus answering positively the first research question. This is achieved with some degree of manufacturing complexity, which imposes a relatively high cost of the designed marker.

The resilience of the proposed marker design was considered during the design process. However, final resilience and performance tests after being subjected to the divertor's environmental conditions were not conducted. Future work might focus on conducting irradiation and temperature variation experiments on the developed markers to ensure that their functioning is according to what has been established.

The alternative corner cube array retro reflector option was abandoned due to

manufacturing challenges as manufacturing processes are out of the scope of the thesis. A hollow corner cube RRM is an attractive option as it would make up a sturdier construction, not requiring the use of glass. Future work could be to explore this option further. It would be interesting to determine in what range retro reflector parameters, divergence and angularity, can be manipulated within that design while taking into account manufacturability considerations.

The second part of the work focused on the development of computational methods to detect and identify markers in low resolution camera images. A general framework for marker detection and identification was proposed that can be easily applied to different constellation geometries. The developed solution has been demonstrated to be moderately resilient to occlusions despite using a relatively small number of uncoded marks. In answering RQ2, the work demonstrated and analysed the applicability of the proposed solution to different use cases: the knuckle of the cassette locking system and the bridging link, using retro reflective markers, and the pipe flange using laser engravings.

The performance of the marker detection and identification strategy is satisfactory in the cases using RRM. However, detectability could be improved in the use case using laser engravings. Diffuse marks are considerably more challenging to distinguish from non-marker background elements. Further work might focus on achieving higher detectability rates with diffuse marks. This could be pursued by improving the visual uniqueness of marks. However, the limit to the size of the marks and camera resolution might invalidate this solution. Alternatively, the detection approach could be improved. Morphological features of blobs might not be enough in some cases to distinguish between diffuse marks and background elements. Perhaps more discriminative features could be obtained from the image patch containing the marker or its neighbourhood. Another path worth exploring might be the use of the contextual information on how marker locations relate to each other at the candidate filtering stage. A more complex identification strategy that can handle a larger amount of missing marks might also be a path worth pursuing. A more in depth analysis of the performance interpretability trade-off considerations in this use case might be particularly relevant as well.

In the last part of the work, a methodology was devised to estimate the end-to-end accuracy of the overall pose estimation strategy on a mock-up prototype. The developed metric effectively and realistically measures the absolute accuracy of the

system, thus answering the third research question. It is worth mentioning that the proposed approach cannot discern between the contributions of different estimation processes, such as camera calibration, hand-eye calibration, manual positioning of alignment tool and errors in the estimate provided by the robot CS to the estimated result. The results show that, even in the worst case scenario, performance is in line with the requirements for the system. The proposed method outperforms the earlier solution, and has a very comparable degree of accuracy and precision as the stereo based or multi view based options. Marker based tracking was demonstrated to be a suitable, robust approach for pose estimation at ITER.

Many considerations are still underway in the ITER project on how operational pieces are to be designed and how RH systems are to be organized. Some of the operational pieces we have considered in this work are still being redesigned and reconsidered. On-site repairing of components might still be found to be a more suitable option rather than removal, transportation and in - hot cell repairing. Remote operations in DEMO and in the power plants of the future might look rather different than what is currently expected, based on what is found from the experiments of ITER. However, there is strong evidence that accurate, reliable localization of human made targets might be one of the key technical challenges in RH. Marker based pose estimation, for its characteristics discussed throughout the thesis, is the most viable approach to solve that problem.

Overall, the thesis constitutes a consistent contribution towards the development of a unified framework for marker based pose estimation at ITER. It integrates knowledge of different areas, such as computer vision, robotics and optics, and touches upon material and mechanical engineering considerations to come up with realistic solution for a key technical challenge. Some of the results might be applicable to other applications where temperature and radiation are a main concerning factor, such as in the nuclear industry or in space applications.





## REFERENCES

- [1] S. C. Cowley, “The quest for fusion power,” *Nature physics*, vol. 12, no. 5, pp. 384–386, 2016. DOI: 10.1038/nphys3719.
- [2] R. Buckingham and A. Loving, “Remote-handling challenges in fusion research and beyond,” *Nature Physics*, vol. 12, no. 5, pp. 391–393, 2016. DOI: 10.1038/nphys3755.
- [3] G. D. Dubus, “From plain visualisation to vibration sensing: Using a camera to control the flexibilities in the ITER remote handling equipment,” Ph.D. dissertation, Tampere University of Technology, 2014.
- [4] D. T. Hamilton, A. Tesini, R. Ranz, and H. Kozaka, “Progress in standardization for ITER remote handling control system,” *Fusion Engineering and Design*, vol. 89, no. 9-10, pp. 2409–2414, 2014. DOI: 10.1016/j.fusengdes.2014.01.003.
- [5] K. Leinemann, F. Katz, H. Knüppel, W. Olbrich, and D. Maisonnier, “Synthetic viewing: Comprehensive work representation, making remote work clearer to the operator,” *Fusion Engineering and Design*, vol. 29, pp. 317–323, 1995. DOI: 10.1016/0920-3796(95)80036-W.
- [6] L. Niu, L. Aha, J. Mattila, A. Gotchev, and E. Ruiz, “A stereoscopic eye-in-hand vision system for remote handling in ITER,” *Fusion Engineering and Design*, vol. 146, pp. 1790–1795, 2019. DOI: 10.1016/j.fusengdes.2019.03.036.
- [7] L. Niu, S. Smirnov, J. Mattila, A. Gotchev, and E. Ruiz, “Robust pose estimation with a stereoscopic camera in harsh environments,” in *Proceedings of International Symposium on Electronic Imaging*, Society for Imaging Science and Technology, 2018. DOI: 10.2352/ISSN.2470-1173.2018.09.IRIACV-126.

- [8] A. Tesini and J. Palmer, “The ITER remote maintenance system,” *Fusion Engineering and Design*, vol. 83, no. 7-9, pp. 810–816, 2008. DOI: 10.1016/j.fusengdes.2008.11.029.
- [9] I. Ribeiro, C. Damiani, A. Tesini, S. Kakudate, M. Siuko, and C. Neri, “The remote handling systems for ITER,” *Fusion Engineering and Design*, vol. 86, no. 6-8, pp. 471–477, 2011. DOI: 10.1016/j.fusengdes.2011.01.138.
- [10] C. Neri, P. Costa, M. F. De Collibus, *et al.*, “ITER in vessel viewing system design and assessment activities,” *Fusion Engineering and Design*, vol. 86, no. 9-11, pp. 1954–1957, 2011. DOI: 10.1016/j.fusengdes.2011.02.047.
- [11] V. Komarov, H. Heidl, R. Tivey, and J. Palmer, “Design progress of the ITER divertor cassette-to-vacuum vessel locking system,” *Fusion Engineering and Design*, vol. 82, no. 15-24, pp. 1866–1870, 2007. DOI: 10.1016/j.fusengdes.2007.07.020.
- [12] V. Lyytikäinen, P. Kinnunen, J. Koivumäki, *et al.*, “Divertor cassette locking system remote handling trials with WHMAN at DTP2,” *Fusion Engineering and Design*, vol. 88, no. 9-10, pp. 2181–2185, 2013. DOI: 10.1016/j.fusengdes.2013.02.114.
- [13] J. Lyytinen, P. Tikka, T. Määttä, T. Avikainen, and S. Rantala, “Development of the remote handling connector for ITER divertor diagnostic system,” *Fusion Engineering and Design*, vol. 165, p. 112 243, 2021. DOI: 10.1016/j.fusengdes.2021.112243.
- [14] A. Vale, R. Ventura, P. Lopes, and I. Ribeiro, “Assessment of navigation technologies for automated guided vehicle in nuclear fusion facilities,” *Robotics and Autonomous Systems*, vol. 97, pp. 153–170, 2017. DOI: 10.1016/j.robot.2017.08.006.
- [15] J. Ferreira, A. Vale, and R. Ventura, “Optimizing range finder sensor network coverage in indoor environment,” in *7th IFAC Symposium on Intelligent Autonomous Vehicles*, 16, vol. 43, Elsevier, 2010, pp. 282–287. DOI: 10.3182/20100906-3-IT-2019.00050.
- [16] J. Ferreira, A. Vale, and I. Ribeiro, “Localization of cask and plug remote handling system in ITER using multiple video cameras,” *Fusion Engineering and Design*, vol. 88, no. 9-10, pp. 1992–1996, 2013. DOI: 10.1016/j.fusengdes.2012.10.008.

- [17] M. Fiala, "Designing highly reliable fiducial markers," *IEEE Transactions on Pattern analysis and machine intelligence*, vol. 32, no. 7, pp. 1317–1324, 2009. DOI: 10.1109/TPAMI.2009.146.
- [18] J. Soares, A. Vale, and R. Ventura, "A multi-purpose rescue vehicle and a human–robot interface architecture for remote assistance in ITER," *Fusion Engineering and Design*, vol. 98, pp. 1656–1659, 2015. DOI: 10.1016/j.fusengdes.2015.06.148.
- [19] R. Horaud, M. Hansard, G. Evangelidis, and C. M  nier, "An overview of depth cameras and range scanners based on time-of-flight technologies," *Machine Vision and Applications*, vol. 27, no. 7, pp. 1005–1020, 2016. DOI: 10.1007/s00138-016-0784-4.
- [20] E. T. Jonasson, J. Boeuf, S. Kyberd, *et al.*, "Reconstructing JET using LIDAR-vision fusion," *Fusion Engineering and Design*, vol. 146, p. 110 952, 2019. DOI: 10.1016/j.fusengdes.2019.03.069.
- [21] G. Dubus, A. Puiu, P. Bates, C. Damiani, R. Reichle, and J. Palmer, "Progress in the design and r  d of the iter in-vessel viewing and metrology system (ivvs)," *Fusion Engineering and Design*, vol. 89, no. 9-10, pp. 2398–2403, 2014. DOI: 10.1016/j.fusengdes.2014.01.012.
- [22] E. T. Jonasson, L. R. Pinto, and A. Vale, "Comparison of three key remote sensing technologies for mobile robot localization in nuclear facilities," *Fusion Engineering and Design*, vol. 172, p. 112 691, 2021. DOI: 10.1016/j.fusengdes.2021.112691.
- [23] V. Goiffon, S. Rizzolo, F. Corbi  re, *et al.*, "Total ionizing dose effects on a radiation-hardened CMOS image sensor demonstrator for ITER remote handling," *IEEE Transactions on Nuclear Science*, vol. 65, no. 1, pp. 101–110, 2017. DOI: 10.1109/TNS.2017.2765481.
- [24] T. Liu, A. W. Burner, T. W. Jones, and D. A. Barrows, "Photogrammetric techniques for aerospace applications," *Progress in Aerospace Sciences*, vol. 54, no. 5, pp. 1–58, 2012. DOI: 10.1016/j.paerosci.2012.03.002.
- [25] G. Burgess, M. Shortis, and P. Scott, "Photographic assessment of retroreflective film properties," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 5, pp. 743–750, 2011. DOI: 10.1016/j.isprsjprs.2011.07.002.

- [26] M. San Biagio, C. Beltran-Gonzalez, S. Giunta, A. Del Bue, and V. Murino, "Automatic inspection of aeronautic components," *Machine Vision and Applications*, vol. 28, no. 5-6, pp. 591–605, 2017. DOI: 10.1007/s00138-017-0839-1.
- [27] W. A. Hoff, L. B. Gatrell, and J. R. Spofford, "Machine-vision-based teleoperation aid," *Telematics and Informatics*, vol. 8, no. 4, pp. 403–423, 1991. DOI: 10.1016/S0736-5853(05)80062-0.
- [28] T. A. Clarke, "Analysis of the properties of targets used in digital close-range photogrammetric measurement," vol. 2350, pp. 251–262, 1994. DOI: 10.1117/12.189137.
- [29] P. Bender, D. Wilkinson, C. Alley, *et al.*, "The lunar laser ranging experiment," *Science*, vol. 182, no. 4109, pp. 229–238, 1973. DOI: 10.1126/science.182.4109.229.
- [30] Q. Feng, B. Zhang, and C. Kuang, "A straightness measurement system using a single-mode fiber-coupled laser module," *Optics & Laser Technology*, vol. 36, no. 4, pp. 279–283, 2004. DOI: 10.1016/j.optlastec.2003.09.016.
- [31] O. Nakamura, M. Goto, K. Toyoda, N. Takai, T. Kurosawa, and T. Nakamata, "A laser tracking robot-performance calibration system using ball-seated bearing mechanisms and a spherically shaped cat's-eye retroreflector," *Review of scientific instruments*, vol. 65, no. 4, pp. 1006–1011, 1994. DOI: 10.1063/1.1145104.
- [32] M. Dong, L. Xu, J. Wang, P. Sun, and L. Zhu, "Variable-weighted grayscale centroiding and accuracy evaluating," *Advances in Mechanical Engineering*, vol. 5, p. 428 608, 2013. DOI: 10.1155/2013/428608.
- [33] M. R. Shortis and J. W. Seager, "A practical target recognition system for close range photogrammetry," *The Photogrammetric Record*, vol. 29, no. 147, pp. 337–355, 2014. DOI: 10.1111/phor.12070.
- [34] A. Lundvall, F. Nikolajeff, and T. Lindström, "High performing micromachined retroreflector," *Optics Express*, vol. 11, no. 20, pp. 2459–2473, 2003. DOI: 10.1364/OE.11.002459.

- [35] J. Yuan, S. Chang, S. Li, and Y. Zhang, "Design and fabrication of micro-cube-corner array retro-reflectors," *Optics Communications*, vol. 209, no. 1-3, pp. 75–83, 2002. DOI: 10.1016/S0030-4018(02)01630-9.
- [36] M. Garcia-Matos, A. Morono, and E. Hodgson, "KU1 quartz glass for remote handling and LIDAR diagnostic optical transmission systems," *Journal of Nuclear Materials*, vol. 283, pp. 890–893, 2000. DOI: 10.1016/S0022-3115(00)00195-1.
- [37] E. Hecht, *Optics, 4th edition*. Addison-Wesley, San Francisco, 2002.
- [38] Lambda Research Corp., Littleton, MA, *Software OSLO version 20.2.0*. [Online]. Available: <https://www.lambdaresearch.com/oslo/>.
- [39] M. Kalaitzakis, B. Cain, S. Carroll, A. Ambrosi, C. Whitehead, and N. Vitzilaios, "Fiducial markers for pose estimation," *Journal of Intelligent & Robotic Systems*, vol. 101, no. 4, pp. 1–26, 2021. DOI: 10.1007/s10846-020-01307-9.
- [40] A. C. Rice, R. K. Harle, and A. R. Beresford, "Analysing fundamental properties of marker-based vision system designs," *Pervasive and Mobile Computing*, vol. 2, no. 4, pp. 453–471, 2006. DOI: 10.1016/j.pmcj.2006.07.006.
- [41] E. Olson, "AprilTag: A robust and flexible visual fiducial system," in *2011 IEEE international conference on robotics and automation*, IEEE, 2011, pp. 3400–3407. DOI: 10.1109/ICRA.2011.5979561.
- [42] J. Wang and E. Olson, "AprilTag 2: Efficient and robust fiducial detection," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2016, pp. 4193–4198. DOI: 10.1109/IROS.2016.7759617.
- [43] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014. DOI: 10.1016/j.patcog.2014.01.005.
- [44] M. Fiala, "ARTag, a fiducial marker system using digital techniques," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, vol. 2, 2005, pp. 590–596. DOI: 10.1109/CVPR.2005.74.

- [45] A. Xu and G. Dudek, “Fourier tag: A smoothly degradable fiducial marker system with configurable payload capacity,” in *Canadian Conference on Computer and Robot Vision (CRV)*, IEEE, 2011, pp. 40–47. DOI: 10.1109/CRV.2011.13.
- [46] J. Sattar, E. Bourque, P. Giguere, and G. Dudek, “Fourier tags: Smoothly degradable fiducial markers for use in human-robot interaction,” in *Canadian Conference on Computer and Robot Vision (CRV)*, IEEE, 2007, pp. 165–174. DOI: 10.1109/CRV.2007.34.
- [47] J. DeGol, T. Bretl, and D. Hoiem, “Chromatag: A colored marker and fast detection algorithm,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 1472–1481. DOI: 10.1109/ICCV.2017.164.
- [48] L. Egri, H. Nabati, and J. Y. Yu, “RainbowTag: A fiducial marker system with a new color segmentation algorithm,” in *2022 International Conference on Connected Vehicle and Expo (ICCVE)*, IEEE, 2022, pp. 1–6. DOI: 10.1109/ICCVE52871.2022.9743123.
- [49] F. Bergamasco, A. Albarelli, E. Rodola, and A. Torsello, “Rune-tag: A high accuracy fiducial marker with strong occlusion resilience,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 113–120. DOI: 10.1109/CVPR.2011.5995544.
- [50] L. Calvet, P. Gurdjos, C. Griwodz, and S. Gasparini, “Detection and accurate localization of circular fiducials under highly challenging conditions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 562–570. DOI: 10.1109/CVPR.2016.67.
- [51] W. A. Hoff, K. Nguyen, and T. Lyon, “Computer-vision-based registration techniques for augmented reality,” vol. 2904, pp. 538–548, 1996. DOI: 10.1117/12.256311.
- [52] J. Yu, W. Jiang, Z. Luo, and L. Yang, “Application of a vision-based single target on robot positioning system,” *Sensors*, vol. 21, no. 5, p. 1829, 2021. DOI: 10.3390/s21051829.

- [53] F. Bergamasco, A. Albarelli, and A. Torsello, “Image-space marker detection and recognition using projective invariants,” in *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, 2011, pp. 381–388. DOI: 10.1109/3DIMPVT.2011.55.
- [54] Bergamasco, Filippo and Albarelli, Andrea and Torsello, Andrea, “Pi-Tag: A fast image-space marker design based on projective invariants,” *Machine vision and applications*, vol. 24, no. 6, pp. 1295–1310, 2013. DOI: 10.1007/s00138-012-0469-6.
- [55] T. Krajník, M. Nitsche, J. Faigl, *et al.*, “A practical multirobot localization system,” *Journal of Intelligent & Robotic Systems*, vol. 76, no. 3, pp. 539–562, 2014. DOI: 10.1007/s10846-014-0041-x.
- [56] T. Krajník, M. Nitsche, J. Faigl, T. Duckett, M. Mejail, and L. Preucil, “External localization system for mobile robotics,” pp. 1–6, 2013. DOI: 10.1109/ICAR.2013.6766520.
- [57] D. Claus and A. W. Fitzgibbon, “Reliable automatic calibration of a marker-based position tracking system,” vol. 1, pp. 300–305, 2005. DOI: 10.1109/ACVMOT.2005.101.
- [58] Claus, David and Fitzgibbon, Andrew W, “Reliable fiducial detection in natural scenes,” in *European Conference on Computer Vision*, 2004, pp. 469–480. DOI: 10.1007/978-3-540-24673-2\_38.
- [59] J. Meza, L. A. Romero, and A. G. Marrugo, “Markerpose: Robust real-time planar target tracking for accurate stereo pose estimation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1282–1290. DOI: 10.1109/CVPRW53098.2021.00141.
- [60] L. B. Gatrell, W. A. Hoff, and C. W. Sklair, “Robust image features: Concentric contrasting circles and their image extraction,” in *Cooperative Intelligent Robotics in Space II*, vol. 1612, 1992, pp. 235–244. DOI: 10.1117/12.56761.
- [61] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003. DOI: 10.1017/CBO9780511811685.020.
- [62] J. Mallon and P. F. Whelan, “Which pattern? biasing aspects of planar calibration patterns and detection methods,” *Pattern recognition letters*, vol. 28, no. 8, pp. 921–930, 2007. DOI: 10.1016/j.patrec.2006.12.008.

- [63] D. Hu, D. DeTone, and T. Malisiewicz, “Deep charuco: Dark charuco marker pose estimation,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8436–8444. DOI: 10.1109/CVPR.2019.00863.
- [64] D. DeTone, T. Malisiewicz, and A. Rabinovich, “Superpoint: Self-supervised interest point detection and description,” in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 224–236. DOI: 10.1109/CVPRW.2018.00060.
- [65] M. Ye, M. Kanski, D. Yang, *et al.*, “Deeptag: An unsupervised deep learning method for motion tracking on cardiac tagging magnetic resonance images,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021, pp. 7261–7271. DOI: 10.1109/TPAMI.2022.3174603.
- [66] J. Chai, H. Zeng, A. Li, and E. W. Ngai, “Deep learning in computer vision: A critical review of emerging techniques and application scenarios,” *Machine Learning with Applications*, vol. 6, p. 100 134, 2021. DOI: 10.1016/j.mlwa.2021.100134.
- [67] W. Liu, D. Anguelov, D. Erhan, *et al.*, “SSD: Single shot multibox detector,” in *European conference on computer vision*, Springer, 2016, pp. 21–37. DOI: 10.1007/978-3-319-46448-0\_2.
- [68] R. Girshick, “Fast R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448. DOI: 10.1109/ICCV.2015.169.
- [69] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, “Superglue: Learning feature matching with graph neural networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4938–4947. DOI: 10.1109/CVPR42600.2020.00499.
- [70] K. M. Yi, E. Trulls, Y. Ono, V. Lepetit, M. Salzmann, and P. Fua, “Learning to find good correspondences,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2666–2674. DOI: 10.1109/CVPR.2018.00282.



- [71] M. Donoser and H. Bischof, "Efficient maximally stable extremal region (MSER) tracking," in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, vol. 1, 2006, pp. 553–560. DOI: 10.1109/CVPR.2006.107.
- [72] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001. DOI: 10.1023/A:1010933404324.
- [73] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000. DOI: 10.1109/34.888718.
- [74] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics*, vol. 2, no. 2, pp. 164–168, 1944. DOI: 10.1090/QAM/10666.
- [75] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963. DOI: 10.1137/0111030.
- [76] O. Rodriguez, "Des lois geometriques qui regissent les déplacements d'un systeme solide dans l'espace et de la variation des coordonnees provenant de déplacements consideres independamment des causes qui peuvent les produire," *J Mathematiques Pures Appliquees*, vol. 5, pp. 380–440, 1840.
- [77] M. Shah, "Solving the robot-world/hand-eye calibration problem using the kronecker product," *Journal of Mechanisms and Robotics*, vol. 5, no. 3, p. 031 007, 2013. DOI: 10.1115/1.4024473.
- [78] I. Ali, O. J. Suominen, E. R. Morales, and A. Gotchev, "Multi-view camera pose estimation for robotic arm manipulation," *IEEE Access*, vol. 8, pp. 174 305–174 316, 2020. DOI: 10.1109/ACCESS.2020.3026108.



# PUBLICATION I

## **Robust vision using retro reflective markers for remote handling in ITER**

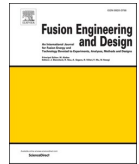
L. G. Ribeiro, O. J. Suominen, S. Peltonen, E. R. Morales, and A. Gotchev

Fusion Engineering and Design, vol. 161, p. 112 080

DOI: 10.1016/j.fusengdes.2020.112080

**Publication reprinted with the permission of the copyright holders.**





# Robust vision using retro reflective markers for remote handling in ITER

Laura Gonçalves Ribeiro<sup>a,\*</sup>, Olli J. Suominen<sup>a</sup>, Sari Peltonen<sup>a</sup>, Emilio Ruiz Morales<sup>b</sup>,  
Atanas Gotchev<sup>a</sup>

<sup>a</sup> Faculty of Information Technology and Communication, Tampere University, 33720 Tampere, Finland

<sup>b</sup> Fusion for Energy (F4E), ITER Delivery Department, Remote Handling Project Team, 08019 Barcelona, Spain

## ARTICLE INFO

### Keywords:

ITER  
Remote handling  
Retro-reflective markers  
Eye-in-hand  
Pose estimation

### MSC:

00-01  
99-00

## ABSTRACT

ITER's working environment is characterized by extreme conditions, that deem maintenance and inspection tasks to be carried out through remote handling. 3D Node is a hardware/software module that extracts critical information from the remote environment during fine alignment tasks using an eye-in-hand camera system and updates the models behind the virtual reality-based remote handling platform. In this work we develop a retro-reflective marker-based version of 3D Node that estimates the pose of a planar target, the knuckle of the divertor cassette locking system, using the markers attached to its surface. We demonstrate a pin-tool insertion task using these methods. Results show that our approach works reliably with a single low-resolution camera and outperforms the previously researched stereo depth estimation based approaches. We conclude that retro-reflective marker-based tracking has the potential to be a key enabler for remote handling operations in ITER.

## 1. Introduction

ITER is a worldwide research undertaking that aims at demonstrating the feasibility of fusion as a large-scale source of energy. ITER's experimental machine, the tokamak, has, at its core, a vacuum chamber where fusion occurs. Below the vacuum chamber is the divertor that extracts the heat and ash produced by the fusion reaction and protects the surrounding walls. During the lifetime of the reactor, the activated components of the divertor have to be changed, inspected and repaired. Since operator access is restricted in the activated areas, maintenance is performed through remote handling (RH). In particular, the 54 cassette assemblies that comprise the divertor need to be installed and replaced by a remotely controlled robot manipulator that is attached to and carried in by a transporter (cask). After removal, the divertor cassettes are placed in the transporter and taken to another facility, the Hot-Cell. To remove and replace cassettes, it is necessary to operate their locking mechanism, the knuckle, using several tools: a wrench tool, a pin tool, and a jack [1]. The tools must be driven into the respective slots in the surface of the knuckle with narrow gaps and tight tolerances.

ITER's RH is based on tele-operation, which requires the operator to be, at every stage, in the loop and fully in control. In order to perform appropriate actions, the operator has to be presented with all relevant information that portrays accurately the current status of the system. With this purpose, the environment integrates an eye-in-hand camera

system to enable augmenting the existing virtual reality models with updated information of the environment. Due to the high radiation levels in the divertor, conventional cameras cannot be used and studies are ongoing on the development of adequate sensors and optical elements to operate in the ITER environment [2,3]. As it is not yet clear what will be the exact characteristics of ITER's cameras, we assume they will likely provide a low resolution output and a low signal to noise ratio, that degrades over the lifetime of the camera [4], since those are the characteristics of the current analog radiation tolerant cameras. The poor visibility conditions, associated with the lack of some of the natural depth cues, deem a manual alignment based on the camera feeds unreliable. It is necessary to augment the operator's perception during fine alignment tasks [5], for which a reliable, automatic estimate of the alignment parameters is required.

To address the challenges described herein, the 3D node system was proposed in earlier works [6–8]. This solution used the information provided by a stereo camera pair to estimate the geometry of the knuckle through stereo matching. The alignment between the sensed and reference point clouds was obtained using an iterative closest point (ICP) algorithm. Even though the results were promising, the solutions were not robust enough. Some outlier positions were reported, for which the estimated poses were completely in disagreement with the acquired data. Although the reported repeatability of the system is in the  $\leq 3$  mm error range for operating distances of 500 to 1000 mm, the absolute

\* Corresponding author.

E-mail address: [laura.goncalvesribeiro@tuni.fi](mailto:laura.goncalvesribeiro@tuni.fi) (L. Gonçalves Ribeiro).

accuracy of the system was not reported. In summary, even though the poses provided by the 3D node seem to be fairly consistent with each other after outlier removal, they do not provide an estimate of what is the agreement between the estimated and the real value of the measures. Furthermore, it has been recently assessed that due to space constraints, the system must be simplified to a single camera and operate in the 300–500 mm distance range with an absolute accuracy of  $\pm 3$  mm. For the aforementioned reasons, there is a need to develop the 3D node further and find a methodology to estimate its absolute accuracy.

