Real Time Fiducial Marker Localisation System with Full 6 DOF Pose Estimation

Jiří Ulrich
Faculty of Electrical Engineering, Czech Technical University in Prague
Czech Republic
jiri.ulrich@fel.cvut.cz

Jan Blaha
Faculty of Electrical Engineering, Czech Technical University in Prague
Czech Republic
jan.blaha@fel.cvut.cz

Ahmad Alsayed
Department of Mechanical, Aerospace and Civil Eng.
University of Manchester
United Kingdom
ahmad.alsayed@manchester.ac.uk

Tomáš Rouček
Faculty of Electrical Engineering, Czech Technical University in Prague
Czech Republic
tomas.roucek@fel.cvut.cz

Farshad Arvin
Swarm & Computational Intelligence Laboratory
University of Durham
United Kingdom
farshad.arvin@durham.ac.uk

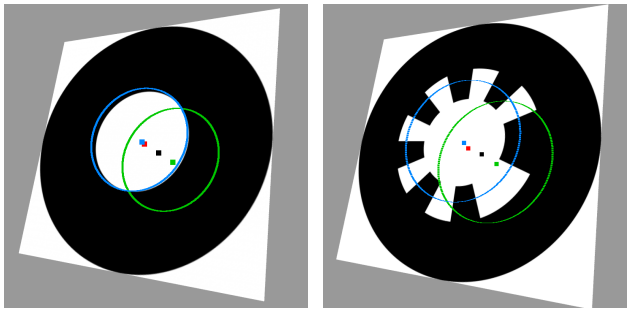
Tomáš Krajník
Faculty of Electrical Engineering, Czech Technical University in Prague
Czech Republic
tomas.krajnik@fel.cvut.cz

ABSTRACT

The ability to reliably determine its own position, as well as the position of surrounding objects, is crucial for any autonomous robot. While this can be achieved with a certain degree of reliability, augmenting the environment with artificial markers that make these tasks easier is often practical. This applies especially to the evaluation of robotic experiments, which often require exact ground truth data containing the positions of the robots. This paper proposes a new method for estimating the position and orientation of circular fiducial markers in 3D space. Simulated and real experiments show that our method achieved three times lower localisation error than the method it derived from. The experiments also indicate that our method outperforms state-of-the-art systems in terms of orientation estimation precision while maintaining similar or better accuracy in position estimation. Moreover, our method is computationally efficient, allowing it to detect and localise several markers in a fraction of the time required by the state-of-the-art fiducial markers. Furthermore, the presented method requires only an off-the-shelf camera and printed tags, can be quickly set up and works in natural light conditions outdoors. These properties make it a viable alternative to expensive high-end localisation systems.

CCS Concepts

•**Computing methodologies** → **Vision for robotics;**



(a) WhyCon (b) WhyCode

Figure 1: Two ambiguous poses are obtained while localising a fiducial marker, outlined in blue and green. The white segment's centre is marked in red. The apparent centre of the entire marker is in black.

Keywords

Fiducial Markers; Swarm Robotics; Visual Tracking

1. INTRODUCTION

As the field of robotics is experiencing rapid expansion, more and more robots appear in every area of our lives. Deployment of robots in domestic and natural environments is enabled by the never-ending advances in hardware platforms, which have become smaller and more capable. The advances in machine learning kept pace with hardware development, and together, they led to versatile methods deployable in robots operating in both industrial and natural environments. However, the holy grail of long-term autonomy, lifelong learning and self-sustainable control has not been achieved yet. Deficiencies in perception, motion con-

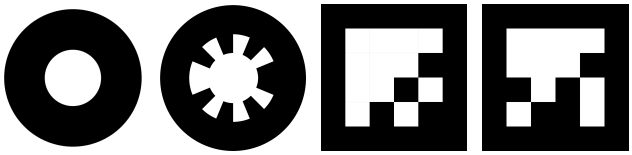
trol, path planning and human-robot interaction still result in robots struggling to be accepted and integrated into society [19, 24]. Thus, the introduction of truly intelligent robots into domestic environments is yet to come.

An essential capability of a mobile robot is autonomous, safe and socially conscious navigation [45, 39, 19]. To achieve that, a robot has to be able to precisely and reliably estimate its position relative to the desired path and objects in its vicinity. Wrong estimation of the robot's orientation or position typically results in erratic behaviour, often leading to damage. Precise and reliable position estimation relies on the accuracy of the measurements provided by the robots' sensors.

The robots are usually equipped with active or passive sensors that utilise infrared or visible light spectrum, respectively. One of the most popular passive sensors is a colour camera, which can be used not only for mapping and localisation but also for object detection or person recognition. The popularity of cameras has risen even more because of the qualitative leap in machine vision methods driven by the advances in artificial neural networks (ANN). Despite the advances in deep learning, engineered methods based on projective geometry and photogrammetry are still popular due to their computational efficiency and explainability. These methods are still being used to detect specific patterns, landmarks, and shapes and calculate their positions relative to each other or the camera. This is a particular case of fiducial marker detection, where the task is to process an image to identify and localise artificial markers in the scene captured by the camera. These fiducial markers are specifically designed for reliable detection and localisation. Most fiducial markers were originally developed for augmented reality applications [27, 12, 47], but their versatility brought them into robotics. Here, they serve both as a reliable source of position information to evaluate experiments [36, 5, 3, 1, 17], for a closed-loop formation control of robotic swarms [42, 43, 50], or to indicate mission-critical objects, such as charging stations [18], entrances [38], objects to interact with [40, 28] or autonomous drone landing [46]. The core advantages of the fiducials are their intuitive use, the negligible price of the tags themselves and the low prices of the cameras and computational equipment.

Despite the fact that the markers are tailored for easy detection and localisation, there are situations where the performance of the marker detection methods needs to be improved. In typical scenarios, camera resolution, physical marker size limitations and lighting conditions restrict the range where the markers can be reliably detected, identified and localised. This can be partially alleviated by carefully choosing the size and material of the markers depending on the particular setup. As the operational conditions and position estimation requirements depend on a particular scenario, there is no universal fiducial fit for every application. This makes the field of fiducial marker detection worth studying.

We present a vision-based method, capable of reliable detection and full six-degree-of-freedom localisation of a circular black-and-white fiducial 'WhyCode' marker, originally presented in [23, 30, 29]. Our method achieves high robustness of detection, reliable identification, and precise localisation at meagre computational costs. In contrast to the previous systems [23, 30, 29], our method explicitly addresses the ambiguity in the pose estimation, see Figure 1. The solution to this ambiguity problem significantly improves the precision of the marker position and orientation estimation. To evaluate the impact of the extension above, we collected a real-world dataset representing a typical application of fiducial markers in swarm robotics. Then, we used the dataset to compare our method to the most popular state-of-the-art algorithms.



(a) WhyCon **(b) WhyCode** **(c) ArUco** **(d) Apriltag**

Figure 2: Commonly used fiducial markers in the field of swarm and mobile robotics.

2. RELATED WORK

In robotics, the *fiducial markers* mainly serve as reference points, indicating the positions of the robots or critical environmental cues. Their detection properties make them popular in scenarios of external or self-localisation when placed on the robots or in the environment, respectively. While their shape and colour are virtually unrestricted, the most popular fiducial tags are planar markers with apriori known geometric characteristics and contrastive colours, see Figure 2. Their shape and colour are tailored so that computer vision methods can robustly detect them and precisely estimate their position and orientation in space. While the planarity of the markers increases the complexity of pose estimation, it allows them to be printed on paper and easily attached to most objects. Applications of the tags supporting localisation robustness and navigation precision can be found in [18], and their deployment in autonomous vehicles is described in [21].

AprilTag is a square black-and-white marker with an embedded 2D binary code. The associated detection method can distinguish between several markers and estimate their position, and orientation [37]. In the detection phase, the image is first binarised, and straight lines are extracted. These lines are then filtered so that only those that form closed contours with four corners remain. The algorithm is set to prevent situations where a marker is missed. Therefore, after this step, a set of candidate detections typically contains many false positives. To remove the false positives, the line batches are tested for further properties, line the maximum corner angle or enclosed area size. Then, the inside of the marker is divided into a grid according to the used coding size and type. The thresholded grid cells form a lexicographical binary code, which offers a scalable tradeoff between the number of encoded IDs and the false-positive rate. In case of a low number of IDs, the encoding scheme provides certain robustness to partial marker occlusions. An extension to AprilTag, which supports more diverse shapes and custom content inside of the marker, was proposed in [25].

