Universiteit Leiden
The Netherlands

Opleiding Artificial Intelligence

Improving STag's occlusion resilience for Augmented Reality

Ayush Kandhai

Supervisors:
Erwin M. Bakker & Michael S. Lew

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

20/06/2023

Abstract

Fiducial Markers are certain specific distinguishable patterns/shapes (markers) which can easily be detected, identified and localized in any image/video using a detection algorithm. This can be used as a robust localization framework for various purposes like Augmented Reality (AR) using pose estimation or for a vision and localization system for robots. There are a good amount of fiducial markers with different strengths and weaknesses catered to their purposes. One obvious weakness of most of the state-of-the-art fiducial markers (ARTag, AprilTag, ArUco) is occlusion. With the slightest occlusion, these markers become undetectable. The biggest break-through when it comes to occlusion has been STag. But even that has some occlusion limitations. The objective of this research project is to enhance the occlusion resilience of the STag marker by mitigating its known occlusion limitations. We aim to compare the performance of the improved STag marker, referred to as STag2, against the original STag marker and evaluate its performance in relation to state-of-the-art markers. While significant efforts were made to minimize the performance degradation compared to the original STag marker, our proposal did not achieve consistent improvements in occlusion resilience. This paper presents a detailed analysis of the challenges encountered and provides valuable insights for future research in the field of marker design and occlusion resilience enhancement.

Contents

1 Introduction **1**

2 Related Work **2**
ARToolkit . . . 2
ARTag . . . 2
ARToolkit Plus . . . 3
ArUco . . . 3
AprilTag . . . 3
CCC . . . 4
CCTag . . . 4
RUNE-Tag . . . 4

3 Fundamentals **5**
3.1 Camera Calibration . . . 5
3.2 Pose estimation . . . 7

4 Baseline method: STag **8**
4.1 Encoding . . . 9
4.2 Detection algorithm . . . 10
4.2.1 Candidate detection . . . 10
4.2.2 Candidate validation . . . 12
4.2.3 Homography refinement . . . 13

5 STag2 **14**
5.1 STag2: Marker design . . . 14
5.2 STag2: Detection algorithm . . . 15
5.2.1 Candidate validation changes . . . 15

6 Experimental setup **17**
6.1 Occlusion experiments . . . 21

7 Experimental results **22**
7.1 Baseline results . . . 23
7.1.1 STag2 vs STag . . . 24
7.1.2 STag2 (black square) vs the state-of-the-art . . . 26
7.1.3 Further analysis of STag vs STag2 results . . . 29
7.2 STag2: Occlusion results . . . 31
7.2.1 STag2 (black square): Single corner occlusion . . . 31
7.2.2 STag2 (white rhombus): Half occlusion . . . 34

8 Conclusions and Further Research **37**

References **40**

1 Introduction

Augmented Reality (AR) has become a rapidly growing field in recent years. From games, like Pokemon Go [PKA+17], to GPS navigation [Sha20]. One of the main concepts of Augmented Reality is robust localization of the environment (in 3d space) in a given image or video feed. Fiducial markers are a great way of achieving this.
Fiducial markers are easily identifiable and distinguishable objects (patterns, shapes, colors, etc.) that can be detected and decoded (if needed) in a given image or video sequence using a detection algorithm. Most fiducial markers used in Augmented Reality have a specific encoding to minimize incorrect marker detections. In most cases this encoding is a binary code. This means that there can be a lot of different markers with different ids (binary codes). This collection of possible binary codes is called a library. In some cases a fiducial marker can have multiple libraries (STag [BTA17], AprilTag [Ols11]).
Aside from Augmented Reality, fiducial markers get used mostly in robotics as a vision, navigation and interaction system for robots [YCC+20] [ZPC21]. The most prominent application used in these implementations of fiducial markers is pose estimation.
There are a few main factors that make a good fiducial marker (for pose estimation):

Accuracy is absolutely necessary for pose estimation. If the fiducial marker does not get located accurately, we cannot estimate the pose accurately.

Stability goes in combination with accuracy. The marker detections need to stay accurate. This means that no jittering should occur. Therefore, we need a stable marker detection method.

Speed is an important factor when it comes to Augmented Reality or anything that requires real-time detections and pose estimation. For Augmented Reality, minimizing detection and pose estimation lag is essenial.

Occlusion-resilience is a fairly unexplored characteristic of fiducial marker detections and pose estimation. This makes sense, since it is very hard to detect something that is occluded. However, allowing some degree of occlusion can improve the effectiveness and usability a fiducial marker significantly.

There are very few markers that have occlusion-resilience for real-time detections. The only state-of-the-art occlusion-resilient marker we have as of now is **STag** [BTA17]. And even STag has its limitations when it comes to occlusions. Only one corner of a STag marker can be covered to still get a somewhat reliable detection.
In this paper, we introduce STag2, a fiducial marker system based on STag with several implemented proposals to enhance its occlusion-resilience. We then quantitatively compared our proposed fiducial marker to selected state-of-the art fiducial markers (ArUco [GJMSMCMJ14], AprilTag [Ols11] and RUNE-Tag [BART11]) to asses its general performance. Furthermore, we compare its performance in the presence of occlusion to the original STag.

2 Related Work

Research on fiducial markers has been an active area of investigation for several decades. Over the years, numerous studies have focused on the development and application of fiducial markers, aiming to improve their detection accuracy, robustness to occlusion, computational efficiency, and overall performance. In this section, we review the existing literature and explore the advancements made in the design, detection, and tracking of fiducial markers, highlighting key contributions and identifying current challenges and research gaps. By understanding the evolution of fiducial marker technologies, we can gain valuable insights into the potential avenues for further improvement and innovation in this field.

