2021 The 21st International Conference on Control, Automation and Systems (ICCAS 2021)
Ramada Plaza Hotel, Jeju, Korea, Oct. 12∼15, 2021

ArUcoE: Enhanced ArUco Marker

Oguz Kedilioglu[1*], Tomás Marcelo Bocco[2], Martin Landesberger[3], Alessandro Rizzo[2] and Jörg Franke[1]

[1]Institute for Factory Automation and Production Systems, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, 91058, Germany, ({oguz.kedilioglu, joerg.franke}@faps.fau.de) * Corresponding author
[2]Department of Electronics and Telecommunications, Politecnico di Torino, Torino, 10129, Italy ({tomas.bocco, alessandro.rizzo}@polito.it)
[3]Heinz Maier-Leibnitz Zentrum (MLZ), Technical University of Munich, Garching, 85748, Germany (martin.landesberger@frm2.tum.de)

Abstract: This paper presents a novel fiducial marker type called ArUcoE. It is obtained from a standard ArUco marker by enhancing it with a chessboard-like pattern. With our approach the pose estimation accuracy of any ArUco marker can easily be increased. Further methods to increase the accuracy are analyzed. By applying a subpixel algorithm to the corner regions we are able to locate the corner points within a pixel and overcome the restriction of pixel-level accuracy. A deep-learning-based super-resolution method is used to artificially increase the pixel density in the same regions. Additionally, the effect of using a single and a stereo camera setup on the accuracy is shown.

Keywords: ArUco, fiducial marker, 6DoF pose estimation, subpixel, super-resolution, deep learning, visual servoing, simulation

1. INTRODUCTION

Machines are getting immersed deeper and deeper into the world. That's why spatial understanding of their environment becomes ever more important. To enable a robot to manipulate objects, it has to be able to separate that object from it's background and know it's pose, which is constituted of position and orientation. Fiducial markers are still relevant for the task of pose estimation, even though more and more deep-learning-based end-to-end models [1–3] are published, which estimate the pose of an object directly from raw images without the need for any purposefully placed features. The predefined patterns with easy to detect features of fiducial markers help to increase the accuracy and robustness of pose estimation. Unlike many deep-learning-based approaches, fiducial markers don't need to be trained with application-specific data and are therefore more flexible and independent of the individual use case.

The case we applied our approach to is the neutron diffractometer STRESS-SPEC at FRMII [4] at Garching, Germany. The neutron beam goes through a sample and is diffracted. The pattern of diffraction reveals material properties of the sample. In order to get precise measurements, the sample has to be positioned very accurately at the neutron beam focus point. The neutrons are generated in a fission process which runs continuously for 60 days. The goal is to maximize the measurement time by minimizing the sample setup time as much as possible with the help of a higher degree of automation. Using fiducial markers instead of a laser tracker to track the robot flange increases the instrument flexibility and enables the measurement of more complex samples, e.g. additively manufactured parts. High-presicion visual servoing is also relevant not only for neutron diffractormeters, but also for other fields of application. Especially the demand for medical robots, which also need to be controlled very accurately and robustly, will increase in the future, because of the aging population and rising wages.

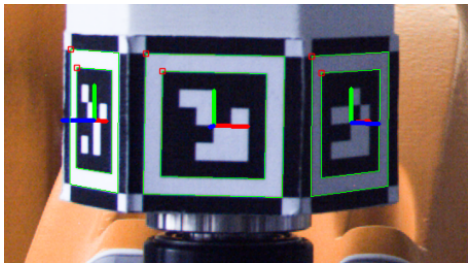

Fig. 1. 6DoF pose estimation of a robot endeffector with ArUcoE

Several fiducial marker types have already been proposed in the past. [5–7] give a good overview. In 2014 the fiducial marker called ArUco was proposed by Garrido-Jurado et al. [8]. It offers a way to automatically create a whole set of markers such that the inter-marker distance is maximized in order to increase the detection robustness. The corners of the squared ArUco marker are used for the pose estimation and the binary pattern within the square is used to identify the marker. Wang et al. [9] proposed an improved version of the ArUco marker. They employ circular patterns at the corners of the original ArUco marker to improve the accuracy of the pose estimation. Our approach, as illustrated in figure 1, applies rectangular enhancement patterns around the original ArUco marker, which is located within the inner green square. In contrast to Wang et al. [9], we additionally utilize a subpixel algorithm [10] to localize the chessboard-like corner patterns of our enhancement. This allows us to estimate the pose of the marker more accurately within the boundaries of a pixel. Ziegler [11] proposed also a very interesting approach which is based on a Data Matrix code within a circle, whereby the Data Matrix code is

978-89-93215-21-2/21/$31.00

used for identification and the circle for localization. The accuracy he achieved doesn't reach our results, but the general idea is worth considering. It could be extended to ArUco markers by placing the binary ArUco pattern within a circle instead of a square.

