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

Laura Gonçalves Ribeiro*, Olli J. Suominen*, Sari Peltonen*, Emilio Ruiz Morales†, and Atanas Gotchev*
E-mails:{laura.goncalvesribero, olli.j.suominen, sari.peltonen, atanas.gotchev}@tuni.fi, {Emilio.Ruiz}@f4e.europa.eu

* Faculty of Information Technology and Communication, Tampere University, 33720 Tampere, Finland
† Fusion for Energy (F4E), ITER Delivery Department, Remote Handling Project Team, 08019 Barcelona, Spain

***Abstract*—Fiducial marker-based tracking is an effective method for pose estimation in coffined 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, optical markers, fiducial markers, retro refletive markers, marker detection, marker identification, remote handling, ITER**

I. Introduction

Fiducial markers are artificial elements that can be added to a scene to help establish robust correspondences between captured image points and 3D models of targets. Marker based pose estimation is a particularly attractive strategy in applications where the targets are human-made structures in confined environments and the accuracy requirement is high. The approach takes full advantage of the constraints in such scenarios 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 an application. ITER constitutes one of the world's most hostile environments, where operators need to perform tight tolerance alignment tasks remotely and in poor visibility conditions. High temperature and radiation levels restrict the equipment that can be used to sense the environment and the quality of the acquired images. In this context, there is a need to develop reliable and accurate methods for pose estimation of specific targets. The estimates can aid the operators in conducting remote handling tasks.

Earlier work has addressed marker based pose estimation in ITER [1], with the development of dedicated markers and of computational methods to detect and identify them and estimate the pose. That work considers the localization of a specific target: 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 millimeter level accuracy. However, the strategies for marker detection and identification struggle with robustness, particularly in cases where markers are occluded or missing. Furthermore, there is a need to apply marker based pose estimation to different targets and operational cases in ITER. The 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 for ITER's retro refletive (RR) markers [2]. It should be able to reliably identify circular, uncoded marks with a moderate level of resilience to occlusion and minimal manual parameter adjustment. It should rely only on a relatively small number of markers and 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, diffuse 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.

This paper is organized as follows: In Section I we have introduced the research question and motivated our work. In Section II we overview and analyse existing solutions. In Section III we detail our proposed approach. In Section IV we evaluate the performance of our proposed solutions and in Section V we draw conclusions and outline future development prospects.

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.

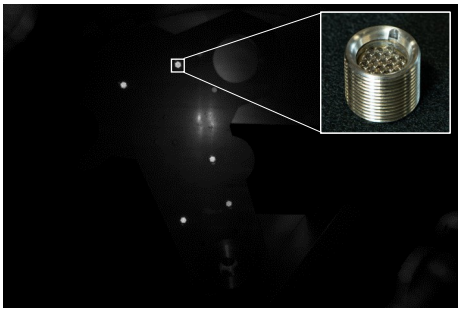

Fig. 1. Example image of an operational piece (knuckle of ITER divertor's cassette locking system) fitted with RR markers [2]. The image emulates the expected greyscale, low resolution output of a radiation tolerant camera and is purposely underexposed to facilitate the segmentation of the RR elements.

II. Background

Visual marker systems have been extensively studied in the literature [3], [4]. However, the large majority of existing approaches is not directly applicable to the ITER use case. Marks occupying a large area or with very fine details are not feasible to be applied in this context. ITER's RR pattern resembles a dot-like structure, as seen in Figure 1. Therefore, we focus our overview on point-based or concentric circle-based systems with a relatively small number of markers. Table I shows the visual appearance of selected approaches. Concentric contracting circles (CCC) [5], small white circles whose centre is coincidental with a larger black circle, are popular alternatives to circular markers. The advantage in using CCCs resides in the unlikelihood that random contrasting intensity regions have nearly coincident centroids.

TABLE I
Appearance of relevant fiducial marker systems

WhyCon [6]–[8]	
Pattern of 5 CCC [9], [10]	
Circular ring marker [11]	
Pi-Tag [12], [13]	
4 Dots [14], [15]	
MarkePose [16]	

Most systems identify markers by arranging them in a distinctive geometric pattern and identifying them based on a set of rules [9]–[13]. Heuristics for marker identification are often based on projective invariants, such as collinearity of points. This type of approach is very likely to fail if one of the marks is occluded or missing. In Pi-Tag [12], [13] the use of a larger number of marks (12 in total) creates some redundancy and provides a moderate robustness to occlusion. The main challenge in applying of Pi-Tag to our targeted use case is 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 WhyCon [6]–[8] the pose is calculated based on a single mark, based on the estimated parameters of a fitted ellipse. 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.