Figure 1: ARToolkit marker examples.

ARToolkit [KB99a] was the first fiducial marker system, introduced in 1999, for Augmented Reality purposes. Its main purpose was pose estimation, estimating the camera's pose relative to a set of (ARToolkit) markers. Fundamentally, the ARToolkit marker is a black square border, where the inside of the marker can be encoded using different (user-defined) patterns/images. The default encoding of a ARToolkit marker shows the word "Hiro", which refers to the name of its creator Hirokazu Kato. The ARToolkit detection algorithm works by looking for potential marker candidates, i.e., detecting a square border, and matching these against a predefined set of marker templates. If the marker is valid, it then proceeds to decode the pattern inside the image and assigns an ID to it. The performance of the ARToolkit markers depends on the size and complexity of the (user-defined) marker library. We opted not to utilize this marker system in our experiments, since various advancements and improvements were made to this marker system, leading to ARTag.

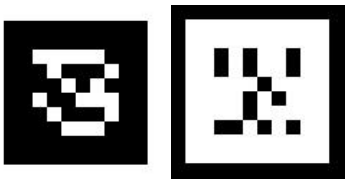


Figure 2: ARTag marker examples.

ARTag [Fia05], inspired by ARToolkit, is the first fiducial marker system to incorporate edge detection and binary encoding into the marker detection process. ARTag, like ARToolkit, also has a border, but the inside is encoded using a binary code of black and white pixels. ARTag can use either a black border or a white border depending on the marker environment. Each of these contains 1001 different binary combinations, amassing a library of 2002 ARTag marker IDs. ARTAg improves the false positive rate of ARToolkit markers, meaning ARTag would have less occurrences of incorrect marker detections. This is mostly due to the encoding method. ARTags are encoded with a 36 bit string of black and white pixels representing 1s and 0s. The first 10 bits are used to decode the marker, while the rest of the bits are for correcting orientation errors. ARTag also has improved identification performance compared to ARToolkit, since there is no need to compare the inside of the marker to the defined marker patterns in the library. We decided to exclude this marker system from our experiments, due to the existence of a similar fiducial marker system, known as , that is more efficient, well-implemented and as a result more popular to use.

ARToolkit Plus [KB99b] is an extension to ARToolkit, inspired by ARTag. ARToolkit Plus mostly works the same as ARToolkit, but takes inspiration from ARTag in marker encoding and decoding (identification) by using a binary encoding instead of a user-defined image/pattern. ARToolkit Plus markers look very similar to ARTag markers. The only differences being the detection algorithm and the encoding process (marker generation). Similarly to ARTag, we decided not to use this marker system in our experiments, due to the existence of ArUco.

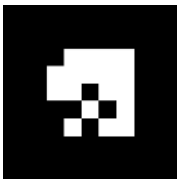


Figure 3: ARToolkit Plus marker example.

ArUco [GJMSMCMJ14] markers are part of OpenCV, which is a popular computer vision library, and therefore ArUco markers are one of the most known and most used fiducial markers to date. Its identification libraries use lexicodes, meaning they are greedily generated with good error correction properties. These get generated using a stochastic lexicode generation algorithm. ArUco is known for its performance, as it uses multiple (smaller) libraries. These libraries are just sections of the complete library of 1024 markers used by ARTag. Like ARTag, the library gets split into smaller libraries based on the size of the binary code (4x4, 5x5, 6x6, 7x7), but unlike ARTag, it also gets split based on the amount of markers needed (50, 100, 250, 1000 or all 1024). The library gets split by only including the specified amount of markers with the highest Hamming distance and is more computationally intensive the higher the number of markers in the library. We used this marker for our performance comparison, because it is the most popular state-of-the-art fiducial marker system.

Figure 4: ArUco marker example.

AprilTag [Ols11] is a lesser known fiducial marker system than ArUco, but is also very popular and more versatile. AprilTag is catered to be exact and efficient in the detection of the position, orientation and identity of a marker. AprilTag looks almost exactly like ArUco/ARTag markers, but can have a wider range of use cases due to having markers with varying bit-counts (from 4 to 36 bits), where having smaller bit counts increases efficiency and having larger bit counts increases the amount of marker IDs (library). AprilTag also excels in offering different kinds of unique marker families for different purposes:

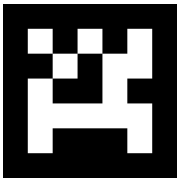


Figure 5: AprilTag marker example.

- Markers that can have more bits outside of the marker border.

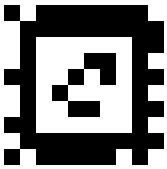


Figure 6: Apriltag Marker with 52 bits.

- Circular markers, that are made to be used in circular areas (where it may be difficult to use square markers) or for aesthetic purposes.

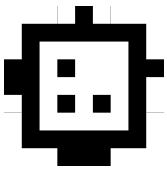


Figure 7: Circular Apriltag Marker.

- Markers with holes in the center, which can inhabit another marker. This can be useful for close-ups, where the marker borders are outside of the camera's viewfinder.

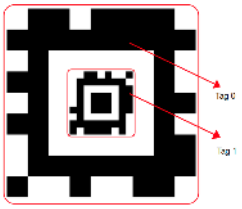


Figure 8: Recursive Apriltag Marker.