This paper addresses the challenges of the earlier system and proposes a more robust marker-based solution. Markers are often used in photogrammetry for accurate and reliable identification and matching of features [9]. They provide higher contrast than most naturally occurring structures and are particularly relevant when dealing with large uniform areas and specular surfaces. Specular surfaces differ in appearance with viewpoint and pose a significant challenge to feature detection. The use of retro-reflective markers can further enhance contrast under certain light conditions. These elements have the advantage of offering a particularly high signal to noise ratio, since they reflect most incident light back to the source for a wide range of incident directions. We hypothesize that the introduction of retro-reflective marker tracking in 3D node will greatly increase its robustness and accuracy.

This paper is organized as follows: Section 2 introduces the architecture of 3D node 2.0. In Section 3, we describe our marker based approach. In Section 4 we evaluate the performance of our implementation against the former solution. In Section 5 we draw conclusions and outline future development prospects.

## 2. System architecture

The architecture of the 3D node has been adapted to include marker-based tracking and is presented in Fig. 1. The new system architecture can obtain an estimate of the alignment using a single camera. 3D node 2.0 receives images in raw format from the camera system through a GigE vision interface. It also receives the information of the pose of the robot end-effector in the coordinate system of the robot base from the manipulator control system (CS). The calibration module provides the required calibration parameters, which are divided into two groups: hand-eye calibration and camera system calibration. Hand-eye calibration parameters correspond to the rigid body transformation between the tool center point (TCP) of the robotic manipulator and the optical axis of the camera. Camera calibration parameters can be divided into intrinsic and extrinsic parameters. Intrinsic parameters are the internal parameters of each camera (focal length, principal axis, and lens

distortion parameters). Extrinsic parameters correspond to the rigid body transformation between the optical axes of the two cameras. The calibration of the system is performed offline, outside of the operating environment, in the Hot-Cell facility. The calibration module is composed of the automatic calibration image capture stage and estimation stage. In the capturing stage, calibration images and corresponding TCP poses are stored in memory. At estimation stage, the information is loaded and hand-eye and camera calibration are computed independently.

The pipeline of 3D Node 2.0 is divided into sequential stages: image linearization, marker detection and identification, and pose estimation. The model of the target (in particular, the relative positions of the marker objects) is a required input to the position estimation step. The estimated target pose and the captured images are used to provide feedback to the operator through the Graphical User Interface (GUI). Likewise, the operator provides the manipulation commands to the robot CS through the GUI.

## 3. Methodology

### 3.1. Retro-reflective marker prototype

In this section, we describe the developed retro-reflector marker design that complies with the requirements of the ITER operating environment. These requirements can be summarized as follows:

- Operating temperature up to 200 degrees Celsius.
- Acceptance for vacuum quality classification 1A and 1B.
- Tolerant to the radiation in the environment. Furthermore, the radiation in the environment should not alter the properties of the material in such a way that compromises its intended functionality.
- Size of the marker should be sufficiently small, such that mounting multiple markers into the cassette is feasible. Moreover, the maximum allowed embedment depth of markers into the cassette is 10 mm.

The operating environment is also characterised by the presence of dust. However, dust is not expected to settle significantly on the observed surfaces or camera lens, since these surfaces are mostly oriented vertically. Therefore, we do not consider in this work dust as one of the environment requirements.

In our market study, we found no commercially-available marker that satisfied these requirements and had a suitable performance. Marker performance was evaluated empirically in a practical test setup similar to what would be used in the final version of the system. As an

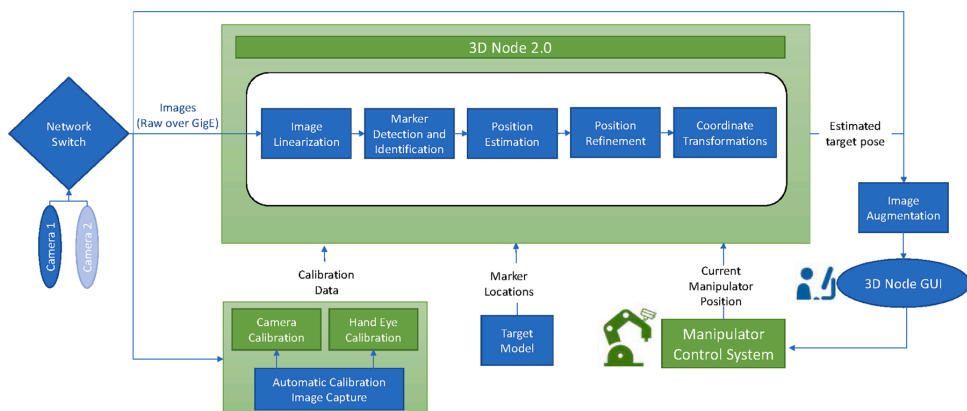


Fig. 1. Architecture of the 3D node 2.0.

alternative approach, we develop a custom made marker object, consisting of an array of KU1 quartz glass beads held by a stainless steel structure. Stainless steel is the main material component in ITER and satisfies all requirements and KU1 quartz glass (fused silica) is approved as a material for all vacuum classes. Moreover, according to the study in [10], gamma radiation should not induce significant changes to the refractive index of KU1 glass in the relevant wavelengths (380–700 nm). The reported changes do not influence our application since our measurements are solely intensity-based and are not affected by minor intensity loss.

The supporting metal structure is a hollow cylinder with 7 mm depth, 14 mm diameter and M14 threads on the outside surface (Fig. 2b). The front surface is a sieve-like structure with 12 holes with 1.5 mm diameter. This design provides attachment and protects the glass beads while simultaneously providing a means of attaching the marker to the cassette. Eight corresponding M14 threads have been incorporated in the cassette design and their relative positions can be seen in (Fig. 2a). The unified outer shape allows the necessary mounting holes in the cassette to be pre-determined, while allowing future marker development and improvement. Furthermore, the proposed design solution makes it possible to easily change the markers in case any of them sustain damage during operation.

In this section, we have demonstrated that it is possible to manufacture a working retro-reflective marker using ITER compliant materials. The described marker was shown empirically to work for the intended purpose, i.e. the contrast between the marker and the surroundings was considered satisfactory in the conditions of the test setup. However, a more dedicated study would be required to maximize the brilliancy of the markers and contrast in the images.

### 3.2. 3D node 2.0 pipeline

The pipeline of 3D node 2.0 starts with the compensation for lens distortion of the captured images. At the undistortion stage, camera calibration parameters provided by the calibration module are used to correct the lens's optical distortions.

The following stage in the pipeline is marker detection. Image pixels are segmented into the categories of *marker* or *not a marker* by adaptive thresholding. The applied thresholding method [11] can compensate for some illumination variations across the image by comparing each pixel to the values of its neighbours. Due to the specular nature of the

knuckle's surfaces, some background regions are classified as *marker* at this stage, a problem that is addressed later by morphological filtering. The discernible small white elements of each marker array are merged into a single elliptic shape by image processing operations. We apply morphological closing and flood-fill operations [12] consecutively, followed by determination of the binary convex hull image. The blobs in the resulting image are identified by connected component labelling [13]. False blobs are excluded based on their geometric properties (such as area and eccentricity). We find the coordinates of the centroid of each detected blob. Assuming that the target and camera have been roughly aligned by the operator allows us to make the approximation that the centroid of the detected blobs roughly corresponds to the centre of the original circles.

Next we must establish a one-to-one correspondence between the known relative marker positions in the world coordinate frame and the detected image coordinates. Markers are ordered and identified according to their relative Euclidean distances. This methods works only for small rotations of the target, since the distance between image points is not a perspective invariant.

Once the planar point correspondences have been established, we use a homography decomposition approach [14] to determine the initial estimate of the rigid body transformation between the camera and the target. At this stage, we check the reprojection error provided by the current estimate. If it is higher than a few pixels, we go back to the marker identification and ordering stage and ask the operator to intervene in the selection of the thresholding and filtering parameters. The initial pose estimate is further optimized through minimization of the reprojection error using the Levenberg-Marquardt algorithm [15,16]. In the case of stereo cameras, we take as initial estimate the mean value of the estimates provided by the two cameras and optimize the results based on the point correspondences from both images.

The possibility of optimizing over several images taken by the same camera at different poses is studied in another work [17]. In this work two multiview estimation methods have been considered: (1) using the poses reported by the robot's CS and (2) not using those. Further on, we provide a comparison with those methods, referred to as "Multiview 1" and "Multiview 2", respectively.

The calculated target pose in the coordinate system of the camera must be converted to valid commands in the coordinate system of the robotic arm. For this, an estimation of the hand-eye transformation is required. A detailed survey of hand-eye calibration methods is presented

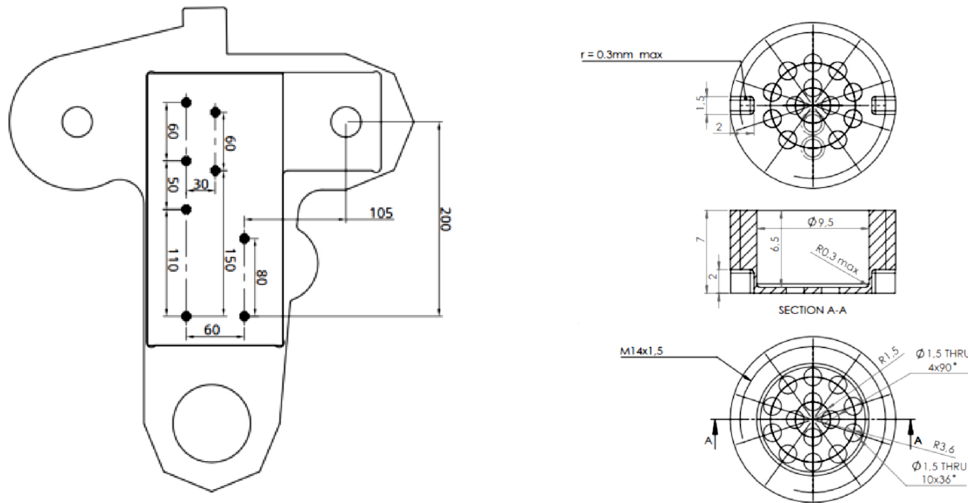


Fig. 2. Model of target prototype.

in [18]. From the surveyed methods we have utilized here the method of Shah [19].

#### 4. Experiments and results

##### 4.1. Equipment and setup

The KUKA KR 16 L6-2 industrial robot is used in our experimental setup to simulate ITER's robotic arm, which is still under development. In our laboratory, the robot is fixed to a static platform, contrary to the ITER design, in which the arm is fixed to a structure that moves on rails. KUKA has a lower payload capacity, requiring the use of lighter weight mock-up versions of the tools. A version of the pin-tool was created for demonstration purposes, where a pipe segment with the corresponding diameter represents the pin. On the right side of Fig. 3 we show the end-effector of the manipulator and the prototype pin-tool.

As a target, we use a 1:1 scale 3D printed replica of the Cassette Locking System (CLS). Since the conditions in the laboratory environment do not pose the same strict material restrictions as the application, a simplified version of the retro reflective markers and their holders was made for testing. In the simplified version, the bead holding arrays are laser cut into a 1.0 mm thick sheet of metal, as shown in Fig. 2a, making up a single holding plate. Prototype beads are made of borosilicate glass. This material does not comply with the ITER environment requirements, but is used to simulate the optical behaviour of fused silica at a cheaper cost. The metal sheet is attached to a back plate made of polycarbonate that simulates the knuckle geometry and is fitted onto the CLS replica. An image of the prototype is shown on the left side of Fig. 3.

A stereo pair of industrial Basler acA1920-48gc cameras was rigidly attached to KUKA's end effector, as shown on the right side of Fig. 3. The cameras have a native resolution of 1920x1200 pixels with an RGB Bayer filter to produce colour images. To emulate the use of radiation tolerant cameras, we convert the images to grayscale and downsample the resolution by a factor of three to 640 × 400 pixels. The cameras are rigidly attached to an aluminium holding piece with a baseline of 200 mm. The baseline is dictated by the size of the lighting elements and the construction of the mock-up tool holder. The design of the mounting is not indicative of the one to be used in the ITER environment, and a different baseline can be designed as needed. The lenses are high quality, low distortion wide angle lenses with a nominal 6 mm focal length. With these particular cameras, the lenses produce an opening angle of roughly

82 degrees. Two Smart Vision RC130 industrial ring lights were fitted to the camera pair, as closely as possible to the camera lenses. These lamps create a relatively uniform light pattern that covers the entire target for an acceptable variety of camera poses.

##### 4.2. Accuracy and precision of 3D node 2.0

The ground-truth values for assessing the accuracy of the system are obtained by manually positioning the robot's end effector at the target origin. This allows us to calculate, for each test pose, the ground-truth transformation between the camera and the target, based on the values reported by the robot's CS.

Given a rigid body transformation,  $T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$ , and the corresponding ground-truth transformation,  $T^{GT} = \begin{bmatrix} R^{GT} & t^{GT} \\ 0 & 1 \end{bmatrix}$ , where  $R$  represents the  $3 \times 3$  rotation matrix and  $t$  represents the  $3 \times 1$  translation vector, we calculate the translation error vector as

$$t^e = t - t^{GT} \quad (1)$$

We calculate absolute translation error ( $e_t$ ) in millimetres and the absolute rotation error ( $e_R$ ) in degrees as follows

$$e_t = \sqrt{(t_x^e)^2 + (t_y^e)^2} \quad (2)$$

$$e_R = |\text{angle}(R) - \text{angle}(R^{GT})| \quad (3)$$

As seen in Eq. (1), we evaluate translation error as the Euclidean distance in each of the two axes that are parallel to the target plane. We disregard the depth axis, since it is not relevant for the alignment task.

In Eq. (3),  $\text{angle}()$  represents the conversion from a rotation matrix to axis-angle representation and consideration of only the attitude quaternion.

We evaluate performance over a set of images captured from 20 random poses within the following limits:

- Distance of 300 to 500 mm from the target.
- Horizontal displacement from the centre of the target smaller than 50 mm.
- Vertical displacement from the centre of the target smaller than 100 mm.

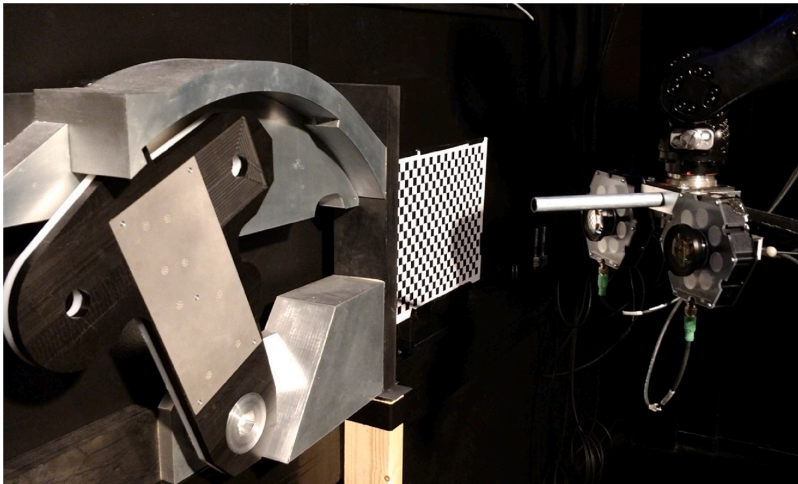


Fig. 3. Illustration of the experimental setup. On the left side, the mock-up target object. On the right side, the camera system, ring lamps and mock-up pin-tool, attached to the end effector of the KUKA robot manipulator.



- Rotations in relation to the aligned pose smaller than 10 degrees.
- The target is fully visible by both cameras.

The  $x$  and  $y$  components of the error vector ( $t^e$ ) are shown in Fig. 4 for the proposed marker-based monocular and stereoscopic approaches. In our visual representations, the red dashed circle is centred in the mean error value and its radius corresponds to the maximum distance between the observations and the center point. The black circle is centred in (0,0) and its diameter corresponds to the maximum allowed error. The requirements of the application have been fulfilled when the dashed red circle is contained within the black circle.

Analysing the results in Fig. 4, we see that the application requirements have been fulfilled and errors are distributed within a fairly narrow region even though we tend to consistently misestimate translation by about 2 mm in both the  $X$  and  $Y$  directions.

#### 4.3. Markerless approach

In the earlier version of 3D Node [7] the pose estimation problem was tackled by fitting a stereoscopic depth estimate to a known CAD model using an ICP algorithm. To compare with this approach, we captured the same sequence of robot poses with the same scene object, but with increased lighting to enable the stereo cameras to see the objects. We give the ground truth pose with some added noise as an initial estimate to the algorithm to make sure it converges on the correct solution, so we can reliably measure its accuracy without concern about parameter tuning.

We observe that with small initial offsets, it converges closely to the expected values (Fig. 5, left side) though it results in a spread of  $\pm 3$  mm, barely within the specification. With larger amounts of offset (Fig. 5, right side), the convergence deteriorates quickly and no longer ends up with correct estimates.

#### 4.4. Marker-based multiview pose estimation

To be able to establish a comparison with the marker-based multiview estimation methods of [17] (Multiview 1 and Multiview 2), for each image in the dataset described in Section 4.1, we record an extra set of 5 random poses.

Results are presented in Fig. 6. We observe that both methods work quite consistently, despite method 1 having one outlier, for which the error is significantly higher.

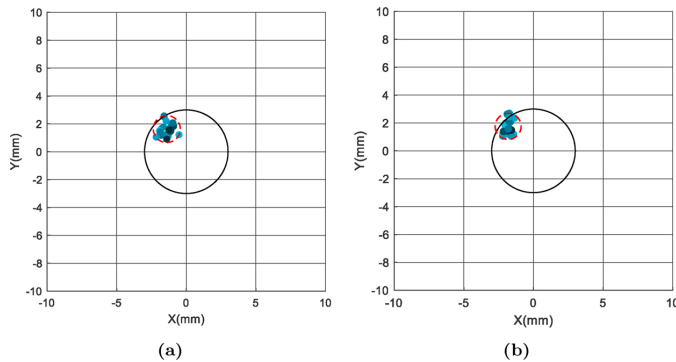


Fig. 4. Representation of absolute error in each of the  $X$  and  $Y$  axis for the estimation provided by left most camera (a) and stereo (b). Each colour point represents a different pose of the camera in relation to the target. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

#### 4.5. Comparative analysis

Comparative results are presented in Tables 1 and 2. In these tables, precision refers to the comparison of an estimated value with the average of estimates, while accuracy refer to the comparison to the ground truth. We present the average and maximum error values among the 20 observations for all evaluated approaches. To compute the statistics in the tables we use the results of ICP with smaller initial offsets (corresponding to the left side of Fig. 5) and exclude the outlier seen in the left side of Fig. 6.

From the analysis of the information in the tables, we see that both marker-based methods, either with one or two cameras, outperform the earlier version of the system. The error of both single and stereo marker-based methods is about three times lower than ICP's. The earlier version has a slightly lower maximum angular error compared to the single camera marker, but only to the scale of 0.1–0.2 degrees.

In summary, the new approach produces much more consistent results than the previous one. Though not a part of this numerical analysis, the new marker based method is also much more robust in finding the correct alignment, and does not in fact require any initial estimates to function properly.

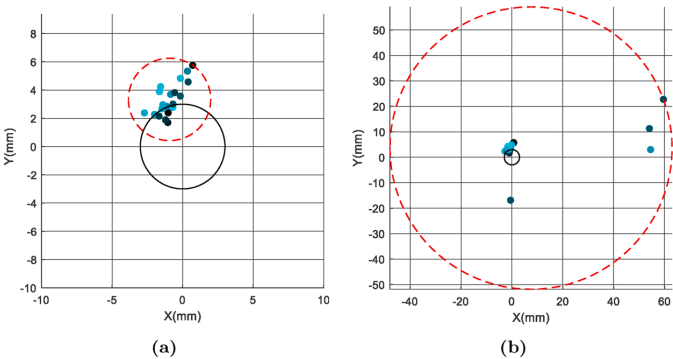
Stereoscopic estimation with markers performs slightly better on the metrics of orientation estimation, but does not perform overall significantly better than monocular and it actually loses to monocular in average translation accuracy.

Regarding the approaches using multiple shots taken with the same camera, both approaches outperform monoscopic estimation in maximum angular error. Method 1 estimates rotation consistently better than Method 2. However, Method 1 shows the highest precision of all methods in estimating translation, although the difference is not very significant.

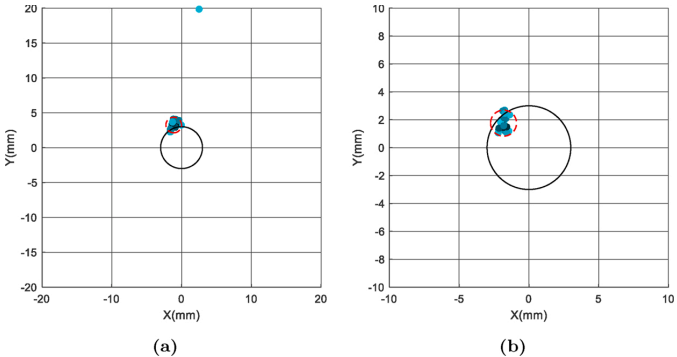
We believe that the offset of about 2 mm in the  $x$  and  $y$  axes that can be seen in the results is due to errors in the poses reported by the robotic arm's CS. This is likely due to the fact that the robot was uncalibrated. These hypotheses were not confirmed, and further study is required.

#### 4.6. Pin-tool insertion demonstration

We set up a proof of concept demonstration of the insertion of the pin-tool in the respective slot in the cassette surface based on the estimate provided by the single-camera method. In the demonstration, the manipulator moves autonomously to the aligned pose, based on the estimate provided by 3D Node 2.0 (Fig. 7b). The aligned pose is such



**Fig. 5.** Representation of absolute error in each of the X and Y axis for the estimation provided by the stereo ICP based method from earlier work, with different amounts of initial pose deviation from ground truth. Offset values of  $|\Delta a| = 5$  deg,  $|\Delta t| = 25$  mm (a). Offset values of  $|\Delta a| = 10$  deg,  $|\Delta t| = 50$  mm (b). Each colour point represents a different pose of the camera in relation to the target. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)



**Fig. 6.** Representation of absolute error in each of the X and Y axis for the estimation provided by the multiview method 1(a) and 2 (b). Each colour point represents a different pose of the camera in relation to the target. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**

Average and maximum error of estimate translations obtained through the comparison with the average of estimates (precision) and ground truth values (accuracy). The values in bold signify the lowest error in each category.

$e_t$ (mm)	Accuracy		Precision	
	Average	Maximum	Average	Maximum
Monoscopic	<b>2.1944</b>	3.0295	0.5597	0.9691
Stereoscopic	2.2924	<b>2.9813</b>	0.5033	0.9707
Stereoscopic ICP [7]	3.6017	5.7915	1.2287	2.9162
Multiview 1	3.4751	4.2541	0.6178	1.1204
Multiview 2	2.5606	3.2318	<b>0.4797</b>	<b>0.9304</b>

that the tool and target planes are aligned, but the tool is still at a distance from the target. After this movement is completed, the operator confirms the approximation to the target and the tool is inserted (Fig. 7c). The demonstration is performed for the same set of 20 randomly chosen poses as described above. The insertion of the tool was successful for all the poses.

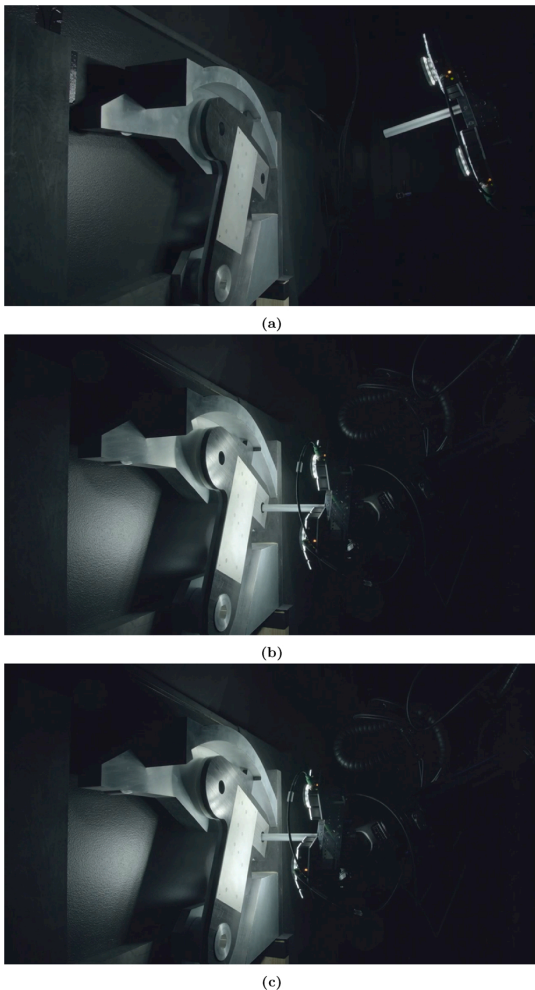
**Table 2**

Average and maximum error of estimate rotations obtained through the comparison with the average of estimates (precision) and ground truth values (accuracy). The values in bold signify the lowest error in each category.

$e_R$ (°)	Accuracy		Precision	
	Average	Maximum	Average	Maximum
Monoscopic	0.4105	1.6886	0.2421	1.2781
Stereoscopic	<b>0.3291</b>	<b>0.7950</b>	<b>0.1563</b>	<b>0.4659</b>
Stereoscopic ICP [7]	0.5367	1.5510	0.3098	1.0143
Multiview 1	0.4735	1.0074	0.2292	0.5339
Multiview 2	0.5381	1.2164	0.2590	0.6783

**5. Conclusions and future work**

In this work, we have demonstrated a retro-reflective marker design that complies with both vacuum class and radiation requirements of ITER's operating environment. We showed that the proposed retro-reflective marker-based implementation of 3D node works more reliably and provides more accurate results than its earlier version. The overall performance of 3D node 2.0 is within the required tolerance ( $< \pm 3$  mm) both in the single camera and stereo camera modes. In our



**Fig. 7.** Demonstration of pin-tool insertion. Random initial pose (a), aligned pose (b), final pose (c). These images are screenshots from the demonstration video. For access to the demonstration video, the reader is referred to the web version of this article.

results, for the tested viewing range (300–500 mm) and baseline value (200 mm), there is no major difference in performance between the monocular, stereoscopic and multi-view approaches. We can conclude that the proposed marker-based approach in any of its modes shows potential to be a key enabler for RH in ITER.

Further study is needed on the refinement and optimization of all elements in the system, in particular: (1) optimization of the retro reflective marker design to maximize brilliancy for the observation angle, and (2) use of more advanced approaches for marker identification. Further work could also quantify the comparative performance of retro-reflectors and standard diffuse markers in the ITER environment, once characteristics of the environment, such as existence of ash, dust or other particles, has been determined. Further work could also address how the calculated alignment (translation vector and rotation matrix) can be optimally presented to the operator and integrated within the existent augmented and virtual reality solutions.

## Authors' contributions

**Laura Gonçalves Ribeiro:** Conceptualization, Methodology, Software, Validation, Writing – original draft.

**Olli J. Suominen:** Conceptualization, Methodology, Project Administration, Supervision, Writing – review & editing.

**Sari Peltonen:** Supervision, Writing – review & editing.

**Emilio Ruiz Morales:** Project Administration, Writing – review & editing.

**Atanas Gotchev:** Supervision, Writing – review & editing.

## Conflict of interest

The authors declare no conflict of interest.

## Declaration of Competing Interest

The authors report no declarations of interest.

## Acknowledgements

The work in this paper was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska Curie grant agreement No. 764951, Immersive Visual Technologies for Safety-Critical Applications and by Fusion for Energy (F4E) and Tampere University under the F4E grant contract F4E-GRT-0901. The research infrastructure of Centre for Immersive Visual Technologies (CIVIT) at Tampere University provided the robotic manipulator, sensors and laboratory space for conducting the experiments. This publication reflects the views only of the authors, and Fusion for Energy cannot be held responsible for any use which may be made of the information contained herein.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.fusengdes.2020.112080>.