Another square fiducial, **ArUco**, also stores its ID in an enclosed binary matrix [15]. Similarly to the AprilTag, it first binarises the image, searches for the contours with an edge detector and then identifies the marker by extracting the cells of the enclosed black-and-white matrix. Contrary to AprilTag, Aruco does not extensively interpolate the detected lines to form enclosed shapes. To deal with occlusions, the AruCo authors suggest using a board with several redundant markers instead of a single one. Recently proposed improvements to ArUco account for spatiotemporal constraints of the marker position and size, achieving faster detection speeds [41]. Naturally, utilisation of these contractions is applicable only in continuous video streams. The work of [21] points out that the unstable detection time of the square markers is their primary deficiency. This applies especially when the markers are used in vision-in-the-loop systems, where real-time performance is crucial.

WhyCon addresses the issue of computational complexity at the expense of the ability to distinguish individual markers. The WhyCon is a circular fiducial marker consisting of two concentric circles of known parameters [23]. Typically, the outer circle is black, and the inner one is white. Due to the marker symmetry, the method can, in theory, estimate only the marker's 5 DOF - 3D position and two axes of orientation. The computational efficiency of the WhyCon makes it capable of tracking hundreds of markers within tens of milliseconds while achieving accuracy in the order of millimetres. Unlike the square markers, no edge detection is performed. Rather, the detection is based on on-demand thresholding, local flood fill segmentation, and a series of on-the-fly tests leading to early rejection of false positives. The detection produces parameters of a conic section, which are subsequently used for pose estimation. In some scenarios, the method's computational efficiency depends only on the number of pixels occupied by the markers and not on the resolution of the entire image. This makes WhyCon a popular choice in systems with constrained computational power, and memory size [13]. The inability to determine full, 3 DOF orientation is addressed by the markers derivatives, the WhyCode and SwarmCon.

WhyCode marker uses the same detection core as WhyCon but extends the system with a versatile identification method [30, 29]. Its segmentation procedure and projection model are the same as WhyCon; thus, it achieves comparable performance in detection and localisation. However, WhyCode incorporates uniquely identifiable 'Necklace' code [9], which also breaks the marker symmetry, allowing to estimate the marker's rotation around its normal. The code is inscribed into the boundary between the white and black marker segments. Therefore, it is one-dimensional, and its length scales only linearly with the marker size or camera resolution compared to square markers, where the marker size to code-length ratio is quadratic. On the other hand, even when the WhyCode marker is too far away for the code to be reliably readable, it can still be detected and localised. Moreover, the identification phase produces a negligible computational overhead. Thus, the fiducial remains as computationally efficient as the original marker while it adds the identification and 3D orientation estimation.

SwarmCon is a different branch of the WhyCon marker meant for ground-based swarm robots [3]. In swarm experiments, the robots often need to be distinguished from each other. However, they typically move in a plane, which is parallel to the image plane of the overhead camera, used to localise them. Thus, SwarmCon enriches the marker with a distinguishable identification and planar orientation estimation. Similarly to WhyCode, the SwarmCon uses the WhyCon core methods and achieves a similar performance. In contrast to WhyCon, which uses two concentric circles, SwarmCon tags consist of two ellipses with slightly offset centres. The orientation of the marker is established by calculating the major axis of the black outer ellipse, with the ambiguity resolved by checking the inner ellipse offset direction. The ID of the marker is then encoded in the dimensions of the major and minor axes of the inner, white ellipse. The SwarmCon system retains the ability to detect circular markers. These are placed in the corners of the experimental area so that the SwarmCon system can automatically establish the transformation between the image coordinates and the coordinate system of the experimental arena. This results in a quick and convenient setup of the localisation system providing marker coordinates directly in the desired coordinate system. However, SwarmCon is usable only for 2d planar localisation as pose estimation of an ellipse at a general pose results in too many ambiguous solutions.

There are many planar black-and-white fiducial markers with various capabilities with a typical application area inclined towards augmented reality. These markers were introduced to enable precise localisation and tracking of a camera position in constrained spaces. One of the most influential square fiducials was ARToolKit [22]. ArToolKit was primarily used to identify the individual markers and neglected the robustness and precision of the marker pose estimation. While ARTag [14] and ARToolKit+ [49] improved the pose estimation performance, they were still tailored for augmented reality rather than for robotic applications. However, the advances in the hardware, leading to ever-increasing camera resolutions and computational power of off-the-shelf systems, the fiducial markers evolved into highly complex shapes, and colours capable of representing plenty of information [11, 6, 8, 25, 7, 16, 31].

As the vision-based systems became more and more capable of recognising common-day objects and using them as natural landmarks, robots do not require to perceive fiducial markers all the time to perform self-localisation. However, issues of lighting, perceptual aliasing, landmark deficiency, appearance variations and environment changes hamper the ability of the robotic systems to operate reliably over long periods of time. Therefore, methods that improve the performance of visual localisation and mapping systems through the introduction of sparsely-placed fiducial markers in the environment have been proposed [32, 33]. In these systems, the authors of UcoSLAM and SPM-SLAM use square-shaped markers as reliable artificial landmarks to support the localisation and map building in repetitive or otherwise tricky environments. Similarly, the autonomous robots meant to operate for weeks to months used the WhyCon markers to reset their position at specific locations in their operational environments [18]. While the above systems can rely exclusively on natural landmarks or artificial markers, they are designed to use a combination of both.

3. MARKER LOCALISATION METHOD

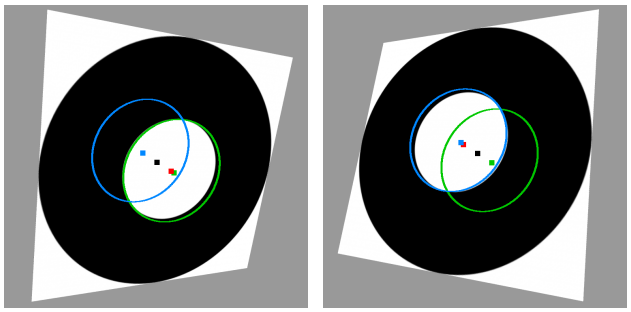
In this section, we will present an advanced fiducial marker localisation system that originates and builds on top of the WhyCon and WhyCode fiducial markers. The system is capable of both online and offline tracking, possesses robustness to motion blur and can adapt to variations in the scene ilumination. Also, the marker detection can run in real time while maintaining high precision. The WhyCon and WhyCode markers have been used and tested in a various applications and typical use cases, such as human-to-robot interaction, robotic swarm and heterogenous robot cooperation as well as low-cost tool for providing reference position for robotic experiments evaluation [50, 36, 3, 34].

The WhyCon class of fiducials feature similar detection algorithm, whose original and extensive description is presenter in the paper [23]. Therefore, the system can achieve accuracy in millimetres and detect and localise multiple markers in real-time. WhyCon and WhyCode markers are camera-based localisation systems that can estimate the position and orientation of up to thousands of markers in an image. At the same time, it is sufficient to use an off-the-shelf web camera and a low-cost printer [13, 30].

In addition, the WhyCode and WhyCon marker detection methods have low resource requirements, which makes them suitable for low-powered or resource-restricted platforms. The marker's circular pattern detection is achieved through flood-fill segmentation, followed by consecutive rules evaluation of segment properties with increasing complexity. The segmentation benefits from on-demand thresholding, which ensures a low number of pixels to be processed. The rules applied to the elementary image segments allow a flexible detection success rate based on the required spatial and image properties of the marker pattern. The segments are skipped whenever any rule is not satisfied. The characteristics of individual segments are cached and examined on the following image frame to initialise the detection because one can assume the scene does not change significantly between the frames. Thus, the tracking can take full advantage of evaluating only a few pixels compared to the whole image.

In the paper [30], the authors have enhanced the original marker system by introducing a flexible binary encoding to create distinguishable and identifiable markers. This special encoding is not affected by changes in orientation, which means that the angle of revolution around the surface normal can be estimated. The used encoding is based on the Necklaces [9]. The newly used identification code is located on the border of the white and black segments. The encoding is wrapped around the circle and enhanced by the Manchester encoding for better robustness. Using the unique feature of Necklace encoding, the embedded ID can be extracted and decoded from anywhere along the perimeter because the binary code is then circularly shifted to settle on the lowest value. Based on the required amount of code shifts, the revolution around the normal can be established.