We decided to use this marker for our performance comparison, because of its increased detection speed and efficiency as a state-of-the-art marker system compared to ArUco.

CCC [GHS92] (Concentric contrasting circle) is fairly different compared to the previously mentioned fiducial markers. Unlike the previously mentioned fiducial markers, CCC is the most basic *circular* fiducial marker consisting of a black ring with a white background. CCC detects black and white centroids of a grayscale image and compares them to each other. Each black and white centroid pair within a certain threshold of each other can be a CCC. CCC has no encoding and is not complex, which can cause a high amount of false positive detections. We did not use this marker system in our experiments, due to its lack of an ID encoding.

Figure 9: CCC marker example.

CCTag is similar to CCC in the sense that it uses concentric black and white rings. The main purpose of CCTag is reliable detection and identification, especially in the presence unidirectional motion blur. CCTag is also very occlusion-resilient and works remarkably well at long distances. Since CCTag doesn't have a simple shape, it is also less prone to missclassification. Contrary to its appearance, CCTag does not have an encoding, but this can be implemented according to the authors. We decided not to use CCTag in our experiments, as the encoded possibilities are limited compared to RUNE-Tag.

Figure 10: CCTag marker example.

RUNE-Tag [BART11] is a circular fiducial marker encoded using one or multiple rings of black circles representing a code. The radius of the encoding circles is decided by the circular layer it falls on. The purpose of this method of encoding is to ensure that the marker is detected in the correct orientation during the decoding phase of the detection process. RUNE-Tag offers stable pose estimation due to the detailed design of the marker. Since the pose is estimated by using the coding circles as point correspondences, having a good amount of circles in the marker makes for stable pose estimation. However, this detailed design can also cause marker decoding problems in certain unforeseen lighting conditions. We chose this marker for our performance comparison purely based on its occlusion-resilience.

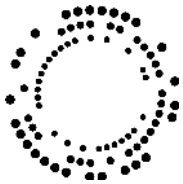


Figure 11: Rune-Tag marker example.

3 Fundamentals

In this section, we will discuss some fundamental topics that are crucial for fully understanding the remainder of this paper. These topics include two key elements of Augmented Reality: pose estimation and camera calibration. We will provide an overview of both of these topics, as well as their importance in the context of Augmented Reality.

3.1 Camera Calibration

One of the first and most obvious uses of fiducial markers is Camera Calibration. Cameras work very similarly to the human eyes. Looking at a simple pinhole camera model:

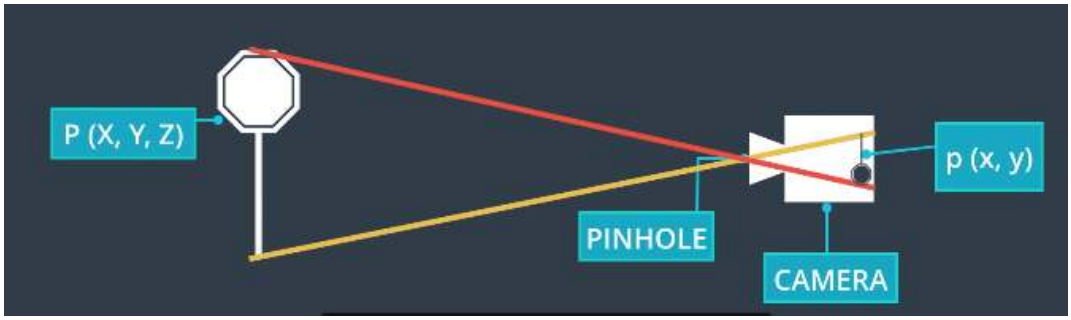


Figure 12: Diagram of the workings of a simple pinhole camera model. [Kum19]

The pinhole of the camera focuses light, reflected off of real-world 3d objects, onto the sensor of the camera forming a 2d image. This transformation from 3d to 2d can mathematically be calculated using a transformative matrix called the **camera matrix**. Using a pinhole to capture light can be a quite lengthy process due to the limited amount of light passed through the pinhole, which is why most cameras use wider aperature lenses. These form images faster, but can create some distortion based on the type of lens used [Kum19]. There are 2 main types of distortions:

1. **Radial distortion** is a type of distortion that distorts the image in such a way that straight lines seem curved.

Figure 13: Radial distortion example. [Kum19]

2. **Tangential distortion** is a type of distortion caused by the lens not being parallel with the image plane, making it difficult to estimate the distance of an object in the image.

Figure 14: Tangential distortion example. [Kum19]

Camera calibration is essential to correct any such distortion that may occur. By using a fiducial marker with a recurring pattern, we can estimate the different distortion coefficients and obtain an estimation of the camera distortion matrix. Additionally, we can derive rotation and translation vectors, which provide an estimation of the camera's position in 3D space. With this information, we can either undistort the image or estimate the object's distance or location in 3D space, which is crucial for many applications, including Augmented Reality [Kum19].
There are several fiducial marker boards that can be used for camera calibration [Brab]:

1. **The chessboard/checkerboard** is the simplest fiducial marker with repeating patterns of corners that can be detected accurately. However, this requires the entire chessboard to be in view, and is prone to misdetections.

2. **The ArUco board** uses ArUco markers for camera calibration as they can be detected and localized quickly and do not require all markers to be visible. However, the accuracy of ArUco markers can be limited by the size and quality of the markers used, as the corners of the markers may not be located as accurately as the chessboard.

