Towards Fast Fiducial Marker with full 6 DOF Pose Estimation

Jiří Ulrich
Faculty of Electrical Engineering
Czech Technical University in Prague
Czech Republic
ulricji1@fel.cvut.cz

Ahmad Alsayed
Swarm & Computational Intelligence Laboratory
The University of Manchester
United Kingdom
ahmad.alsayed@manchester.ac.uk

Farshad Arvin
Swarm & Computational Intelligence Laboratory
The University of Manchester
United Kingdom
farshad.arvin@manchester.ac.uk

Tomáš Krajník
Faculty of Electrical Engineering
Czech Technical University in Prague
Czech Republic
krajnt1@fel.cvut.cz

ABSTRACT

This paper proposes a new method for the full 6 degrees of freedom pose estimation of a circular fiducial marker. This circular black-and-white planar marker provides a unique and versatile identification of individual markers while maintaining a real-time detection. Such a marker and the vision localisation system based on it is suitable for both external and self-localisation. Together with an off-the-shelf camera, the marker aims to provide a sufficient pose estimation accuracy to substitute the current high-end localisation systems. In order to assess the performance of our proposed marker system, we evaluate its capabilities against the current state-of-the-art methods in terms of their ability to estimate the 2D and 3D positions. For such purpose, a real-world dataset, inspired by typical applications in mobile and swarm robotics, was collected as the performance under the real conditions provides better insights into the method's potential than an artificially simulated environment. The experiments performed show that the method presented here achieved three times the accuracy of the marker it was derived from.

CCS CONCEPTS

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

KEYWORDS

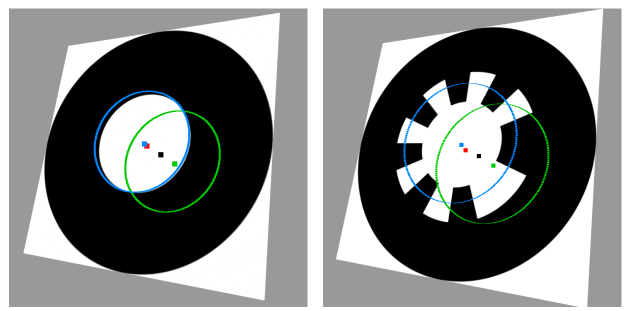
Fiducial Markers; Swarm Robotics; Visual Tracking

ACM Reference Format:
Jiří Ulrich, Ahmad Alsayed, Farshad Arvin, and Tomáš Krajník. 2022. Towards Fast Fiducial Marker with full 6 DOF Pose Estimation. In *The 37th ACM/SIGAPP Symposium on Applied Computing (SAC '22), April 25–29, 2022, Virtual Event,* . ACM, New York, NY, USA, Article 4, 8 pages. https://doi.org/10.1145/3477314.3507043

SAC '22, April 25–29, 2022, Virtual Event,

ACM ISBN 978-1-4503-8713-2/22/04…$15.00
https://doi.org/10.1145/3477314.3507043

1 INTRODUCTION



(a) WhyCon **(b) WhyCode**

Figure 1: Fiducial marker pose estimation ambiguity. In blue and greeen, there are two possible position candidates. The red and black dots are the perspective centers of the white inner and overall segments respectively.

The field of robotics has undergone a significant expansion, and nowadays the robots can be found in almost every area of our lives. The broad range of possible robotic deployments would not be possible without the neverending development of the hardware, which is smaller and more versatile every few years. With changes in the robotic platforms, the necessary control procedures have to adjust accordingly to whether we operate a heavy-machinery soldering robot or deploy a service robot interacting with people and its surroundings. Even though the overall dream is to achieve full autonomy and self-sustainable control, the industry is far from it. The robots are still not easily accepted by society [15], and when a disturbance in their behaviour occurs, they are considered non-functional. Therefore, their activities are usually restricted to a safe and consistent environment.

One of the most important abilities of mobile robots is to find and follow a path to achieve a given goal. Thus, navigation and localisation within an environment is essential to completing such tasks correctly. Wrong estimation of the robot's state might mislead the navigation and cause improper assumptions about the surroundings resulting in even damaging itself or others. In order to avoid undefined behaviour, the robots have to rely extensively on the accuracy of the inputs from sensors.

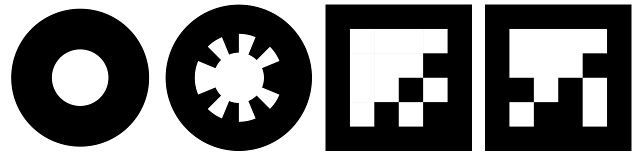
The robots are usually equipped with vison sensors measuring various parts of the light spectrum providing projections of the space to observe the environment. The most common sensor is an RGB camera, which became more popular for labelling, object detection, and classification, especially with the rise of convolutional neural networks. However, one does not have to always turn to a higher abstraction level of the image frame because plenty of machine vision techniques have been developed since the beginning of digital image capture, projective geometry, and photogrammetry. In the image scene, one can equally detect important groups of pixels, patterns, landmarks, and shapes and calculate the mutual relations within each other or the camera. To ease the process of visual localisation, one can enrich the scene by simple artificial unique patterns, *fiducial markers*, which have predetermined characteristics allowing simple detection and pose estimation. Hand in hand with the spread of robotics, the applications of the fiducials evolve mainly originating as support for augmented reality, they became a reliable source of position information used in multi-robot and swarm robotics [29, 30, 32]. They can be even used for the evaluation of robotic experiments where they can provide the ground truth [1, 3, 5, 26]. The main advantage of the fiducials is undoubtedly their extremely low cost and the need for only a calibrated camera.