Although the typical application of the fiducial markers was to provide only the planar 2D position of robotic swarms, estimate the plane orientation with respect to the optical axis, or distinguish markers from each other, the problem of estimating the full 3D pose was suppressed. Thus, there is an unresolved ambiguity in the position and orientation. When examining the ambiguity, the origin can be found in the later phase of estimation, as the position and orientation equation always yields two possible solutions. To address this issue, the authors further address the lacking feature in the system to ensure accurate 3D pose estimation.



(a) θ = [0, -0.3, 0.5] rad (b) θ = [0, 0.3, -0.5] rad

Figure 3: WhyCon pose ambiguity. Two solutions to pose estimation are displayed in blue and green. The markers provide the same possible solutions while having opposite orientations.

3.1 Addressing Ambiguity of Solutions

The image segment output by the detection pipeline is ready to be further evaluated to establish the position and orientation of a marker. At first, we start by calculating the fundamental geometric properties of the found segments. The segment's centre point corresponds to the perspective centre of the marker. Further, the ellipse corresponding to the perspective projection of the circular marker is estimated by eigen analysis of a covariance matrix formed from the segment points. The transformation to a camera-centred coordinate system is applied to the calculated characteristics which are then used to express the conic section of the sought ellipse whose points $\mathbf{X} = (u'v'1)^T$, where u', v' are the transformed coordinates form the conic section $\mathbf{Q}$ as $\mathbf{X}^T\mathbf{Q}\mathbf{X} = 0$. The marker position $\mathbf{x}$ and normal vector of the marker's plane $\mathbf{n}$ are found by decomposing the formed conic section as presented in [51, 30].

$$\mathbf{x} = s_3 \frac{r}{\sqrt{-\lambda_0\lambda_2}} \left(s_1\lambda_2\mathbf{q_0}\sqrt{\frac{\lambda_0-\lambda_1}{\lambda_0-\lambda_2}} + s_2\lambda_0\mathbf{q_2}\sqrt{\frac{\lambda_1-\lambda_2}{\lambda_0-\lambda_2}} \right) \tag{1}$$

$$\mathbf{n} = s_1\lambda_2\mathbf{q_0}\sqrt{\frac{\lambda_0-\lambda_1}{\lambda_0-\lambda_2}} + s_2\lambda_0\mathbf{q_2}\sqrt{\frac{\lambda_1-\lambda_2}{\lambda_0-\lambda_2}} \tag{2}$$

where the radius of a marker is r, the eigen decomposition of the conic $\mathbf{Q}$ provides the eigenvectors $\mathbf{q_0}, \mathbf{q_1}, \mathbf{q_2}$ and the corresponding eigen values is the following form $\lambda_0 \geq \lambda_1 > 0 > \lambda_2$. The symbols s_1, s_2, s_3 represent undetermined signs.

The total number of eight possible outcomes from the equations can be reduced by applying spatial restrictions on the marker with respect to the camera. We can assume that the marker is only visible when placed in front of the camera. Also, we can derive a similar restriction about the marker's surface normal vector as the marker is expected to be shown to the camera rather than facing away from it. Thus, we can

formulate the following conditions that have to be satisfied

$$\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \mathbf{n} < 0 \tag{3}$$

$$\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \mathbf{x} > 0 \tag{4}$$

Two out of the three undetermined signs can be derived based on the restrictions above. Therefore, we can obtain two pairs of position and surface normals estimates, $\mathbf{x_1}, \mathbf{n_1}$ and $\mathbf{x_2}, \mathbf{n_2}$. A unique solution is also possible when $\lambda_0 = \lambda_1$; however, it does not affect the further steps. Those two pairs of solutions are the sought source of ambiguity in pose estimation, and one cannot choose from the two options arbitrarily. In [51], the authors suggest incorporating additional measurements from robotic sensors to select the correct solution. Unfortunately, additional sensors are not always the available option. Fortunately, the marker pattern contains enough data to resolve the ambiguity. The WhyCon and WhyCode do not fully share the same pattern, so they require different steps to resolve the final pose estimation.

3.1.1 *WhyCon's Solution*

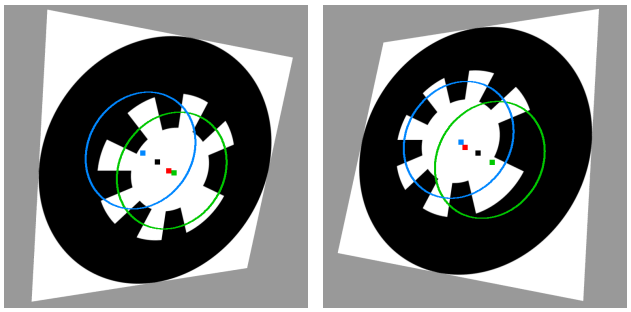
The WhyCon marker is a circular fiducial marker consisting of two concentric circles, one black and one white. The concentricity of the circles provides additional information that can be exploited for improved accuracy. Currently, during detecting and localising the marker, the inner white circle is evaluated along with the black ellipse, and the characteristics of the whole pattern are established. However, the already computed characteristics of the white circle are not used in the subsequent pose estimation procedure.

The white segment of the marker is crucial for resolving the ambiguity in pose estimation because it provides an additional source of perspective information that is tightly related to the projection of the overall marker. Specifically, the white circle's perspective centre can be assumed to be closely related to the real projected overall centre of the marker. We first convert the 3D position vector back to the pixel coordinates to resolve the ambiguity in pose estimation. This yields two points that are equidistant from the perspective centre of the marker. The correct position vector is then chosen as the one that is closer to the white centre. The distance between the reprojected points and the white segment centre serves as the selection condition for resolving ambiguity. It is important to note that the perspective has a more negligible effect on smaller objects in the scene. This allows us to assume that the perspective centre of the white circle will not drift from the projected centre of the whole marker.

Figure 3 demonstrates the presented ambiguity resolution procedure. The two images contain two orientations of a marker which share the same pair of ambiguous positions and normal vectors. As they vary only in orientation, the white segment hints the which pose estimate to choose. By utilising the concentric circles in the WhyCon marker and the inner white circle, we can significantly improve the accuracy and stability of the pose estimation in robotics and computer vision applications.

3.1.2 *WhyCode's Solution*

The WhyCode fiducial marker differs in structure, which



(a) θ = [0, -0.3, 0.5] rad **(b) θ = [0, 0.3, -0.5] rad**

Figure 4: WhyCode pose ambiguity. Two solutions to pose estimation are displayed in blue and green. The markers provide the same possible solutions while having opposite orientations.

makes it harder to determine the correct solution for its pose estimation. Moreover, correctly identifying the marker is closely tied to choosing the right solution. The white inner area of the marker cannot be used the same way as in the WhyCon marker due to the binary encoding between the black and white circles. This results in an uneven segment with a shifted centre in contrast to a simple circle, making it useful only for parameter testing in the detection phase. This effect is more pronounced when a low number of bits is encoded because the wider white teeth affect the averaged centre to drift in their direction. The bias of the white centre is reduced when using a high number of bits for encoding ID or when the perspective of a specific viewing angle makes the white sections less significant. A possible approach to this problem would be to increase the number of encoding bits, but this would come at the cost of losing the marker's benefit of high detection range and robustness.

The detection pipeline already processes the circular code, making it suitable for exploiting it to decide on the solution. A brightness signal is extracted by considering a given number of points around the marker's perspective centre to extract the binary code. Originally, the signal was processed to remove the gradient, and then the maximum was found, which was assumed to be the middle of the widest white section. The encoded binary code is extracted by under-

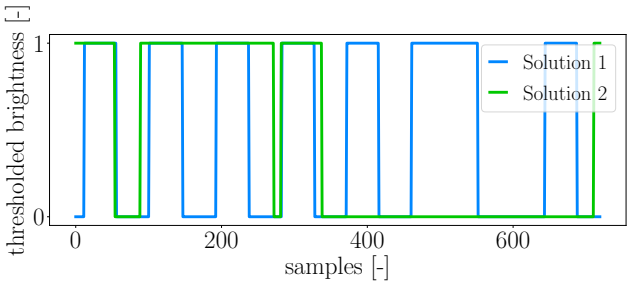


Figure 5: Binary signal extracted by sampling along the perimeters highlighted in Figure 4b. The signal of the correct solution is in blue, and the uneven signal in green is of the other solution from the ambiguity pair.

sampling and binarising the signal with a tooth-wide step starting at the found maximum point. However, this identification method has a disadvantage because the brightest point does not have to be located in the middle of a section; instead, it is dependent on the scene illumination. Therefore, it can lead to misalignment of the resampling step and produce an incorrect binary code.