## References

- [1] V. Lyytikäinen, P. Kinnunen, J. Koivumäki, J. Mattila, M. Siuko, S. Esque, J. Palmer, Divertor cassette locking system remote handling trials with WHMAN at DTP2, *Fusion Eng. Des.* 88 (9–10) (2013) 2181–2185, <https://doi.org/10.1016/j.fusengdes.2013.02.114>.
- [2] V. Goiffon, S. Rizzolo, F. Corbière, S. Rolando, S. Bounasser, M. Sergeant, A. Chabane, O. Marcelot, M. Estribeau, P. Magnan, et al., Total ionizing dose effects on a radiation-hardened CMOS image sensor demonstrator for ITER remote handling, *IEEE Trans. Nucl. Sci.* 65 (1) (2017) 101–110, <https://doi.org/10.1109/TNS.2017.2765481>.
- [3] T. Allanche, P. Paillet, V. Goiffon, C. Muller, M. Van Uffelen, L. Mont-Casellas, O. Duhamel, C. Marcandella, S. Rizzolo, P. Magnan, et al., Vulnerability and hardening studies of optical and illumination systems at MGy dose levels, *IEEE Trans. Nucl. Sci.* 65 (1) (2017) 132–140, <https://doi.org/10.1109/TNS.2017.2783187>.
- [4] R. King, D. Hamilton, Augmented virtualised reality-applications and benefits in remote handling for fusion, *Fusion Eng. Des.* 84 (2–6) (2009) 1055–1057, <https://doi.org/10.1016/j.fusengdes.2009.02.031>.
- [5] Z. Ziaei, A. Hahto, J. Mattila, A. Gotchev, E. Ruiz, Semeraro, Real-time markerless augmented reality for remote handling system in bad viewing conditions, *Fusion Eng. Des.* 86 (9–11) (2011) 2033–2038, <https://doi.org/10.1016/j.fusengdes.2010.12.082>.
- [6] L. Niu, L. Aha, J. Mattila, A. Gotchev, E. Ruiz, A stereoscopic eye-in-hand vision system for remote handling in ITER, *Fusion Eng. Des.* 146 (2019) 1790–1795, <https://doi.org/10.1016/j.fusengdes.2019.03.036>.
- [7] L. Niu, S. Smirnov, J. Mattila, A. Gotchev, E. Ruiz, Robust pose estimation with a stereoscopic camera in harsh environments, *Electron. Imaging* 2018 (9) (2018) 126–131, <https://doi.org/10.2352/ISSN.2470-1173.2018.09.IRIACV-126>.
- [8] L. Niu, O. Suominen, M.M. Aref, J. Mattila, E. Ruiz, S. Esque, Eye-in-hand manipulation for remote handling: experimental setup, *IOP Conference Series: Materials Science and Engineering*, vol. 320 (2018), <https://doi.org/10.1088/1757-899X/320/1/012007>.
- [9] T.A. Clarke, Analysis of the properties of targets used in digital close-range photogrammetric measurement, in: S.F. El-Hakim (Ed.), *Videometrics III*, vol.

- 2350, International Society for Optics and Photonics, SPIE, 1994, pp. 251–262, <https://doi.org/10.1117/12.189137>.
- [10] M. Garcia-Matos, A. Morono, E. Hodgson, KU1 quartz glass for remote handling and LIDAR diagnostic optical transmission systems, *J. Nucl. Mater.* 283 (2000) 890–893, [https://doi.org/10.1016/S0022-3115\(00\)00195-1](https://doi.org/10.1016/S0022-3115(00)00195-1).
- [11] D. Bradley, G. Roth, Adaptive thresholding using the integral image, *J. Graphics Tools* 12 (2) (2007) 13–21, <https://doi.org/10.1080/2151237X.2007.10129236>.
- [12] P. Soille, *Morphological Image Analysis: Principles and Applications*, Springer Science & Business Media, 2013, <https://doi.org/10.1007/978-3-662-05088-0>.
- [13] R.M. Haralick, L.G. Shapiro, *Computer and Robot Vision*, vol. 1, Addison-Wesley, 1992, pp. 28–48.
- [14] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2003, <https://doi.org/10.1017/CBO9780511811685.020>.
- [15] K. Levenberg, A method for the solution of certain non-linear problems in least squares, *Q. Appl. Math.* 2 (2) (1944) 164–168, <https://doi.org/10.1090/qam/10666>.
- [16] D.W. Marquardt, An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Ind. Appl. Math.* 11 (2) (1963) 431–441, <https://doi.org/10.1137/0111030>.
- [17] I. Ali, O.J. Suominen, E.R. Morales, A. Gotchev, Multi-view camera pose estimation for robotic arm manipulation, *IEEE Access* 8 (2020) 174305–174316, <https://doi.org/10.1109/ACCESS.2020.3026108>.
- [18] I. Ali, O. Suominen, A. Gotchev, E.R. Morales, Methods for simultaneous robot-world-hand-eye calibration: a comparative study, *Sensors* 19 (12) (2019) 2837, <https://doi.org/10.3390/s19122837>.
- [19] M. Shah, Solving the robot-world/hand-eye calibration problem using the kronecker product, *J. Mech. Robot.* 5 (3) (2013), <https://doi.org/10.1115/1.4024473>.

# PUBLICATION II

## **Retro-reflective-marker-aided target pose estimation in a safety-critical environment**

L. G. Ribeiro, O. J. Suominen, A. Durmush, S. Peltonen, E. Ruiz Morales, and  
A. Gotchev






Applied Sciences, vol. 11, no. 1, p. 3  
DOI: 10.3390/app11010003

**Publication reprinted with the permission of the copyright holders.**



## Article

# Retro-Reflective-Marker-Aided Target Pose Estimation in a Safety-Critical Environment

Laura Gonçalves Ribeiro <sup>1,\*</sup> , Olli J. Suominen <sup>1</sup> , Ahmed Durmush <sup>1</sup> , Sari Peltonen <sup>1</sup> , Emilio Ruiz Morales <sup>2</sup> and Atanas Gotchev <sup>1</sup> 

<sup>1</sup> Faculty of Information Technology and Communication Sciences, Tampere University, 33720 Tampere, Finland; olli.j.suominen@tuni.fi (O.J.S.); ahmed.durmush@tuni.fi (A.D.); sari.peltonen@tuni.fi (S.P.); atanas.gotchev@tuni.fi (A.G.)

<sup>2</sup> Fusion for Energy (F4E), ITER Delivery Department, Remote Handling Project Team, 08019 Barcelona, Spain; Emilio.Ruiz@f4e.europa.eu

\* Correspondence: laura.goncalvesribeiro@tuni.fi

**Featured Application:** Remote handling of ITER's divertor cassette locking system.

**Abstract:** Visual technologies have an indispensable role in safety-critical applications, where tasks must often be performed through teleoperation. Due to the lack of stereoscopic and motion parallax depth cues in conventional images, alignment tasks pose a significant challenge to remote operation. In this context, machine vision can provide mission-critical information to augment the operator's perception. In this paper, we propose a retro-reflector marker-based teleoperation aid to be used in hostile remote handling environments. The system computes the remote manipulator's position with respect to the target using a set of one or two low-resolution cameras attached to its wrist. We develop an end-to-end pipeline of calibration, marker detection, and pose estimation, and extensively study the performance of the overall system. The results demonstrate that we have successfully engineered a retro-reflective marker from materials that can withstand the extreme temperature and radiation levels of the environment. Furthermore, we demonstrate that the proposed marker-based approach provides robust and reliable estimates and significantly outperforms a previous stereo-matching-based approach, even with a single camera.

**Keywords:** safety critical; teleoperation; eye-in-hand; optical tracking; retro-reflective markers; marker detection; stereoscopic; pose estimation



**Citation:** Ribeiro, L.G.; Suominen, O.J.; Peltonen, S.; Ruiz Morales, E.; Gotchev, A. Retro-Reflective-Marker-Aided Target Pose Estimation in a Safety-Critical Environment. *Appl. Sci.* **2021**, *11*, 3. <https://dx.doi.org/10.3390/app1101003>

Received: 1 December 2020

Accepted: 17 December 2020

Published: 22 December 2020

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Visual technologies have been increasingly integrated into industrial contexts with the intent of increasing work turnover, repeatability, or operator safety [1,2]. In safety-critical contexts, where operator access is restricted and tasks need to be performed through teleoperation, such as space, underwater exploration, mining, and nuclear reactors [3], these technologies have a particularly indispensable role. This is true especially in large-scale applications with many dynamic elements, where the current status of the system cannot be modeled accurately.

For safe and effective task performance, operators should have full situational awareness of the current status of the remote environment, especially during fine alignment tasks, where the operator has to position the manipulator in a certain position and orientation in relation to a target object with a high degree of accuracy. These tasks can represent a particular challenge for teleoperation due to the lack of some of the natural depth cues in standard image or video feedback [4,5].

In this context, machine vision approaches can be used to extract and refine mission-critical information from the images, such as the relative position between the tool and the target, which can be used as a teleoperation aid. This type of teleoperation aid has

been successfully used for maintenance and assembly tasks in extra-vehicular space applications [4]. In the aforementioned work, the computed six degrees of freedom (DOF) transformation between the current and desired end-effector positions is overlaid on the camera feed showed to the operator as either graphics or text.

The application that motivated our work is remote handling of ITER's (<https://www.iter.org/>) divertor cassette locking system. ITER constitutes one of the world's most hostile remote handling environments and is characterized by the unpredictable nature of its maintenance situations and tasks [6]. In ITER, all tasks are to be performed using a human-in-the-loop telemanipulation approach where the operator is always in control, although remotely located to minimize the risk of ionizing radiation exposure [6]. Operation in this environment differs from a common industrial environment, since high radiation levels and temperatures and strong magnetic fields strictly limit the materials that can be used in the activated areas, including the hardware that can be used for sensing the environment. This deems necessary the use of radiation-tolerant cameras, which are often characterized by a low resolution, gray-scale output, and high level of noise, which degrades with exposure to radiation over the lifetime of the camera. The operation of the divertor cassette locking system requires the alignment of several tools to their respective slots. Both the tools and the manipulator are carried in by a transporter that moves on rails throughout the facility. The tolerances for error in tool-slot alignment are small, with  $\pm 3$  mm of maximum allowed error in translation. Taking into account the difficulty in making a visual alignment in this range, there is a need to develop an accurate machine-vision-based aid that can comply with the strict requirements of the application.

Our aim here is to develop a system that calculates, using a set of one or two cameras attached to the wrist of the robotic manipulator, the transformation between the robot end-effector and the target (cassette). This information is intended to be used to update the models behind the virtual-reality-based remote handling platform [7,8].

Most man-made structures in the ITER environment, like the divertor cassettes, have large, flat, textureless surfaces, which are often specularly reflective and look different from different perspectives. Such surfaces pose a challenge to conventional 3D reconstruction algorithms that require the establishment of robust feature correspondences across multiple images. In a previous work [9,10], some of these challenges have been circumvented by extracting edge information from images provided by an array of two cameras as a prior step to stereo matching. The system obtains a relatively sparse 3D representation of the scene that is, at a later stage, matched with the known 3D model of the cassette to calculate the six-DOF transformation between the camera and the target. We consider the aforementioned approach to be sub-optimal for the application, since it does not explicitly exploit the prior knowledge of the structure of the environment at the feature detection and geometry reconstruction stages. As a result, this earlier solution is not as robust or accurate as required for the application at hand. We hypothesize that more reliable results can be achieved by detecting the known features of the cassette, or, even better, through the introduction of marker objects in the cassette surface that can be tracked by cameras unambiguously and accurately. Furthermore, there is a need, due to space constraints, to reduce the number of cameras to just one while still maintaining the targeted high-accuracy target pose estimation.

Marker-based tracking has been used in human motion analysis [11], medicine [12], and augmented reality [13–15] applications, localization of wearables [16], and in unmanned aerial vehicle and robot navigation systems [17–20]. It has also been used in industrial contexts [11], such as component inspection and validation [21,22]. It can obtain significantly superior results compared to tracking natural features, particularly when objects in the scene have limited contrast features, have few identifiable features, or change appearance with viewpoint [22,23]. Marker objects used for tracking can be LEDs, white diffuse markers, light projected targets, or retro-reflective markers [22].

A retro-reflector is an optical element that returns light to its source for a wide range of incident directions. They are best known for their use in traffic safety, where they improve



visibility under poor light conditions by returning incoming headlamp light in a small cone that covers the viewpoint of the driver. The intensity of retro-reflected light in the direction of the source can be up to hundreds to even thousands of times higher than that of a diffuse target [24]. They have also been used together with a laser for making distance measurements in space [25], industrial [26], and robotics applications [27].

In photogrammetry, retro-reflective markers allow the production of an almost binary image by controlling image exposure in such a way that the background is under-exposed and largely eliminated [28]. This process has advantages in target recognition and extraction, high-precision locating, and noise suppression [24].

As the above-mentioned characteristics deem retro-reflective markers optimal for the ITER application, in this paper, we aim to design a retro-reflective marker setting that complies with the strict requirements of such environment and can be attached to the surface of the cassettes for their effective tracking. We develop and test a system that uses said markers to reliably estimate the alignment between a pin-tool and the respective slots in the cassette surface.

The concept has previously been presented on a system level to the fusion engineering community [29]. Here, we detail it from a machine vision perspective with the following contributions:

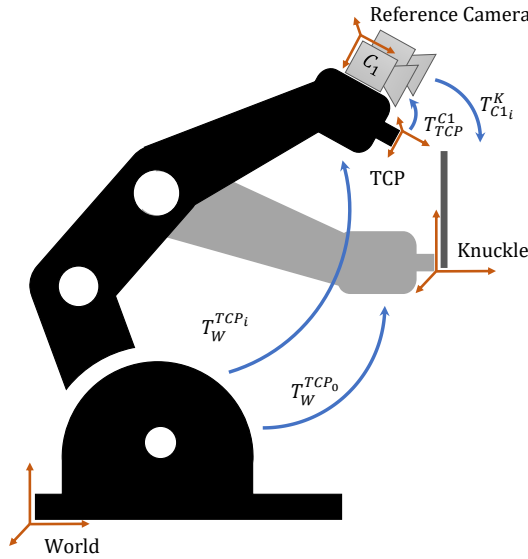
- A marker design that complies with the extremely strict requirements set by the operating environment.
- A methodology for marker detection and correspondence and optimization of the estimated pose.
- An approach for remote calibration of radiation-activated cameras.
- A study of the performance of monocular versus stereoscopic pose estimation on synthetically generated data with varying baselines.
- A comprehensive prototype implementation of the system and thorough evaluation of its overall performance.

The paper is organized as follows: In Section 2, we detail our proposed methods. In Section 3, we present the results of an extensive study of the performance of the system and a comparison to the earlier solution. In Section 4, we present our concluding remarks and outline future development prospects.

## 2. Methods

In the terminology of this paper,  $T$  represents a homogeneous rigid body transformation between two coordinate frames, which can be written as follows:  $T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$ , where  $R$  represents the  $3 \times 3$  rotation matrix and  $t$  represents the  $3 \times 1$  translation vector. The main coordinate frames and transformations that will be referenced throughout the work are represented in Figure 1.

Here,  $T_W^{TCP_i}$  represents the transformation from the robot-base coordinate frame to the robot's tool center point (TCP) that is given at each position  $i$  by the robot's control system (CS).  $T_{TCP}^{C1}$  represents the transformation between the robot's TCP and the attached reference camera that is independent of the position of the manipulator and is obtained through a hand-eye calibration procedure.  $T_{Ci}^K$  represents the transformation between the reference camera and a reference point in the world (e.g., origin of a target) at each position  $i$ .  $T_W^{TCP_0}$  represents the transformation between the robot base and the TCP when the TCP and target coordinate systems coincide, i.e., the aligned position.



**Figure 1.** Relevant coordinate frames and rigid body transformations.

### 2.1. Retro-Reflective Marker Design

Retro-reflectors can be characterized by four main properties: retro-reflectance (ratio of retro-reflected to incident light), brilliancy (reflectance for a specific observation angle), divergence (maximum angular distribution of retro-reflected light), and angularity (brilliancy as a function of the incident angle) [30]. In many applications, the desired marker behavior resembles the ideal retro-reflector, which allows large entrance angles (high angularity) and returns as much light as possible in as small a cone of light as possible (high reflectance, high brilliancy, and low divergence). However, in applications where there is a displacement between the light source and the detector, higher divergence might be acceptable or even required. As an example, in traffic applications, where there is a displacement between the car's headlamps and the driver's eyes of about 60 cm, the useful retro-reflecting angle has been quantified to be from nearly 0 to approximately 3 degrees for distances of 12 to 122 m [31]. In this case, the aim is to optimize brilliancy at the observation angle rather than at the incident angle. This is often achieved by using a retro-reflector with higher divergence, though at the cost of some intensity loss.

The same principle applies to photogrammetry and dictates the first functional requirement of an effective retro-reflector for our application. The useful retro-reflecting angle can be easily quantified for a specific setup as a function of the imaging distance ( $H$ ) and the displacement between the camera and the light source ( $D$ ) as follows:

$$\delta = \tan^{-1} \frac{H}{D}. \quad (1)$$

For the setup used in this work, we calculate a maximum useful retro-reflecting angle of  $\pm 18^\circ$  for the shortest working distance and the maximum possible displacement between the camera and the source.

The second functional requirement of a successful retro-reflector is the acceptance angle. We consider that an acceptance angle of  $\pm 10^\circ$  is sufficient for our application, since an operator can, with very little effort, make a preliminary target alignment within this range. Furthermore, there is no minimum specification for the retro-reflectance, but the aim is to maximize its value.

There is a relatively long list of environment requirements dictated by the ITER application when it comes to materials that stay in the divertor for a considerable amount of time. These include an operating temperature of up to 200 °C and adequate radiation tolerance. In practice, these restrictions deem unfeasible the use of most common materials, except for stainless steel (which makes up most of the other structures in the divertor) and fused silica glass (which is planned to be used in diagnostic windows and fiber optics). To the best of our knowledge, there is no adhesive material that can be used in this environment. Furthermore, the size of the developed marker should be rather small, such that the mounting of several specimens in a 200 × 200 mm area is feasible and the embedment depth should be rather small so as not to compromise the structural integrity of the cassette.

Retro-reflection is often achieved using one of two designs: a corner cube or a lens and a mirror. Corner-cube retro-reflectors (CCR) rely on the reflection of light on three mutually orthogonal surfaces. Generally, CCRs are manufactured either as a truncated cube corner of a transparent material (i.e., prism) or as flat reflective surfaces surrounding empty space. Cat's eye retro-reflectors consist of a primary lens with a secondary mirror located at its focus [30]. These can be easily implemented using a sphere of a transparent material. Spherical retro-reflectors generally have higher divergence due to spherical aberration and lower retro-reflectivity. On the other hand, corner-cube retro-reflectors usually have higher retro-reflectivity, while their fabrication is more difficult [30,32].

A retro-reflective marker can consist of a single shape or an array of retro-reflective shapes of smaller dimensions. The advantage of an array-based pattern is that the depth of the marker can be smaller and the shape of the overall retro-reflective region can be tailored. Furthermore, whilst the shape of the retro-reflective area changes with viewing angle for single-shape retro-reflectors, for those composed of arrays of small elements, the shape is independent of the orientation of the marker. On the other hand, such a construction might be more difficult to implement, as it requires a more precise manufacturing process with tighter tolerances.

In a preliminary study conducted to determine the viability of off-the-shelf retro-reflectors for use in ITER, we found that the only viable solution belonged to the single-element prism category (Edmund Optics BK-7 glass prism). No ready-made markers were found in the corner-cube array or bead array categories that could be used in our application.

Therefore, we propose a custom-made solution consisting of an array of glass beads. Such beads are often used within paint or attached via adhesive to other materials (e.g., paper or plastic). Due to the restriction of using adhesives and paints in ITER environment, the main challenge of using a bead array in our application is their attachment to the surface of the cassette. For this purpose, we propose a custom sieve-like metal holding structure that can hold the beads in place and can be easily mounted into the cassette surface, as illustrated in Figure 2.

Another potential challenge to this approach is that the retro-reflectance of glass beads is quite sensitive to their refractive index ( $n$ ) [33], and the refractive index of fused silica ( $\text{SiO}_2$ ) is relatively low when compared to other glass compositions ( $n_{\text{SiO}_2} = 1.4585$ ). In fact, if we consider the simplest case of a single specular internal reflection from the back of the silica sphere, we can calculate a divergence value of  $\pm 25$  degrees as follows [34]:

$$\theta = 2 \sin^{-1} \left( \frac{h}{r} \right) - 4 \sin^{-1} \left( \frac{h}{n * r} \right), \quad (2)$$

where  $h$  represents the incident height, as shown in the left side of Figure 3, and  $r$  represents the radius of the glass bead. In the right side of Figure 3, we show the calculated divergence as a function of the  $h$  value.

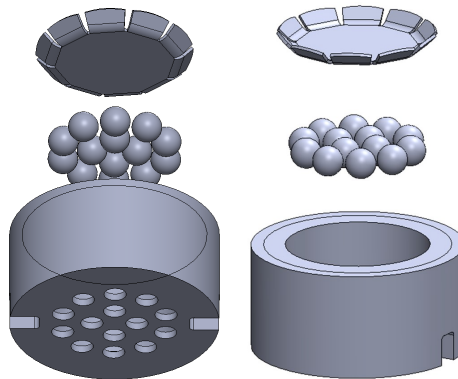


Figure 2. Retro-reflective bead array marker housing.

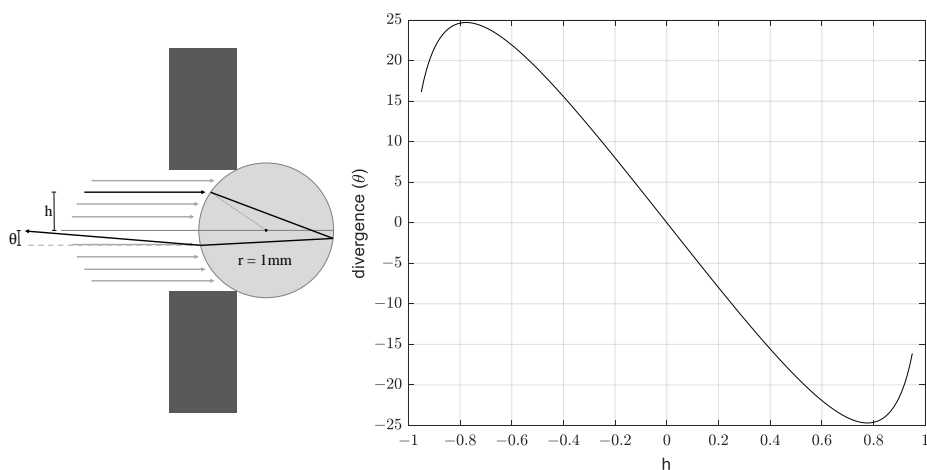


Figure 3. Incident height ( $h$ ) and divergence ( $\theta$ ) of light in a retro-reflective bead.

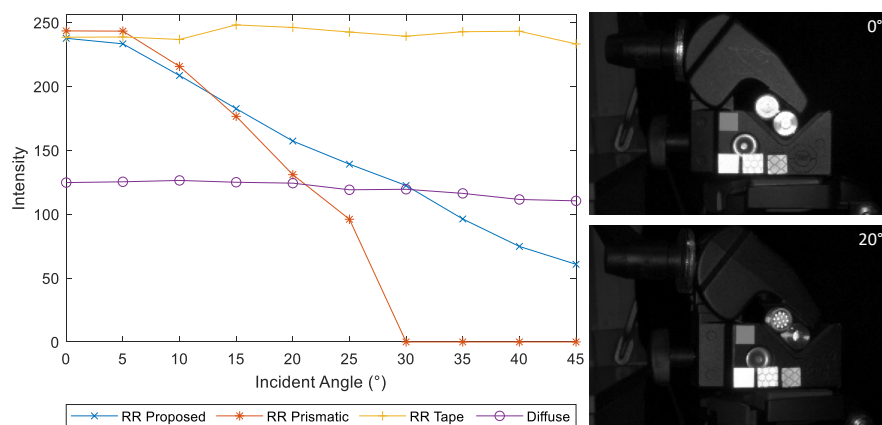
This divergence is higher than required, leading to some loss of intensity. Therefore, this design serves the purposes of the application, even if it does not fully optimize brilliancy at the observation angle. In Figure 4, we illustrate our proposed bead array marker design side by side with the out-of-the-box glass prism, a white diffuse element, and the commonly used bead-based retro-reflective tape that cannot be used for the application, but serves as a reference for a behavior that would be desirable.

The alternative of manufacturing a stainless steel hollow corner-cube array could technically be implemented and would have the advantage of being made from a single piece and a potentially larger reflective area.

However, while many methods are described in the literature for manufacturing corner-cube arrays [32] on both the micro and macro scale, the commercial availability of such methods is still limited and a suitable manufacturer could not be found. Some customized manufacturing methods were considered, mainly consisting of producing a strong enough inverse tool that could emboss the desired pattern onto a stainless steel piece using a hydraulic press. The advantage of this approach would be the flexibility of producing multiple marker specimens at a fixed cost and the fact that the markers could be a single piece made from a single material. However, a suitable method for manufacturing the pressing tool proved hard to find. We considered manufacturing

by electrical discharge machining (EDM). However, the surface finish of the tool was quite rough, and the manufacturing tolerances reduced the sharpness of the corners to a curvature with a diameter of approximately 2.0 mm. Furthermore, when pressing the tool onto the metal to create the retro-reflective pattern, the excess material tends to accumulate between the borders of the corner-cube indentation, leading to relatively large non-retro-reflective flat regions. An alternative option would be to create a single-shape corner-cube inverse tool, which would have to be stamped into the piece multiple times to create an array-like structure. Such a tool can be manufactured with a different technique that allows sharper corners. However, this brings additional practical issues related to the alignment of the tool that diminish the usability of this approach.

We concluded that the most viable and cost-effective solution for the application is the use of the bead-array retro-reflective marker array. The experiments presented further on in this paper were conducted using this design. We create a marker constellation design for the cassette piece, where we aim to maximize the number of markers while keeping the drilling into the piece at an acceptable level. We partially utilize holes that were already in the design of the cassette surface for other purposes. There is no possibility to mount markers at different depths, so all markers are at the same depth plane. Furthermore, we opt not to use coded markers in order to optimize the available bright marker area and due to the fact that the low resolution makes it hard to detect features of smaller dimensions.



**Figure 4.** Illustration of our proposed marker, the off-the-shelf glass prism, retro-reflective tape, and a white diffuse element. Data on the left side of the figure are obtained by averaging the pixel values within the marker area in images taken at several incident angles.

## 2.2. Automatic Offline Calibration Routine

During operation inside the reactor, it is expected that cameras' calibration parameters will be affected by shocks, vibration, temperature changes, or other unforeseeable environment factors. Even though there are not yet experimental data on how volatile these parameters will be (since the reactor is under construction), there is a definite need to develop reliable methods for calibration of activated cameras (cameras that have been exposed to radiation and, therefore, cannot be in the same environment as the operators at the same time). The developed calibration procedure should happen offline in the hot-cell facility (dedicated area outside of the operating environment to process activated components) and without manual intervention. The method of choice for camera calibration throughout our work is Matlab's implementation [35] of Zhang's widely known algorithm for photogrammetric camera calibration [36] using the distortion model of [37]. This method relies on a manual calibration image capture step, which means that the operator would have to move the manipulator and the attached cameras in the vicinity of a calibration pattern and capture a minimum of three calibration images at different camera orientations. This is an

undesirable, time-consuming, and repetitive procedure with a low level of repeatability, since different manipulator positions provide slightly different calibration results. Some methods have been proposed in the literature for automatizing the image capturing step through the display of virtual targets on a screen [38]. These methods are, however, not applicable to our application due to the difficulty of introducing additional electronics in the hot-cell facility. Therefore, we developed a calibration routine to be performed autonomously by the robotic arm that produces fairly repeatable results. As a target, we used a checkerboard pattern that was placed in the hot-cell facility prior to the start of the calibration routine. Even though the same procedure could be used to re-calibrate the cameras in the reactor, based on the known relative position of the retro-reflective markers, we did not explore this option, mostly due to the tight space constraints in the environment.