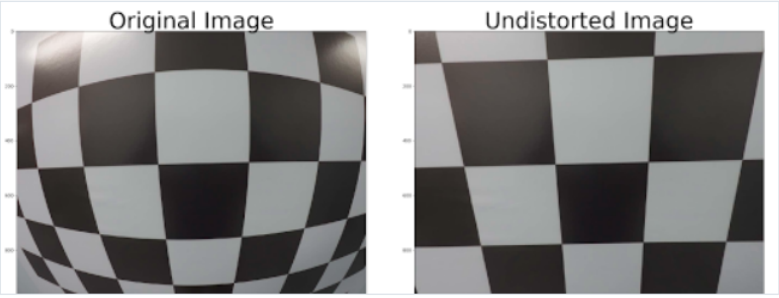


Figure 15: Example of undistorting an image after camera calibration. [Kum19]

3. **The ChArUco board** combines the chessboard with the ArUco board to provide the best of both worlds, resulting in fast marker detections, accurate corner detections and some occlusion resilience.

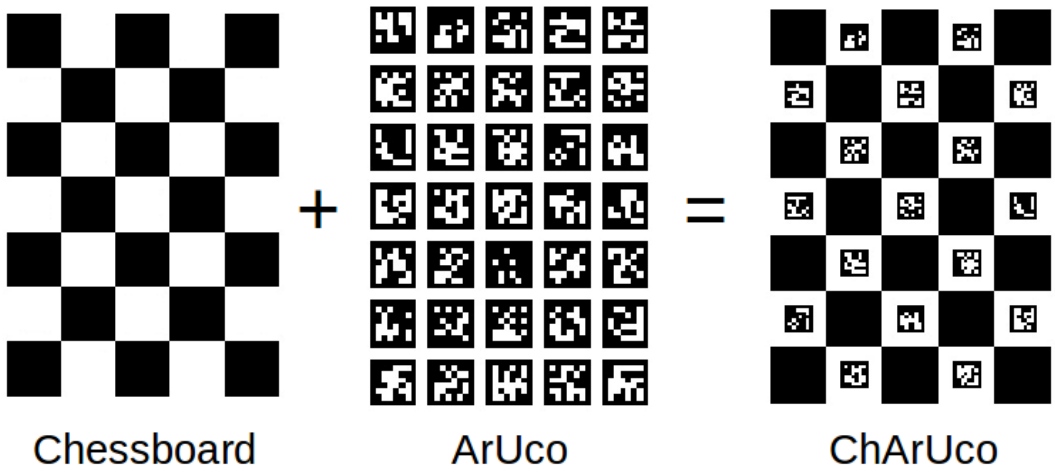


Figure 16: Fiducial markers used for camera calibration. [Kum19]

To calibrate the camera used in our experiments, we utilized the ChArUco board due to its efficiency, accuracy and flexibility compared to the checkerboard and ArUco board.

3.2 Pose estimation

Pose estimation refers to the process of estimating the position and orientation of a fiducial marker in 3D space. This is achieved by determining the six degrees of freedom that describe the marker's pose, which are the three degrees of position (x, y, z) and the three degrees of rotation (roll, pitch, yaw).
The process of estimating the pose of a marker involves using the data obtained from camera calibration and the detection of the corners of the marker in 2D space. The camera matrix, obtained from calibration, is used to map the 2D marker detection coordinates to 3D space. This information,

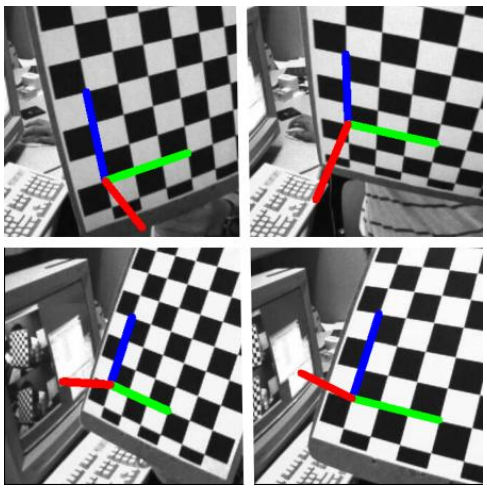


Figure 17: Pose estimation using checkerboard [Brac].

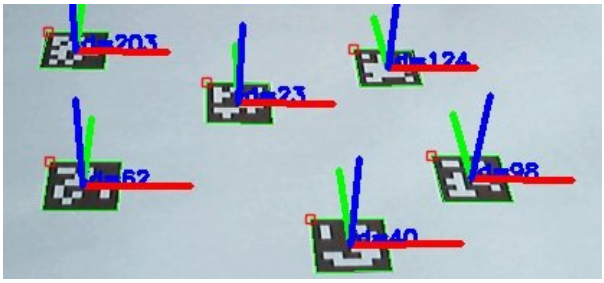


Figure 18: Pose estimation using ArUco marker [Braa].

combined with the distortion coefficients obtained during calibration, allows for accurate estimation of the marker's pose in 3D space [Brac].

4 Baseline method: STag

Before exploring improvement possibilities for STag, it is important to have a comprehensive understanding of how the current STag marker functions. In this section, we will examine the encoding process, as well as the detection process of STag.
STag is a square marker that contains an inner circle with (binary) encoded information. The square border allows for efficient detection, decoding, and estimation of the marker's homography matrix. The homography matrix (H) of a fiducial marker represents the geometric transformation that maps points in the marker's reference frame to corresponding points in the image (see Fig. 19). The inner circle serves to refine this homography estimation and improve the accuracy of the marker's localization.