The markers are not all-saving because everything based on measurements and sensors is imperfect. There are many aspects which can decrease the performance. The maximum resolution restricts the information from a camera, and the more challenging the light conditions are, the more noise and possible image artefacts occur. One must also consider the materials used to produce the markers because the reflection and the colour rendering might cause false segmentation or classification. Therefore, it is necessary to study how the methods work and make them precise and robust to become a reliable and trustworthy source for robot localisation.

In this paper, we present an enhanced marker localisation system capable of the full 6 DOF pose estimation based on the original circular black-and-white fiducial marker WhyCon and WhyCode, which feature real-time performance, detection robustness and effective encoding of unique identification [18, 20, 21]. Our proposed marker extends those systems with previously unaddressed solving of ambiguity in the pose estimation, see Figure 1. We collected a real-world robotic dataset representing a typical application of fiducial markers in swarm robotics. Then, we evaluated our new method together with the most popular state-of-the-art methods on the dataset to show it achieves comparable accuracy while keeping previously declared computational effectiveness.

2 RELATED WORK

The observable environment can be enhanced by artificial objects, *fiducial markers*, which have such properties that they can be used as reference points to support robotic localisation and navigation. Those artefacts can be relied on and incorporated in demanding scenarios of ego-localisation and external localisation. Even though the fiducials can be of any shape and colour, the most popular are passive planar markers which have apriori known geometric characteristics and simple colours, see Figure 2. Therefore, various computer vision methods can detect the markers and even estimate their position and orientation in space. The planarity increases the complexity for pose estimation; however, it widely broadens the possible applications as the markers can be placed on basically anything. Applications supporting robot motion or localisation can be found in [14] while the deployment in autonomous vehicles is thoroughly examined in [16].



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

Figure 2: The most popular fiducial markers currently used in the mobile and swarm robotics applications.

AprilTag is a black-and-white square marker with a 2D binary code inside. The marker is able to estimate the full 6 DOF and unique identification [27]. The detection phase starts with binarising the image and extracting all lines; then, the lines are filtered, so only such enclosing and forming a four-corner object remains. This initial phase produces a high number of false positives and a low amount of false negatives, which might be beneficial in not leaving any marker behind. Then, the line batches are tested to satisfy allowed pattern characteristics involving the maximum corner angle or enclosed area size. The inside of the marker is later sampled by a grid corresponding to the chosen coding size of the identification matrix. The ID encoding is based on lexicographical binary code, which offers variable adjustment of false-positive rate together with the scalable amount of patterns. Markers can be detected and identified even when slightly covered, but they have to be generated with an appropriately robust coding size. AprilTag has been extended [19] to support more shapes and even allow custom content within the marker that is not related to detection nor identification.

ArUco represents another fiducial based on the square shape, which stores the identification information within a binary matrix [13]. The method processes the image similar to AprilTag, and so it starts with binarising the image where it searches marker contours with an edge detector. Then, it proceeds with the identification matrix extraction for marker identification. Contrary to AprilTag, the marker does not interpolate the found lines as much to form enclosed shapes; therefore, ArUco is less resistant to occlusion. The suggested approach uses a marker board occupied with many redundant markers to provide an alternative foundation for the pose estimation instead of the occluded markers. More recently introduced improvements in ArUco [28] increases the detection speed by considering temporal information about marker position and size; thus, faster line segmentation can be applied. However, the tracking is only beneficial in a continuous video stream, as, in the case of individual image evaluation, there would not be the possibility to reuse the previous parameters. The unstable detection time of the square-based markers is one of the deficiencies described in [16] because it might hold back the whole dependent processing pipeline.

WhyCon is a real-time circular fiducial marker consisting of two concentric circles of known parameters [18]. The outer circle is black, and the inner one is white. They share a common centre, and their diameters are known in advance. The method can estimate the marker's 5 DOF as the missing degree of freedom is the rotation around the surface normal. Such capability of pose estimation is especially useful in applications where only the 3D position is required and not the orientation. WhyCon can perform detection and tracking of hundreds of markers within tens of milliseconds while achieving accuracy in the order of a couple millimetres. The core of the detection is based on on-demand thresholding and local flood fill segmentation. Within a series of segment verification tests, the pattern characteristics are calculated on the fly and passed to construct the conic section, which is used to pose estimation. The marker excels in the computational efficiency and the low amount of required memory. The solution of the lacking ability to determine the complete orientation or identification has been proposed by derived markers, WhyCode, and SwarmCon.