This issue can be solved by binarising the signal after it is sampled using a threshold value corresponding to the mean brightness. Section edges can then be easily detected and form the Manchester-encoded Necklace code. While this approach might potentially reduce the robustness to scene illumination differences, it significantly improves the identification performance.

After modifying the decoding process, we can reliably use it to resolve the pose ambiguity. This involves obtaining both pairs of positions and normal vectors and then transforming them back into the image coordinates. We then extract the circular code as described earlier, resulting in two thresholded brightness signals. If the signal was sampled around the correct centre, the black and white sections would be of similar length, or, in the case of polarity change, the length would be twice as much. Therefore, the signal with sections of two particular lengths only indicates the correct pose, while the signal with variable and unevenly long sections indicates an incorrect pose estimate. Thus, the variance of the position of the teeth edges along the circular code is calculated for both signals. The corresponding pose estimate to the signal with a smaller variance is then considered the sought ambiguity solution. However, the pose estimation becomes more complex as the encoding has to be evaluated twice, and the parameters of the ambiguity test have to be obtained. Nevertheless, it does not impact real-time performance.

In Figure 4, it is not difficult to distinguish the correct solution, although their orientations are related through ambiguity. One can also notice the importance of correctly resolving the pose as decoding the embedded circular code relies on it. The signal evaluation phase is presented in Figure 5, which is the thresholded brightness of the marker in Figure 4b.

4. DATASETS

In order to evaluate the pose estimation quality of the individual state-of-the-art markers, we decided to test their performance first on simulated artificial data and then on real-world data. As the fiducial markers can provide very accurate estimations and are sometimes used for reference measurement in robotic experiments, we had to approach the testing carefully. Therefore, we started with creating a simulated world with only a marker and a camera, so we would be able to evaluate the methods in perfect and repeatable conditions. The actual applications are by nature more realistic in terms of sensor data and not so repeatable, and thus it is necessary to know the performance of the methods under real-world conditions. We also used the same generated markers across all datasets as it can influence the detection capabilities. WhyCode was using 8-bit encoding, which yields 30 markers, the ArUco dictionary

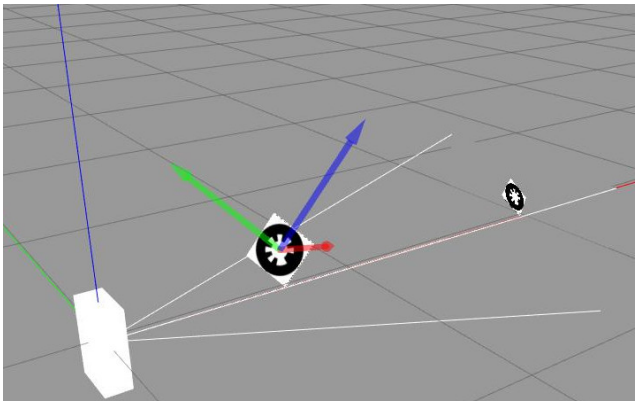


Figure 6: Overview of the Gazebo world with the included models of the 3D external localisation scenario.

was set to 4×4 and to contain 50 different markers, and AprilTag markers were generated from the 16h5 encoding family.

4.1 Artificial Dataset

To generate the artificial dataset, we chose the Gazebo framework, which can be directly used with the Robot operating system. Gazebo offers an extensive amount of options and community support, together with a highly precise and trustworthy physics engine and graphics rendering. In the framework, a simple yet sufficient artificial world was prepared to represent one of the most common scenarios - external localisation in 3D. The camera parameters were set not to perform any noise and distortion so we could achieve a perfect image. The camera's internal parameters were also chosen to represent an ideal camera. The prepared simulated world had also modified the light source by allowing only ambient light at the highest intensity so the marker would be illuminated evenly. The commonly used and standardised camera plugin *gazebo_ros_camera* was set to capture the images at Full HD resolution. For a fair comparison, we generated the markers to encode a similar number of IDs.

This scenario represents a typical usage of a fiducial marker, so pose estimation of a marker in the camera-centred coordinated system. The prepared artificial world contains only a single marker and a camera, as shown in Figure 6. For our convenience, the camera is located at the origin of the world coordinate system. The marker is positioned and rotated randomly within the camera's observable space when running the simulation. In order to gain a better understanding of the methods' performance, the size of a marker's edge or marker's diameter was set to $20\,\text{cm} \times 20\,\text{cm}$. It is around the maximum that can still fit on an A4 paper while keeping a white separation border around the perimeter of markers.

The position is uniformly sampled in each direction from appropriate ranges. The distance can vary from 3 m to 5 m, and the left-right position can reach up to 1.5 m in both ways, and the up-down displacement shares the same range and the horizontal one. The process of selecting the orientation was also random. The rotation around the normal vector was sampled uniformly from the range of a full rev-

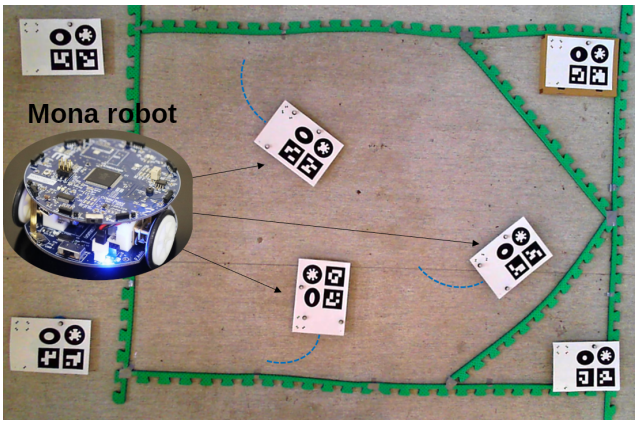


Figure 7: Overview of the swarm arena with three moving robots and four stationary in the corners.

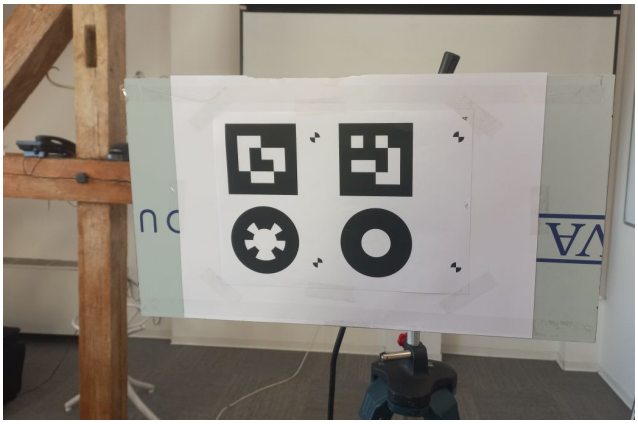


Figure 8: Camera's point of view of the pan-tilt unit with the marker board at the joint home configuration.

olution, 0 rad to 2π rad. Then, we decided to separate the generated data into two categories based on the size of the remaining two angles. The *small-angle* data represents a random draw of the angles from the interval $[-\pi/6, \pi/6]$ rad. While the *large-angle* data were obtained by sampling the angles from $[-\pi/3, -\pi/6] \cup [\pi/6, \pi/3]$ rad. Therefore, we were able to generate images of the marker's pose at extreme angles and change position quickly without the use of specialised and high-end equipment, which would be required for real-world data.

4.2 Real-world Datasets

Two real-world datasets were collected to verify and compare findings based on the artificially simulated data. Also, without using real sensory data, it is hard to rely on the performance of the individual methods under realistic conditions. Therefore, the following datasets focus on evaluating the methods' position and orientation capabilities.

4.2.1 Swarm Arena

This dataset is one of the typical use of such fiducial markers as they are used for the localisation of swarms in an arena acting as a cheap motion capture system. The benefit of requiring only a single camera without specialised and expensive hardware led to the popularity of markers in the field of swarm robotics [4, 34, 35]. It is often that the specialised systems use infrared emitters, which can, unfortunately, blind many basic sensors attached to swarm robots. Thus the passive markers are preferred.

The arena with swarm robots is displayed in Figure 7. The arena has dimensions of two by three meters, and the observing off-the-shelf camera is located three meters over it. The camera was recording at Full HD resolution at 30 FPS. The reference pose estimation was done by the Vicon motion capture system, which is an active system based on infrared retroreflective markers achieving submillimeter precision.