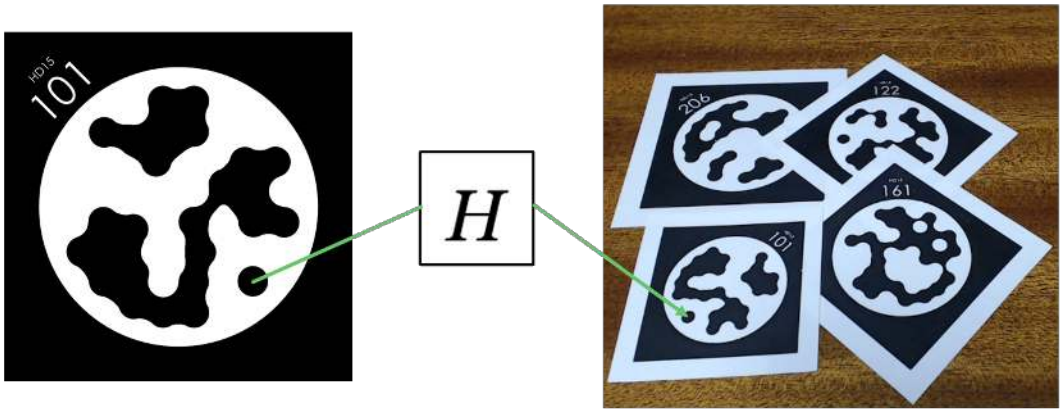


Figure 19: visual representation of workings of homography matrix H.

4.1 Encoding

STag markers are encoded using 48 black circles inside of the inner circular border of the marker. A simulated annealing method is used to efficiently pack these 48 black circles inside of the inner circular border. The encoding process involves using a lexicode generation algorithm, similar to

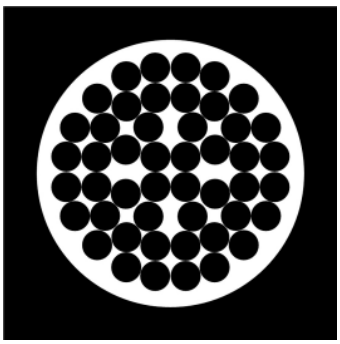


Figure 20: 48 circle layout in STag. [BTA17]

ArUco and AprilTag, that performs a two-step hierarchical exhaustive search of an n-dimensional binary space of code-words,where n corresponds to the length of the code-words. The search process focuses on selecting code-words that exhibit a Hamming distance greater than a pre-defined minimum requirement between each other. Since the marker has a circular encoding, the circular permutations of the code-words are also involved in the search. To limit the time complexity, this search algorithm generates 48-bit code-words in two steps:

1. Generate 12-bit code-words.

2. Generate complete 48-bit code-words as 4 combinations of these 12-bit code-words.

To accommodate different needs and applications, libraries with different sizes can be generated based on the chosen Hamming distance. A smaller library can be created for faster decoding, while a larger library can be generated to provide a larger number of unique markers.

(min.) HD	11	13	15	17	19	21	23
Library size	22,309	2,884	766	157	38	12	6

Table 1: Library sizes based on various (min.) Hamming distances (HD).

After encoding the code-word into the black circles of the marker, resulting in a specific identification, morphological erosion and dilation are performed a number of times to close the gaps between certain circles and create a unique shape for decoding.

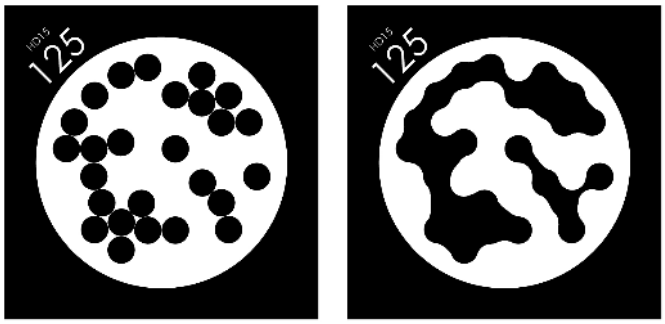


Figure 21: morphological transformations to encoding of STag. [BTA17]

The purpose of these morphological transformations is to create a unique enough shape to increase the Bit Error Ratio (BER). BER represents the proportion of error bits allowed to still be able to decode the marker, so a marker with a BER of 0.1 can be decoded with at most 10% of the code being read incorrectly. In other words, the marker can be decoded with up to 10% of the marker's encoding being occluded. [BTA17]

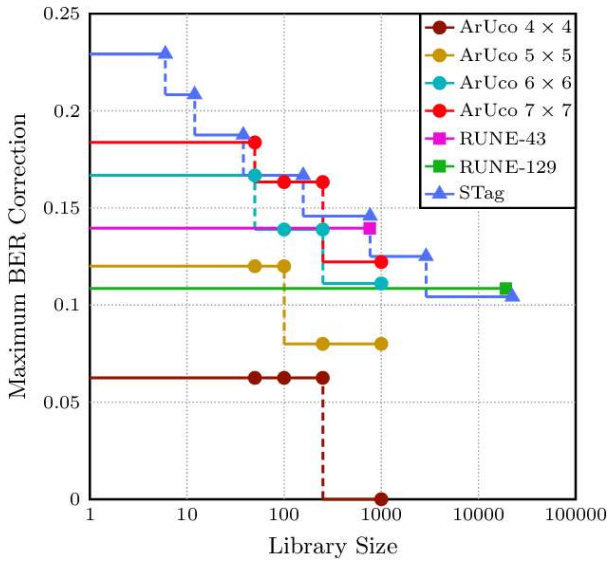


Figure 22: Maximum BER correction capabilities of a variety of marker systems. [BTA17]

4.2 Detection algorithm

The STag detection algorithm comprises three main parts: candidate detection, candidate validation, and homography refinement, as shown in Fig. 23.