WhyCode marker is a descendant of the WhyCon system, which relies on the same core algorithm while extending the system with versatile identification [20, 21]. It shares the same thresholding, segmentation, and projection model; thus, it achieves comparable performance. WhyCode introduces uniquely identifiable encoding, which helps with the missing angle estimation as in WhyCon. The used encoding allows scalable and variable unique marker generation with a binary code between the white and black segments of the overall pattern. The binary code represents mathematical Necklaces that are invariant to rotation; therefore, they can be rotated until reaching the minimum. The rotation required is equivalent to the sought rotation around the marker's surface normal vector. The identification phase affects the evaluation time insignificantly, and so, the fiducial remains as computationally efficient as the original marker while it adds the identification and 3D orientation estimation. Nevertheless, the marker is restricted to shorter distances and smaller angles because the binary code has to be appropriately visible to identify.

SwarmCon represents a different evolution branch of the WhyCon marker, which instead of introducing new features, only modifies the already existing ones [3]. Therefore, it again relies on the internals and performance of WhyCon, but enriches the marker with a distinguishable identification system and planar orientation. The significant difference is the change from concentric circles to centre-offset ellipses. The offset and individual shape of the ellipses matter because based on the mutual size and (co-)vertices of both ellipses, the ID is encoded. Also, considering the slight offset of the centres, the base orientation can be established and later on estimate the orientation as the rotation angle of the (co-)vertices. The main disadvantage is that it can be deployed only in planar environments because pose estimation of an ellipse at a general pose leads to many ambiguous solutions. SwarmCon offers an application-dependent function of creating and transforming the results into a custom swarm arena frame.

Many other planar black-and-white fiducial markers have been presented with various capabilities; however, their area of use has been oriented towards augmented reality. They were introduced in the early beginning of the camera vision techniques for localisation and tracking. The foundational and influential square passive marker was ARToolKit [17], focused on the AR tracking, which was mainly used for the possibility to uniquely identify the markers with only a little focus on the actual quality of the pose estimation. The later evolved markers ARTag [12] and ARToolKit+ [31] further improved the performance but were still more oriented towards augmented reality than fulfilling the needs of the robotics community in the estimation accuracy and computational requirements. However, with the increase in computational power of robotic platforms and camera resolution, the fiducial markers evolved into highly complex shapes and colours capable of representing plenty of information bits [6–8, 10, 19].

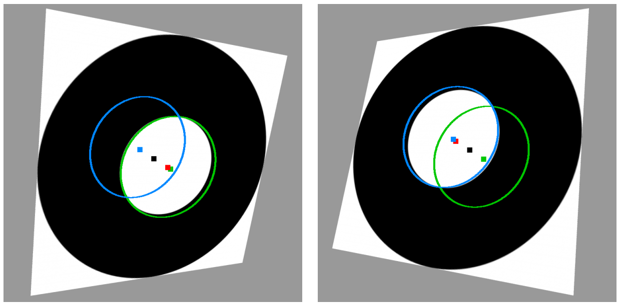
Recently introduced methods extend the incorporation of the fiducials even more as they combine the natural and artificial landmark localisation systems to overcome the volatility and repetitions of natural landmarks, which are often experienced by most of the visual SLAMs [22, 23]. The authors of UcoSLAM and SPM-SLAM use a square-based marker as a reliable artificial landmark to support the localisation and map building in repetitive environments. The systems can use only the marker, keypoint features, or a combination of them; thus, it maintains the versatility of the traditional SLAM approaches. The marker supports the determination of the map scale as its dimensions are a priory known. In terms of the features, the methods are designed to incorporate any type of image descriptors, which provide enough reliable information.

3 FIDUCIAL ESTIMATION METHOD

In this section, we introduce the enhanced fiducial marker localisation system based on the WhyCon and WhyCode fiducial markers, which are capable of online tracking, robustness to motion blur and changing light condition, and real-time detection performance. The markers have been deployed in wide range of scenarios involving multi-robot formation synchronization, human-to-robot interation, and low-cost ground-truth of robotic experimets [3, 24, 26, 32].

The whole WhyCon family of related markers share the same detection core as presented in detail in [18], meaning close to milimetre accuracy and the ability to process many markers in real-time. WhyCon and the further evolved WhyCode fiducials are open-source vision-based localisation systems with high accuracy which needs only a simple web camera to detect and localise up to thousands of markers per image frame [11, 21]. The fiducials benefit from computational efficiency compared to Apriltag or ArUco, which allows them to be use even in power-restricted robotic platforms [21]. Flood-fill segmentation, cascaded tests with increasing complexity and on-demand thresholding results in fast detection of the circular pattern. The simple segments are tested for complying with the sought geometric properties and they are rejected the first time they fail. The properties are recorded for each marker and reused for later searches in a new frame as the marker is expected not to displace in the image drastically; and thus, it allows fast tracking while reducing the amount of segments and parameters to be processed.

The authors in [21] added to the previous system the versatile binary encoding to form uniquely indetifiable markers. The encoding is on top it also invariant to rotation; and so, it is possible to recover the rotation around the surface normal. The place for the newly used marker ID is in between the black and white circles where



(a) $\theta = [0, -0.3, 0.5]\mathrm{rad}$ **(b)** $\theta = [0, 0.3, -0.5]\mathrm{rad}$

Figure 3: WhyCon pose ambiguity. In blue and greeen, there are two possible position candidates. Both markers are at exact opposite orientation.

we can find the circular Manchester-encoded binary Necklaces [9], resembling *teeth*. Thanks to the properties of the Necklace code, we can read the circular binary code from any starting point; and then only rotate to the right to settle with on smallest value. The amount of rotations neccessary estimates the rotation around the surface normal.