Three robots drove randomly in the arena, following only simple collision avoidance rules. The robots covered the whole arena, which was also staked out by the four stationary robots located in the corners. They aimed to establish a reference coordinate frame between the motion capture system and the camera. Each robot carried a board with all the evaluated methods' patterns so the observation conditions would be equal. The edge length, as well as the circle diameter, had the same size 8.4 cm. The platform used was the MONA [2], which is an open-source and open-hardware swarm robot.

The recorded data contain more than seven thousand images and more than sixty thousand poses from the Vicon system. The reference measurements by the Vicon system were recorded around eight times faster than the camera stream. Such a property is common across high-end motion capture frameworks. Thus, aligning and matching the individual images with the corresponding pose estimation was necessary. At first, the key moments detectable in both streams were localised, the start and end of a movement. Then, starting at the key point, we found the timewise nearest Vicon data to an image frame. The alignment outputted 5716 pairs of images and Vicon measurements.

4.2.2 Manipulator Interaction

The following dataset represents another important use case of the fiducial markers, the orientation estimation. The orientation estimation is especially critical in applications involving an unmanned aerial vehicle autonomous landing, robotic docking and interaction with an environment [26]. The orientation is especially challenging to estimate in situations when the marker's plane is closely parallel to the image plane. However, those situations are often critical for delicate interaction, manipulation and motion control.

Therefore, we prepared another board with individual fiducial markers and attached it to a pan-tilt unit. The unit, Flir PTU D-46, was used to automatically and precisely rotate the markers by a given step in both controllable axes. In the default home configuration, the unit directly faced the camera as in Figure 8. The pan movement corresponded to the rotation around the board's normal, and the rotation range was -90 to 90 degrees. The consecutive tilt movement ranged from -40 to 30 degrees. Thus, there were 555 different states of the unit joint configuration.

The board was observed by the Intel RealSense D415 camera. The camera captured each orientation state of the board at the resolution of 1280×720. The marker's edge or diameter size was again set to 8.4 cm. The pan-tilt unit was positioned one, two and three metres from the camera. Thus, we recorded three datasets for orientation error evaluation with increasing distance.

5. EXPERIMENTS

The presented experiments aim to thoroughly examine and stress the capabilities of the state-of-the-art methods and our proposed improvements to the WhyCode method. We used for this purpose both artificially generated datasets and real-world datasets. As any sensor is not perfect, the camera produces noise, and its lens causes the image to bend and deform. The deformation can be overcome by carefully calibrating the camera to find the internal camera parameters and the distortion parameters of the commonly used plumb bob model. For such purposes, we used the well-known OpenCV calibration toolbox.

The swarm arena dataset also include image sequences when the robots in the swarm dataset move quickly and cause a motion blur effect in the markers, which further provides real-world data one can encounter in practice. Even though the computer simulation could provide some overview of the performance, the findings could not be straightforwardly expected under natural conditions.

The experimental scenarios are motivated by the typical applications of the fiducial markers, thus estimating the marker's pose in unrestricted 3D space and also in the restricted planar arena-like setup. The deployment of the markers to provide the external localisation for a robotic arena is expected in the swarm community as the markers serve as a cheap tracking system without the need for specialised hardware. To ensure the same conditions for the individual markers, they processed the same video recordings with the same intrinsic and rectification parameters.

The methods were evaluated based on the average error metrics specified below. To also reach a statistical conclusion about the relative performance of the individual methods, we exploited the nature of the experiments and used non-parametric methods developed for complete block design studies, namely the Quade test [10] with Šidák's correction [44] and tested at the 5% significance level. For this, the library scikit-posthocs [48] was used.

5.1 Artificial Dataset

The primary goal of the experiments on the simulated artificial data was to thoroughly test the general task of external localisation of a marker in 3D. The capabilities of the individual methods to estimate the full 6 degrees of freedom were intensively evaluated even on edge cases of spacial configuration. Even though those special cases are not typical, as they can be easily filtered out in practice, it is essential not to omit them. Therefore, this experiment consists of two similar evaluations based on the random angle distribution ranges. The small-angle experiment is more challenging for ambiguity resolution and has the angles upper bounded by $\pm\pi/6$ rad. Contrary, the large-angle experiment stresses the ability to detect and estimate the pose under large observation angles from $\pm\pi/6$ rad up to $\pm\pi/3$ rad, respectively. In order not to lose the big picture of the overall performance, the fiducial markers must deliver satisfactory results in all situations. There are several stages of successful pose estimation, beginning at the image thresholding and segmentation, then selecting correct segments, calculating their position and orientation, and finally extracting the embedded identification information. Despite the fact that in this experiment, we measure just the pose error, the individual stages and their errors affect the resulting pose estimation.

In order to express and measure the error of orientation estimation, we applied the metric suggested by the authors of [20]. They review several methods and recommend using the following metric for measuring the difference between

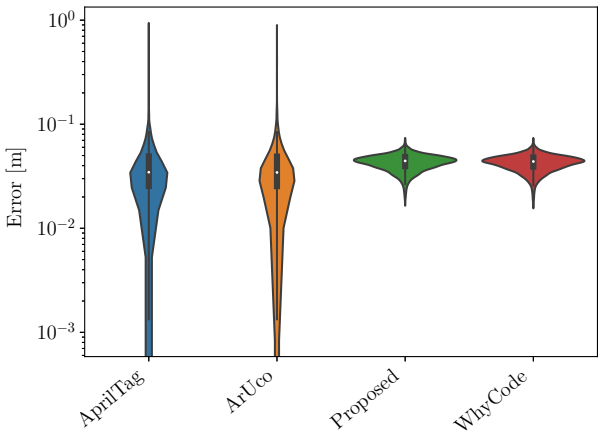


Figure 9: Position estimation error distribution on the artificial small-angle dataset. The error axis is in the logarithmic scale.

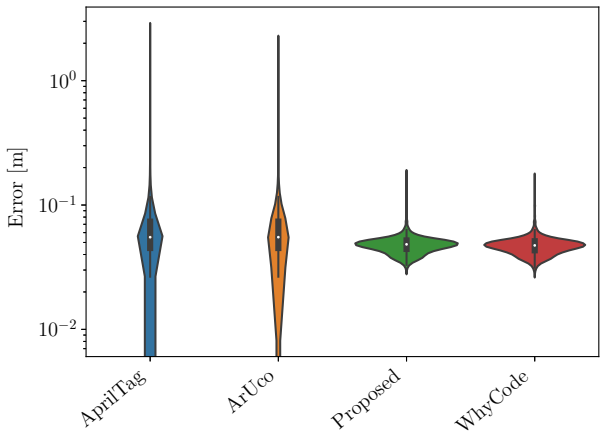


Figure 10: Position estimation error distribution on the artificial large-angle dataset. The error axis is in the logarithmic scale.

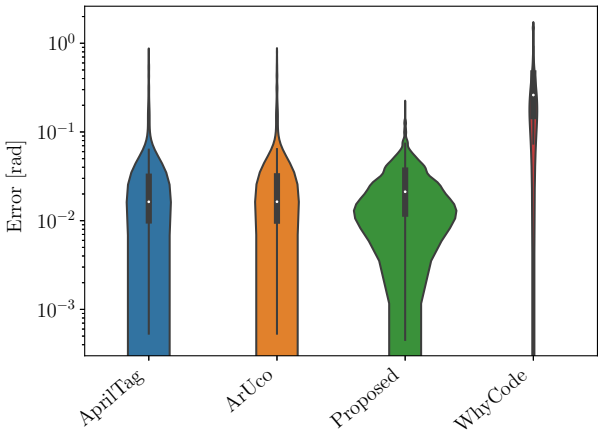


Figure 11: Orientation estimation error distribution on the artificial small-angle dataset. The error axis is in the logarithmic scale.