4.2.1 Candidate detection

As we can see in Fig. 23, the candidate detection algorithm consists of 4 major stages:

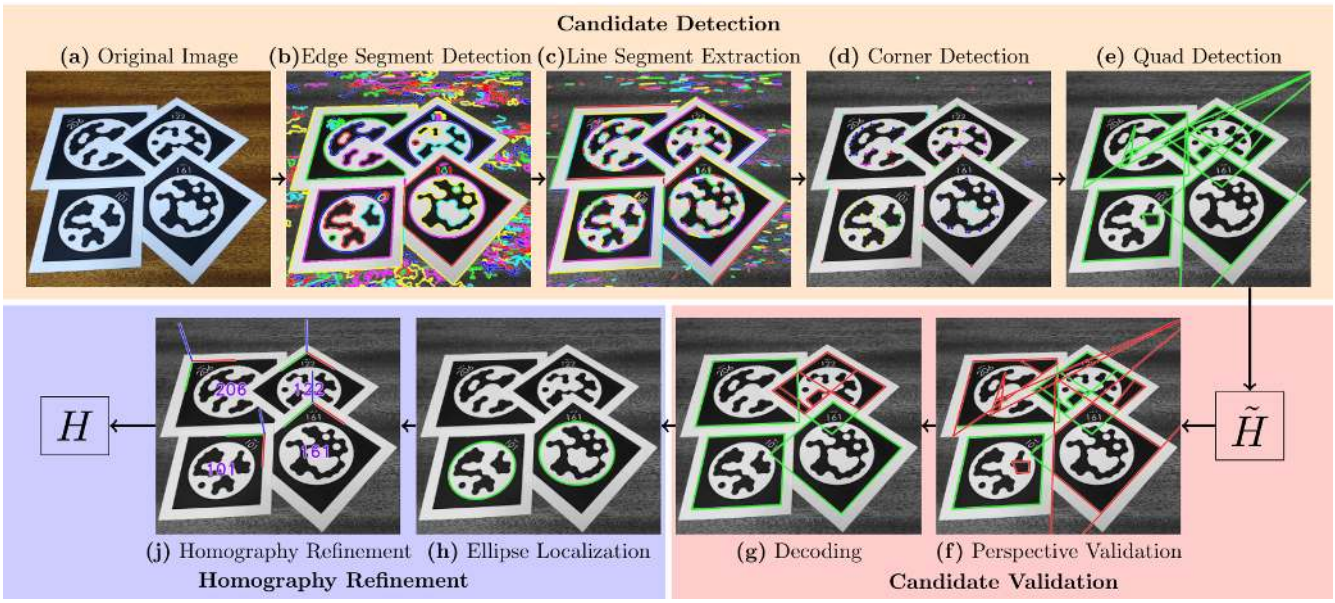


Figure 23: STag detection algorithm flowchart [BTA17], where $\tilde{H}$ represents the estimated homography (before Homography Refinement) and H represents the final homography (after Homography Refinement).

1. **Edge segment detection** (see Fig. 23b): The EDPF algorithm [CC12] is used to detect edge segments as pixel arrays by determining certain anchor points, and grouping these based on the total gradient response along the segment's path. The EDPF algorithm then uses the Helmholtz principle [hel08] by eliminating edge segments with a high probability of occuring randomly to ensure that the edge segments are valid and well-localized.

2. **Line segment extraction** (see Fig. 23c): The EDLines algorithm [AT11] is used to extract line segments from the edge segments. This is achieved by fitting a line to the first detected pixel of an edge segment and extending it until the pixels are no longer on the line. This is then repeated from the next pixel of the edge segments until all pixels of every edge segment have been processed.

3. **Corner detection** (see Fig. 23d): Corners are extracted by getting the intersections of the line segments per edge segment.

4. **Quad detection** (see Fig. 23e): Three consecutive corners of an edge segment are used to detect all the quads in the image. Using Three consecutive corners allows for the occlusion of one corner. To address potential occlusion of each corner in a quad, the detection process generates four separate detections for every quad, each starting from a different corner. This approach ensures that occlusion of any individual corner is accounted for and improves the overall robustness of the detection algorithm.

4.2.2 Candidate validation

The candidate validation process consists of 2 steps:

1. **Perspective validation** (see Fig. 23f) eliminates as many false candidates as possible based on the imposed perspective distortion, which we can estimate using the following formula:

$$\alpha_{rel} = \frac{\alpha_{max}}{\alpha_{min}} \tag{1}$$

, where α_{rel} is the relative depth or distance from a point to the camera, which is the ratio of the largest depth (α_{max}) compared to the smallest depth (α_{min}).
α_{max} and α_{min} can be calculated by extending both of the opposing parallel edges of a square. These will intersect in 2 different points at infinity, where there exists a line passing through both of these intersections at infinity (l_∞). After projecting these intersections, the image/projection of these intersections become finite (see Fig. 24).