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 exhaustively classifies image patches. The data is annotated using a simple detection algorithm that is reset manually upon failure. 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.

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 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 [9] and [10]. 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 tasks 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.

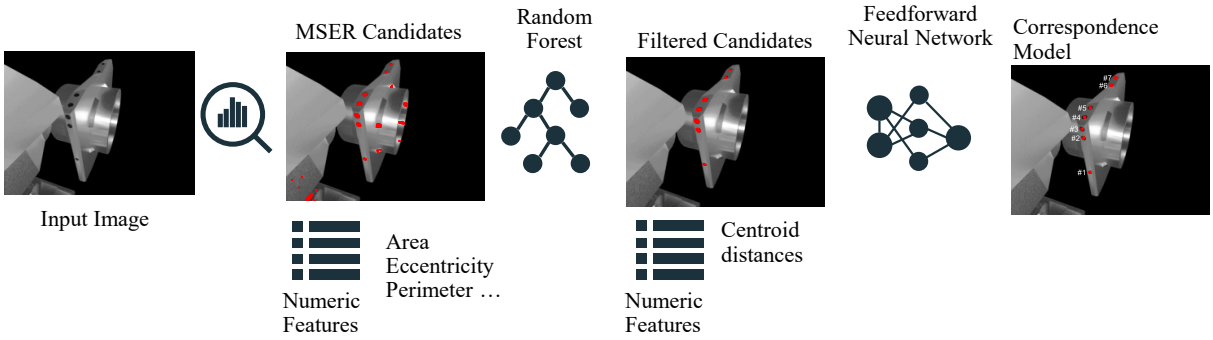

Fig. 2. Proposed marker detection and recognition pipeline.

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 in the literature [1], [9]–[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 2. 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 user 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 [20] as we found it provides fairly good results with a relatively small amount of training data. The following features are extracted from the candidate regions: Normalized Area, Circularity, Solidity, Extent, Mean Greyscale Intensity of Area and Centroid.

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 [21]. The normalization step is a common strategy to make optimization better behaved [22]. Further it removes the dependence of features on 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 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 [23]. This type of network is considered a "universal approximator". It should be able to approximate any continuous function, given that it has enough neurons in the hidden layer. 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. We found that using more than 30 neurons does not significantly improve performance. The network is trained with scaled conjugate gradient back-propagation. 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

	Output Matrix				
	0.2080	0.2080	0.0000	0.0032	0.0004
	0.1983	0.1983	0.0001	0.9968	0.0055
	0.1414	0.1414	0.0000	0.0000	0.0172
	0.2533	0.2533	0.9996	0.0000	0.0001
	0.1990	0.1990	0.0002	0.0000	0.9768
Max Value	4	4	4	2	5
Proposed	3	1	4	2	5

Permutation of Markers	Summed Probabilities
"54321"	1.4495
"54312"	0.4609
"54231"	0.4528
"54113"	0.4578
"31425"	3.3226
…	…

Fig. 3. Example output matrix. Chosen classes according to 'max row' strategy which leads to duplicate classifications, and our proposed approach

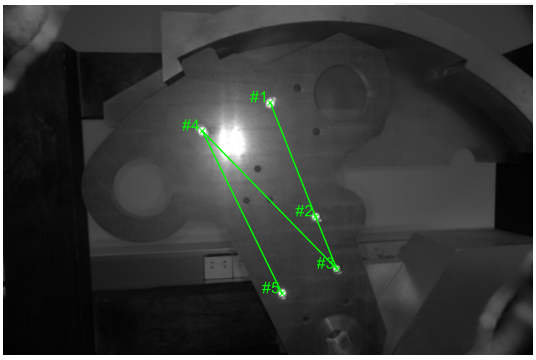

Fig. 4. Example successful detection and identification - knuckle use case

that of Figure 3 (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 3 (right side).

IV. Performance Evaluation

In testing the success of out methodology we consider two very significantly different use cases in ITER's divertor: the knuckle of the cassette locking system [24] (using RR markers) and divertor pipe flange [25] (using laser engravings). Realistic true scale prototypes are built for these two use cases. We also consider the localization of the bridging link in-vessel connector [26]. We analyse and evaluate the last case using synthetically generated data based on 3D models of the environment. The geometry of the pieces can be inferred from the example images in Figures 4, 5 and 6.