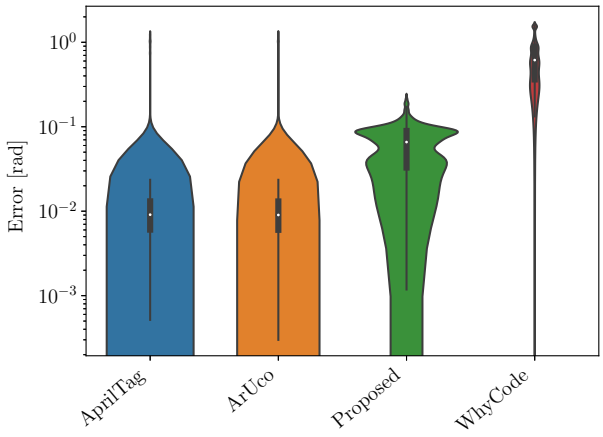


Figure 12: Orientation estimation error distribution on the artificial large-angle dataset. The error axis is in the logarithmic scale.

two quaternions.

$$\begin{aligned} &\Phi\colon S^3 \times S^3 \to \mathbb{R}^+ \\ &\Phi(\mathbf{q_1}, \mathbf{q_2}) = \arccos |\mathbf{q_1} \cdot \mathbf{q_2}|, \end{aligned} \tag{5}$$

where $S^3 = \{\mathbf{q} \in \mathbb{R}^4 | \|\mathbf{q}\|^2 = 1\}$, $\cdot$ is a dot product, and since Φ has to be non-negative, we restrict the range of values to $[0, \pi/2]$ radians. The metric can be interpreted as calculating the smallest rotation angle between the quaternions.

We let the simulation generate 3000 random poses and the respective camera images for both angles ranges options. After, the methods were evaluated on the data. Wrongly identified markers or situations when the marker was not detected at all were dropped. In the small-angle experiment, the methods detected and correctly identified the marker in all images. However, the second simulation with large angles was challenging for WhyCode as it wrongly identified three images, and in 334 situations, it could not detect the marker. The AprilTag and ArUco presented no false IDs and could not detect the marker in 72 and 66 images, respectively. In total, 348 images and estimations had to be skipped as not all of the markers provided a pose estimate.

Figures 9-12 show the performance metric of individual methods under individual scenarios described in the previous section. Starting with the position estimation, the first notable thing is the presence of very extreme outlier values in the errors by square-based methods AprilTag and ArUco. As to be expected, the performance of these methods also differs significantly between the scenario with low angle and high angle under which the marker is being detected. While the average performance of AprilTag and ArUco surpasses by a little the accuracy of both the WhyCode and Proposed method, these give much more robust output. Based on statistical testing, we identified the order of methods—with their respective average error in the brackets and with a less-sign where a statistically significant difference was found—to be AprilTag (4.19 cm) < WhyCode (4.34cm) < ArUco (4.37 cm) < Proposed (4.39 cm) for the small angle scenario and WhyCode (4.78 cm) < Proposed (4.88 cm) < AprilTag (6.65 cm), ArUco (6.99cm) for the large viewing angle scenario.

The errors in the orientation estimation were computed using the aforementioned metric over three degrees of freedom specifying quaternions. When the viewing angle (Figure 11) was low, the best-performing method was the new Proposed method. The whole order of performance was Proposed (0.026 rad) < AprilTag (0.0580 rad), ArUco (0.0588 rad) < WhyCode (0.413 rad). On the other hand, where the angle was large (Figure 12), the square-based methods slightly outperformed the Proposed method with the full order being AprilTag (0.044 rad), ArUco (0.048 rad) < Proposed (0.065 rad) < WhyCode (0.615 rad).

5.2 Swarm Arena

This general evaluation of individual methods examines the position estimation performance in 3D. The distributions of errors in position estimation of individual methods are shown in Figure 13. Based on the produced errors over the whole dataset, the evaluated methods show significant differences compared to each other. The ArUco fiducial marker achieved the smallest average estimation error of 17 millimetres and became the most accurate one in this testing scenario. Our method Proposed 2D achieved comparable results with an average of 18 millimetres, followed by its full 3D version at 31 millimetres, comparable with AprilTag performance of 35 millimetres of average error. Evaluating the performance of our proposed marker, it managed to estimate the position more accurately with smaller error than the AprilTag marker. The distribution of position estimation error of the Proposed method does not resemble unimodal distribution and is likely caused by undistinguishable situations in ambiguity resolution. Overall, the Proposed method achieved comparable or smaller error in position estimation than the other evaluated methods, especially compared to the original WhyCode method, and it seems to be more robust, looking at the extreme error values.

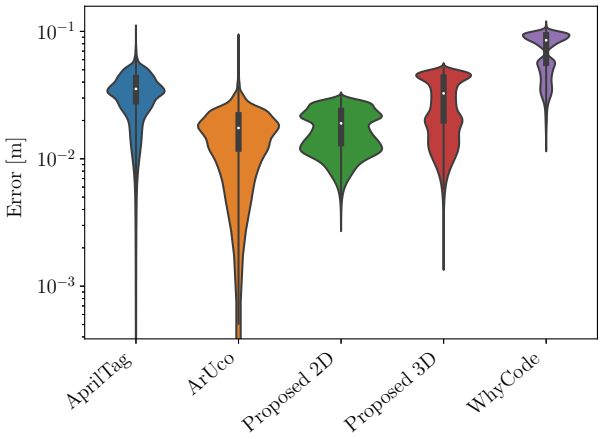


Figure 13: Position estimation error distribution on the Swarm arena dataset. The error axis is in the logarithmic scale.

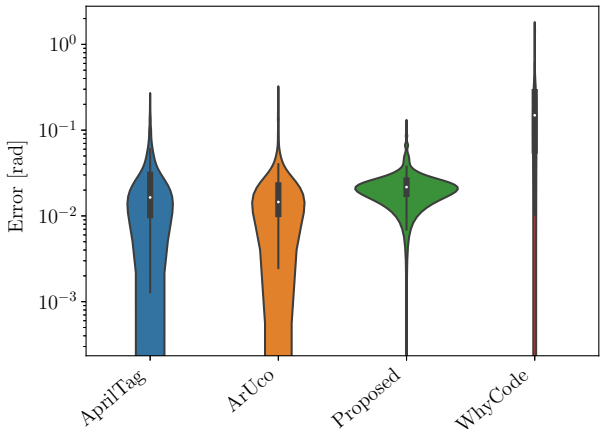


Figure 14: Orientation estimation error distribution on the Manipulator interation dataset. In this case the marker board is positioned approximately 1 meter from the camera. The error axis is in the logarithmic scale.

One of the typical deployments of fiducial systems is providing a planar position of robots in a swam arena. Therefore, we decide to evaluate the performance of estimating the 2D position. One can notice a difference compared to the previous evaluation. It is caused by the fact that the proposed marker, when hinted, can take advantage of estimating only a planar position and orientation instead of estimating an unrestricted 3D pose. The performance boost of the 2D estimation with respect to the 3D estimation is clearly noticeable. By leveraging the additional information, the performance shifted to that of the ArUco marker. The overall error distribution parameters of our marker improved noticeably. Despite the improvements, one can see that the error distribution remains slightly bimodal but more compact. One of the explanations could be not fully resolving the ambiguity in edge cases of pose estimation. Nevertheless, the comparison of the individual methods shows that our proposed method is, in this scenario, comparable with ArUco rather than with AprilTag as it was in the full 3D estimation experiment.

5.2.1 Execution Time

The performed modifications of the WhyCode marker could cause a slowdown of the localisation system. Therefore, we also measured the methods' runtime of detection and localisation. The execution time was measured on the swarm dataset and averaged over all 7297 image frames. During the evaluation, we were only interested in the amount of time the method's function call takes, assuming the input data are already in the required form. The testing hardware was a laptop with the Intel Core i7-8550U processor and 16GB of system memory. In Table 1, the average execution time is presented, and based on that, we can conclude that the presented changes to the system did not affect the run time significantly.

Table 1: Execution time of the individual methods. The speedup is relative to the AprilTag.

Marker	Time [ms/frame]	Speedup [%]
AprilTag	30.3	N/A
ArUco	30.2	0.2
Proposed marker	0.9	3465.4

5.3 Manipulator Interaction

The motivation of this experiment is the precision of the orientation estimation, as it is crucial in many robotic applications involving manipulator grasping and interaction or autonomous drone landing. In those situations, an erroneous orientation estimation can lead to task failure or even collision with the environment. That is why we tested the

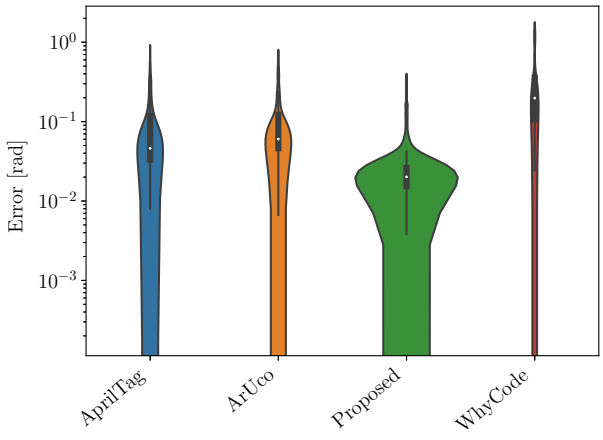


Figure 15: Orientation estimation error distribution on the Manipulator interation dataset. In this case the marker board is positioned approximately 2 meter from the camera. The error axis is in the logarithmic scale.