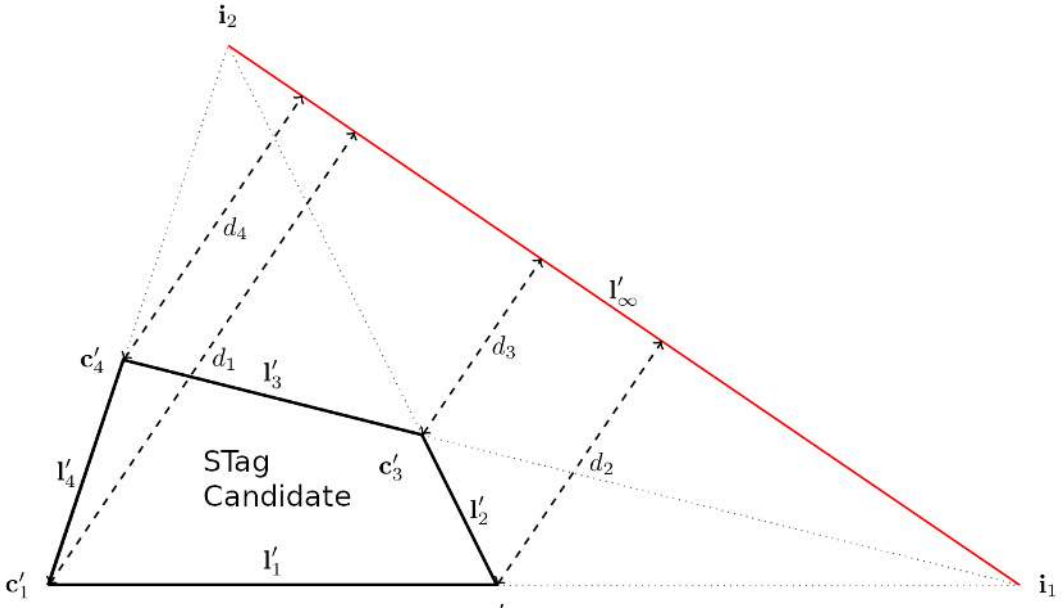


Figure 24: STag candidate with corners $\mathbf{c'_i}$ and vertices $\mathbf{l'_i}$, opposing parallel line intersections $\mathbf{i_1}$ and $\mathbf{i_2}$ with the projection of l_∞, $\mathbf{l'_\infty}$ going through both intersections. distances $\mathbf{d_i}$ are shown from c'_i to l'_∞. [BTA17]

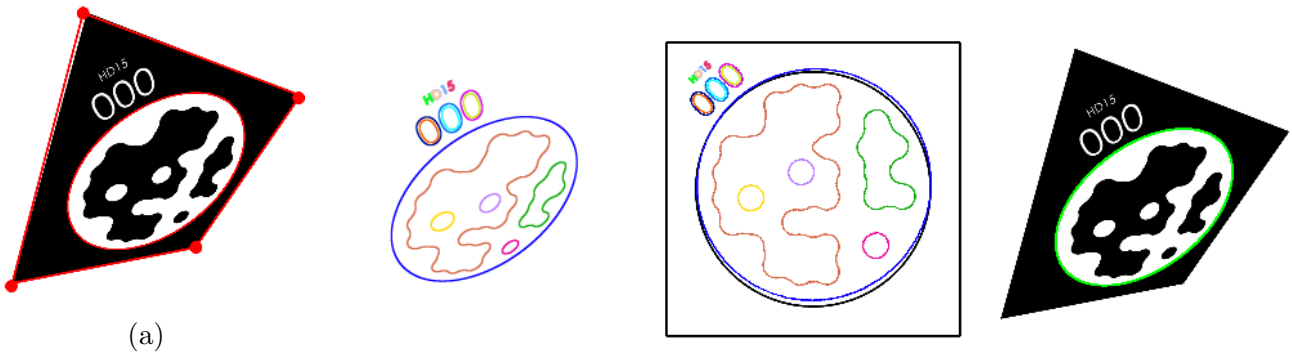
The distances (d_i) between l'_∞ and the corners c'_i have a linearly negative relationship with the depth (α) of the corresponding point, therefore:

$$\alpha_{rel} = \frac{\alpha_{max}}{\alpha_{min}} = \frac{kd_{min}^{-1}}{kd_{max}^{-1}} = \frac{d_{max}}{d_{min}} \tag{2}$$

where k is a scalar constant.
α_{rel} represents the degree of perspective distortion experienced by a square to obtain the projected quad. It is compared against a predetermined threshold, and any quads with a higher α_{rel} are discarded, thereby reducing the false positive rate. Although this approach may potentially reduce the detection rate, when an appropriate threshold is chosen, it offers a valuable tradeoff in terms of minimizing false positives.

2. **Decoding** (see Fig. 23g) eliminates false candidates by first estimating the homography matrix (H) of a candidate using its corners. For this, an arbitrary order of the corners is used. The homography is then used to sample the projections of the encoding to get the marker's code-word.
We maintain a library of all code-words and a list of their circular rotations. When a marker is detected, the code-word is read and compared to each element in the list using XOR and population count operations. If the Hamming distance between the read code-word and a code-word in the list is within the maximum number of bits that can be corrected (BER), we associate the candidate with the corresponding rotation and ID. If the rotation chosen during the detection process is found to be incorrect, we adjust the homography accordingly. Otherwise, we eliminate that marker after decoding.

4.2.3 Homography refinement



(a) Marker with incorrect corner localizations and expected ellipse location.

(b) Detected inner marker edge segments.

(c) Backprojected edge segments.

(d) Ellipse fitted to the closest edge segment.