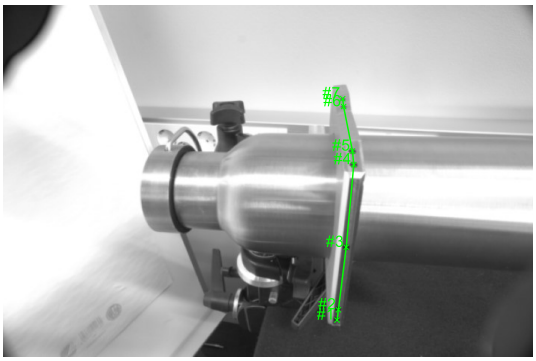

Fig. 5. Example successful detection and identification - pipe flange use case

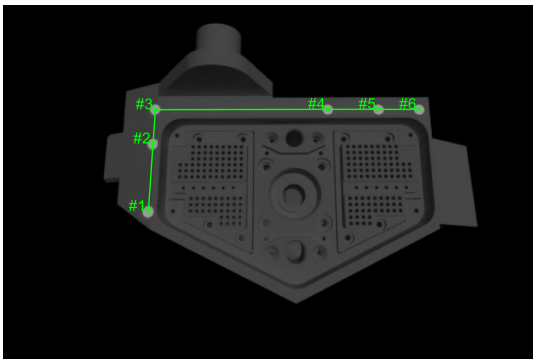

Fig. 6. Example successful detection and identification - bridging link use case

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

[1]Over whole image set. [2]Over the test set
[3]Over the set of images where all markers were detected

TABLE III
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%	28.4%
#8 Missing	N/A	44%	N/A

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 II. 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. There is a much higher rate of false candidates in the second usecase. This is partially due to the small size of the marks and the similarity of their visual appearance to other dark, diffuse elements in the environment. Detection rates are considerably higher in use case III due to the less challenging characteristics of the synthetic images.

Table III quantifies the performance of the identification stage when specific markers are occluded/ not detected. Results demonstrate that the algorithm is moderately robust to occlusion and is often able to recover from cases where one of the marks is missing. We note that identification is more likely to fail if specific marks are missing (for example marker number 7 in use case II and markers number 5 or 6 in use case III). In such cases the missing marks lead different constellation orientations to appear visually similar. This effect can be minimized to a certain extent if occlusion is considered when building the constellations.

B. Pose Estimation Performance

For the sake of completeness we present estimates of the performance of pose estimation based on detected marks. 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 significant variations in the accuracy of pose estimation, as our work has focused on improving the robustness of marker detection strategies. Table IV shows the translational error and Table V the rotational error.

TABLE IV
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

V. CONCLUSIONS

We have developed and tested a novel framework for uncoded marker detection and identification. Detection rates are high for the use case using RR markers. In detecting laser engravings the approach struggles to differentiate marks

TABLE V
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

from other small dark background elements. Future work might explore the use of more advanced candidate filtering methods or the inclusion of more discriminative features. The use concentric circles rather than dark circles or other similar strategies might help in differentiating the diffuse marks from background elements. However, practical considerations such as available space and resolution need be taken into account in using more detailed marks.

The developed marker correspondence methods 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. However, when specific marks are missing, constellations might become indifferentiable. Marker constellations should be built to avoid such redundancies whenever possible. More advanced methods or the use of unique coded marks would be required for full occlusion resilience.

We demonstrated that the proposed approach can be straightforwardly applied to three use cases in ITER, with considerably different geometries and limitations on the areas where marks can be placed. 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. Further work might test the applicability and performance of the proposed framework in other example use cases. It would be interesting to compare the performance of the proposed approach against more complex machine learning architectures and perform a more thorough analysis of the complexity performance tradeoff.

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] 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.

[4] Kalaitzakis, Michail and Cain, Brennan and Carroll, Sabrina and Ambrosi, Anand and Whitehead, Camden and Vitzilaios, Nikolaos, "Fiducial markers for pose estimation," *Journal of Intelligent & Robotic Systems*, vol. 101, no. 4, pp. 1–26, 2021.

[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] 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.

[7] 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.

[8] M. Nitsche, T. Krajnik, P. Cizek, M. Mejail, T. Duckett *et al.*, "Whycon: an efficient, marker-based localization system," 2015.

[9] 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.

[10] 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.

[11] 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.

[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, "Pi-tag: 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 (mser) tracking," in *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, vol. 1. Ieee, 2006, pp. 553–560.

[20] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[21] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[22] 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.

[23] E. Belilovsky, M. Eickenberg, and E. Oyallon, "Shallow learning for deep networks," 2018.

[24] V. Lyytikäinen, P. Kinnunen, J. Koivumäki, J. Mattila, M. Siuko, S. Esque, and J. Palmer, "Divertor cassette locking system remote handling trials with WHMAN at DTP2," *Fusion Engineering and Design*, vol. 88, no. 9-10, pp. 2181–2185, 2013.

[25] R. E. Shuff, "Development of remote handling pipe jointing tools for iter," Ph.D. dissertation, Tampere University of Technology, 2012.

[26] 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.