Due to the nature of the encountered marker applications, there was no necessity to fully deal with the 6 DOF pose estimation. The typical required usage was to estimate the planar 2D position in case of swarms, obtain the tilt of the marker's normal, or identify individual markers. However, as there was no need for 3D orientation, there is an unaddressed instability in the pose estimation. Tracking down the source of the instability, one notices the unsolved process of choosing the correct pose solution as there are always two position vectors and normal vectors to choose from. Then, an arbitrary pair of vectors was outputed and considered as the correct estimation. It causes only relatively small difference in position precition, but a major error in an orientation. The missing features have been thoroughly addressed by [16].

3.1 Addressing ambiguity of solutions

Once the detected and verified image segment achieves the end of the pattern test cascade, it is passed to the shared pose estimation core. The projective center of the sought marker can be find easily as an arithmetic mean of all the coordinates of the segment pixels. Then, we construct the covariance matric and decompose it to the eigenvalues and eigenvectors to obtain the ellipse description of the projected circle. The (co-)vertices and center are transformed to the standard canonical camera coordinates, which later allow us to formulate the ellipse as a conic section where all the ellipse points have to satisfy $\mathbf{X}^T\mathbf{Q}\mathbf{X} = 0$, where $\mathbf{Q}$ is the conic section of the segment border and $\mathbf{X} = (u'v'1)^T$ is an ellipse point in the canonical form and homogeneous coordinates.

The conic is then decomposed to calculate the surface normal $\mathbf{n}$ and position of the marker $\mathbf{x}$ as thoroughly explained [21, 33]

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

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

where r is the outer circle radius, eigenvalue analysis of $\mathbf{Q}$ yeilds eigenvalues $\lambda_0 \geq \lambda_1 > 0 > \lambda_2$ and their respective normalised eigenvectors $\mathbf{q_0}, \mathbf{q_1}, \mathbf{q_2}$, and s_1, s_2, s_3 are unknown signs.

We can then propose restrictions on the choice of the undetermined signes because the fiducial is detected if and only if it is placed in front of the camera. Similarly, the surface normal vector have to face towards the camera. So, the constrains can be expressed as below

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

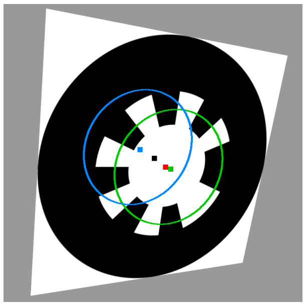
$$\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \mathbf{x} > 0 \tag{4}$$

The inequalities reduces the uncertainty to only one unknown sign which yields two possible solutions $\mathbf{x_1}, \mathbf{x_2}$ and $\mathbf{n_1}, \mathbf{n_2}$. We do not consider the special case when there is only one solution in case of $\lambda_0 = \lambda_1$. Here comes the unsolved ambiguity as only the restriction on the position is applied, which actually results in even more arbitrary normal vectors. Therefore, one has to use look for supporting information to choose the proper solution as in [33], where the authors used an auxiliary sensor. However, we can find the neccessary missing information within the markers, but as WhyCon and WhyCode have quite different structer, each marker requires an individual approach as described below.

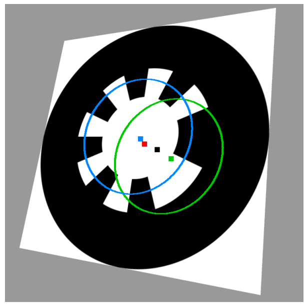
3.1.1 WhyCon's solution. WhyCon marker are just two concentric black and white circles, and the concentricity is the actual additional information we can benefit from. Currently, during the cascade of characteristics tests, we segment also the inner white area and together with the black ellipse the parameters are calculated. Then, only the overall undiferentiated segment is passed to the pose estimation procedure while the inner circle is examined just for verification of the elliptical nature of the segment like the area ratio of the segments, or the radii ratio. All the already computed descriptions of the white circle is then not used at all.

The white segment is the key to the resolution because suddenly there are two tightly related sources of the ellipse projection and perspective. One can observe that the perspective has less influence on smaller objects than on larger. So, we can assume the perspective center of the white circle will not shift away from the projected center of the black's one. Together with the cocentricity of the two circles, we can understand the mentioned white center as an approximation of the real projected common center. The ambiguity decition criterion is then the distance between the individual solution candidates to the white center. We at first transform the position vector from the camera coordinates back to the image coordinates which gives us two points equally separated from the overall center. Then, the sought correct vector is the closer one to the white center.

In Figure 3, we can examine the ambiguity decision introduced above. Both figures are related throught the third undetermined



(a) $\theta = [0, -0.3, 0.5]$rad



(b) $\theta = [0, 0.3, -0.5]$rad

Figure 4: WhyCode pose ambiguity. In blue and greeen, there are two possible position candidates. Both markers are at exact opposite orientation.

sign mentioned in the equations for the calculation of position and normal vectors. The markers poses differ only in the orientation; and thus, we can decide on the correct solution base on the white inner circle.