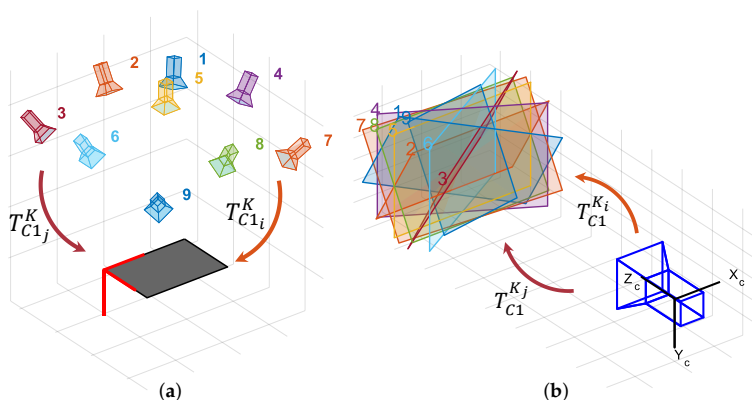
As a preliminary step, we summarize what is established as a good set of calibration images: Even though it is possible to obtain a calibration estimate from a smaller number of images [36], it is often recommended to acquire at least 10 images of the calibration pattern at different orientations relative to the camera. The images should be captured at a distance that corresponds roughly to the operating distance so that the pattern is in focus. The positions of the pattern shall be such that:

(1) The checkerboard is at an angle lower than 45 degrees relative to the camera plane. According to the original work [36], when tested on synthetic data, the calibration method obtains most accurate estimates when the angle between the calibration target and image plane is higher. However, in real imaging conditions, where blur and distortion must be taken into account, angles close to 45 degrees might compromise the accuracy of corner detection, depending on the size of the checkerboard and the depth of field of the camera-lens system.

(2) Feature points should be detectable on as much of the image frame as possible for adequate estimation of the lens distortion parameters.

In our scenario, additional practical considerations need to be taken into account, such as the work-space limitations of the robotic arm and singularity regions of its joints.

To fulfill the above-mentioned requirements, the developed automatic capture routine moves the camera in discrete positions within a spherical surface centered in the calibration pattern center, as shown in Figure 5. The radius of the spherical surface corresponds, approximately, to the average operating distance.



**Figure 5.** Camera movement in a dome around the centre of the checkerboard pattern seen from a pattern-centric perspective (a) and from a camera centric perspective (b).

The camera rotations are such that the center of the pattern is kept in the center of the image plane. The described movement is equivalent to the rotation of the calibration pattern around its center in each of the three rotation dimensions (yaw, pitch, and roll) in a

camera-centric interpretation of the problem. For the sake of simplicity, we calculate the desired camera-to-target transformations ( $T_{C1_i}^K$ ) using the camera-centric interpretation:

$$T_{C1_i}^K = T_{C1}^{K_i} = T_{C1}^{K_0} T_{K_0}^{K_i}. \quad (3)$$

The base-to-end-effector transformations that bring the camera to the desired calibration positions are calculated in the following manner:

$$T_W^{TCP_i} = T_W^{TCP_j} T_{TCP}^{C1} T_{C1_j}^K T_{C1_i}^K{}^{-1} T_{TCP}^{C1}, \quad (4)$$

where  $T_{TCP}^{C1}$  represents a rough estimate of the hand–eye calibration.  $T_W^{TCP_j}$  represents the initial position reported by the robot’s CS and  $T_{C1_j}^K$  represents the initial transforming between the camera and the target and is estimated from a single image based on nominal values of the camera calibration parameters. The calibration procedure runs automatically. All the operator needs to do is to move the manipulator so that the checkerboard is within the field of view of the camera for the initial pose. At the end of the procedure, the operator will receive a visual representation of the re-projection error indicating whether the procedure was successful. The procedure can be used to capture any desired number of images, and the steps within the range of movement are adapted accordingly.

The captured images are used to calibrate the intrinsic parameters of the cameras and the extrinsic parameters in the case where a stereo pair is used and serve as data for a hand–eye calibration routine. A hand–eye calibration is required for this application in order to align the coordinate systems of the robotic manipulator and the vision system. The details of such a calibration are outside the scope of this paper, and are presented in detail in [39]. The calibration procedure shall be repeated as needed based on the volatility of the estimated parameters in the real operating conditions.

### 2.3. Marker Detection and Identification

The preliminary step to position estimation is the determination of correspondences between 2D image features and known 3D world points. The developed marker detection and identification approach is shown in Figure 6.

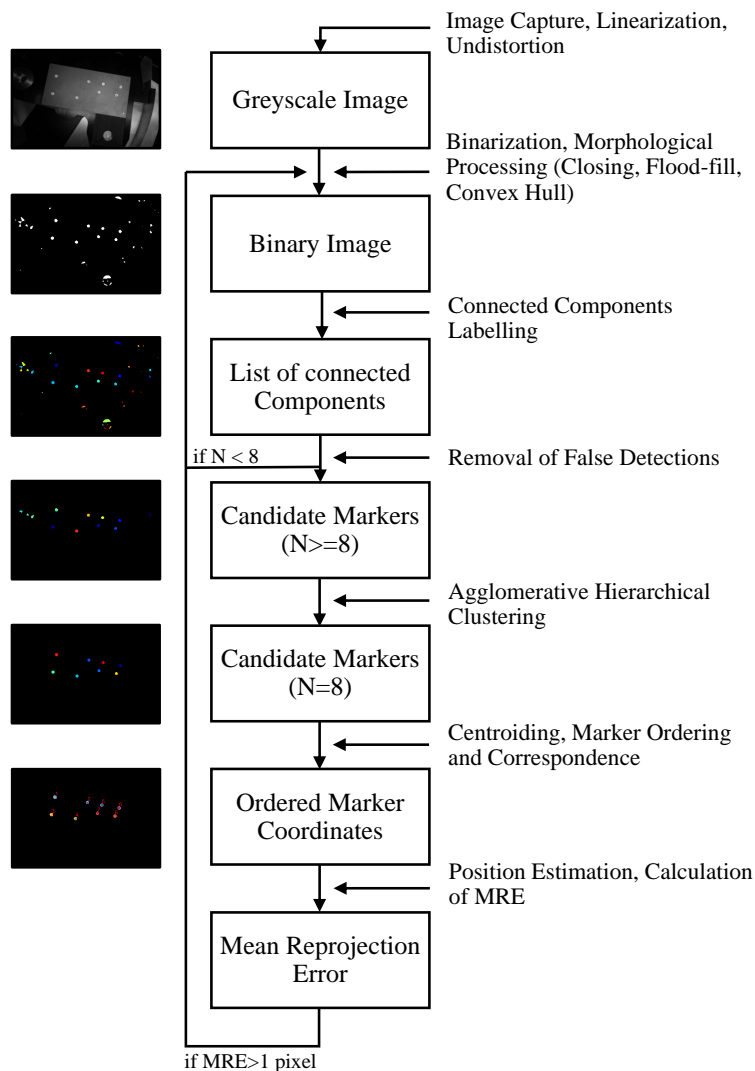
It relies on the prior knowledge of the characteristics of the markers and their layout with respect to each other. We perform marker detection and identification independently for each camera image to keep the approach flexible and suitable for the monocular case.

The workflow of marker detection starts with image capture, linearization, and undistortion. Images are segmented into foreground (markers) and background through adaptive thresholding as follows:

$$I_{x,y}^{BW} = \begin{cases} 1, & \text{if } I_{x,y} > s \text{ med}_{x,y} + T \\ 0, & \text{otherwise} \end{cases}. \quad (5)$$

Here,  $I_{x,y}^{BW}$  represents the value of the resulting binary image at pixel coordinates (x,y), while  $I_{x,y}$  represents the value of the original gray-scale image at the corresponding coordinate points.  $\text{med}_{x,y}$  represents the median of the  $w \times h$  surrounding pixels at coordinate (x,y). Adaptive thresholding preserves hard contrast lines while ignoring soft gradient changes. Our approach resembles the methods of [40]. However, it compares each image pixel to the median of the  $w \times h$  surrounding pixels, instead of their average, since this approach effectively removes outliers and better preserves details and edges. The value  $s$  is a scale factor that contributes to the classification of more or fewer pixels as foreground. We have modified the original method by adding a constant value,  $T$ , to the threshold matrix, with the intent of exploiting the fact that the intensity difference between the markers and the background is more significant than in the diffuse marker case. Even though the retro-reflective markers are considerably brighter than their surroundings, the large specular reflections of the cassette surface and some degree of inhomogeneity of the lighting

conditions deem advantageous the use of a local thresholding method. However, this has a tendency to increase the number of false blobs detected in the background, which will be handled by morphological filtering. In the following processing steps, several techniques are utilized to determine the suitability of marker candidates. If these point to an incorrect segmentation, the software returns to the thresholding step and the operator is asked to intervene by modifying the threshold parameters in small incremental steps. Decreasing  $s$  and/or  $T$  will lead to a higher number of pixels being classified as foreground. While manually refining the parameters, the operator is presented with the thresholded image and the result of the overall marker detection pipeline.



**Figure 6.** Workflow of image processing operations leading to marker detection and identification.

After segmentation, each marker is represented by an array of blobs that are merged into a single elliptical shape by applying morphological closing and flood-fill operations [41], consecutively followed by determination of the binary convex hull image. The blobs in the resulting image are identified by connected component labeling [42] and



non-markers are excluded through the comparison of several properties of the connected compound elements to appropriate thresholds. We consider that markers often differ from the remaining blobs with respect to their size (area) and/or and shape (circularity and eccentricity). Circularity is defined by the following ratio:

$$Circularity = \frac{4\pi Area}{Perimeter^2} \quad (6)$$

and has a minimum value of 1 for a circular disk, and is greater than 1 for all other geometric features. For an ellipse, it increases with the imaging angle monotonously and has a value of 1.5 at 60 degree angle [43]. Eccentricity describes the ratio of the distance between the foci and major axis length of the ellipse that has the same second-moments as the region. It is closer to 0 for a circle-like object, and closer to 1 for a line-like object.

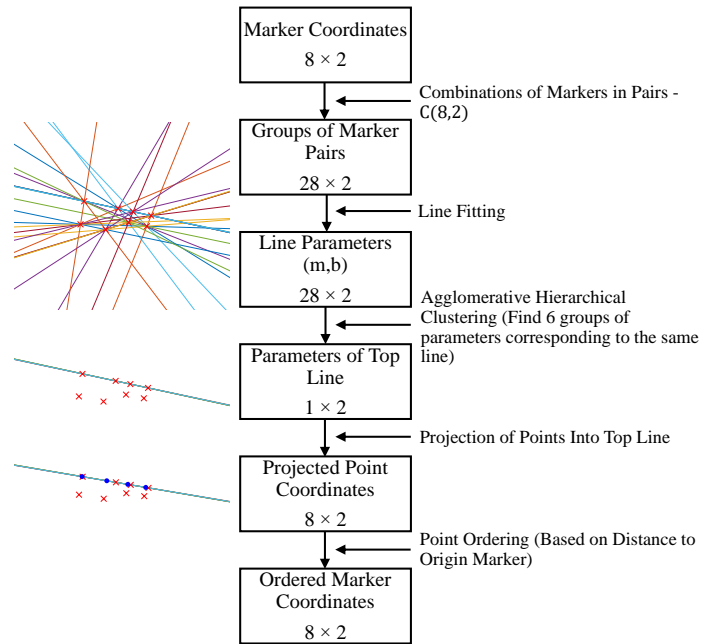
If after the elimination step, the number of identified markers is less than expected, the procedure returns to the segmentation step and prompts the operator to adjust the segmentation parameters.

If after the elimination step, the number of candidate markers is higher than expected, markers are grouped into clusters according to their features, creating an agglomerative hierarchical cluster tree. Selected features are: area, perimeter, circularity, solidity, convex area, equivalent diameter, extent, minimum Feret diameter and maximum, and minimum intensity of the grayscale blob. Solidity describes the extent to which a shape is convex or concave. It corresponds to the ratio between the area and the convex area (area enclosed by a convex hull). The equivalent diameter corresponds to the diameter of a circle with the same area as the region. Extent refers to the ratio of pixels in the region to pixels in the total bounding box. The minimum Feret diameter expresses the minimum distance between any two boundary points on the antipodal vertices of the convex hull that encloses the object. Data values are normalized before computing similarity and the Euclidean distance is used as a distance metric. The hierarchical tree is pruned at the lowest branch, where a class of eight elements can be found. A single feature point is found within each marker by calculating the centroid coordinates  $(x_0, y_0)$ :

$$x_0 = \frac{\sum_{i=1}^w \sum_{j=1}^h x_i \cdot I_{i,j}^{BW}}{\sum_{i=1}^w \sum_{j=1}^h I_{i,j}^{BW}}, \quad y_0 = \frac{\sum_{i=1}^w \sum_{j=1}^h y_i \cdot I_{i,j}^{BW}}{\sum_{i=1}^w \sum_{j=1}^h I_{i,j}^{BW}}, \quad (7)$$

where  $w, h$  represent the size of the window in which the centroid is being computed and  $I_{i,j}^{BW}$  represents the value of the binary image at coordinate  $(i, j)$ . By making the approximation that the centroid of the detected markers corresponds to the center of the original circles, we disregard the perspective bias [44] introduced by perspective projection. We assume that the camera and target have been roughly aligned and, consequently, that the value of the perspective bias is rather small.

In order to establish a one-to-one correspondence between the known relative marker positions in the world coordinate frame and the detected image coordinates, the detected markers must be correctly ordered. The proposed marker ordering approach is shown in Figure 7 and finds the line passing through four of the proposed markers. This is achieved by computing all possible combinations of the markers in groups of two (28 groups), fitting a line to each marker pair and finding the six groups of parameters corresponding to the same line using agglomerative hierarchical clustering. The hierarchical cluster tree is pruned at the lowest point, where there is a class composed of five elements. After the line has been determined, the remaining four points can be projected onto the line, and markers can be discriminated from their distance along the line to the origin marker.



**Figure 7.** Workflow of marker ordering.

### Pose Estimation

Pose estimation refers to the calculation of the rigid body transformation between the camera and the target ( $T_{C1_i}^K$ ) based on a set of detected 2D image points and 3D known world point correspondences.

Pose estimation has been extensively studied in the literature, and many iterative and non-iterative methods have been proposed to solve the problem under different conditions, depending on the number of available point correspondences, presence of outliers, how noisy their coordinates are, whether the problem has to be solved online or offline, and if the camera parameters are known or estimated simultaneously [45–47].

In our application, control points are known to be co-planar and are detected with a relatively high degree of accuracy. We use the pose estimate provided by the direct linear transform (DLT) homography estimation method [36,48] as an initial estimate to start an iterative approach that minimizes the re-projection error as a non-linear least squares problem.

The minimization of the re-projection error is solved using the Levenberg–Marquardt algorithm [49,50] and can be written for the single-camera case in the following manner:

$$\arg \min_{r,t} \sum_{i=1}^N [m_i - P(M_i, K, r, t)]^2, \quad (8)$$

where  $P$  describes the projection of world points  $(X, Y, Z)$  into image points  $(x, y)$ :

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \lambda K \begin{bmatrix} R & t \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (9)$$

$M_i$  and  $m_i$  represent world and image point coordinates, respectively.  $\lambda$  represents a scaling factor. The three-dimensional vector  $r$  represents the axis-angle representation of the rotation matrix,  $R$ , obtained using the Rodriguez formula [51].

The intrinsic matrix,  $K$ , describing the internal parameters of the camera is computed at the calibration stage and is written as function of the camera focal length ( $f_x, f_y$ ) and principle point ( $c_x, c_y$ ) as follows:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \quad (10)$$

We solve the stereo camera case in a similar manner. Here, the homography estimation and decomposition approach is used to compute an estimate of the transformation between each camera and the world reference frame ( $T_{C1}^K, T_{C2}^K$ ). The known extrinsic camera parameters describing the rigid body transformation between the auxiliary and reference camera ( $T_{C1}^{C2}$ ) are used to relate the estimate provided by the auxiliary camera to the frame of the reference camera:

$$T_{C1}^K = T_{C1}^{C2} T_{C2}^K. \quad (11)$$

The average value of the estimate provided by each camera is taken as the initial estimate, which is minimized in the following manner:

$$\arg \min_{r,t} \sum_{i=1}^N [m_i^l - P(M_i^l, K^l, r, t)]^2 + [m_i^r - P(M_i^r, K^r, r, t)]^2. \quad (12)$$

### 3. Results

In this section, we evaluate the performance of the proposed system and, when possible, of its individual components, on synthetic and real data, and establish a comparison between the single and stereo estimation methods. In all the upcoming experiments, the threshold parameters for marker detection (as defined in Section 2.3) have the following values: initial proposal of thresholding parameters  $s$  and  $T$  are 1.18 and 0.03, respectively. These values are set based on the lighting conditions and should be refined for each specific setup. They can also be tuned by the operator if the search for the markers is unsuccessful. The neighborhood size for adaptive thresholding is  $11 \times 11$  pixels. The structural element for the morphological closing operation is a disk with radius of three pixels. The flood fill operation and connected component labeling use a pixel connectivity of 4. For the elimination of false blobs, the acceptable ranges for area and circularity are [20, 400] pixels and [1, 1.5], respectively. The maximum allowed value for eccentricity is 0.9. These values are set based on the relative sizes of the markers and the image for the specified operating distances.

#### 3.1. Synthetic Data

The main synthetic dataset is composed of 20 images corresponding to different relative poses of the camera and the target. The poses were generated randomly, while assuring that the target was fully visible and within the following limits:

(1) Distance of 300 to 500 mm from the camera to the target, corresponding to the working range established by the application.

(2) Horizontal and vertical deviations of  $-50$  to  $50$  mm and  $-100$  to  $100$  mm from the central position in the  $x$  and  $y$  axis. Rotations of  $-10$  to  $10$  degrees in relation to the aligned position in each rotation axis. The operator is expected to be able to easily make a preliminary alignment within these limits based on the camera feeds.

The images of the main synthetic dataset are used to evaluate the performance of single-camera approaches and correspond to the left-most camera in approaches using a stereo camera pair. In order to study the effect of different baseline values in estimations using stereo vision, five additional sub-datasets were generated, each composed of 20 im-

ages corresponding to the right-most camera for a certain baseline value. In addition to being translated by the baseline value, the right-most camera is rotated horizontally in the direction of the other camera to increase the shared field of view of the stereo pair. Sets of values for baseline and rotation are 100 mm and  $10^\circ$ , 150 mm and  $15^\circ$ , 200 mm and  $20^\circ$ , 250 mm and  $25^\circ$ , and 300 mm and  $30^\circ$ .

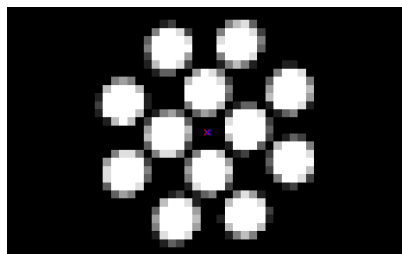


**Figure 8.** Example synthetic images of the checkerboard (a) and the bead array (b) targets, at the same relative position.

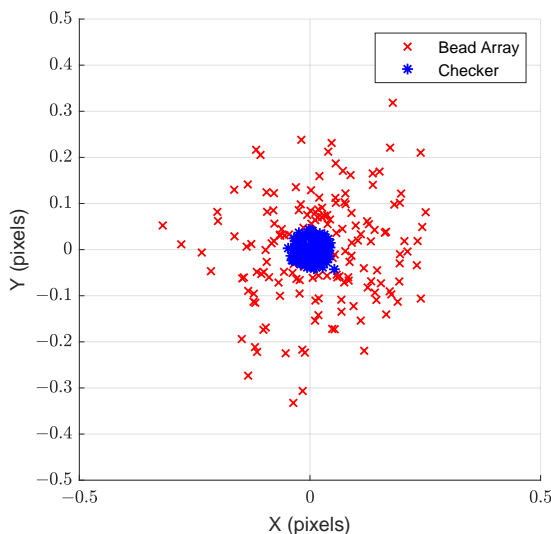
We generated synthetic images of both a checkerboard and our developed bead-array target (Figure 8). Synthetic images were generated using our own software, in which the retro-reflective beads are modeled as flat white diffuse circles on the black target plane. Ground truth checkerboard corners and bead-array center values are extracted from the rendering software and are used as ground truth for evaluating the accuracy of feature detection. These datasets also have associated ground truth intrinsic and extrinsic camera calibration parameters. The rigid body transformation between the left-most camera and the origin of the target is extracted as the ground truth value for evaluating the accuracy of pose estimation methods.

### 3.1.1. Evaluating the Performance of Marker Detection

The performance of the marker detection algorithm was evaluated by comparing the ground truth and detected feature center points for each of the 20 images of the main synthetic dataset, as shown in Figure 9. The difference in image coordinates between the detected and ground truth points is shown in Figure 10 in the  $x$  and  $y$  dimensions. This representation shows that error in feature detection is well in the sub-pixel range for the studied range of positions. For comparison, results of the gold-standard algorithm for checkerboard detection with sub-pixel refinement of [52] are included.



**Figure 9.** Cropped section of a synthetic image of the bead array target. The center of the marker detected from the image is shown in red, while the ground truth value is shown in blue.



**Figure 10.** Error in the X and Y axes corresponding to the detection of marker centroids (x) and checkerboard corners (\*) in synthetic data. Each data point represents an image feature point.

### 3.1.2. Evaluating the Performance of Pose Estimation

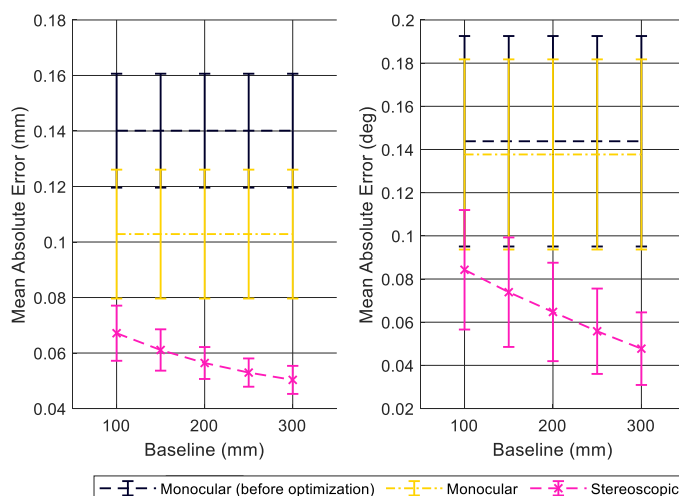
Performance of position determination was evaluated by comparing the calculated reference-camera-to-target transformations ( $T_{C1_t}^W$ ) to the ground truth. We calculate the translation error ( $e_t$ ) between the estimated and ground truth rigid body transformations as the Euclidean distance in each of the two axes that are parallel to the target plane. Since it is not relevant for the alignment task, we disregard the depth axis.

The rotation error ( $e_R$ ) is calculated as the absolute difference between the magnitude of the angle around the rotation axis (as defined in the axis-angle representation). At this stage, we use the ground truth feature point coordinate values. We evaluate the performance of monocular and stereoscopic estimates for baseline values of 100, 150, 200, 250, and 300 mm. We observed that using ground truth feature points, errors are in the range of  $10^{-5}$  mm in translation and  $10^{-5}$  degrees in rotation for all approaches, which is considered to be insignificant. Therefore, in the following experiments, random noise with a uniform distribution and several range values was added to the feature point coordinate values, and the results were averaged over 20 tests.

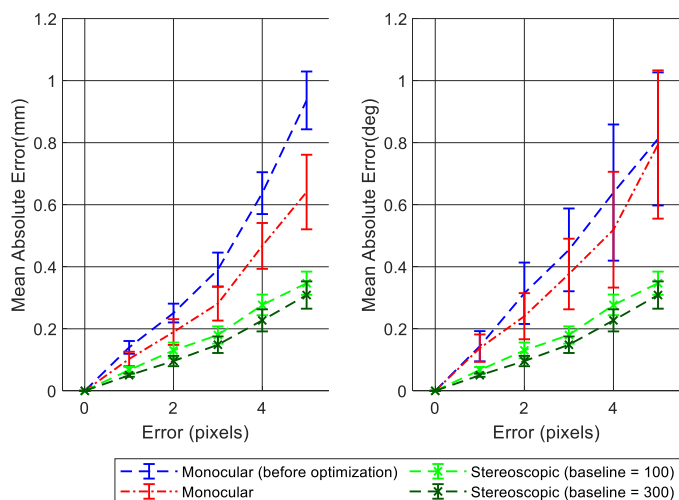
In Figure 11, we present position estimation errors  $e_t$  and  $e_R$ , averaged over the 20 positions of the synthetic dataset for added noise in the range of  $-0.5$  to  $0.5$  pixels. We can see that stereoscopic estimation produces the best results and performs better as the baseline increases. We can also see that standard deviation follows the same trend, showing that more consistent results are found with stereo, particularly for higher baseline values.

In Figure 12, we compare the performance of monocular and stereoscopic approaches for the smallest and largest considered baselines as a function of the added noise level. We see that stereoscopic estimation performs best and is the most robust to added noise. We observe that the error for approaches using stereo vision becomes lower as the cameras' baseline increases for all considered noise levels.

We can consider in the application a baseline range of 100–300 mm and a maximum expected corner noise level of two pixels. In these conditions, the difference in translation and rotation error between monocular and stereoscopic estimates is less than 0.1 mm and 0.1 degrees.



**Figure 11.** Translation (left side) and rotation (right side) error of position estimation over the images of the synthetic dataset using ground truth feature points with added random noise within the interval of  $[-0.5, 0.5]$  pixels. Data points in the graph represent the average error over the 20 images of the dataset and error bars represent the standard deviation.



**Figure 12.** Translation (left side) and rotation (right side) error of position estimation over the images of the synthetic dataset using ground truth feature points for several added noise levels. Data points in the graph represent the average error over the 20 images of the dataset and error bars represent the standard deviation.

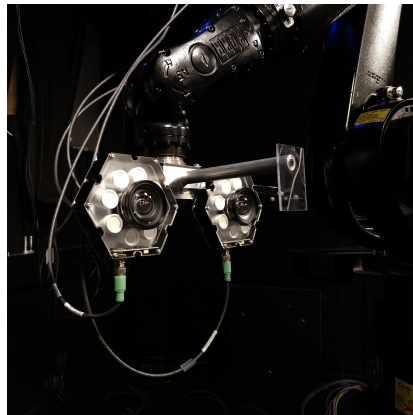
### 3.2. Real Data

#### 3.2.1. Experimental Setup

Our experimental setup is shown in Figure 13. The KUKA KR 16 L6-2 industrial robot was used to simulate the manipulator that would be used in the application. The manufacturing of the entire target assembly, as described in the methods section, is costly and time-consuming; therefore, it was approximated by a prototype of similar characteristics. The glass beads are made of borosilicate glass ( $n = 1.48$ ) instead of fused silica ( $n = 1.46$ )

and the sieve-like holding structure is made of a single 1 mm thick metal piece attached to a prototype of the target.