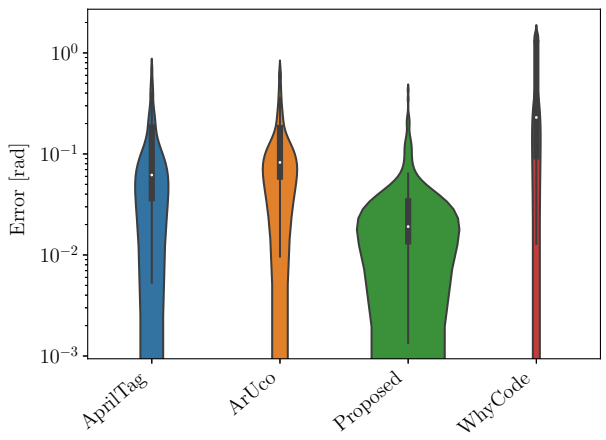


Figure 16: Orientation estimation error distribution on the Manipulator interation dataset. In this case the marker board is positioned approximately 3 meter from the camera. The error axis is in the logarithmic scale.

methods' capabilities with increasing distance from the camera to gain an overview of their performance. There were three datasets with the difference of the board with markers being approximately one, two and three meters away from the camera. This also allowed us to present the mean error evolution of the individual markers and the standard deviation. The orientation error was measured using the same metric as introduced before.

The performance of individual methods with varying detection distances is shown in Figures 14- 16. Looking closely at the error distributions, the relative performance of the methods is very similar across the distances, but the overall performance diminishes with the distance. This is expected as the markers have fewer and fewer pixels in the images, which makes them more volatile. However, unlike both square-based methods and the original WhyCode, our Proposed method maintains an outstanding performance of 0.042 rad even when the marker reaches a 3-meter distance (Figure 17), and the variance in the errors grows much less than for other markers (Figure 18).

At the distance of 1m ArUco, with an average error of 0.0241 rad, slightly outperforms Proposed with 0.0243 rad. This is followed by AprilTag with an average error of 0.028 rad and WhyCode with 0.32 rad, with the whole order statistically significant. At 2 and 3 meters, the ordering of the methods is the same and complete, with Proposed coming first, then AprilTag, then ArUco and finally WhyCode.

6. CONCLUSION

In this paper, we introduced a novel approach to fiducial marker localisation, which can provide real-time performance, accurate identification, and complete 6 degrees of freedom estimation. The proposed method builds upon the popular WhyCode and WhyCon marker, which uses a circular marker with a circular binary Necklace-based ID pattern. To overcome the limitation of resolving ambiguity in the pose estimation of the original method, we introduced a resolution procedure which improved the estimation of the marker's position and orientation, as the original estimation was not stable and precise. These improvements also target the accuracy of identifying individual markers, which is directly related to determining their rotation around their surface normals.

At first, the proposed method estimates two possible poses of a marker. Then, it evaluates both pose candidates with the pattern properties related to their back projection to the image space. However, the same resolution procedure could not be applied to both underlying fiducial marker algorithms. Therefore, two individual approaches were presented for each marker. We evaluated our method on real-world datasets inspired by typical mobile and swarm robotics applications to test our improvements.

Apart from comparing with the foundational methods, the

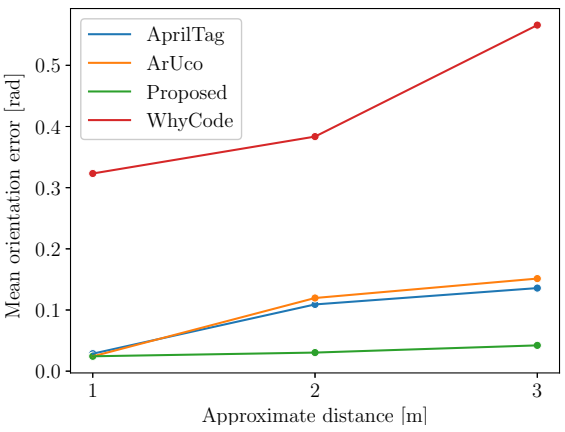


Figure 17: The evolution of the mean orientation estimation error with the distance from the camera.

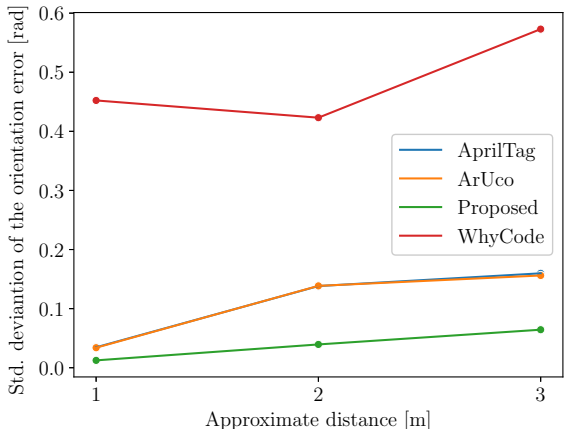


Figure 18: The evolution of the standard deviation of the orientation estimation error with the distance from the camera.

performance of our fiducial marker method was compared with the most commonly used state-of-the-art markers in scenarios of swarm arena and manipulator interaction as well as in an artificial robotic simulator. After analysing the results, we established that our fiducial marker method offers a comparable performance while retaining the real-time processing efficiency of the original WhyCode marker. The outcome of the comparison demonstrates that our method achieved a threefold increase in position accuracy compared to the original WhyCode, and it even outperformed AprilTag and ArUco in all evaluation scenarios in terms of robustness. Based on the orientation error evolution with respect to distance from the camera, we see a clear advantage of our method in larger but still realistic distance from the camera as it maintains lower mean error as well as standard deviation. In the position estimation experiments, our method achieved comparable performance with the square-based fiducial markers; however, the exact order of the methods was experimental condition dependent.

The next feature the fiducial marker could possess is providing the probability of correctly decoding the embedded ID. Such a feature could be helpful for continuous tracking even in adverse conditions. Also, one could transform the method to instantaneous localisation instead of tracking. However, the such act involves a change of the detection core.

To promote reproducibility of our research, we have disclosed all the codes and datasets at `https://github.com/jiriUlr/whycon-ros/wiki/2023-ACM-SIGAPP-ACR-23-1`.

Acknowledgments

The presented work was sponsored by the European Union's Horizon 2020 research and innovation FET Open programme under grant agreement No.964492 project 'RoboRoyale'.

7. REFERENCES

[1] F. Arvin, A. Attar, A. E. Turgut, and S. Yue. Power-law distribution of long-term experimental data in swarm robotics. In *International Conference in Swarm Intelligence*, pages 551–559. Springer, 2015.

[2] F. Arvin, J. Espinosa, B. Bird, A. West, S. Watson, and B. Lennox. Mona: an affordable open-source mobile robot for education and research. *Journal of Intelligent & Robotic Systems*, 94(3-4):761–775, 2019.

[3] F. Arvin, T. Krajník, A. E. Turgut, and S. Yue. Cosϕ: artificial pheromone system for robotic swarms research. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 407–412. IEEE, 2015.

[4] F. Arvin, A. E. Turgut, T. Krajník, S. Rahimi, I. E. Okay, S. Yue, S. Watson, and B. Lennox. ϕ clust: Pheromone-based aggregation for robotic swarms. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4288–4294. IEEE, 2018.

[5] A. Basit, W. S. Qureshi, M. N. Dailey, and T. Krajník. Joint localization of pursuit quadcopters and target using monocular cues. *Journal of Intelligent & Robotic Systems*, 78(3):613–630, 2015.

[6] F. Bergamasco, A. Albarelli, L. Cosmo, E. Rodola, and A. Torsello. An accurate and robust artificial marker based on cyclic codes. *IEEE transactions on pattern analysis and machine intelligence*, 38(12):2359–2373, 2016.

[7] T. Birdal, I. Dobryden, and S. Ilic. X-tag: A fiducial tag for flexible and accurate bundle adjustment. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 556–564. IEEE, 2016.

[8] 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*, pages 562–570, 2016.