3.1.2 WhyCode's solution. WhyCode is structurally different which makes the ambiguity decision more difficult. Also, the choice of the correct solution is tightly connected to the correct identification. The white inner area cannot be exploited the same because of the binary encoding in between the circles which results in an uneven and asymmetrical segments with different centers compared to a simple circle. Thus, it can be only used in the parameter validaion during the detection phase. The effect is especially noticeable when a low number of bits is encoded because the white teeth are larger and shift the averaged center more to towards them. However, when the number of encoded bits is high enough or the marker is observed under particular angle, the shifting effect of the dominant white teeth is weaker. The strightforward attempt of the solution would be to only use more encoding bits, but such approach would decrease the current detection range and robustness to camera imperfections.

The circular code can be analyzed for the ambiguity decision because the detection pipeline has to process it anyway. The extraction of the binary code starts with sampling around the marker perspective center obtaining a brightness signal. The authors of WhyCode firstly smoothed out the gradient to equalise the signal; and then, the position with the highest brightness is found and considered as the approximated center of a white tooth. Then, the signal is resampled and thresholded with tooth-wide step beginning from the found point to obtain the binary code. This process is actually one of the identification flaws as the strongest point can be found almost anywhere along the signal depending on the illumination. Therefore, the resulting binary code can be misaligned with the original signal, so completely wrong.

The described drawback has a solution in a different analysis of the brightness signal. The signal can be binarised right after sampling it with the threshold of the overall average value. Teeth edges are then found easily and converted into the Manchester-encoded Necklace code. This change in decoding decreases the ability to handle significant light changes; however, it highly increases the identification reliability.

Once the decoding procedure is improved, we can used it in ambiguity decition phase. Similarly to WhyCon approach, both solution pairs are estimated and then reprojected into the original pixel frame. Then, we continue with identification of the circular code as we have just described above; thus, we will have two binary brightness signals. We assume there will be a noticeable deformation of teeth within the signal connected to the wrong overall projected center; because, if the signal had been measured based on the correct center, the black and white teeth would have the same or doubled size. Based on the above, we can calculate the variance of the teeth position on the sampling rig. The solution tight to the projected center with lower variance can be assumed to be the correct pose estimation. However, this decision approach complicates the detection process by doubling the identification step and calculating the variance statistics. On the other hand, the real-time performance has not been changed insignificantly.

Visualizing the ambiguity pair, see Figure 4, we can immedeately judge which solution is the correct one; even though, the markers face oposite directions. Both sampling circles are drawn to demonstrate the dependency of the identification on choosing proper solution. The described decision based on binarized brightness signal of marker in Figure 4b is presented in Figure 5

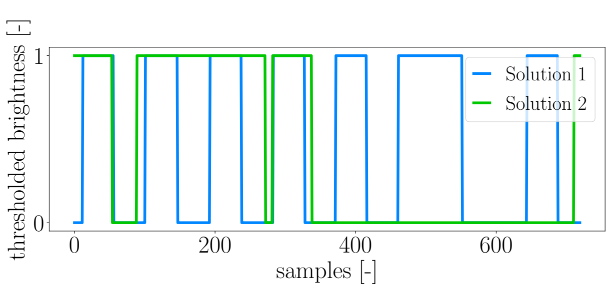


Figure 5: Tresholded brightness signals obtained when sampling the marker in Figure 4. The green signal shows the shifted and inaccurate sampling while the blue one is the correct binary signal.

4 DATASETS

To compare the newly presented fiducial marker method with the state-of-the-art fiducials, they were evaluated on a real-world dataset. All the markers achieve up to a few millimetres accuracy, and some of them were primarily designed to provide ground truth for mobile robot experiments. The methods were evaluated on a real-world dataset to demonstrate their actual performance under naturally occurring noise and imperfect image capturing process.

The real-world dataset represents the external localisation application of the markers in swarm robotics, providing a low-cost ground-truth system. This application of the fiducials is popular [4, 24, 25] as the swarms can then operate in various conditions without the need for setting up complex localisation systems of multiple cameras and synchronisation units. The critical aspect is the usage of infrared light emitters and receivers within the swarm community as an essential means of communication among the

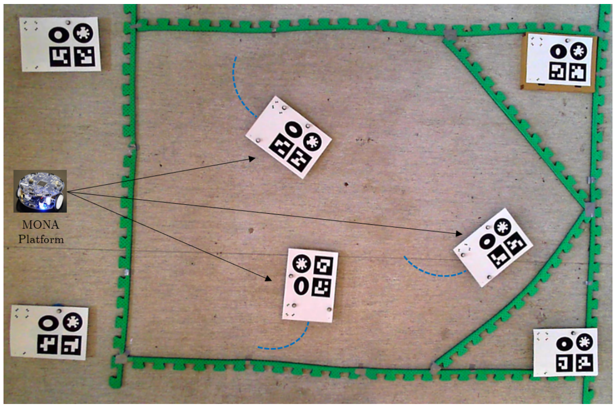


Figure 6: Image frame from the real-world dataset displaying the three robots with marker boards on top and another four marker boards in the corners.

nearby robots. However, the high-end motion capture systems interfere with those sensors and restrict them as the systems are primarily based on markers reflecting or emitting infrared light [3]. Therefore, in many applications, it is favourable to deploy those passive visual fiducials.