We use two Basler acA1920-50gc industrial machine vision cameras with native resolution of  $1920 \times 1200$ . The cameras have an RGB Bayer filter to produce color images. To emulate the use of radiation-tolerant (RADTOL) cameras, the images were converted to grayscale in-camera. The original resolution was down-sampled by a factor of three using a bicubic kernel on  $4 \times 4$  neighborhoods. The resulting effective resolution ( $640 \times 400$ ) is close to that of a RADTOL camera that we considered as a reference, the Visatec RC2 ( $586 \times 330$ ), producing a fairly realistic approximation of the image quality of the application. We used lenses with 6 mm focal length that had a horizontal field of view of approximately 82 degrees. The stereo camera pair was attached to a metal plate and had a fixed baseline of 210 mm. A set of radial illuminators (Smart Vision RC130) were mounted as closely as possible to the camera lenses. Each illuminator was constituted of eight high-intensity LEDs that create a near collimated, uniform light pattern that has a diameter of 280 mm at a distance of 500 mm from the lamp. In Figure 14, we show an example image taken in these conditions. For calibration of the cameras, we used a set of 20 images of a  $24 \times 17$  checkerboard pattern.



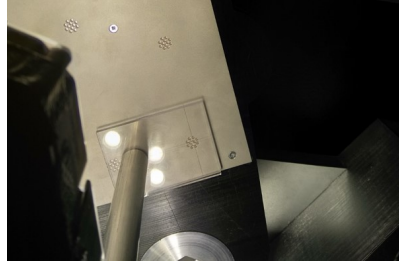
**Figure 13.** Experimental setup.



**Figure 14.** Example image. The bead arrays can be seen as significantly brighter spots in an otherwise underexposed background.

### 3.2.2. Reference Values

The reference values for assessing the accuracy of the system were obtained by using an alignment piece, shown in Figure 15, made of a transparent polymer that had been engraved with alignment axes. Reference values were obtained by manually positioning the alignment piece in the aligned position and acquiring the transformation provided by the manipulator's CS,  $T_W^{TCP_0}$ .



**Figure 15.** Image of part of the target and the developed alignment piece in the aligned position.

Provided that the robot's TCP has been assigned to the cross-hair, the transformation between the TCP and the target origin is, in the aligned position, the identity matrix. The definition of the TCP position is a built-in function of the KUKA robot controller, where the transformation between the robot's flange and the alignment piece is estimated through the XYZ Four-Point TCP Calibration Method. This method involves bringing the tool center to a known position and recording several points as the position is approached from different orientations.

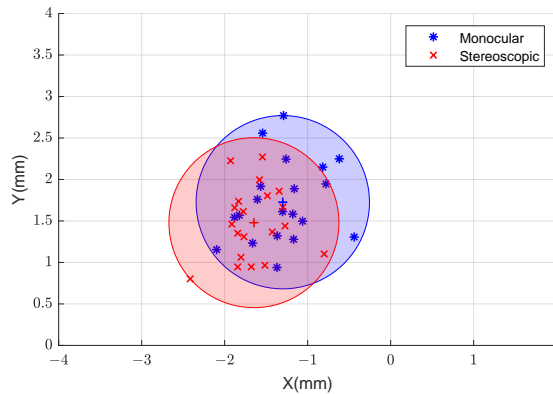
For a given test position,  $i$ , the rigid body transformation between the optical axis of the camera and the coordinate system of the target,  $T_{C1_i}^K$ , is given as follows:

$$T_{C1_i}^K = (T_{TCP}^{C1})^{-1} (T_W^{TCP})^{-1} T_W^{TCP0} T_{TCP0}^K, \quad (13)$$

where  $T_W^{TCP}$  is provided by the robot's CS at position  $i$  and  $T_{TCP}^{C1}$  is the calibrated hand-eye transformation. The transformation between the tool center point,  $T_{TCP0}^K$ , is the identity matrix. In our experiments, for each test position, we estimate  $T_{C1_i}^K$  and compare it with the reference value described above.

### 3.2.3. Evaluation of Overall System Performance

The overall performance of the proposed system is evaluated in a set of 20 initial camera positions, similar to those described in Section 3.1. In Figure 16, we show the translation errors between the estimated and ground truth transformations in each of the two axes that are parallel to the target plane.



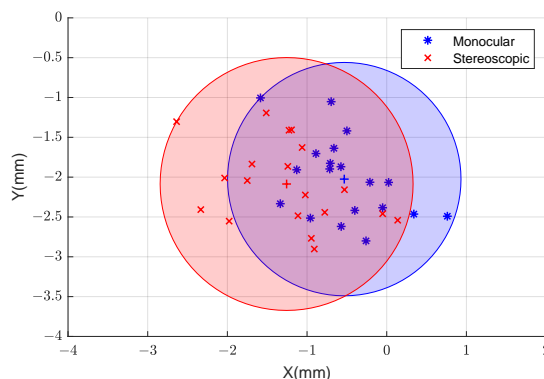
**Figure 16.** Error in each of the X and Y axes for monocular (blue) and stereoscopic (red) estimates using our proposed approach.



### 3.2.4. Evaluation of Position Estimation

The performance of the position determination step was evaluated by using a checkerboard pattern as a target and full-resolution images. We used a state-of-the-art checkerboard detection algorithm [52], which is considered to be very accurate. In this manner, we attempted to limit, to the extent possible, the effect of the developed target and marker detection methods on the results.

This experiment was run on approximately the same camera poses as described above. The distance between the camera and the target was increased to fit the larger checkerboard target in the cameras' fields of view. The results are presented in Figure 17.

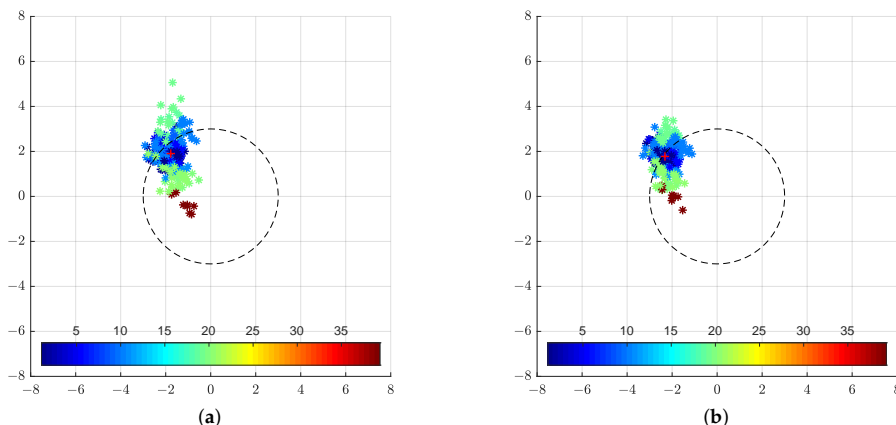


**Figure 17.** Error in each of the X and Y axes for monocular (blue) and stereoscopic (red) estimates using corner values provided by the checkerboard detection algorithm in full-resolution images.

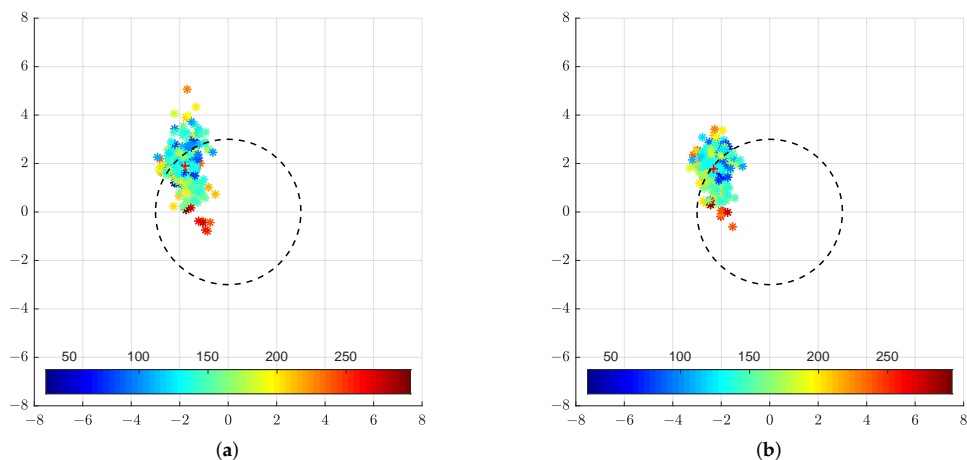
### 3.2.5. Study of the Effect of Camera Slant in the Original Position on the Overall System Performance

For the sake of completeness, an extensive analysis was conducted over a more comprehensive set of positions outside of the limits established in Section 3.1. We analyzed a set of 296 images taken at random initial camera positions and orientations.

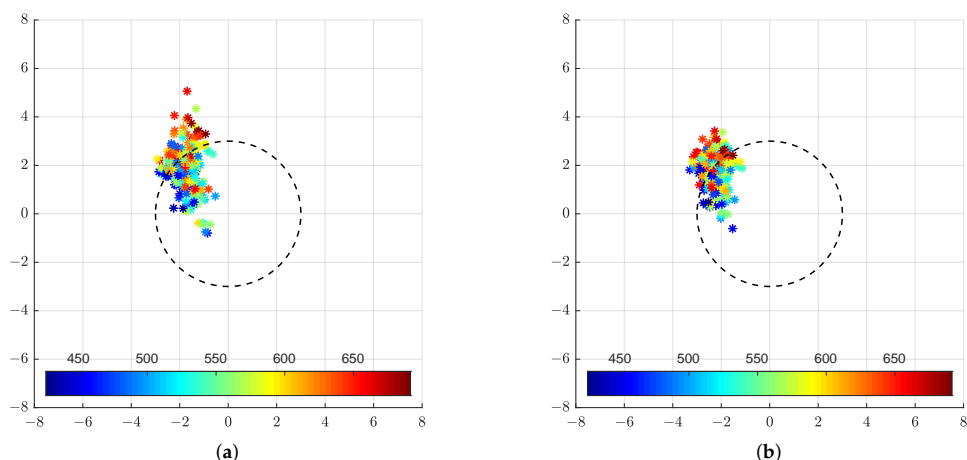
In Figures 18–20, we present the results as a function of the characteristics of the initial position: total rotation between camera and target and the horizontal and vertical deviation (Euclidean distance) from the central position and distance between the camera and the target. In these figures, the red cross signifies the mean of the observations and the dashed black circle represents the error tolerance of the application.



**Figure 18.** Error in each of the X and Y axis for the estimation provided by the monocular (a) and stereoscopic (b) approaches. The colours represent the total rotation between the camera and target planes.



**Figure 19.** Error in each of the X and Y axis for the estimation provided by the monocular (a) and stereoscopic (b) approaches. The colours represent the horizontal and vertical deviation from the from the central position.



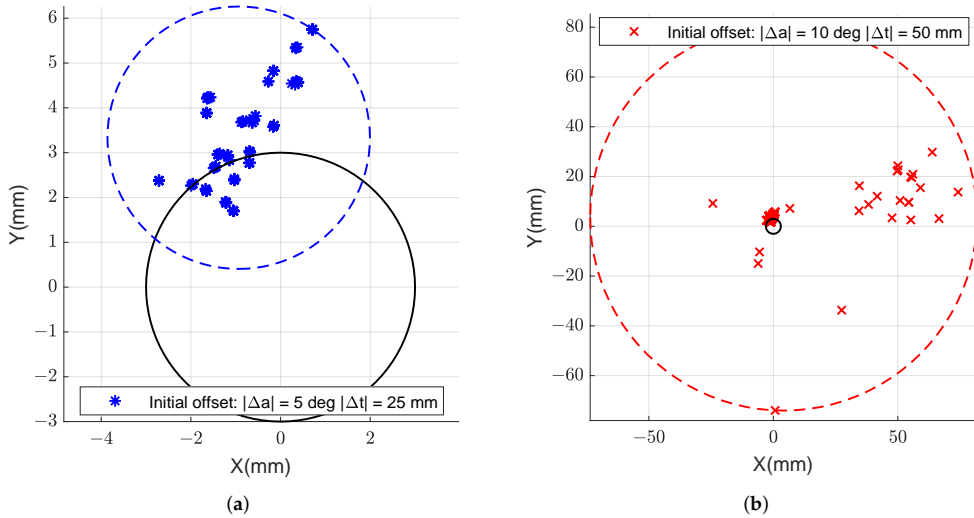
**Figure 20.** Error in each of the X and Y axis for the estimation provided by the monocular (a) and stereoscopic (b) approaches. The colours represent the distance between the camera and target.

### 3.2.6. Comparison to the Earlier Solution

In order to establish a comparison between the newly developed marker-based method and the approach of [10], we ran corresponding data through their method that consists of estimating a 3D representation of the scene by stereo matching and aligning it to the reference CAD model using an iterative closest point (ICP) algorithm. To obtain corresponding data, we captured the same sequence of robot poses with the same scene object, but with different lighting conditions. Since the stereo reconstruction relies on Lambertian reflectivity of the target object, it requires more light than the new marker-based approach.

The earlier approach relies on a number of parameters, which should be tuned to the current scene properties. This is needed to avoid falling into local minima and converging on false position estimates. To avoid the time-consuming manual tuning step to adapt to this particular setup, we gave an initial estimate to the algorithm based on the known true target position with an added offset. For the objective of measuring the accuracy (not the

robustness) of the earlier solutions, this approach is justified for the sake of a comparison. We found that adding up to 25 mm of position error and five degrees of angular error still results in the algorithm converging as close to the correct position as the data allow (Figure 21, left side). Larger amounts (such as 50 mm and 10 degrees) make it difficult for ICP to find the correct neighborhood to converge, and the pose estimates start having large amounts of outliers (Figure 21, right side).



**Figure 21.** Error in each of the X and Y axis for the estimation provided by the stereo ICP based method from earlier work, with different amounts of initial position deviation from ground truth. Offset values of  $|\Delta a| = 5$  deg,  $|\Delta t| = 25$  mm (a). Offset values of  $|\Delta a| = 10$  deg,  $|\Delta t| = 50$  mm (b). Each colour point represents a different position of the camera in relation to the target. The continuous black circle represents the error tolerance of the application.

### 3.2.7. Analyses of Overall System Performance Using an Image-Based Metric

An alternative, image-based metric was designed to evaluate the overall performance of the system without relying on the manual alignment procedure and alignment tool calibration (described in Section 3.2.2) to provide a reference value, allowing us to rule out errors in those processes. However, this metric expresses not only the accuracy of our system and the ability of the robot's CS to report its position correctly, but also its ability to reach the target position. Therefore, we can expect the overall errors to be potentially higher.

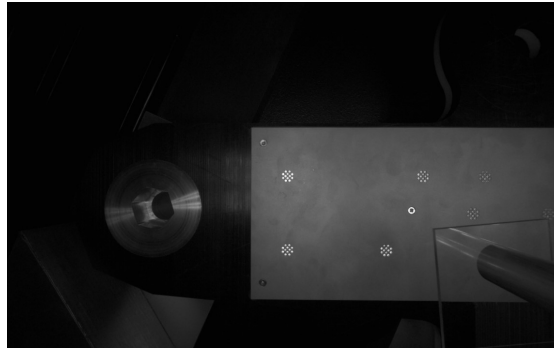
The procedure starts by capturing an image at the initial position and calculating the transformation between the left camera and the target,  $T_{C1_i}^K$ , using our developed methods. As a second step, the base-to-end-effector transformation that brings the camera to the desired position,  $T_W^{TCP_j}$ , is determined by:

$$T_W^{TCP_j} = T_W^{TCP_i} T_{TCP}^{C1} T_{C1_i}^K T_{C1_j}^{K^{-1}} T_{TCP}^{C1}, \quad (14)$$

where  $T_{C1_j}^K$ , the transformation between the camera and the target in the desired position, has the value:

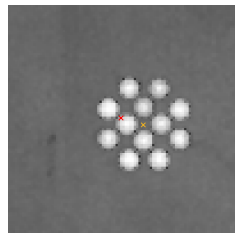
$$T_{C1_j}^K = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 250 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (15)$$

which corresponds to aligning the camera and target plane at a distance of 250 mm from the target. The command is given to the robot controller, which moves the camera to the calculated position, where a full-resolution image of the target is taken (Figure 22).



**Figure 22.** Example image taken at the aligned position. If the estimate is correct, the center of the reference marker should correspond to the center of the image.

If the alignment is correct, the center of the reference marker is expected to coincide with the center of the image. Therefore, an alignment error can be measured in pixels, i.e., as the difference between the center of the marker and the center of the image (Figure 23). The measured error can be converted into millimeters, since the dimensions of the target are known (e.g., we know that the diameter of one bead corresponds to 1.95 mm).



**Figure 23.** Example of a cropped image taken at the aligned position. The yellow cross represents the center of the marker and the red cross the center of the image. A combined overall performance error can be calculated, in pixels, as the difference between the two points.

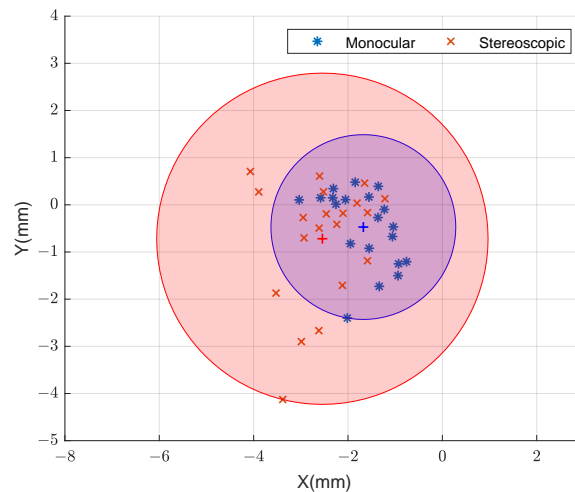
In Figure 24, we show the calculated error using this method for the set of 20 initial camera positions that we have been analyzing so far.

### 3.2.8. Discussion

Through the study of the overall performance of the system, we conclude that both the monocular and stereoscopic approaches satisfy the requirements of the application for the studied range of positions, as all data points are within the  $\pm 3$  mm tolerance (Figure 16). Furthermore, the results are considered to be quite precise, as the errors are distributed within a circle of approximately 1 mm radius around the mean value. There are no significant performance differences between the monocular and stereoscopic approaches. This indicates that the system can be simplified to its monocular version without significant loss of accuracy or precision. This is a clear improvement over methods based on 3D reconstruction from multiple cameras, since the physical space on the manipulator is limited.

When using the reference method for marker detection (checkerboard target), we observe a similar trend in the results (Figure 17). However, precision is lower, with errors

distributed in a circle of around 1.5 mm radius around the mean value, which is likely related with the increased distance between the camera and the target.



**Figure 24.** Error in each of the X and Y axes for monocular (blue) and stereoscopic (red) estimates provided by the image-based method.

In all experiments, regardless of the evaluation metric, we observe an offset in the estimated errors. When relating the local coordinate systems of the targets with the physical setup, it becomes clear that the offset has a strong directional component that coincides with the gravity vector. Furthermore, the remaining directional components of the offset tend to follow the direction in which the end effector extends from the base. The offset can be traced back to an error in the positions reported by the control system of the robot, and it is higher the further the arm extends from the robot base. This data leads us to conclude that the offset is related to a deflection of the robot base under the manipulator's weight.

This is a known problem in robotics, often tackled through non-kinematic calibration of the manipulator. Non-kinematic calibration of robot manipulators is of particular interest to the ITER application due to the fact that the high payloads can cause considerable deflections of the robot base [53]. However, the investigation of such methods falls out of the scope of this work, and in our results, no methods have been applied to compensate for this deviation.

Therefore, in the upcoming discussion on how several characteristics of the initial pose affect the quality of our estimate, we consider better estimates those that are nearest the mean error value.

In this setting, it is clear from Figure 18 that the overall system provides better estimates for less slanted initial camera positions. This is likely influenced by the characteristics of the developed marker detection algorithm, which assumes that the centroid of the detected ellipses in the images corresponds to the center of the circular markers. This approximation is better when the camera plane and the object plane are parallel. We also note that the stereoscopic estimation seems to have a more advantageous performance for higher angles, reducing the spread of the results and increasing the precision of our estimation.

In Figure 19, we see that a higher distance from the central position is similarly associated with worst estimates, although the difference is not as obvious as in the previous case.

As Figures 18 and 19 indicate, images from initial poses with lower angle and smaller distance generally lead to consistent estimates closest to the mean error, i.e., have the best precision. In Figure 20, we see a slightly different pattern. Higher distance from the target is associated with higher error, while poses closer to the target have error closer to 0. We can

conclude that there is a direct influence of the distance to the target (and degree to which the robot arm is extended) on the reported error values.

As the comparison between Figures 16 and 21 shows, our proposed approach significantly outperforms, as hypothesized, the state-of-the-art method for pose estimation in the ITER environment, even while using a single camera. Furthermore, our approach is significantly more tolerant to the variation of light conditions, needing far less fine-tuning to be adapted to a new setting. The maximum deviation from the mean is, in the best case, around 3 mm, contrasting with the 1 mm of the proposed approach and consuming all the available tolerance.

The experiment that uses an image-based metric (Figure 24) provides a similar distribution of error to those of the other experiments, although both the precision and accuracy are lower due to the inclusion of the alignment task.

#### 4. Conclusions

Our studies demonstrate that marker-based tracking is a suitable option for localization of the elements of the divertor remote handling system of the ITER, as well as for other similar applications. Even though characteristics of these environments might invalidate the use of retro-reflective markers of the kind conventionally used in photogrammetry, such as retro-reflective tape, we have shown that it is possible to develop a custom made retro-reflective marker out of materials that are commonly used in industrial environments (stainless steel and fused silica). The results of the overall system performance are within the targeted range of error and the system was demonstrated to work quite robustly, even without compensating for the deviations in the positions reported by the manipulator's control system. Furthermore, the proposed approach was demonstrated to be considerably better than the earlier stereo-matching-based solution. It was demonstrated that the system can be simplified to a single camera without significant loss of accuracy or precision.

Further developments of this work should include the optimization of the brilliancy of the retro-reflective markers, as well as the development of marker detection and correspondence methods to allow handling of marker occlusion. Further, it is particularly relevant that future studies consider how the errors caused by the deflection of the robot base under the manipulator's weight can be estimated and compensated, taking into account the high and variable payloads in the ITER application. How the information computed by the proposed system can be optimally presented to the operator and the effect of the availability of this information on teleoperation performance should also be subject to a dedicated study in the future.

**Author Contributions:** Conceptualization, L.G.R. and O.J.S.; Methodology, L.G.R. and O.J.S.; Project administration, O.J.S. and E.R.M.; Resources, A.D.; Software, L.G.R.; Supervision, O.J.S., S.P., and A.G.; Validation, L.G.R.; Writing—original draft, L.G.R.; Writing—review and editing, O.J.S., S.P., and A.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** The work in this paper was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska Curie grant agreement No. 764951, Immersive Visual Technologies for Safety-Critical Applications and by Fusion for Energy (F4E), and Tampere University under the F4E grant contract F4E-GRT-0901. This publication reflects the views only of the authors, and Fusion for Energy cannot be held responsible for any use which may be made of the information contained herein. The research infrastructure of the Center for Immersive Visual Technologies (CIVIT) at Tampere University provided the robotic manipulator, sensors, and laboratory space for conducting the experiments.

**Conflicts of Interest:** The authors declare no conflict of interest. This publication reflects the views only of the authors, and Fusion for Energy cannot be held responsible for any use which may be made of the information contained herein.

## References

- Pérez, L.; Rodríguez, Í.; Rodríguez, N.; Usamentiaga, R.; García, D.F. Robot guidance using machine vision techniques in industrial environments: A comparative review. *Sensors* **2016**, *16*, 335. [\[CrossRef\]](#) [\[PubMed\]](#)
- Malamas, E.N.; Petrakis, E.G.; Zervakis, M.; Petit, L.; Legat, J.D. A survey on industrial vision systems, applications and tools. *Image Vis. Comput.* **2003**, *21*, 171–188. [\[CrossRef\]](#)
- Lichiardopol, S. *A Survey on Teleoperation*; DCT Rapporten; Vol. 2007.155; Technical Report; Technische Universiteit Eindhoven: Eindhoven, The Netherlands, 2007.
- Hoff, W.A.; Gatrell, L.B.; Spofford, J.R. Machine vision based teleoperation aid. *Telemat. Inform.* **1991**, *4*, 403–423. [\[CrossRef\]](#)
- Miyanishi, Y.; Sahin, E.; Makinen, J.; Akpinar, U.; Suominen, O.; Gotchev, A. Subjective comparison of monocular and stereoscopic vision in teleoperation of a robot arm manipulator. *Electron. Imaging* **2019**, *2019*, 659–1–659–6. [\[CrossRef\]](#)
- Tesini, A.; Palmer, J. The ITER remote maintenance system. *Fusion Eng. Des.* **2008**, *83*, 810–816. [\[CrossRef\]](#)
- Sanders, S.; Rolfe, A.; others. The use of virtual reality for preparation and implementation of JET remote handling operations. *Fusion Eng. Des.* **2003**, *69*, 157–161. [\[CrossRef\]](#)
- Heemskerck, C.; De Baar, M.; Boessenkool, H.; Graafland, B.; Haye, M.; Koning, J.; Vahedi, M.; Visser, M. Extending Virtual Reality simulation of ITER maintenance operations with dynamic effects. *Fusion Eng. Des.* **2011**, *86*, 2082–2086. [\[CrossRef\]](#)
- Niu, L.; Aha, L.; Mattila, J.; Gotchev, A.; Ruiz, E. A stereoscopic eye-in-hand vision system for remote handling in ITER. *Fusion Eng. Des.* **2019**. [\[CrossRef\]](#)
- Niu, L.; Smirnov, S.; Mattila, J.; Gotchev, A.; Ruiz, E. Robust pose estimation with the stereoscopic camera in harsh environment. *Electron. Imaging* **2018**, 1–6. [\[CrossRef\]](#)
- Burgess, G.; Shortis, M.; Scott, P. Photographic assessment of retroreflective film properties. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 743–750. [\[CrossRef\]](#)
- Schauer, D.; Krueger, T.; Lueth, T. Development of autoclavable reflective optical markers for navigation based surgery. In *Perspective in Image-Guided Surgery*; World Scientific: Singapore, 2004; pp. 109–117. [\[CrossRef\]](#)
- Bhatnagar, D.K. *Position Trackers for Head Mounted Display Systems: A Survey*; University of North Carolina: Chapel Hill, NC, USA, 1993; Volume 10.
- Vogt, S.; Khamene, A.; Sauer, F.; Niemann, H. Single camera tracking of marker clusters: Multiparameter cluster optimization and experimental verification. In Proceedings of the International Symposium on Mixed and Augmented Reality, Darmstadt, Germany, 1 October 2002; pp. 127–136. [\[CrossRef\]](#)
- Bergamasco, F.; Albarelli, A.; Torsello, A. Pi-tag: A fast image-space marker design based on projective invariants. *Mach. Vis. Appl.* **2013**, *24*, 1295–1310. [\[CrossRef\]](#)
- Nakazato, Y.; Kanbara, M.; Yokoya, N. Localization of wearable users using invisible retro-reflective markers and an IR camera. In Proceedings of the Stereoscopic Displays and Virtual Reality Systems XII, International Society for Optics and Photonics, San Jose, CA, USA, 22 March 2005; Volume 5664, pp. 563–570.
- Sereewattana, M.; Ruchanurucks, M.; Siddhichai, S. Depth estimation of markers for UAV automatic landing control using stereo vision with a single camera. In Proceedings of the International Conference Information and Communication Technology for Embedded System, Ayutthaya, Thailand, 23–25 January 2014.
- Faessler, M.; Mueggler, E.; Schwabe, K.; Scaramuzza, D. A monocular pose estimation system based on infrared leds. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–5 June 2014; pp. 907–913. [\[CrossRef\]](#)
- Hattori, S.; Akimoto, K.; Fraser, C.; Imoto, H. Automated procedures with coded targets in industrial vision metrology. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 441–446.
- Tushev, S.; Sukhovilov, B.; Sartasov, E. Architecture of industrial close-range photogrammetric system with multi-functional coded targets. In Proceedings of the 2017 2nd International Ural Conference on Measurements (UralCon), Chelyabinsk, Russia, 16–19 October 2017; pp. 435–442. [\[CrossRef\]](#)
- San Biagio, M.; Beltran-Gonzalez, C.; Giunta, S.; Del Bue, A.; Murino, V. Automatic inspection of aeronautic components. *Mach. Vis. Appl.* **2017**, *28*, 591–605. [\[CrossRef\]](#)
- Liu, T.; Burner, A.W.; Jones, T.W.; Barrows, D.A. Photogrammetric techniques for aerospace applications. *Prog. Aerosp. Sci.* **2012**, *54*, 1–58. [\[CrossRef\]](#)
- Clarke, T.A. Analysis of the properties of targets used in digital close-range photogrammetric measurement. In Proceedings of the Videometrics III, International Society for Optics and Photonics, Boston, MA, USA, 6 October 1994; Volume 2350, pp. 251–262. [\[CrossRef\]](#)
- Dong, M.; Xu, L.; Wang, J.; Sun, P.; Zhu, L. Variable-weighted grayscale centroiding and accuracy evaluating. *Adv. Mech. Eng.* **2013**, *5*, 428608. [\[CrossRef\]](#)
- Bender, P.; Wilkinson, D.; Alley, C.; Currie, D.; Faller, J.; Mulholland, J.; Siverberg, E.; Plotkin, H.; Kaula, W.; MacDonald, G. The Lunar laser ranging experiment. *Science* **1973**, *182*, 229–238. [\[CrossRef\]](#)
- Feng, Q.; Zhang, B.; Kuang, C. A straightness measurement system using a single-mode fiber-coupled laser module. *Opt. Laser Technol.* **2004**, *36*, 279–283. [\[CrossRef\]](#)