[9] W. Y. Chen and J. D. Louck. Necklaces, mss sequences, and dna sequences. *Advances in applied mathematics*, 18(1):18–32, 1997.

[10] W. J. Conover. *Practical nonparametric statistics*, volume 350. john wiley & sons, 1999.

[11] 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*, pages 1472–1481, 2017.

[12] A. Fages, C. Fleury, and T. Tsandilas. Understanding multi-view collaboration between augmented reality and remote desktop users. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–27, 2022.

[13] J. Faigl, T. Krajník, J. Chudoba, L. Přeučil, and M. Saska. Low-cost embedded system for relative localization in robotic swarms. In *2013 IEEE International Conference on Robotics and Automation*, pages 993–998. IEEE, 2013.

[14] M. Fiala. Artag, an improved marker system based on artoolkit. *National Research Council Canada, Publication Number: NRC*, 47419:2004, 2004.

[15] 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*, 47(6):2280–2292, 2014.

[16] C. Getschmann and F. Echtler. Seedmarkers: Embeddable markers for physical objects. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 1–11, 2021.

[17] L. Halodová, E. Dvořáková, F. Majer, J. Ulrich, T. Vintr, K. Kusumam, and T. Krajník. Adaptive image processing methods for outdoor autonomous vehicles. In *Modelling and Simulation for Autonomous Systems: 5th International Conference, MESAS 2018, Prague, Czech Republic, October 17–19, 2018, Revised Selected papers*, pages 456–476. Springer, 2019.

[18] N. Hawes, C. Burbridge, F. Jovan, L. Kunze, B. Lacerda, L. Mudrova, J. Young, J. Wyatt, D. Hebesberger, T. Kortner, et al. The strands project: Long-term autonomy in everyday environments. *IEEE Robotics & Automation Magazine*, 24(3):146–156, 2017.

[19] D. Hebesberger, T. Koertner, C. Gisinger, and J. Pripfl. A long-term autonomous robot at a care

hospital: A mixed methods study on social acceptance and experiences of staff and older adults. *International Journal of Social Robotics*, 9(3):417–429, 2017.

[20] D. Q. Huynh. Metrics for 3d rotations: Comparison and analysis. *Journal of Mathematical Imaging and Vision*, 35(2):155–164, 2009.

[21] P. Irmisch. Camera-based distance estimation for autonomous vehicles. Master's thesis, Technische Universität Berlin, 2017.

[22] H. Kato and M. Billinghurst. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*, pages 85–94. IEEE, 1999.

[23] 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*, 76(3-4):539–562, 2014.

[24] T. Krajník, T. Vintr, G. Broughton, F. Majer, T. Rouček, J. Ulrich, J. Blaha, V. Pěčonková, and M. Rektoris. Chronorobotics: Representing the structure of time for service robots. In *Proceedings of the 2020 4th International Symposium on Computer Science and Intelligent Control*, pages 1–8, 2020.

[25] M. Krogius, A. Haggenmiller, and E. Olson. Flexible layouts for fiducial tags. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1898–1903. IEEE, 2019.

[26] W. Kwon, J. H. Park, M. Lee, J. Her, S.-H. Kim, and J.-W. Seo. Robust autonomous navigation of unmanned aerial vehicles (uavs) for warehouses' inventory application. *IEEE Robotics and Automation Letters*, 5(1):243–249, 2019.

[27] S. Kyian and R. Teather. Selection performance using a smartphone in vr with redirected input. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*, pages 1–12, 2021.

[28] P. Lightbody, P. Baxter, and M. Hanheide. Studying table-top manipulation tasks: a robust framework for object tracking in collaboration. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 177–178, 2018.

[29] P. Lightbody, T. Krajník, and M. Hanheide. An efficient visual fiducial localisation system. *ACM SIGAPP Applied Computing Review*, 17(3):28–37, 2017.

[30] P. Lightbody, T. Krajník, and M. Hanheide. A versatile high-performance visual fiducial marker detection system with scalable identity encoding. In *Proceedings of the Symposium on Applied Computing*, pages 276–282, 2017.

[31] X. Liu, L. Wang, J. Xiong, C. Lin, X. Gao, J. Li, and Y. Wang. Uqrcom: Underwater wireless communication based on qr code. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(4):1–22, 2023.

[32] R. Munoz-Salinas, M. J. Marin-Jimenez, and R. Medina-Carnicer. Spm-slam: Simultaneous localization and mapping with squared planar markers. *Pattern Recognition*, 86:156–171, 2019.

[33] R. Muñoz-Salinas and R. Medina-Carnicer. Ucoslam: Simultaneous localization and mapping by fusion of keypoints and squared planar markers. *Pattern Recognition*, 101:107193, 2020.

[34] S. Na, Y. Qiu, A. E. Turgut, J. Ulrich, T. Krajník, S. Yue, B. Lennox, and F. Arvin. Bio-inspired artificial pheromone system for swarm robotics applications. *Adaptive Behavior*, page 1059712320918936, 2020.

[35] S. Na, M. Raoufi, A. E. Turgut, T. Krajník, and F. Arvin. Extended artificial pheromone system for swarm robotic applications. In *Artificial life conference proceedings*, pages 608–615. MIT Press, 2019.

[36] M. Nitsche, T. Pire, T. Krajník, M. Kulich, and M. Mejail. Monte carlo localization for teach-and-repeat feature-based navigation. In *Conference Towards Autonomous Robotic Systems*, pages 13–24. Springer, 2014.

[37] E. Olson. Apriltag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*, pages 3400–3407. IEEE, 2011.

[38] M. Petrlík, T. Báča, D. Heřt, M. Vrba, T. Krajník, and M. Saska. A robust uav system for operations in a constrained environment. *IEEE Robotics and Automation Letters*, 5(2):2169–2176, 2020.

[39] A. Poibrenski, M. Klusch, I. Vozniak, and C. Müller. Multimodal multi-pedestrian path prediction for autonomous cars. *ACM SIGAPP Applied Computing Review*, 20(4):5–17, 2021.

[40] F. Rekabi-Bana, M. Stefanec, J. Ulrich, E. E. Keyvan, T. Rouček, G. Broughton, G. B. Y., Ö. Şahin, A. E. Turgut, E. Şahin, T. Krajník, T. Schmickl, and F. Arvin. Mechatronic design for multi robots-insect swarms interactions. In *IEEE International Conference on Mechatronics - ICM*, 2023.

[41] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer. Speeded up detection of squared fiducial markers. *Image and vision Computing*, 76:38–47, 2018.

[42] M. Saska. Mav-swarms: unmanned aerial vehicles stabilized along a given path using onboard relative localization. In *2015 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 894–903. IEEE, 2015.

[43] M. Saska, J. Chudoba, L. Přeučil, J. Thomas, G. Loianno, A. Třešňák, V. Vonásek, and V. Kumar. Autonomous deployment of swarms of micro-aerial vehicles in cooperative surveillance. In *2014 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 584–595. IEEE, 2014.

[44] Z. Šidák. Rectangular confidence regions for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62(318):626–633, 1967.

[45] J. R. Souza, D. O. Sales, P. Y. Shinzato, F. S. Osório, and D. F. Wolf. Template-based autonomous navigation and obstacle avoidance in urban environments. *ACM SIGAPP Applied Computing Review*, 11(4):49–59, 2011.

[46] J. Springer and M. Kyas. Autonomous drone landing:

Marked landing pads and solidified lava flows. *arXiv preprint arXiv:2302.00786*, 2023.

[47] B. Steuerlein and S. Mayer. Conductive fiducial tangibles for everyone: A data simulation-based toolkit using deep learning. *Proceedings of the ACM on Human-Computer Interaction*, 6(MHCI):1–22, 2022.

[48] M. Terpilowski. scikit-posthocs: Pairwise multiple comparison tests in python. *The Journal of Open Source Software*, 4(36):1169, 2019.

[49] D. Wagner and D. Schmalstieg. Artoolkitplus for pose trackin on mobile devices. In *12th Computer Vision Winter Workshop 07*, pages 139–146. ., 2007.

[50] A. Weinstein, A. Cho, G. Loianno, and V. Kumar. Visual inertial odometry swarm: An autonomous swarm of vision-based quadrotors. *IEEE Robotics and Automation Letters*, 3(3):1801–1807, 2018.

[51] S. Yang, S. A. Scherer, and A. Zell. An onboard monocular vision system for autonomous takeoff, hovering and landing of a micro aerial vehicle. *Journal of Intelligent & Robotic Systems*, 69:499–515, 2013.