The swarm arena, see Figure 6, in the dataset is of rectangular proportions with sides of two and three meters. It is captured by an off-the-shelf USB camera with low lens distortion and it is positioned approximately three meters above the arena. The video stream has a resolution of 1920×1080 pixels and a frame rate of approximately 30 frames per second with occasional frame dropping while recording the stream. In order to obtain the ground-truth measurements, we decided to use the Vicon localisation and motion capture system, providing a submillimeter resolution.

The recorded data consists of pseudorandomised robot movements. There are three robots moving to cover most of the arena, and there are four stationary robots in each corner, providing a reference frame for individual fiducial markers. The robots are equipped with a board on top with fiducials, AprilTag, ArUco, and our proposed marker. All the markers have the same size of 8.4 cm and have been generated with their reasonably small sets of IDs. Thus, ArUco binary matrix IDs are from the 4×4 dictionary, WhyCode markers have 8-bit encoding, and AprilTag identifies markers with a $16h5$ ID set. The used swarm robotic platform is MONA [2] as it is a small open-source, open-hardware, and affordable platform for research and education. It is based on the microcontroller ATmega 328, which is the same as in Arduino Mini/Pro; thus, there is a large compatibility of libraries and sensors.

The total dataset length is a little over four minutes consisting of 7297 image frames and 60099 Vicon measurements. The Vicon system, which measures the ground truth, provides measurements at a higher frequency, approximately eight times more than the camera frame rate. High frequency data is a feature shared among high precision systems; therefore, we first had to match the ground truth with the frames. To synchronise the high-frequency data with the video, we used the beginning and the end of the movements of the robots as key points which are straightforward to find in both data streams. By aligning the beginning and the end of the data, we iterated over the video frames and assigned them the closest Vicon measurements. In order to make the data comparable, we also have to reduce the number of samples, so there will be corresponding pairs for each video and Vicon frame resulting in 5716 frames forming approximately three minutes.

5 EXPERIMENTS

To assess the newly presented fiducial marker performace, we evaluate the position estimation accuracy on the real-world dataset. Under real-world conditions, most of cameras and any sensors suffer from noise and a particular deformation of the sensed data. Therefore, we thoroughly calibrated the camera to obtain the intrinsic camera matrix and also the rectification coefficients for the plumb bob distortion camera model. The calibration was estimated using the widely known OpenCV calibration tools. In the dataset, there also occur another common phenomenon in the image processing, the motion blur, which is caused by fast moving and spinning robots. Thus, tested methods are evaluated on image frames which aim to reflect the real conditions in robotics applications. Although, one could only test the fiducials in an artificial simulator, it would not provide reliably transferable results in practice.

We propose the criteria below to compare the results generated by the methods. There are two main use cases of the fiducials applications, the first one is the general full pose estimation in 3D, and the second is only considering the 2D planar position. The choice is based on real robotic and experiment-evaluation applications. The method's application is an external localisation of moving robots because the fiducials can serve as an alternative to the expensive motion tracking systems. The tests were evaluated on an identical video stream with a known camera matrix and distortion parameters. We compare the resulting marker errors to see if the difference is statistically significant by testing them with the Student's t-test. The confidence level was set to five percent through all evaluations. During all t-test evaluations the p-values were strictly close to zero.

5.1 General 3D pose estimation

This first test aimed to evaluate and present the ability of the fiducials to precisely estimate the full 6 DOF. We measured the position estimation errors and expresed them as a histogram, see Figure 7. In Table 1, we extracted the key accuracy indicators, which describe the error distribution. Evaluating the results, the performance of individual markers is straightforward and they significantly differ from each other. The marker with highest accuracy is ArUco reaching an average error below 17 millimeters. Comparing our marker with AprilTag, we managed to outperform it and achieve a lower error. Based on the error evolution, one could suggest that our error appears to be bimodal and could probably be improved in a future investigation. Nevertheless, our position estimation error is comparable and lower than the state-of-the-art method.

5.2 Planar 2D pose estimation

The following test examines the error of planar pose estimation as it is one of the most popular applications in swarm robotics. The results vary from the previous evaluation because our marker can

	AprilTag	ArUco	Proposed marker		WhyCode
			2D	3D	
Mean	35.53	17.26	18.68	31.43	75.44
Median	35.35	17.45	18.98	32.61	85.22
Std. dev.	11.56	7.66	6.15	12.81	21.96

Table 1: Position estimation error [mm]

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

Table 2: Comparison of the execution time of the compared methods. The speedup is relative to the AprilTag