Our contribution consists of

- a novel fiducial marker type called ArUcoE,
- analysis of the effects of subpixel and deep learning based super-resolution methods and
- evaluation of the pose estimation accuracy with a mono and a stereo camera setup.

2. METHODOLOGY

Our methodology to increase the pose estimation accuracy of ArUco markers consists of enhancing the original ArUco marker with additional rectangular patterns (figure 2), and using subpixel and super-resolution approaches. The black rectangular enhancement patterns are located around the original marker, but they don't cover the whole outer square, thus leaving the corners white. This creates a chessboard-like pattern, which is easy to detect. First, the inner ArUco marker is detected using standard techniques. After the marker's identification, the corners of the outer frame can be easily found and used for the final pose estimation. The region of interest to estimate the pose of the new marker ArUcoE is shown with a yellow square in figure 2.

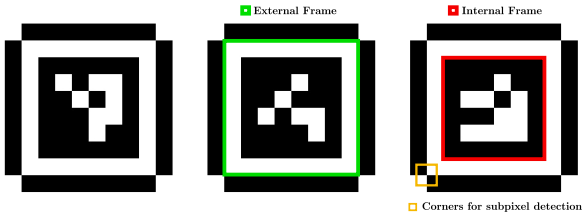

Fig. 2. Enhanced ArUco marker (ArUcoE). The red square contains the original ArUco marker. Our enhancement pattern is outside of the green square.

We detect these corners with the subpixel function from OpenCV [12] called cornerSubPix. That allows us to locate the corner points within a pixel, allowing us to go beyond the size limit of a pixel, as shown in figure 3. Thereby, the gradient of the neighboring pixel values is used to estimate the location of the corner within a pixel.

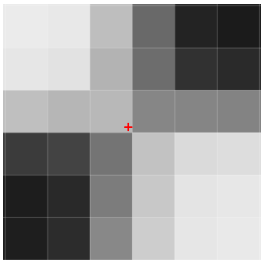

Fig. 3. Subpixel corner detection.

Another method to increase the localization accuracy of the ArUcoE corners is super-resolution, which artificially increases the image resolution. The deep learning models for super-resolution are called Super-Resolution Neural Networks (SRNN). Our SRNN, shown in figure 4, was inspired by [13–15] and consists of a convolutional neural network with 11 layers. The first layer has six kernels with size 6x6 and the other ones have each six kernels with size 3x3. The model was implemented with TensorFlow [16] and was trained with a set of 44832 corner images, which was devided into 404 batches. Mean square error (MSE) was used as the loss function. We apply super-resolution only to the regions of interest consisting of 64x64 pixels each, to increase the effiency of the method. The 64x64 image patches serve as ground truth data, i.e. as the desired output, during training. They are blurred with a kernel from [17], manipulated with random noise and then downsampled to 16x16 pixels to create input data for the training.

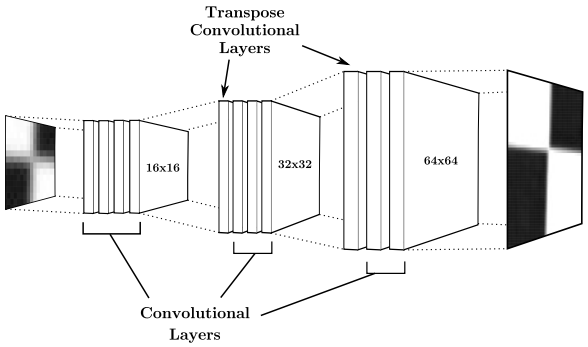

Fig. 4. SRNN architecture consisting of 11 layers to artificially increase the image resolution.

3. EVALUATION

We evaluated our methods in a simulation environment created with Blender [18] and with real hardware. The effect of our ArUco marker enhancement, the subpixel and super-resolution approaches and of using a stereo setup instead of a mono setup on the pose estimation accuracy, are shown.

The simulated cameras, shown in figure 5, have a focal length of 50 mm, a resolution of 1080x1080 pixels and a sensor size of 14.2x14.2 mm. The cameras are located 1000 mm away from an array of markers. In the stereo setup both cameras are 300 mm apart from each other and are tilted at an angle of 7.5° around the vertical axis. Both of them are virtually calibrated with a chessboard pattern located at the same distance as the markers.