27. Nakamura, O.; Goto, M.; Toyoda, K.; Takai, N.; Kurosawa, T.; Nakamata, T. A laser tracking robot-performance calibration system using ball-seated bearing mechanisms and a spherically shaped cat's-eye retroreflector. *Rev. Sci. Instrum.* **1994**, *65*, 1006–1011. [\[CrossRef\]](#)
28. Shortis, M.R.; Seager, J.W. A practical target recognition system for close range photogrammetry. *Photogramm. Rec.* **2014**, *29*, 337–355. [\[CrossRef\]](#)
29. Ribeiro, L.G.; Suominen, O.; Peltonen, S.; Morales, E.R.; Gotchev, A. Robust Vision Using Retro Reflective Markers for Remote Handling in ITER. *Fusion Eng. Des.* **2020**, *161*, 112080. [\[CrossRef\]](#)
30. Lundvall, A.; Nikolajeff, F.; Lindström, T. High performing micromachined retroreflector. *Opt. Express* **2003**, *11*, 2459–2473. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Stoudt, M.; Vedom, K. Retroreflection from spherical glass beads in highway pavement markings. 1: Specular reflection. *Appl. Opt.* **1978**, *17*, 1855–1858. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Yuan, J.; Chang, S.; Li, S.; Zhang, Y. Design and fabrication of micro-cube-corner array retro-reflectors. *Opt. Commun.* **2002**, *209*, 75–83. [\[CrossRef\]](#)
33. Shin, S.Y.; Lee, J.I.; Chung, W.J.; Cho, S.H.; Choi, Y.G. Assessing the refractive index of glass beads for use in road-marking applications via retroreflectance measurement. *Curr. Opt. Photonics* **2019**, *3*, 415–422. [\[CrossRef\]](#)
34. Yongbing, L.; Guoxiong, Z.; Zhen, L. An improved cat's-eye retroreflector used in a laser tracking interferometer system. *Meas. Sci. Technol.* **2003**, *14*, N36. [\[CrossRef\]](#)
35. MATLAB. Version 9.6.0 (R2019a); The MathWorks Inc.: Natick, MA, USA, 2019.
36. Zhang, Z. A flexible new technique for camera calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [\[CrossRef\]](#)
37. Heikkilä, J.; Silven, O. A four-step camera calibration procedure with implicit image correction. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, 17–19 June 1997; pp. 1106–1112. [\[CrossRef\]](#)
38. Tan, L.; Wang, Y.; Yu, H.; Zhu, J. Automatic camera calibration using active displays of a virtual pattern. *Sensors* **2017**, *17*, 685. [\[CrossRef\]](#)
39. Ali, I.; Suominen, O.; Gotchev, A.; Morales, E.R. Methods for Simultaneous Robot-World-Hand-Eye Calibration: A Comparative Study. *Sensors* **2019**, *19*, 2837. [\[CrossRef\]](#)
40. Bradley, D.; Roth, G. Adaptive thresholding using the integral image. *J. Graph. Tools* **2007**, *12*, 13–21. [\[CrossRef\]](#)
41. Soille, P. *Morphological Image Analysis: Principles and Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013. [\[CrossRef\]](#)
42. Haralick, R.M.; Shapiro, L.G. *Computer and Robot Vision*; Addison-Wesley: Boston, MA, USA, 1992; Volume 1, pp. 28–48.
43. Ahn, S.J.; Rauh, W.; Kim, S.I. Circular coded target for automation of optical 3D-measurement and camera calibration. *Int. J. Pattern Recognit. Artif. Intell.* **2001**, *15*, 905–919. [\[CrossRef\]](#)
44. Mallon, J.; Whelan, P.F. Which pattern? Biasing aspects of planar calibration patterns and detection methods. *Pattern Recognit. Lett.* **2007**, *28*, 921–930. [\[CrossRef\]](#)
45. Collins, T.; Bartoli, A. Infinitesimal plane-based pose estimation. *Int. J. Comput. Vis.* **2014**, *109*, 252–286. [\[CrossRef\]](#)
46. Schweighofer, G.; Pinz, A. Robust pose estimation from a planar target. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 2024–2030. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Sturm, P. Algorithms for plane-based pose estimation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2000 (Cat. No. PR00662), Hilton Head Island, SC, USA, 13–15 June 2000; Volume 1, pp. 706–711. [\[CrossRef\]](#)
48. Hartley, R.; Zisserman, A. *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, MA, USA, 2003. [\[CrossRef\]](#)
49. Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Q. Appl. Math.* **1944**, *2*, 164–168. [\[CrossRef\]](#)
50. Marquardt, D.W. An algorithm for least-squares estimation of nonlinear parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441. [\[CrossRef\]](#)
51. Rodriguez, O. Des lois geometriques qui regissent les déplacements d'un systeme solide dans l'espace et de la variation des coordonnees provenant de déplacements consideres independamment des causes qui peuvent les produire. *J. Math. Pures Appl.* **1840**, *5*, 380–440.
52. Geiger, A.; Moosmann, F.; Car, Ö.; Schuster, B. Automatic camera and range sensor calibration using a single shot. In Proceedings of the 2012 IEEE International Conference on Robotics and Automation, Saint Paul, MN, USA, 14–18 May 2012; pp. 3936–3943. [\[CrossRef\]](#)
53. Kivelä, T.; Saarinen, H.; Mattila, J.; Hämäläinen, V.; Siuko, M.; Semeraro, L. Calibration and compensation of deflections and compliances in remote handling equipment configurations. *Fusion Eng. Des.* **2011**, *86*, 2043–2046. [\[CrossRef\]](#)



# **PUBLICATION**

## **III**

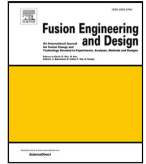
### **Design and manufacturing of an optimized retro reflective marker for photogrammetric pose estimation in ITER**

L. G. Ribeiro, O. J. Suominen, P. Bates, S. Peltonen, E. R. Morales, and  
A. Gotchev

Fusion Engineering and Design, vol. 191, p. 113 557  
DOI: 10.1016/j.fusengdes.2023.113557

**Publication reprinted with the permission of the copyright holders.**





# Design and manufacturing of an optimized retro reflective marker for photogrammetric pose estimation in ITER

Laura Gonçalves Ribeiro<sup>a,\*</sup>, Olli J. Suominen<sup>a</sup>, Philip Bates<sup>c</sup>, Sari Peltonen<sup>a</sup>, Emilio Ruiz Morales<sup>b</sup>, Atanas Gotchev<sup>a</sup>

<sup>a</sup> Faculty of Information Technology and Communication, Tampere University, 33720 Tampere, Finland

<sup>b</sup> Fusion for Energy (F4E), ITER Delivery Department, Remote Handling Project Team, 08019 Barcelona, Spain

<sup>c</sup> Fusion for Energy, c/ Josep Pla, no 2 - Torres Diagonal Litoral - Edificio B3, 08019, Barcelona, Spain

## ARTICLE INFO

### Keywords:

Remote handling  
Pose estimation  
Photogrammetry  
Retro reflector  
Retro reflective marker  
Cat's-eye retro reflector

## ABSTRACT

Retro reflective markers can remarkably aid photogrammetry tasks in challenging visual environments. They have been demonstrated to be key enablers of pose estimation for remote handling in ITER. However, the strict requirements of the ITER environment have previously markedly constrained the design of such elements and limited their performance. In this work, we identify several retro reflector designs based on the cat's eye principle that are applicable to the ITER usecase and propose a methodology for optimizing their performance. We circumvent some of the environmental constraints by changing the curvature radius and distance to the reflective surface. We model, manufacture and test a marker that fulfils all the application requirements while achieving a performance increase of at least 50% over the previous solution in the targeted working range.

## 1. Introduction

A retro reflector is an optical element that returns incoming light in a cone around its source for a wide range of incident directions. Retro reflective (RR) markers have often been used in photogrammetry to provide reliable, high contrasting features in visually challenging environments [1,2]. They are mainly characterized by their *brilliance* (reflectance for a specific observation angle), *divergence* (maximum angular distribution of retro reflected light) and *angularity* (brilliance as a function of the incident angle). An ideal retro reflector has the lowest divergence and highest angularity values. However, in photogrammetry, the light source and the sensor are usually not co-located. Therefore, a good retro reflector for such applications should have a divergence value that covers the viewing angles, without overly exceeding them.

The use of RR markers for pose estimation in ITER has been explored in a previous study [3]. The results of that work demonstrated that RR marker-enabled tracking simplifies the 3D node system [4] by using a single camera only and achieves good performance in terms of accuracy, precision, and robustness. The main objective of the previous study was to engineer and manufacture an RR marker from materials that can withstand the extreme environmental conditions, namely high temperature and radiation, the knuckle of the divertor's cassette locking system is exposed to. The proposed approach consisted of an array of spherical fused silica glass beads with a rear face coating, held by

a stainless steel housing and a sieve-shaped cover. Each glass bead acts as a cat's eye retro reflector, where the primary lens focuses light into a reflective back layer. The design complies with the environment's requirements. Stainless steel and fused silica are already used as material components in the environment. Further, the changes to the relevant optical properties of fused silica, namely its refractive index and absorbance, have been found to be negligible in the targeted wavelengths of 380 to 700 nm [5]. Furthermore, the design avoids the problematic use of adhesives.

The main problem with the earlier approach is that the proposed retro reflector model has only two degrees of freedom: the radius of the sphere ( $r$ ) and the refractive index of the sphere's material ( $n$ ). Given the refractive index of fused silica ( $n_{\text{SiO}_2} = 1.4585$ ), the marker has a divergence of approximately 25 degrees, as illustrated in Fig. 1, left side, while the observation angles in the application can be led to be a few degrees, even at shorter working distances.

The divergence could be optimized by using a type of glass with a higher refractive index, as shown in Fig. 1, upper right corner. However, only a very limited subset of optical glasses are suitable for the high total integrated gamma radiation doses ( $>10$  MGy) in the ITER divertor region. Most optical glasses rapidly become absorbing. Fused silica is overall the best option due to availability, cost, radiation resistance and the fact that it has already been tested [6] and cleared [7]

\* Corresponding author.

E-mail address: [laura.goncalvesribeiro@tuni.fi](mailto:laura.goncalvesribeiro@tuni.fi) (L. Gonçalves Ribeiro).

<https://doi.org/10.1016/j.fusengdes.2023.113557>

Received 9 May 2022; Received in revised form 5 January 2023; Accepted 9 February 2023

Available online 15 February 2023

0920-3796/© 2023 Published by Elsevier B.V.

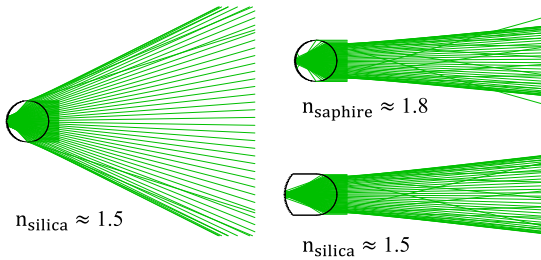


Fig. 1. Retro reflected light distribution of a fused silica glass bead (left side), narrower distribution produced by a type of glass with higher refractive index (upper right corner) and by manipulating the distance between the lens and the reflective surface.

for use in ITER. While using fused silica, it is not possible to optimize brilliancy at the viewing angles within the spherical glass bead design.

In this study, we aim at designing an alternative RR marker with narrower light distribution, which fulfils all the constraints of the application. We hypothesize that this can be achieved having a model with higher number of degrees of freedom, allowing to better optimize brilliancy at the observation angles. More specifically, we focus on retro reflector designs where the distance between the focusing element and the reflective back layer can be manipulated, as shown in Fig. 1, lower right corner.

In Section 2, we present three marker types and propose a methodology for optimizing their performance. We analyse the pros and cons of each type and propose a manufacturing methodology for the most advantageous option. In Section 3, we compare the performance of the developed solutions versus the state of the art using both ray tracing simulations and real data.

## 2. Methodology

### 2.1. Retro reflective marker designs

The retro reflector design proposed in [8] is used as a starting point in our work. It is composed of a front layer of densely packed convex spherical micro lenses, a back surface of densely packed concave spherical micro mirrors and a transparent layer in between. The lens and the mirror have coinciding optical axes, a common aperture diameter and each lens and corresponding mirror form one retro reflecting element. The distance between the primary lens and secondary spherical mirror can be adjusted. This design is essentially based on a biconvex microlens array. We refer to it as “type A” in Fig. 2. It has five degrees of freedom:

- the curvature radius of the focusing lens ( $R_l$ ),
- the curvature radius of the mirror ( $R_m$ ),
- the aperture diameter of the lens and mirror ( $a$ ),
- and the refractive index of the spacer material ( $n$ ).

The main challenge we anticipate with this approach is that the custom manufacturing of biconvex fused silica micro lenses is rather resource consuming. The tooling costs at the prototyping stages are high for any manufacturing techniques we are aware of. Further, the higher volume production is still expensive for the relatively small amount of units that the application requires. As a solution to these impediments, we consider two alternative marker designs where it is still possible to vary the distance,  $d$ , and the curvature radius of the focusing and reflective surfaces,  $R_l$  and  $R_m$ .

Design type B considers the use of a plano convex micro lens array to focus light onto a reflective piece. We expect that an off-the-shelf lens array can potentially be used, which would reduce the cost of the prototyping and manufacturing stages. The convex planar

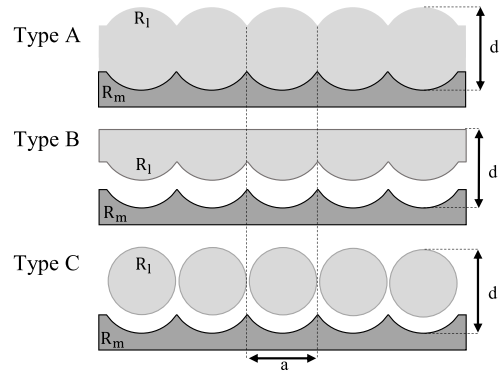


Fig. 2. Retro reflector design alternatives: based on a biconvex micro lens array (type A), a plano convex microlens array (type B), and an array of ball lenses (type C).

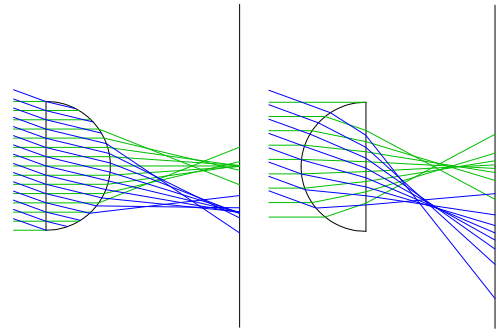


Fig. 3. Refraction of incoming light rays of different orientations by a plano convex (left) and convex plano (right) lens.

option, where the micro lens array is flipped by  $180^\circ$ , would have some desirable characteristics in relation to its counterpart, such as lower comatic and spherical aberration. However, this option suffers from a higher value of astigmatism and was discarded for this reason. We see in Fig. 3 (right side) an illustration of the effects of astigmatism on the tangential plane for the two varieties. For higher values of astigmatism, the focus points fall onto a parabolic surface and do not satisfy the Petzval condition [9].

Design type C considers the use of several spherical glass lenses that are aligned and kept at a given distance to a reflective plate. The main advantage of this approach is the ease of manufacturing and consequent low cost of spherical glass beads and their intrinsically high aperture. The main disadvantage is the increased challenge in manufacturing a housing where each individual lens is held in place and aligned to the corresponding reflective surface. The alignment requires particularly high precision if the ball lenses have small dimensions.

The use of conventionally available reflective coatings (e.g. gold and silver) in the ITER environment adds risk, both due radiation resistance of the coating itself, or the extra materials needed to adhere the coating to the glass surface, as due to concerns about activation of the adhering materials. Therefore, for all marker types, the reflective part shall consist of a polished metal piece that has been indented with the array of curved surfaces.

### 2.2. Design constraints

There is not an absolute requirement on the retro reflectivity of markers at ITER. The main objective in a photogrammetry setup is

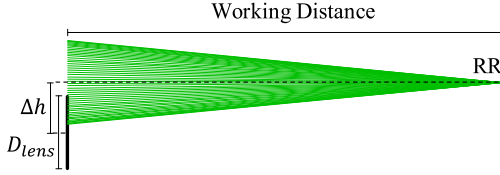


Fig. 4. Plan view of design constraints. We aim at optimizing the number of rays arriving at a camera lens with diameter  $D_{lens}$ , which is displaced from the light source by a fixed distance  $\Delta h$ . The working distance is variable within a certain range.

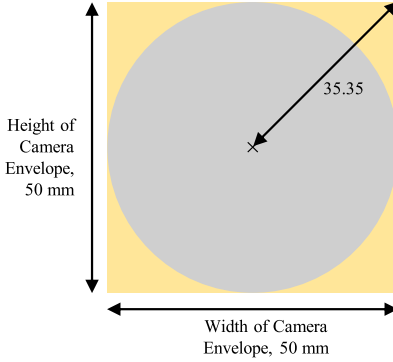


Fig. 5. Plan view of camera and lighting element envelope with maximum possible lens footprint.

to maximize the contrast, as seen in camera images, between the markers and the background. A marker with a high retro reflectivity value makes it possible to underexpose and largely eliminate challenging backgrounds and largely facilitates segmentation. Therefore, while designing an optimized RR marker in our work, the objective is to maximize the percentage of incident light rays arriving at the camera lens. As a starting step, the size of the camera lens, the distance of the camera to the lighting element and the working distance, as seen in Fig. 4, need to be specified.

In our target use case, the camera is to be installed on the wrist of a robotic manipulator, where space is severely limited. The maximum allowable envelope for the camera and lighting is  $50 \times 50$  mm. Within this envelope we can have, at most, a lens radius of 25 mm and maximum distance between the lens and the lamp of approximately 35 mm, as shown in Fig. 5.

Further, the working distance is fixed based on the available space, to a range of 300 to 500 mm. As angularity goes, we aim at maximizing the uniformity of light return for entrance angles in the range of  $-20^\circ$  to  $20^\circ$ .

### 2.3. Optimization of retro reflector performance

The proposed methodology for designing an optimized cat's eye retro reflector consists of the following sequential steps:

**Step 1:** Establishing the size of each individual retro reflecting unit. The size of each individual retro reflecting unit determines what is the maximum possible aperture size. The maximum aperture size can be chosen based on the camera resolution and working distance, such that the individual structures that constitute one marker are indistinguishable in the camera view. This will lead to the marker being more easily perceived, at later stages, as a uniform object. However, there is a trade off between the size of the features and the manufacturing challenges. The smaller the size of the features, the tighter the requirements on the

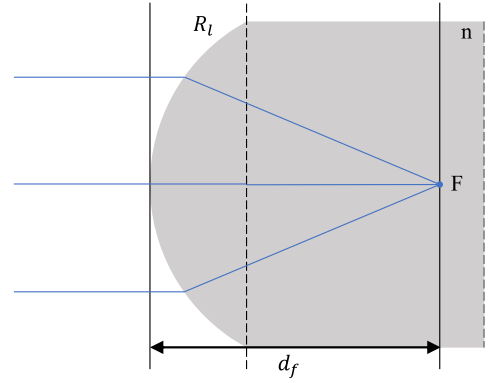


Fig. 6. Generic lens structure composed of one curved medium interface with curvature radius  $R_l$ . The lens has refractive index  $n$  and the surrounding medium has refractive index  $n_{air}$ . Paraxial focus is at the lens medium, at distance  $d_f$  from the first interface.

manufacturing tolerances will be to ensure the designed functionality of the piece, particularly when it comes to the alignment of the optical axes of lens and mirror. As the aperture can only be as large as the curvature radius of the lens, it follows that  $R_l = a_{max}/2$ . In the conditions described in Section 2.2, and considering a typical camera resolution of  $640 \times 480$  pixels, we start with  $a_{max} = 0.4$  mm and  $R_l = 0.2$  mm.

For marker type C, this step corresponds to the selection of the size of the glass spheres. In this case, it is important to consider the additional manufacturing challenges in holding and aligning several independent elements of small dimensions at a fixed distance to the reflective plate. For this reason, the use of larger sized ball lenses might be worth considering, even if it leads to the final RR element not being perceived as fully uniform.

**Step 2:** Establishing the distance between the lens and the reflective surface ( $d$ ), based on the distance at which the paraxial rays converge ( $d_f$ ).

The paraxial focal length of the lens can be calculated, for each design type, as described below [9].

- For type A, i.e. the biconvex lenses case, there is one interaction in one single interface, as shown in Fig. 6:

$$d_f = \frac{nR_l}{n-1}. \quad (1)$$

- For type B, i.e. the plano convex lenses case, the first interface is a plane ( $r_1 = \infty$ ), as shown in Fig. 7:

The power of each individual interface ( $K_1$ ,  $K_2$ ) and the total power of the lens ( $K$ ) are as follows:

$$K_1 = 0, \quad K_2 = -\frac{n-1}{-R_l}, \quad (2)$$

$$K = K_1 + K_2 - K_1 K_2 \frac{t}{n}, \quad (3)$$

$$K = \frac{n-1}{R_l}. \quad (4)$$

The effective focal length is given by:

$$f = \frac{1}{K} = \frac{R_l}{n-1} \quad (5)$$

and the back focal length (BFL) is given by:

$$BFL = f \frac{1}{n} = \frac{R_l}{n(n-1)}. \quad (6)$$

Therefore,

$$d_f = BFL + t = \frac{R_l}{n(n-1)} + t. \quad (7)$$

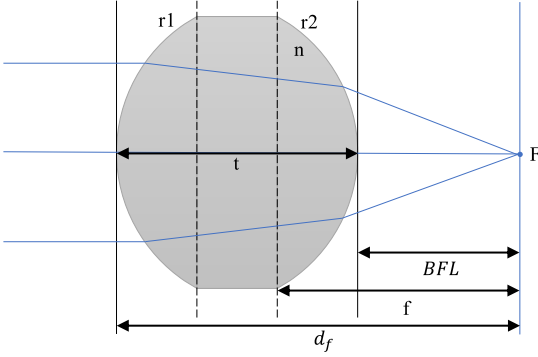


Fig. 7. Generic lens structure composed of two curved medium interfaces with radii  $r_1$  and  $r_2$  and a total thickness  $t$ . The lens has refractive index  $n$  and the surrounding medium has refractive index  $n_{air}$ . Paraxial focus is at the outgoing medium, at distance BFL from the last interface.

- For type C, i.e. the ball lenses case,  $r_1 = r_2 = R_l$ ,  $t = 2R_l$ , as shown in Fig. 7.

The power of each individual interface ( $K_1$ ,  $K_2$ ) and the total power of the lens ( $K$ ) are:

$$K_1 = \frac{n-1}{R_l}, \quad K_2 = -\frac{n-1}{-R_l}, \quad (8)$$

$$\begin{aligned} K &= K_1 + K_2 - K_1 K_2 \frac{2R_l}{n} \\ &= \frac{n-1}{R_l} + \frac{n-1}{R_l} - \left( \frac{n-1}{R_l} \right)^2 \frac{2R_l}{n} \\ &= \frac{2n^2 - 2n - 2n^2 + 4n - 2}{nR_l} \\ &= \frac{2(n-1)}{nR_l}. \end{aligned} \quad (9)$$

The effective focal length is given by:

$$f = \frac{1}{K} = \frac{nR_l}{2(n-1)} \quad (10)$$

and the BFL is given by:

$$BFL = f \frac{2-n}{n} = \frac{R_l(2-n)}{2(n-1)}. \quad (11)$$

Therefore,

$$d_f = BFL + 2R_l = \frac{R_l(2-n)}{2(n-1)} + 2R_l. \quad (12)$$

The distance to the mirror ( $d$ ) can be set equal or different to the focus distance ( $d_f$ ), depending on the desired optical behaviour. In Fig. 8 we show the light ray distributions corresponding to different configurations ( $d = d_f$ ,  $d = d_f - 0.08$ ,  $d = d_f - 0.16$  and  $d = d_f - 0.24$ ) within marker type A. Similar information is presented in Fig. 9 for marker type C. We find that specific configurations with  $d < d_f$  produce a narrower cone of retro reflection than  $d = d_f$ . This allows us to tailor the divergence to the use case to a certain degree. Further, in these same configurations we observe a light ray distribution that favours the periphery of the cone. This type of distribution is advantageous in further optimizing the number of rays arriving at the lens.

Fig. 10 quantifies the percentage of incident rays arriving at the circular surface of the camera lens for several marker configurations within type A. The data in this and in subsequent figures in Section 2, is obtained using a commercial sequential raytracing software (OSLO by Lambda Research Corporation), considering a single light source that is vertically displaced in relation to the circular surface of the lens. The displacement distance is as described in Section 2.2.

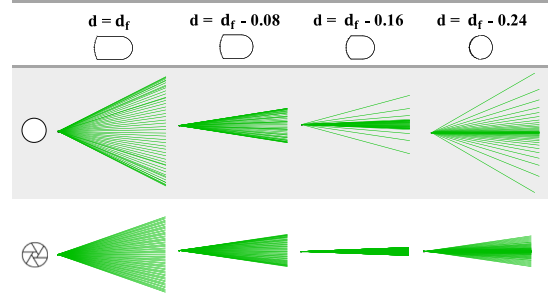


Fig. 8. Light ray distributions for different RR marker configurations of type A. The distance to the reflective surface ( $d$ ) varies in relation to the focus distance ( $d_f$ ). In the top row we show the maximum (0.4 mm), and on the bottom row a smaller aperture.

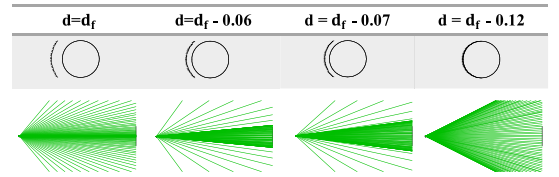


Fig. 9. Light ray distributions for different RR marker configurations of type C. The distance to the reflective surface ( $d$ ) is varied in relation to the focus distance ( $d_f$ ). The spheres have a radius of 0.2 mm.