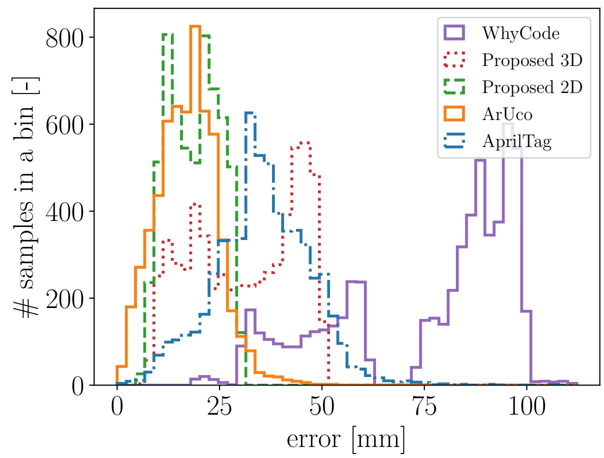


Figure 7: Position error estimation distribution

use and benefit from the information whether it estimates only the planar pose or the general space configuration. The error characteristics are in Table 1 and for the histogram vizualisation see Figure 7. We can immediately conclude that our method provided significantly improved results in the 2D case compared to the 3D. The improvement in the average position error is approximately 18 millimiteres and moves our marker closer to ArUco's performance. Together with the lower mean error, we can notice the error evolution is suddenly closer to unimodal and the error distribution parameters improved, but after detailed evaluation, the error is slightly bimodal. The reason might be still not perfect ambiguity resolution. However, the ordering of the fiducials based on accuracy remains the same, but our marker now became comparable with ArUco instead of AprilTag.

5.3 Execution time

In order to demonstrate that the improvements did not cause a major bottleneck in the performance of our proposed marker, we measured the execution time. We let our marker and the state-of-the-art markers run over the whole dataset, containing 7297 images and measured the time required by the appropriate method entry point. In terms of the hardware used, we evaluated this test on a mobile laptop with Intel Core i7-8550U processor and 16GB of system memory. The resulting performance is summarised in Table 2. Our proposed marker maintains its significant performance based on the measured execution time.

6 CONCLUSION

This paper presented a new fiducial marker localisation method capable of real-time performance, unique identification, and the full 6 DOF estimation. The system is based on the widely used method WhyCode, a black-and-white roundel marker with a binary Necklace-based ID. The original system is extended with steps to resolve previously unaddressed ambiguity in position and orientation estimation; because otherwise, the system was highly unreliable and produced inconsistent results. The extension addresses even the problematic identification of unique markers and the related rotation estimation around the marker's surface normal.

The introduced novelty is the ambiguity decision process when estimating the two possible poses and testing them for the sought pattern. Unfortunately, the same approach could not be used for both foundational fiducials, WhyCon and WhyCode. The new decision criteria allow the marker system to estimate the full 6 DOF because it provides more reliable and stable results compared to an arbitrary solution. In order to verify our assumptions, we recorded and evaluated our method on a real-world dataset representing a typical application in the field of swarm and mobile robotics. The fiducial markers accuracy was tested against the ground-truth generated by high-performance Vicon motion and localisation system.

We compared the capabilities of the methods with the most popular state-of-the-art methods in two different scenarios. Considering the outcome, we showed our fiducial marker achieves a comparable performance while maintaining the aspects of the original markers, mainly the real-time processing efficiency. The results demonstrate the method achieved three times the accuracy of the original WhyCode and it outperformed AprilTag in both test cases, while ArUco sets the next accuracy goal for us.

In future works, one could incorporate a different ellipse segmentation method instead of flood-fill. Thus, the entire image frame would be processed contrary to only tracking given markers. The next aspect to further consider is to improve the identification, so one could choose whether to output even uncertain estimation to maintain the continuity in measurements.

Acknowledgments. The presented work was sponsored by the Czech Science Foundation (GAČR) research project No. 20-27034J and by the European Union's Horizon 2020 research and innovation FET Open programme under grant agreement No.964492 project 'RoboRoyale'.

REFERENCES

[1] Farshad Arvin, Abdolrahman Attar, Ali Emre Turgut, and Shigang Yue. 2015. Power-law distribution of long-term experimental data in swarm robotics. In *International Conference in Swarm Intelligence*. Springer, 551–559.

[2] Farshad Arvin, Jose Espinosa, Benjamin Bird, Andrew West, Simon Watson, and Barry Lennox. 2019. Mona: an affordable open-source mobile robot for education and research. *Journal of Intelligent & Robotic Systems* 94, 3-4 (2019), 761–775.

[3] Farshad Arvin, Tomáš Krajník, Ali Emre Turgut, and Shigang Yue. 2015. COSΦ: artificial pheromone system for robotic swarms research. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*. IEEE, 407–412.

[4] Farshad Arvin, Ali Emre Turgut, Tomáš Krajník, Salar Rahimi, Ilkin Ege Okay, Shigang Yue, Simon Watson, and Barry Lennox. 2018. Φ Clust: Pheromone-Based Aggregation for Robotic Swarms. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 4288–4294.

[5] Abdul Basit, Waqar S Qureshi, Matthew N Dailey, and Tomáš Krajník. 2015. Joint localization of pursuit quadcopters and target using monocular cues. *Journal of Intelligent & Robotic Systems* 78, 3 (2015), 613–630.

[6] Filippo Bergamasco, Andrea Albarelli, Luca Cosmo, Emanuele Rodola, and Andrea Torsello. 2016. An accurate and robust artificial marker based on cyclic codes. *IEEE transactions on pattern analysis and machine intelligence* 38, 12 (2016), 2359–2373.

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

[8] Lilian Calvet, Pierre Gurdjos, Carsten Griwodz, and Simone Gasparini. 2016. Detection and accurate localization of circular fiducials under highly challenging conditions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 562–570.