The real evaluation setup can be seen in figure 6. It consists of two cameras of the type jAi GO-5000C-USB [19] with 5 megapixels and 16 mm focal length attached to a frame 300 mm apart from each other and directed towards the markers located at the endeffector of an industrial robot arm. The pose estimation accuracy is evaluated with the 6DoF laser tracker Leica AT901 [20], which tracks the three reflectors attached above the markers. The measured marker poses are contained within a cube of 250 mm, whose center is 775 mm away from the baseline of the cameras. To be able to evaluate the pose es-

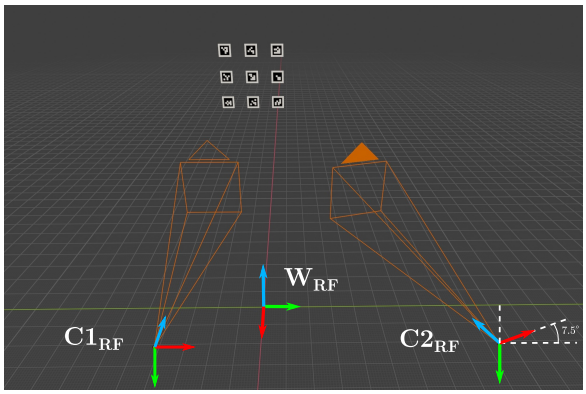

Fig. 5. Virtual setup in Blender. W_{RF} is the world reference frame, $C1_{RF}$ the reference frame of the first camera and $C2_{RF}$ the reference frame of the second camera.

timation accuracy of the camera system with the laser tracker, they have to be set in relation to each other with the help of hand-eye calibration.

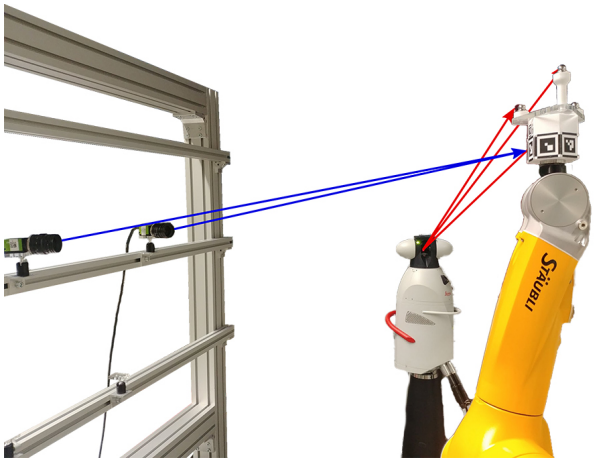

Fig. 6. Real evaluation setup. The blue arrows indicate the measurement of the ArUcoE markers by the cameras, while the red arrows show the measurement of the reflectors by the laser tracker.

Table 1 shows the standard deviation of the pose estimation translation error in the simulated Blender environment. The results of "ArUcoE mono subpixel" and "ArUco mono subpixel" show that our enhancement pattern improves the accuracy by orders of magnitude. Using a stereo setup instead of a mono setup halves the errors. The values of "ArUco mono subpixel" and "ArUco mono" show that the subpixel approach yields a comparatively small improvement.

Table 1. Pose estimation translation error in simulated environment.

Method	σ_x (μm)	σ_y (μm)	σ_z (μm)
ArUco mono	1435.3	1501.8	2752.1
ArUco mono subpixel	1353.0	1356.3	568.5
ArUcoE mono subpixel	18.7	26.5	333.4
ArUcoE stereo subpixel	8.7	4.4	72.0

The standard deviation of the pose error in the real environment is listed in the tables 2 and 3. Here you can compare the effects of the subpixel and super-resolution methods.

Table 2. Pose estimation translation error in real environment.

Method	σ_x (μm)	σ_y (μm)	σ_z (μm)
ArUcoE mono subpixel	713.4	423.5	351.2
ArUcoE stereo super-res.	143.7	141.5	206.4
ArUcoE stereo subpixel	132.1	141.4	197.7

It becomes clear that there is not much difference in accuracy between the methods subpixel and super-resolution. The subpixel method is slightly better, but it's main advantage is its low computational cost compared to the 11-layer deep learning model used for the super-resolution method.

Table 3. Pose estimation rotation error in real environment.