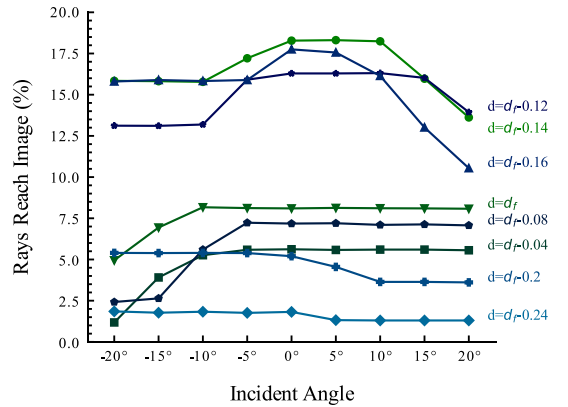
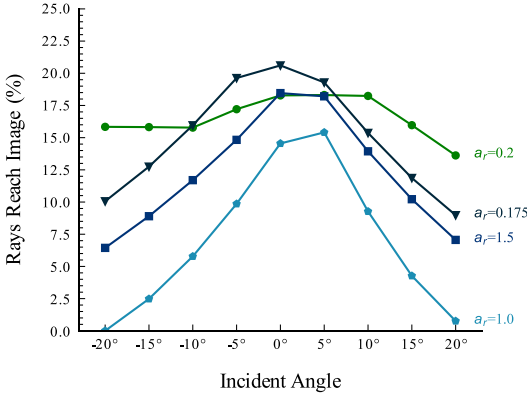


Fig. 10. Behaviour of RR marker configurations of type A with varying distance to the reflective surface. The curvature radius and aperture of the primary lens are constant with  $R_l = 0.2$  and  $a = 0.4$ . The curvature radius of the reflective surface,  $R_m$ , is changed accordingly, to fulfil the condition of retro reflection. Working distance is 300 mm.

A similar analysis can be easily obtained for the other two configurations. From this type of representation the most advantageous configuration can be easily determined. While Fig. 10 shows the percentage of returned rays for a single working distance, in determining the most advantageous configuration the full range of working distances must be considered.

**Step 3: Fulfilling the retro reflection condition.** Due to Petzval field curvature aberration, as we go further away from the optical axis, the focal point does not actually fall on a focal plane at distance  $d_f$ , but rather on a spherical surface. In order to obtain a retro reflector that keeps its RR behaviour over a wide range of incident directions, the reflective surface should accompany this curvature [11]. This ensures



**Fig. 11.** Behaviour of RR marker configurations of type A with varying aperture radius ( $a_r$ ). The curvature radius of the primary lens and reflective surface, and the distance between them are constant with  $R_l = 0.2$ ,  $d = d_f - 0.14$  and  $R_m = 0.3$ . Working distance is 300 mm.

that focus is kept at the same distance relative to the reflective surface regardless of the incident angle.

In the case where the reflective surface is placed at the focus distance, its curvature radius should equal the Petzval radius of the lens. For a generic  $d$ , the condition of retro reflection can be defined as a function of the Petzval radius ( $R_{Petzval}$ ) and the distances  $d$  and  $d_f$ , for each marker type, as described below.

- For type A, the condition of retro reflection is as follows:

$$R_m = d - R_l. \quad (13)$$

- For type B, the Petzval radius is:

$$R_{Petzval} = -\frac{n R_l}{n-1} \quad (14)$$

and the retro reflection condition is:

$$\begin{aligned} R_m &= -\frac{n R_l}{n-1} - BFL + d \\ &= -\frac{n R_l}{n-1} - \frac{R_l}{n(n-1)} + d \\ &= -\frac{R_l(n+1)}{n} + d. \end{aligned} \quad (15)$$

- For type C, the Petzval radius is:

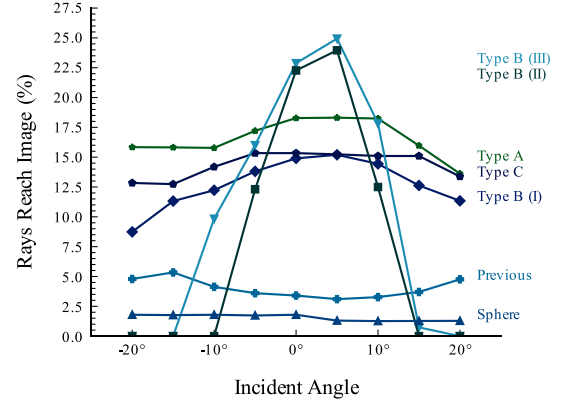
$$R_{Petzval} = \frac{nR}{2(n-1)} \quad (16)$$

and the retro reflection condition:

$$\begin{aligned} R_m &= \frac{nR}{2(n-1)} - BFL + d \\ &= \frac{nR}{2(n-1)} - \frac{r(2-n)}{2(n-1)} + d \\ &= R_l + d. \end{aligned} \quad (17)$$

**Step 4:** Establishing the final aperture value ( $a$ ) within the range  $0 < a < a_{max}$ .

There is an evident trade-off in the establishment of the aperture value. On one hand, smaller aperture values lead to lower divergence through the minimization of the effects of spherical aberration. On the other hand, smaller aperture values are associated with worse angularity. This trade-off is evident in Fig. 11, which quantifies the percentage of incident light rays arriving at the lens surface for several marker configurations of type A.



**Fig. 12.** Behaviour of selected marker configurations of each type. Type A with  $R_l = 0.2$  and  $d = d_f - 0.14$ , type B (I) with  $R_l = 0.2$  and  $d = d_f - 0.03$ , and type C with  $R_l = 0.5$  and  $d = d_f - 0.15$ . Type B (II) and Type B (III) correspond to the two off the shelf varieties. For these types  $R_m$  is changed accordingly to fulfil the retro reflection conditions. “Previous” represents the previous solution, where  $R_l = 1$  and the reflective plate is flat ( $R_m = 0$ ). “Sphere” represents the traditional cat’s eye retro reflector where the reflective plate follows the curvature of the sphere with  $R_l = R_m = 1$ . Working distance is 300 mm.

For this reason, we opt to manipulate the divergence of the RR by changing the distance to the reflective surface (in step 2) and tend to keep the aperture at the highest possible value. This allows us to achieve simultaneously lower divergence and higher angularity values, leading to a higher return in the desired range of incident angles.

In an ideal case, the maximum entrance angle occurs when the paraxial focus strikes the edge of the reflective surface. However, when considering the effects of spherical aberration the outcome becomes less favourable. There may be a considerable percentage of reflected rays that fall out of the aperture of the primary lens, degrading the angular performance. For this reason we consider that the type of analysis in Fig. 11 is more informative than the calculation of the maximum entrance angle.

#### 2.4. Choosing the most advantageous design type

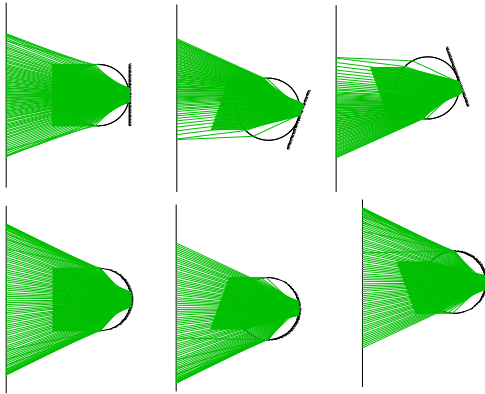
As seen in Fig. 12, we are often able to optimize the retro reflectivity values of type A better than those of types B and C.

However, marker types B and C do not fall far behind and have the potential to significantly improve upon the previous approach. The previous approach differs from a standard cat’s eye retro reflector in the backing plate being flat, rather than accompanying the curvature of the sphere. This creates an asymmetrical retro reflected light distribution, as demonstrated in Fig. 13. This distribution slightly improves performance in our application in relation to the case where the backing accompanies the curvature of the sphere.

We found that it is uncommon to find off-the-shelf varieties of plano convex micro lens arrays with very high effective lens aperture values in comparison to their radius of curvature (also referred to as high lens sag values). In Table 1, we describe the characteristics of the two off-the-shelf lens arrays we found with acceptable dimensions. These types are optimized following steps 2 to 3 in our proposed methodology.

We see in Fig. 12 that of these two varieties, III has, as expected, better angular performance. While the plano convex options have a more limited angular performance than the remaining options, they reach higher values of retro reflectance at smaller viewing angles.

However, there is an additional challenge with this approach that we have not yet considered. Often, for cost efficiency, these arrays are cut from larger sheets and the distribution of lenses in the array is



**Fig. 13.** Light pattern created by the previous marker, where the reflective plate is flat (top), and by a standard cat's eye retro reflector, where the curvature of the reflective surface accompanies the curvature of the sphere (bottom).

**Table 1**

Radius of curvature of the lens ( $R_l$ ), aperture ( $a$ ) and thickness ( $t$ ) of two off-the-shelf plano convex micro lens varieties.

	$R_l$	$a$	$t$
II (Okotech)	0.425	0.111	0.05
III (Suss Microoptics)	0.35	0.25	0.9

not exactly known. This constitutes a problem since the exact distance between each lens and the edge of the array needs to be known to construct the backing plate of corresponding features. Measuring the distance of the first lens to the edge of the array would be possible, although not trivial due to the small size of the lenses. Further, it would be highly inefficient to measure, model and manufacture a marker housing for each individual lens array piece. This led us to discard the use of an off the shelf plano convex lens array from our subsequent studies.

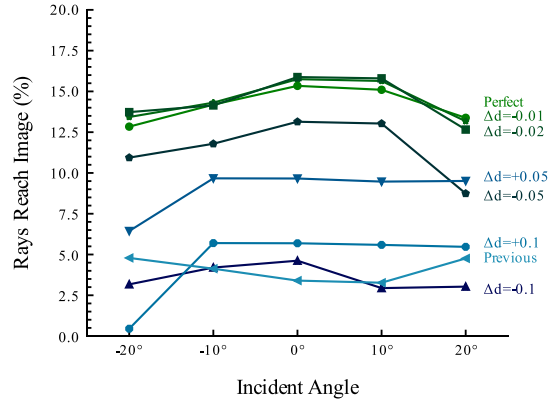
Due to the manufacturability considerations presented above, we deem marker type C to have the best cost-performance trade-off for the targeted use case and, moving forward, we focus on this approach.

We analyse the robustness of the selected marker design to variations of its parameters. We focus primarily on parameters that are influenced by the manufacturing process of the marker housing: the distance to the reflective plate,  $d$ , and the alignment of the optical axis of ball lenses and reflective surfaces. We expect larger tolerances in the manufacturing process in  $d$ . We have empirically found that there are no significant changes for deviations of  $d$  under  $100\ \mu\text{m}$ . In Fig. 14 we see that the design is quite tolerant to errors in this distance. Even for very significant errors ( $\Delta d = 0.1\ \text{mm}$  for  $d = d_f - 0.15\ \text{mm}$ ), the behaviour is not worse than the previous solution. When it comes to the alignment of the optical axis, we have empirically found that deviations in the order of  $10\ \mu\text{m}$  cause no significant changes to the behaviour of the retro reflector. As for radii,  $R_l$  and  $R_m$ , the manufacturing process of spheres is comparatively more controllable and repeatable.

## 2.5. Manufacturing process

In this section we describe the manufacturing process of the retro reflective marker housing, the reflective backing and the required holding structures (Fig. 15).

A tool with inverse features of the reflective backing array (Fig. 16) has been produced by CNC machining and is applied through hydraulic pressing to indent the reflector array into a stainless-steel block (Fig. 15 - part 1). Four holes are added by EDM wire cutting on the same block



**Fig. 14.** Robustness of selected marker configuration of type C ( $R_l = 0.5$ ,  $d = d_f - 0.15$ ,  $R_m = 0.65$ ) to errors in the distance to the reflective surface ( $\Delta d$ ).

— two holes for the guiding pins used to align the assembly (Fig. 15 - part 5), and two holes to act as outgassing channels. Outgassing channels are needed due to the vacuum conditions in ITER. The resulting block is milled to size and threaded on the outer surface.

The distance between the reflective surface and the reflective beads is ensured by a spacer, i.e. a plate with suitably sized holes matching the reflector pattern (Fig. 15 - part 2). This plate is made by EDM wire cutting. Outgassing channels and guiding pin holes are added using the same method. To ensure repeatability, a cylinder with the through holes is first made and then sliced to the specified thickness to create the spacer discs. Considering the expected thermal expansion of materials in the marker (fused silica has a lower thermal expansion coefficient than stainless steel) and to prevent damage of the glass beads resting against sharp edges, the holes' inner edges are bevelled to match the curvature of the beads.

The marker is encapsulated in a stainless-steel cylinder (Fig. 15 - part 4) milled to be hollow and threaded with the corresponding thread to fit the reflector piece. The spacer disc rests on top of the reflector held in place by guiding pins cut from 1 mm thick stainless steel welding wire, and the glass beads (Fig. 15 - part 6) rest on their respective holes on the spacer. In order to hold the glass spheres without using a sieve-like structure, which would reduce the reflective area, or adhesives, which are largely unavailable in the environment, a glass window is introduced in the design of the housing (Fig. 15 - part 7). The topmost feature is a shoulder to keep in place the glass window. An EDM cut spacer ring (Fig. 15 - part 3) is added to avoid pressure of the glass beads on the glass window.

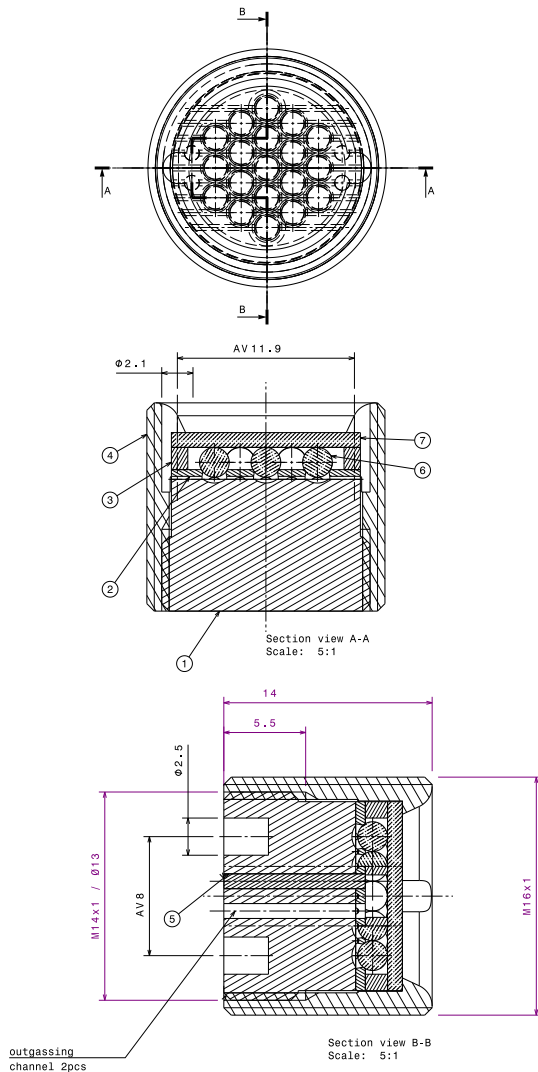
## 3. Results

### 3.1. Raytracing simulations

The analysis done thus far assumes a single perfect specular reflection at the reflective surface of the marker. This is a simplification of the problem at hand. Light can, in fact, get scattered and reflected multiple times at different interfaces, the reflection might not be perfectly specular, and some light might be absorbed by the materials.

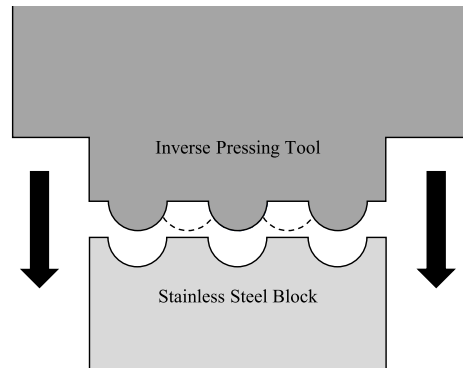
To get a more realistic understanding of the comparative performance of the RR markers, we analysed the performance of the previous and the newly developed marker using a commercial Monte Carlo ray tracing software (TracePro by Lambdare [12]). We analyse the effect of the optical characteristics of the backing plates, particularly their specular, on the performance of RR markers. Fig. 17 shows the simulation results. We see that, while a specularity of 100% would



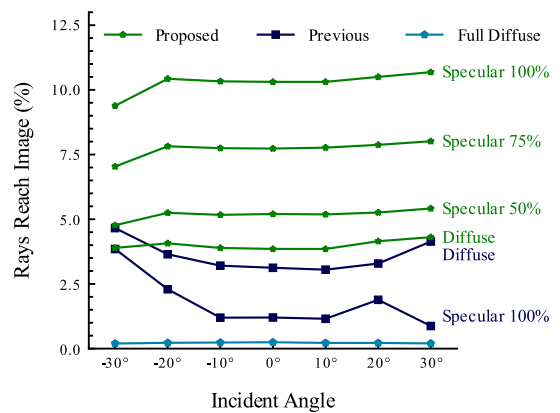


**Fig. 15.** Diagram of full marker assembly. 1- Reflective Base. 2-Spacer Disk. 3- Spacer Ring. 4- Body. 5- Guiding Pins. 6-Spherical Glass Beads. 7-Glass Window.

optimize the behaviour of the proposed marker, a diffuse back plate is the most advantageous for the performance of the previous design. In any case, the proposed design is an improvement over the earlier solution, regardless of the specularity. Let us consider an expected specularity of around 75% in the proposed marker and diffuse in the previous one. In these conditions, the simulation shows an expected performance gain of the proposed solution of 1.5 to 3 times the previous one at the shortest working distance of 300 mm. This simulation also shows the clear performance benefit in the use RR markers over a diffuse marker. Please note that this analysis considers a single light source. When using an assortment of light sources around the camera as shown in Fig. 19 the contributions of opposing light sources are summed and results are symmetric around the 0 angle.



**Fig. 16.** Inverse pressing tool used to indent the array of spherical reflective surfaces into the stainless steel backing.



**Fig. 17.** Behaviour of the previous and proposed markers for different optical characteristics of their respective reflective surfaces according to ray-tracing simulations. “Full Diffuse” refers to a fully diffuse, non RR marker. Working distance is 300 mm.

### 3.2. Experimental evaluation

The angular response of RR markers is measured using a digital camera and a ring light. With this test setup we obtain response curves that are meaningful and directly applicable to photogrammetry use cases. The response is evaluated for increasing entrance and viewing angles by rotating the test samples around their centre points. The experimental setup is depicted in [Fig. 18](#).

The camera has  $1920 \times 1200$  resolution and the lens has a horizontal field of view of 71.2 degrees. The camera resolution corresponds to approximately 0.4 mm per pixel at the sampling distance of 300 mm and 0.5 mm per pixel at the sampling distance of 500 mm.

We compute the response curve at two discrete values (300 mm and 500 mm), corresponding to the limits of the targeted working range.

The setup of light sources is as seen in Fig. 19, with two sets of two diametrically opposite LEDs. The distance from the centre of each LED to the centre of the camera lens is 22 mm. The observation angle (between the centre of the camera and the light source) is, therefore, fixed for a given working distance and is approximately 4 degrees at 300 mm and 3 degrees at 500 mm. Response curves are symmetrical around the 0 degree incident angle, due to the added contributions of opposing light sources.

In designing the RR marker, the optimization procedure described in the methodology section was followed for these conditions. Some

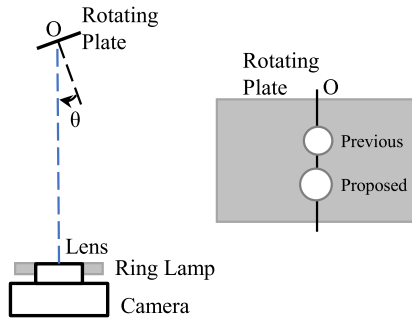


Fig. 18. Experimental setup for RR marker testing (not to scale). Samples are rotated around centre point O by angle  $\theta$ . Camera and LED sources are stationary throughout the experiments.



Fig. 19. Light setup around camera lens composed of four LEDs at the corners of the camera envelope. Distance from the centre of each LED to the centre of the camera lens is 22 mm. Diameter of the camera lens is 30 mm.

adjustments had to be made to accommodate manufacturability aspects. The final design has the following parameters:  $R_l = 0.5$  mm,  $d = 2.39$  mm,  $R_m = 0.625$  mm. The aperture value is 1.5 mm.

The response curves in Fig. 20 are in line with what was anticipated in the theoretical analysis. One can observe a gain of at least 1.5 times the brightness over the previous approach in the relevant range of working distances.

#### 4. Conclusions

In this work, we have proposed novel methods for the design and manufacturing of optimized retro reflective markers. The aim was to circumvent the material restrictions in our intended application and obtain a better response using the allowed fused silica glass material by changing the distance to and the curvature of the reflective surface.

We have identified three alternative design types, based on a custom made biconvex lens array, an off the shelf plano convex lens array and an array of glass spheres. We analysed the options from performance, manufacturing and practicality points of view and decided on the most advantageous type for our application, namely the ball lens option.

We proposed an efficient and scalable method for manufacturing the most advantageous option. We evaluated the performance of the manufactured marker in a photogrammetric setup. This allowed us to effectively access what will be the gain in contrast in the resulting images. The developed marker was shown to increase the acquired image values in 1.5 to 2.5 times. Demonstrably, as predicted by our simulations, the amount of light returned to the camera is effectively much higher and much more uniform in the range of 0 to 20 degree

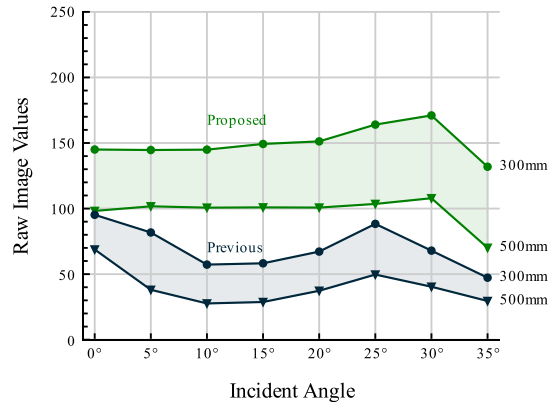


Fig. 20. Experimental evaluation of average intensity of all pixels belonging to the marker, in raw image values.

incident angle range. The marker is effectively easier to discriminate from the surrounding environment than the previous solution, either by automatic segmentation or manual identification by an operator. It has the potential of being a key enabling element of pose estimation in challenging visual conditions.

#### CRedit authorship contribution statement

**Laura Gonçalves Ribeiro:** Conceptualization, Methodology, Validation, Writing – original draft. **Olli J. Suominen:** Conceptualization, Methodology, Project Administration, Supervision, Writing – review & editing. **Philip Bates:** Supervision, Writing – review & editing. **Sari Peltonen:** Supervision, Writing – review & editing. **Emilio Ruiz Morales:** Project administration, Writing – review & editing. **Atanas Gotchev:** Supervision, Writing – review & editing.

#### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Data availability

Data will be made available on request.

#### Acknowledgements

The work leading to this publication has been funded by Fusion for Energy under Grant F4E-GRT-0901. This publication reflects the views only of the authors, and Fusion for Energy cannot be held responsible for any use which may be made of the information contained herein.

#### References

- [1] T. Liu, A.W. Burner, T.W. Jones, D.A. Barrows, Photogrammetric techniques for aerospace applications, *Prog. Aerosp. Sci.* 54 (2012) 1–58, <http://dx.doi.org/10.1016/j.paerosci.2012.03.002>.
- [2] T.A. Clarke, Analysis of the properties of targets used in digital close-range photogrammetric measurement, in: *Videometrics III*, Vol. 2350, International Society for Optics and Photonics, 1994, pp. 251–262, <http://dx.doi.org/10.1117/12.189137>.
- [3] L.G. Ribeiro, O.J. Suominen, S. Peltonen, E.R. Morales, A. Gotchev, Robust vision using retro reflective markers for remote handling in ITER, *Fusion Eng. Des.* 161 (2020) 112080, <http://dx.doi.org/10.1016/j.fusengdes.2020.112080>.

- [4] L. Niu, L. Aha, J. Mattila, A. Gotchev, E. Ruiz, A stereoscopic eye-in-hand vision system for remote handling in ITER, *Fusion Eng. Des.* 146 (2019) 1790–1795, <http://dx.doi.org/10.1016/j.fusengdes.2019.03.036>.
- [5] M. Garcia-Matos, A. Moroño, E. Hodgson, KU1 quartz glass for remote handling and LIDAR diagnostic optical transmission systems, *J. Nucl. Mater.* 283 (2000) 890–893, [http://dx.doi.org/10.1016/S0022-3115\(00\)00195-1](http://dx.doi.org/10.1016/S0022-3115(00)00195-1).
- [6] A. Moroño, E. Hodgson, Radiation induced optical absorption and radioluminescence in electron irradiated SiO<sub>2</sub>, *J. Nucl. Mater.* 258 (1998) 1889–1892, [http://dx.doi.org/10.1016/S0022-3115\(98\)00350-X](http://dx.doi.org/10.1016/S0022-3115(98)00350-X).
- [7] L. Worth, et al., *ITER Vacuum Handbook*, Technical Report ITER\_D2EZ9UM, ITER, Saint-Paul-Lez-Durance, France, 2009.
- [8] A. Lundvall, F. Nikolajeff, T. Lindström, High performing micromachined retroreflector, *Opt. Express* 11 (20) (2003) 2459–2473, <http://dx.doi.org/10.1364/OE.11.002459>.
- [9] E. Hecht, *Optics*, Vol. 2, fourth ed., Addison-Wesley, San Francisco, 2002, p. 3.
- [10] Lamda Research Corp., Littleton, MA, Software OSLO version 20.2.0. URL <https://www.lambdaresearch.com/oslo/>.
- [11] F. Blechinger, B. Aichtner, *Handbook of Optical Systems, Volume 4: Survey of Optical Instruments*, Vol. 4, Wiley-VCH, 2005.
- [12] Lamda Research Corp., Littleton, MA, Software TracePro version 21.1.0. URL <https://lambdaresearch.com/tracepro/>.



# PUBLICATION IV

## **ITER-Tag: An adaptable framework for robust uncoded fiducial marker detection and identification**

L. G. Ribeiro, O. J. Suominen, S. Peltonen, E. R. Morales, and A. Gotchev

In 2022 10th European Workshop on Visual Information Processing (EUVIP), 2022  
DOI: 10.1109/EUVIP53989.2022.9922671

© 2023 IEEE. Reprinted, with permission, from Ribeiro, L. G., Suominen, O. J., Peltonen, S., Morales, E. R., & Gotchev, A. (2022, September). ITER-Tag: An Adaptable Framework for Robust Uncoded Fiducial Marker Detection and Identification. In 2022 10th European Workshop on Visual Information Processing (EUVIP) (pp. 1-6). IEEE.



# ITER-Tag: An Adaptable Framework for Robust Uncoded Fiducial Marker Detection and Identification

<sup>1st</sup> Laura Gonçalves Ribeiro  
Tampere University, Finland  
laura.goncalvesribeiro@tuni.fi

<sup>2nd</sup> Olli J. Suominen  
Tampere University, Finland  
olli.j.suominen@tuni.fi

<sup>3rd</sup> Sari Peltonen  
Tampere University, Finland  
sari.peltonen@tuni.fi

<sup>4th</sup> Emilio Ruiz Morales  
Fusion for Energy (F4E), Spain  
Emilio.Ruiz@f4e.europa.eu

<sup>5th</sup> Atanas Gotchev  
Tampere University, Finland  
atanas.gotchev@tuni.fi

**Abstract**—Fiducial marker-based tracking is an effective method for pose estimation in confined environments, such as the International Thermonuclear Experimental Reactor. In this paper, we propose a novel framework for marker detection and identification that is moderately robust to occlusion, even while using a relatively small number of marks. The proposed approach consists of a hybrid pipeline that extracts marker candidates from images using classical methods and identifies uncoded markers using a shallow feed forward neural network. The method requires minimal parameter tuning, data collection and annotation. The methods can be easily adapted to different use cases with varying number and positions of the marks. We test the robustness of our approach in three different use cases in ITER’s divertor, using either retro reflective markers or laser engravings and achieve high detectability rates. We demonstrate how the proposed approach can be used to accurately and robustly retrieve the six-degree-of-freedom pose of the targets.