[9] William YC Chen and James D Louck. 1997. Necklaces, MSS sequences, and DNA sequences. *Advances in applied mathematics* 18, 1 (1997), 18–32.

[10] Joseph DeGol, Timothy Bretl, and Derek Hoiem. 2017. Chromatag: A colored marker and fast detection algorithm. In *Proceedings of the IEEE International Conference on Computer Vision*. 1472–1481.

[11] Jan Faigl, Tomáš Krajník, Jan Chudoba, Libor Přeučil, and Martin Saska. 2013. Low-cost embedded system for relative localization in robotic swarms. In *2013 IEEE International Conference on Robotics and Automation*. IEEE, 993–998.

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

[13] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (2014), 2280–2292.

[14] Nick Hawes, Christopher Burbridge, Ferdian Jovan, Lars Kunze, Bruno Lacerda, Lenka Mudrova, Jay Young, Jeremy Wyatt, Denise Hebesberger, Tobias Kortner, et al. 2017. The strands project: Long-term autonomy in everyday environments. *IEEE Robotics & Automation Magazine* 24, 3 (2017), 146–156.

[15] Denise Hebesberger, Tobias Koertner, Christoph Gisinger, and Jürgen Pripfl. 2017. 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 (2017), 417–429.

[16] Patrick Irmisch. 2017. *Camera-based distance estimation for autonomous vehicles*. Master's thesis. Technische Universität Berlin.

[17] Hirokazu Kato and Mark Billinghurst. 1999. 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)*. IEEE, 85–94.

[18] Tomáš Krajník, Matías Nitsche, Jan Faigl, Petr Vaněk, Martin Saska, Libor Přeučil, Tom Duckett, and Marta Mejail. 2014. A practical multirobot localization system. *Journal of Intelligent & Robotic Systems* 76, 3-4 (2014), 539–562.

[19] Maximilian Krogius, Acshi Haggenmiller, and Edwin Olson. 2019. Flexible layouts for fiducial tags. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 1898–1903.

[20] Peter Lightbody, Tomáš Krajník, and Marc Hanheide. 2017. An efficient visual fiducial localisation system. *ACM SIGAPP Applied Computing Review* 17, 3 (2017), 28–37.

[21] Peter Lightbody, Tomáš Krajník, and Marc Hanheide. 2017. A versatile high-performance visual fiducial marker detection system with scalable identity encoding. In *Proceedings of the Symposium on Applied Computing*. 276–282.

[22] Rafael Munoz-Salinas, Manuel J Marin-Jimenez, and Rafael Medina-Carnicer. 2019. SPM-SLAM: Simultaneous localization and mapping with squared planar markers. *Pattern Recognition* 86 (2019), 156–171.

[23] Rafael Muñoz-Salinas and Rafael Medina-Carnicer. 2020. UcoSLAM: Simultaneous localization and mapping by fusion of keypoints and squared planar markers. *Pattern Recognition* 101 (2020), 107193.

[24] Seongin Na, Yiping Qiu, Ali E Turgut, Jiří Ulrich, Tomáš Krajník, Shigang Yue, Barry Lennox, and Farshad Arvin. 2020. Bio-inspired artificial pheromone system for swarm robotics applications. *Adaptive Behavior* (2020), 1059712320918936.

[25] Seongin Na, Mohsen Raoufi, Ali Emre Turgut, Tomáš Krajník, and Farshad Arvin. 2019. Extended artificial pheromone system for swarm robotic applications. In *Artificial life conference proceedings*. MIT Press, 608–615.

[26] Matías Nitsche, Taihú Pire, Tomáš Krajník, Miroslav Kulich, and Marta Mejail. 2014. Monte carlo localization for teach-and-repeat feature-based navigation. In *Conference Towards Autonomous Robotic Systems*. Springer, 13–24.

[27] Edwin Olson. 2011. AprilTag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*. IEEE, 3400–3407.

[28] Francisco J Romero-Ramirez, Rafael Muñoz-Salinas, and Rafael Medina-Carnicer. 2018. Speeded up detection of squared fiducial markers. *Image and vision Computing* 76 (2018), 38–47.

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

[30] Martin Saska, Jan Chudoba, Libor Přeučil, Justin Thomas, Giuseppe Loianno, Adam Třešňák, Vojtěch Vonásek, and Vijay Kumar. 2014. Autonomous deployment of swarms of micro-aerial vehicles in cooperative surveillance. In *2014 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 584–595.

[31] Daniel Wagner and Dieter Schmalstieg. 2007. Artoolkitplus for pose tracking on mobile devices. (2007).

[32] Aaron Weinstein, Adam Cho, Giuseppe Loianno, and Vijay Kumar. 2018. Visual inertial odometry swarm: An autonomous swarm of vision-based quadrotors. *IEEE Robotics and Automation Letters* 3, 3 (2018), 1801–1807.

[33] Shaowu Yang, Sebastian A Scherer, and Andreas Zell. 2013. An onboard monocular vision system for autonomous takeoff, hovering and landing of a micro aerial vehicle. *Journal of Intelligent & Robotic Systems* 69, 1-4 (2013), 499–515.