Method	σ_{Rx} (deg)	σ_{Ry} (deg)	σ_{Rz} (deg)
ArUcoE mono subpixel	0.85	0.74	0.42
ArUcoE stereo super-res.	0.16	0.14	0.18
ArUcoE stereo subpixel	0.13	0.14	0.14

4. CONCLUSION

We proposed an easy way to enhance the ArUco marker and showed that it improves the pose estimation accuracy considerably. Furthermore, we investigated subpixel and super-resolution methods and the positive effect of using a stereo setup. Even though a single camera is enough to get the 6DoF pose of the ArUco and ArUcoE markers, it is worthwhile to use more cameras. The slightly better accuracy and much lower computational cost of the subpixel algorithm suggests its superiority over the super-resolution method. But we still think that further investigations into super-resolution methods are reasonable, because they could allow smaller markers to be detected robustly.

ACKNOWLEDGEMENT

The project receives funding from the German Federal Ministry of Education and Research under grant agreement 05K19WEA (project RAPtOr).

REFERENCES

[1] Chen Wang, Danfei Xu, Yuke Zhu, Roberto Martín-Martín, Cewu Lu, Li Fei-Fei, and Silvio Savarese. Densefusion: 6d object pose estimation by iterative dense fusion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3343–3352, 2019.

[2] Chen Song, Jiaru Song, and Qixing Huang. Hybridpose: 6d object pose estimation under hybrid representations. *arXiv preprint arXiv:2001.01869*, 2020.

[3] Wei Chen, Jinming Duan, Hector Basevi, Hyung Jin Chang, and Ales Leonardis. Pointposenet: Point pose network for robust 6d object pose estimation, 2020.

[4] Christian Randau, Heinz-Gnter Brokmeier, Weimin Gan, Michael Hofmann, M. Voeller, William Tekouo, N. Al-hamdany, Gnther A. Seidl, and Andreas Schreyer. Improved sample manipulation at the stress-spec neutron diffractometer using an industrial 6-axis robot for texture and strain analyses. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 794:67–75, 2015.

[5] Oualid Araar, Imad Eddine Mokhtari, and Mohamed Bengherabi. Pdcat: a framework for fast, robust, and occlusion resilient fiducial marker tracking. 2020.

[6] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco J. Madrid-Cuevas, and Rafael Medina-Carnicer. Generation of fiducial marker dictionaries using mixed integer linear programming. *Pattern Recognition*, 51:481–491, March 2016.

[7] Xiang Zhang, S. Fronz, and N. Navab. Visual marker detection and decoding in ar systems: a comparative study. pages 97–106, Darmstadt, Germany, 2002. IEEE.

[8] Sergio Garrido-Jurado, Rafael Muoz-Salinas, Francisco J. Madrid-Cuevas, and Manuel J. Marn-Jimnez. Automatic generation and detection of highly reliable fiducial markers under occlusion. 47:2280–2292, 2014.

[9] Yutao Wang, Zongpeng Zheng, Zhiqi Su, Gang Yang, Zhong Wang, and Yu Luo. An improved aruco marker for monocular vision ranging, 2020.

[10] Wolfgang Förstner and Eberhard Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS intercommission conference on fast processing of photogrammetric data*, pages 281–305. Interlaken, 1987.

[11] Christian Ziegler. *Ganzheitliche Automatisierung mechatronischer Systeme in der Medizin am Beispiel Strahlentherapie*. Bericht aus dem Lehrstuhl fr Fertigungsautomatisierung und Produktionssysteme. Meisenbach, Bamberg, 2013.

[12] OpenCV. Open source computer vision library, 2015.

[13] Minghua Wang and Qiang Wang. Hypergraph-regularized sparse representation for single color image super resolution. *Journal of Visual Communication and Image Representation*, 74:102951, 2021.

[14] Yasser K. Badran, Gouda I. Salama, Tarek A. Mahmoud, Aiman Mousa, and Adel E. Moussa. Single image super resolution based on learning features to constrain back projection, 2019.

[15] Yunfeng Zhang, Qinglan Fan, Fangxun Bao, Yifang Liu, and Caiming Zhang. Single-image super-resolution based on rational fractal interpolation. *IEEE Transactions on Image Processing*, 27(8):3782–3797, 2018.

[16] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[17] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, and Jing-Hao Xue. Deep learning for single image super-resolution: A brief review. August 2018.

[18] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Blender Institute, Amsterdam, 2020.

[19] JAI. Go series go-5000c-usb machine vision area scan camera, 2021.

[20] HEXAGON. Leica absolute tracker at901, 2021.