**Index Terms**—pose estimation, fiducial markers, retro reflective markers, marker detection, marker identification, remote handling, ITER

## I. INTRODUCTION

Marker based tracking is often used in applications where there is a need to track man-made structures with accuracy in a confined environment. This kind of approach takes full advantage of the constraints in such settings and is often able to reach more accurate and reliable results with less complex algorithmic methods.

Remote handling in the International Thermonuclear Experimental Reactor (ITER) is an example of such cases. ITER constitutes one of the worlds most hostile environments, where operators need to perform tight tolerance alignment tasks remotely and in poor visibility conditions. In this setting there is a need to develop reliable and accurate methods for pose estimation of specific pieces that can act as a remote handling alignment aid.

The work leading to this publication has been funded by Fusion for Energy under Grant F4E-GRT-0901. This publication reflects the views only of the authors, and Fusion for Energy cannot be held responsible for any use which may be made of the information contained herein.

The problem of pose estimation in ITER has been addressed in previous work, with the development of specific markers and of computational methods to detect and identify them and estimate the pose [1]. The earlier work addressed one specific use case - the knuckle of the divertor cassette locking system. The end to end accuracy of the overall hardware-software system was evaluated and reached the targeted values. However, the proposed strategies for marker detection and identification struggles with robustness, particularly in cases where markers are occluded or missing. Further, there is a need to apply the developed methods to different targets and operational cases in ITER. The existing methods are not readily adaptable to different use cases and marker constellation geometries.

The objective of the present work is to develop a novel marker detection and identification framework than can be used with ITER’s newly developed retro reflective (RR) markers [2]. It should be able to reliably identify uncoded markers with a moderate level of resilience to occlusion and minimal manual parameter adjustments. It should rely only on a relatively small number of marks and should be easily adaptable to different constellations with varying sizes of the available area and number of marks. Further, as there are use cases where RR markers cannot be used, the approach should be able to handle, even with lower degree of robustness, non RR marks such as laser engravings. Processing speed is not the main constraint as the detection is not expected to run on a video sequence, although it should still be kept within a reasonable time.

## II. BACKGROUND

Visual marker systems have been extensively studied in the literature. A wide range of approaches with different strengths and limitations has been proposed. Detailed overviews can be found in [3] and [4]. We focus our overview on point-based or concentric circle-based systems with a relatively small number of markers. Figure 1 shows the visual appearance of selected approaches.






Name	Marker
Concentric contracting circles (CCC) [5] Whycon [9-11]	
Pattern of five CCC markers [6][7] Circular-ring marker [8]	
Pi-Tag [12] [13]	
4 Dots [14] [15]	
MarkerPose [16]	

Fig. 1. Overview of relevant fiducial marker systems. Focus is on point-based or concentric circle-based systems with a relatively small number of markers.

One of the first and most popular marker geometries is the concentric contracting circle (CCC) [5]. It consists of a small white circle whose centre is coincidental with a larger black circle. In the seminal approach [5], individual CCCs are extracted from images by thresholding based segmentation. The centroids of the white and black regions are computed and compared. The robustness of the method is based on the unlikelihood that random contrasting intensity regions have nearly coincident centroids. Additional filtering steps such as comparing the ratio of areas of the inner and outer circles and comparing the bounding box of the regions can be applied to eliminate false candidates. This detection strategy is not completely invariant to rotations of the circle’s plane. Therefore, there is a limit to the range of observation positions and angles that it can accommodate. In [6] and [7] identical CCC’s are arranged in a distinctive geometric pattern and identified based on a set of rules. The pattern consists of four markers on the edges of a rectangle and one additional marker on one of the sides to remove the roll ambiguity. The rules to identify the pattern consist on finding the three colinear CCC’s by testing each subset of three points for collinearity. Once these have been found, the remaining two point can be identified by their location in relation to this line. This process is very likely to fail if one of the marks is occluded. A similar approach with more complex heuristics and a slightly different marker distribution is followed in [8].

The pattern in WhyCon [9]–[11] is very similar to that of CCC. After being detected, marker candidates are tested for certain properties with increasing complexity: minimum size, first roundness measure, centroid position of the two sections, their area ratio and a more complex circularity measure. If a candidate passes all tests, the ellipse centre and semi-axes will be calculated from the pattern pixel information. The position of the target with respect to the camera is calculated based on

the ellipse parameters. In this type of approach the accuracy of pose estimation heavily relies on the accuracy to which the ellipse centre and semi-axes are estimated and is likely to fail when the circle has low resolution.

Pi-Tag [12], [13] is composed of 12 dots distributed along the sides of a square (4 per side, with shared corners). Marker are detected by ellipse fitting on a thresholded image, filtering and identification based on a build set of rules. Each pi-Tag pattern can be encoded by the configurations of certain dots. The higher amount of marks creates some redundancy and provides a moderate robustness to occlusion.

The main disadvantage of Pi-Tag in our targeted use case are the relatively high number of markers needed. It needs at least 7 markers to function without any occlusion resilience. Further, Pi-tag is limited to occupying a square shaped area and cannot fully utilize the available space in irregularly shaped targets. Further we find that the approach requires a certain level of parameter tuning to work in varying light conditions.

In [14], [15] a learning based approach is proposed to find four dots placed at corners of a square. The strategy consists of collecting a sample of fiducial images under varying conditions and training a classifier that takes as input the 12x12 sub window surrounding each image pixel. It classified each pixel as “marker” or “not a marker”. The training data is obtained by acquiring training videos that are annotated using a simple detection algorithm that is reset manually upon failure. The samples are rotated and lightened or darkened to artificially increase the variation in the training set.

MarkerPose [16] follows a deep learning based approach similar to that of Deep Charuco [17]. Both share the network structure of “Superpoint” [18](learning-based version of the classical SIFT) that estimates key point coordinates and their descriptors. In MarkerPose, circle’s centres and IDs are estimated with a two headed deep neural network that takes full images as an input. The IDs are used to solve the correspondence problem between two stereoscopic views. The marker 2D locations are used to extract square patches from the image containing the ellipses. Another deep neural network (EllipSegNet) is used for ellipse contour estimation and sub-pixel centroid estimation.

ITER’s RR pattern resembles a dot-like structure such as that of Pi-Tag. We cannot expect to obtain exactly the same contrast pattern as a CCC. While the contrast between the inner and outer circles is ensured and is expected to be quite high, the contrast between the outer circle and the surrounding environment is not, as can be seen in the examples in Figure 2.

The shape of the RR pattern varies with viewing poses due to the depth of the marker housing and cannot be expected to be exactly circular. Further, the marker resolution for the established camera resolution and working distances is rather low. Therefore, a reliable pose cannot be estimated based on the parameters of a fitted ellipse, such as in Whycon.

The current approach to find retro reflective (RR) markers in ITER [1] has some resemblances to the classical techniques we have mentioned thus far. Marker detection is done by adaptive



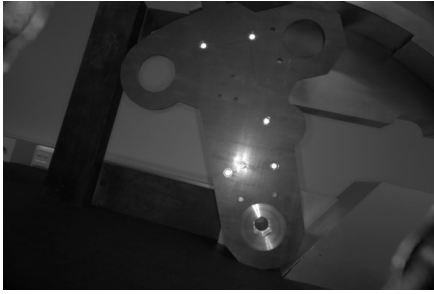


Fig. 2. Example image of a operational piece fitted with ITER's RR markers.

thresholding, filtering is based on calculation and sorting of morphological features of detected blobs and identification follows a set of rules similar to that of [6] and [7]. It was developed for a constellation of eight RR markers and the identification strategy fails if any of the markers is missing. Further, this approach requires a certain degree of manual parameter tuning of segmentation and filtering thresholds.

When applying the marker based pose estimation methods to different use cases in ITER a set of challenges comes up. In different use cases, the geometry of the tracked pieces can be rather different. Further, different pieces may have different limitations on the areas where markers cannot be placed. It is unlikely that one constellation geometry optimally fits all. With the need to adapt the marker detection and identification strategy to different use cases in ITER, comes the challenge of developing an identification strategy than can easily be adapted to different constellation geometries. The previous strategies, where a different set of heuristics needs to be created to address each use case is not a viable one.

### III. METHODOLOGY

Classical approaches for marker identification often rely on many heuristics and manual parameter tuning. They have a limited performance when it comes to occlusion resilience, especially while using a small number of markers, as seen from the examples above [1], [6]–[8], [12], [13]. Convolution neural network-based approaches, such as those of [16], [17] have been applied to a wide variety of computer vision problems in recent years. These approaches take as input full images and output marker's locations and/or IDs in a black box manner, where the algorithm is responsible for identifying the relevant features. They can often achieve very good results, even in very challenging conditions. However, these models are not easily explainable, their training requires large, realistic, labelled datasets and the determination of a potentially high number of hyper-parameters. We hypothesize that the task of identifying RR markers in a relatively controlled environment can be solved with enough accuracy using models and strategies with lower complexity. Therefore, in this work we propose a hybrid approach as illustrated in Figure 3. A list of marker candidate coordinates is generated by classical

blob detection and extraction. This stage is followed by an operation that filters out wrong matches based on the morphological characteristics of the candidates. At the last step, the correspondence is established between each marker and its match in the constellation model. We explore the possibility of using, at this last step, a shallow feed forward neural network for fault tolerant (occlusion resilient) constellation recognition.

#### A. MSER Candidate Detection

At the marker detection stage we use the maximally stable extremal regions (MSER) algorithm [19], as we have empirically found that it is able to consistently generate very suitable candidates with minimal parameter tuning. This approach works by integrating over segmentation thresholds and identifying areas that are most stable throughout the variation. MSER segmentation outputs multiple mutual overlapping regions corresponding to different segmentation thresholds. For each set of overlapping regions we consider only the candidate with smallest area and discard others.

#### B. Candidate Filtering

Once marker candidates are identified in an image we aim to discriminate which correspond to actual markers. In the previous work, a hierarchical clustering strategy was applied to solve this problem. The approach consisted of pruning the hierarchical cluster tree at the lowest point where there is a cluster with  $N$  elements, with  $N$  being the number of markers in the constellation. The assumptions behind this approach is that the cluster with the correct number of elements likely corresponds to the correct one. This approach cannot handle missing or occluded markers since, in these cases, the number of elements in the correct cluster is not known. The expected number of clusters is also not known a priori, since outlier blobs may have highly varying characteristics. This poses a challenge in identifying which cluster corresponds to the class "marker".

For the reasons presented above we opted to formulate the problem as a binary classification rather than a clustering one. This requires a set of training data to be labelled so that for each blob detected in the image there is a label of "marker" or "non marker". We have found that labelling a set of few tens of images is often enough to obtain a suitable filtering performance. Labeling is done through a semi-automatic procedure, where the candidates are presented as highlighted spots in the image and the operator selects which correspond to actual markers. The process is rather straightforward and intuitive. Re-training would only be needed in cases where the operational scenario is significantly changed, such as when working on a different use case or there are significant changes to the expected working range.

The classification problem is solved using a random forest algorithm. The following features are extracted from the candidate regions: Normalized Area, Circularity, Solidity, Extent, Mean Greyscale Intensity of Area and Centroid.

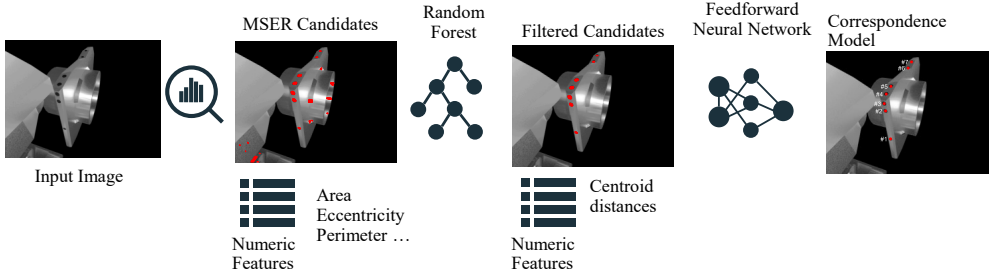


Fig. 3. Proposed marker detection and recognition pipeline.

### C. Marker Identification

We establish the correspondence between detected marker centroids and their known world coordinates using a shallow feed forward neural network. The network takes as inputs centroid distances of filtered markers. One of the advantages of this type of approach is that the expected image position of centroids can be easily computed for a given pose. By generating a set of random poses we straightforwardly generate a realistic dataset to train the network.

*Problem Formulation:* The problem is formulated as a classification task where each detected marker in one image is classified into categories: “1”, “2”, “3”, “4” or “5”. The contextual information of each marker’s centre coordinates in relation to the others is used as input information. The extracted features are the 2D distances of the marker to each of the other  $N-1$  detected markers:  $(dx_i, dy_i)_{i \in [1, N]}$ .  $N$  markers detected in one image generate  $N$  rows of data. Each row is an independent entry to the classification process. Pairs of  $(x, y)$  distances are sorted based on their ascending  $x$ -coordinates to reduce the permutations, while keeping the correspondence between the two  $(x, y)$  dimensions.

*Pre-Processing Steps:* The camera calibration parameters are known at detection time and the intrinsic matrix,  $K$ , is used to normalize the point coordinates to the interval  $[-1, 1]$ . An image point expressed in normalized coordinates is given by:  $x' = K^{-1}x$ , where  $K^{-1}$  is the inverse of the intrinsic matrix and  $x$  the original coordinates [20]. We anticipate that this normalization step will contribute to enhancing the generalization of our approach to variations in the camera parameters.

We process unknown inputs by replacing them with the mean value of their respective row. Unknown inputs happen when not all constellation markers are detected in the image. When this happens, not only will a row be eliminated but all other markers in the same image will have one pair of unknown features.

One-hot encoding is used to represent target and output data. One-hot encoding is a common way of representing categorical data to be presented to algorithms that require their inputs and outputs to be numerical and when there is no order relationship between the classes. We normalize inputs to have

zero mean and unity variance, which has a positive impact on classification output.

*Classification:* We explore, for the task at hand, the use of a shallow neural feed-forward neural network with 1 hidden layer and 30 neurons. One advantage of 1-hidden layer neural networks is that they are easier to interpret and analyse than their more complex counterparts [21].

This type of network is considered a “universal approximator”. It should be able to approximate any continuous function, given that it has enough nodes or neurons. We expect that with increasing number of neurons we will be able to decrease the error up to a point where performance will decrease due to the loss of generalization. The network is trained with scaled conjugate gradient backpropagation. We use cross entropy as a performance function.

The output of the network after softmax consists of values between 0 and 1 for each class, which all sum up to one. These can be inferred as a probability distribution. For a set of  $N=5$  input markers we obtain an output matrix similar to that of Figure 4 (left side). Each matrix column represents a separate network output for each of the markers in one image. In choosing the most likely class for each marker we want to find a method that enforces both of the following conditions:

1. Same marker cannot belong to more than one class (only one match is correct in each column).
2. There cannot be more than one marker of the same class in an image (only one match is correct in each row).

The most obvious approach of choosing the maximum value for each sample/ column enforces condition 1. However, it does not enforce condition 2, which may lead to several markers in the same image being classified as corresponding to the same model point. In order to enforce both conditions we calculate, for each permutation of markers (‘12345’, ‘23451’, ‘24351’, etc.) the sum of its probabilities. The correct combination should be that with the highest value of that sum. See Figure 4 (right side).

## IV. PERFORMANCE EVALUATION

In testing the success of our methodology we consider two very significantly different use cases in ITER’s divertor: the knuckle of the cassette locking system (using RR markers) and divertor pipe flange (using laser engravings) [22]. Realistic

	Output Matrix					Permutation of Markers	Summed Probabilities
	0.2080	0.2080	0.0000	0.0032	0.0004	"54321"	1.4495
	0.1983	0.1983	0.0001	0.9968	0.0055	"54312"	0.4609
	0.1414	0.1414	0.0000	0.0000	0.0172	"54231"	0.4528
	0.2533	0.2533	0.9996	0.0000	0.0001	"54113"	0.4578
	0.1990	0.1990	0.0002	0.0000	0.9768	"31425"	3.3226
Max Value	4	4	4	2	5	...	...
Proposed	3	1	4	2	5		

Fig. 4. Example output matrix. Chosen classes according to 'max row' strategy which leads to duplicate classifications, and our proposed approach

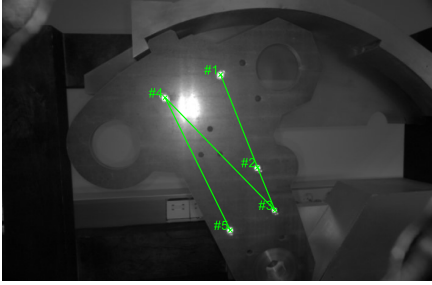


Fig. 5. Example successful detection and identification - knuckle use case

true scale prototypes are built for these two use cases. We also consider the bridging link usecase [23]. We analyse and evaluate the last case on synthetic data. The geometry of the pieces can be inferred from the example images in Figures 5, 6 and 7.

#### A. Robustness of Marker Detection Pipeline

We evaluate the performance of the marker detection pipeline in terms of robustness at each of its steps and as a whole. Results are shown in Table I. Table II quantifies the performance of the identification stage when specific markers are occluded/ not detected.

As demonstrated by the results the proposed methodology succeeds at detecting and identifying RR markers in usecase I at a very high rate. As would be expected, detection of laser engravings is more challenging than detection of RR markers.

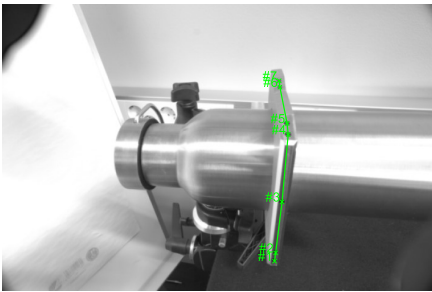


Fig. 6. Example successful detection and identification - pipe flange use case

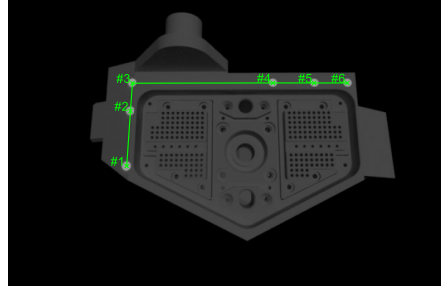


Fig. 7. Example successful detection and identification - bridging link use case

There is a much higher rate of false candidates in the second usecase. Detection rates are considerably higher in use case III due to the less challenging characteristics of the images. Results in Table II demonstrate that the algorithm provides moderate robustness to occlusion of markers.

TABLE I  
ROBUSTNESS MEASURE OF MARKER DETECTION AND IDENTIFICATION AT CONSECUTIVE PIPELINE STAGES - KNUCKLE (I), PIPE FLANGE (II) AND BRIDGING LINK (III) USE CASES

		I	II	III
<b>Candidate Generation<sup>1</sup></b>	Detection Rate	99%	95%	100%
	$N_{detected}/N_{markers}$			
<b>Filtering<sup>2</sup></b>	False Candidate Rate	13%	74%	5%
	$N_{false\ candidates}/N_{candidates}$			
	True Positive Rate	100%	98%	100%
	$N_{true\ positives}/N_{positives}$			
<b>Identification<sup>3</sup></b>	False Positive Rate	2%	3%	0%
	$N_{false\ positives}/N_{positives}$			
	True Identification Rate	100%	100%	100%
<b>Full Pipeline<sup>1</sup></b>	$N_{true\ identification}/N_{markers}$			
	Detection Rate	95%	65%	100%
	$N_{successful\ images}/N_{images}$			

<sup>1</sup>Over whole image set. <sup>2</sup>Over the test set

<sup>3</sup>Over the set of images where all markers were detected

TABLE II  
ROBUSTNESS MEASURE OF CORRESPONDENCE STEP WITH RESPECT TO OCCLUSION OF MARKERS - KNUCKLE (I) AND PIPE FLANGE (II) AND BRIDGING LINK (III) USE CASES

True Correspondence Rate	I	II	III
#1 Missing	100%	73%	74.4
#2 Missing	92%	72%	76.8
#3 Missing	75%	80%	88.8
#4 Missing	95%	74%	65.2
#5 Missing	100%	93%	48.4
#6 Missing	N/A	60%	N/A
#8 Missing	N/A	44%	N/A

#### B. Accuracy of Pose Estimation

For the sake of completeness we present also the results of positional accuracy. The methods to estimate the pose and

the laboratory setup to collect the images and methods to determine the ground truth position are as described in [1]. We do not expect relevant changes in performance in terms of accuracy, since our work has focuses on improving the detectability of the pattern. Table III shows the translational error and Table IV the rotational error.

TABLE III  
AVERAGE AND LOWEST ACCURACY AND PRECISION OF ESTIMATED  
TRANSLATIONS - KNUCKLE USE CASE / PIPE FLANGE USE CASE

	Accuracy		Precision	
	Average	Maximum	Average	Maximum
$e_t(mm)$	1.26 / 0.82	2.50 / 1.44	0.71 / 0.26	1.55 / 0.64
$e_x(mm)$	0.81 / 0.56	2.30 / 1.12	0.49 / 0.17	1.53 / 0.55
$e_y(mm)$	0.89 / 0.57	1.62 / 0.95	0.43 / 0.16	1.49 / 0.46
$e_z(mm)$	2.25 / 0.22	6.48 / 0.53	2.09 / 0.22	5.48 / 0.55

TABLE IV  
AVERAGE AND LOWEST ACCURACY AND PRECISION OF ESTIMATED  
ROTATIONS - KNUCKLE USE CASE / PIPE FLANGE USE CASE

	Accuracy		Precision	
	Average	Maximum	Average	Maximum
$e_R(^{\circ})$	1.84 / 1.50	4.04 / 1.67	0.53 / 0.05	2.20 / 0.17
$e_a(^{\circ})$	1.51 / 0.29	3.83 / 0.54	0.60 / 0.10	2.32 / 0.24
$e_b(^{\circ})$	0.91 / 0.12	1.32 / 0.37	0.18 / 0.06	0.50 / 0.25
$e_c(^{\circ})$	0.25 / 1.46	0.98 / 1.63	0.17 / 0.05	0.73 / 0.16

## V. CONCLUSIONS

We have developed and tested a novel framework for detection and identification of fiducial markers. The robustness of the proposed approach was evaluated at each of the pipeline stages and for the overall process. We showed that the newly developed algorithms offer moderate resilience to occlusion while using a rather small set of markers. The system is very often able to recover the pose when one of the markers is missing. This is a significant improvement over the previous strategy which did not have any mechanism to cope with missing marks. Further, we have demonstrated that the proposed approach can be straightforwardly applied to three use cases in ITER. We have evaluated the end-to-end accuracy and precision of pose estimation and demonstrated that the overall system works accurately and reliably, fulfilling the requirements of the application.

## REFERENCES

- [1] L. G. Ribeiro, O. J. Suominen, A. Durmush, S. Peltonen, E. Ruiz Morales, and A. Gotchev, "Retro-reflective-marker-aided target pose estimation in a safety-critical environment," *Applied Sciences*, vol. 11, no. 1, p. 3, 2020.
- [2] L. G. Ribeiro, O. J. Suominen, P. Bates, S. Peltonen, E. R. Morales, and A. Gotchev, "Design and manufacturing of an optimized retro reflective marker for photogrammetric pose estimation in ITER," *Submitted to Fusion Engineering and Design*, arXiv preprint 2205.05486, 2022.
- [3] Kalaitzakis, Michail and Cain, Brennan and Carroll, Sabrina and Ambrosi, Anand and Whitehead, Camden and Vitzilaos, Nikolaos, "Fiducial markers for pose estimation," *Journal of Intelligent & Robotic Systems*, vol. 101, no. 4, pp. 1–26, 2021.
- [4] A. C. Rice, R. K. Harle, and A. R. Beresford, "Analysing fundamental properties of marker-based vision system designs," *Pervasive and Mobile Computing*, vol. 2, no. 4, pp. 453–471, 2006.
- [5] L. B. Gatrell, W. A. Hoff, and C. W. Sklair, "Robust image features: Concentric contrasting circles and their image extraction," in *Cooperative Intelligent Robotics in Space II*, vol. 1612. International Society for Optics and Photonics, 1992, pp. 235–244.
- [6] W. A. Hoff, K. Nguyen, and T. Lyon, "Computer-vision-based registration techniques for augmented reality," in *Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling*, vol. 2904. SPIE, 1996, pp. 538–548.
- [7] W. A. Hoff, L. B. Gatrell, and J. R. Spofford, "Machine-vision-based teleoperation aid," *Telematics and informatics*, vol. 8, no. 4, pp. 403–423, 1991.
- [8] J. Yu, W. Jiang, Z. Luo, and L. Yang, "Application of a vision-based single target on robot positioning system," *Sensors*, vol. 21, no. 5, p. 1829, 2021.
- [9] T. Krajník, M. Nitsche, J. Faigl, P. Vaněk, M. Saska, L. Přeučil, T. Duckett, and M. Mejail, "A practical multirobot localization system," *Journal of Intelligent & Robotic Systems*, vol. 76, no. 3, pp. 539–562, 2014.
- [10] T. Krajník, M. Nitsche, J. Faigl, T. Duckett, M. Mejail, and L. Přeučil, "External localization system for mobile robotics," in *2013 16th International Conference on Advanced Robotics (ICAR)*. IEEE, 2013, pp. 1–6.
- [11] M. Nitsche, T. Krajník, P. Cizek, M. Mejail, T. Duckett *et al.*, "Whycon: an efficient, marker-based localization system," 2015.
- [12] F. Bergamasco, A. Albarelli, and A. Torsello, "Image-space marker detection and recognition using projective invariants," in *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*. IEEE, 2011, pp. 381–388.
- [13] Bergamasco, Filippo and Albarelli, Andrea and Torsello, Andrea, "Pitag: a fast image-space marker design based on projective invariants," *Machine vision and applications*, vol. 24, no. 6, pp. 1295–1310, 2013.
- [14] D. Claus and A. W. Fitzgibbon, "Reliable automatic calibration of a marker-based position tracking system," in *2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1*, vol. 1. IEEE, 2005, pp. 300–305.
- [15] Claus, David and Fitzgibbon, Andrew W, "Reliable fiducial detection in natural scenes," in *European Conference on Computer Vision*. Springer, 2004, pp. 469–480.
- [16] J. Meza, L. A. Romero, and A. G. Marrugo, "Markerpose: robust real-time planar target tracking for accurate stereo pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 1282–1290.
- [17] D. Hu, D. DeTone, and T. Malisiewicz, "Deep charuco: Dark charuco marker pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8436–8444.
- [18] D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superpoint: Self-supervised interest point detection and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018, pp. 224–236.
- [19] M. Donoser and H. Bischof, "Efficient maximally stable extremal region (msr) tracking," in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, vol. 1. Ieee, 2006, pp. 553–560.
- [20] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [21] E. Belilovsky, M. Eickenberg, and E. Oyallon, "Shallow learning for deep networks," 2018.
- [22] R. E. Shuff, "Development of remote handling pipe jointing tools for iter," Ph.D. dissertation, Tampere University of Technology, 2012.
- [23] J. Lyytinen, P. Tikka, T. Määttä, T. Avikainen, and S. Rantala, "Development of the remote handling connector for iter divertor diagnostic system," *Fusion Engineering and Design*, vol. 165, p. 112243, 2021.