Figure 25: Visual representation of STag homography refinement. [BTA17]

There are some cases where the corners of the markers are not correctly localized (see an exaggerated example, depicted in Fig. 25a) due to various factors such as noise in the image, imperfect edge segment detection and perspective distortion. To address this issue, we apply homography refinement

to improve the localization accuracy of the marker. The refinement process involves localizing the circle inside the marker (see Fig. 25):

1. Focus on the edge segments inside the marker (see Fig. 25b).
2. Back project the inner edge segments onto a square to obtain a normalized representation of the detection, ensuring that the process works equivalently in all poses (see Fig. 25c).
3. The expected circular border (with incorrect corner localizations) is compared to the edge segments using distance to find the edge segment closest to the expected circular border. An ellipse is then fitted to the projection of this edge segment (see Fig. 25d).

This ellipse is then used to adjust the markers homography based on the predetermined position of the circle inside the marker. This improves the localization stability, since the ellipse is more accurately localized than the corners (See [BTA17, section 4.3] for a more detailed explanation). It is worth noting that we disregard homography refinement, if the circle inside the marker is occluded, as this could lead to incorrect circle fitting and worsen the marker's localization stability.

5 STag2

In this section, we will discuss the design and detection algorithm changes made in STag2 to improve the occlusion-resilience of the original STag marker. The main idea behind this improvement is based on the fact that a square can be detected using only three corners. Thus, the addition of another square inside the marker at a 45 degree angle, called a rhombus, provides enough corners for the marker to be occluded close to halfway, while still allowing for detection.

5.1 STag2: Marker design

The main design improvement of STag2 is aimed at increasing its occlusion-resilience. This is achieved by adding another square, rotated 45 degrees, inside the original square marker. The corners of this new inner square form a rhombus which provides additional detection points, ensuring that the marker can still be detected with close to half of its area occluded (see Fig. 26).

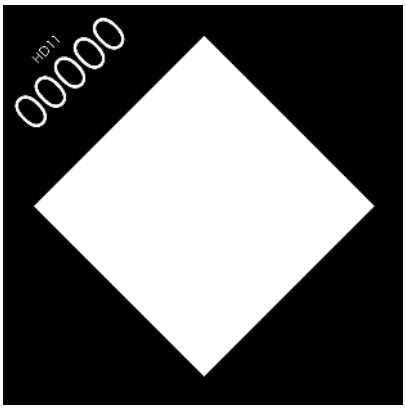


Figure 26: White rhombus inside of marker Square

Other changes were made to maintain the baseline performance of the marker. These include inverting the color of the circular border, blending it with the code circles and fitting these inside of the new square (to retain the homography refinement capabilities of the marker) (see Fig. 27). The encoding of the marker was left unchanged, thereby maintaining the same bit error ratio (BER) as the original STag.

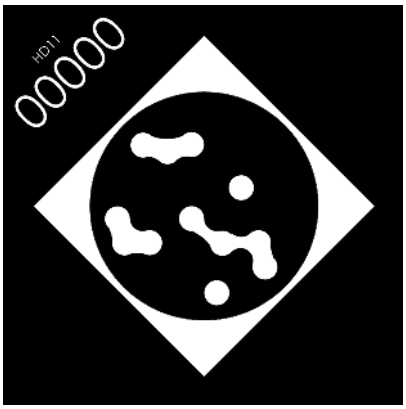


Figure 27: Proposed STag2 marker design with encoding blended with circular border

One of the main challenges introduced by the design changes is the reduction in encoding size due to the inclusion of the white rhombus. While this tradeoff enhances occlusion resilience, it leads to a slight decrease in the maximum detection distance. It is important to note that the design was not optimized specifically for this tradeoff, as dedicated experiments were not conducted to evaluate its impact, and had to be left to future research.

5.2 STag2: Detection algorithm

The STag2 detection algorithm involves some significant changes to the candidate validation process. Other parts of the algorithm have relatively minor adjustments. The candidate detection phase can now also detect white squares with black borders, and the decoding and pose refinement processes are scaled down to fit the new, smaller code circle.

5.2.1 Candidate validation changes

The most significant changes are in the candidate validation process, specifically in the addition of a new step called **Rhombus correction**. This step is inserted between the Perspective validation and Decoding steps and focuses on the white rhombus added to the marker design.

Rhombus correction
After the Candidate detection phase, which now also includes white quads (with black borders), and the Perspective validation step comes Rhombus correction. Rhombus correction can be divided into 2 phases:

1. **Rhombus detection**: The algorithm checks if the detected quad is the STag2 marker's white rhombus. This is determined by deducing the color of the quad using specific points inside

and outside the quad detection. These marker points are then classified as black or white by comparing the marker points with a thresholded representation of the detected quad:

- if **all marker points are correctly classified**, meaning all marker points outside of the quad are black and all marker points inside the quad are white, we can assume that the quad represents the white rhombus of a marker without any occlusions.
- if **at least half of the marker points are correctly classified**, meaning at least half of the marker points outside of the quad are black and at least half of the marker points inside the quad are white, we can still assume that the quad is a marker's white rhombus, but with some occlusions. It is important to note that there exists a tradeoff between the minimum number of correctly classified marker points and the Bit Error Ratio (BER) of the marker encoding. This tradeoff offers an avenue for further investigation in future research.
- **Otherwise**, it is assumed that the detected quad is a black marker.

These assumptions will get tested during the decoding step.

2. **Quad correction**: If the detected quad is classified as a white rhombus, the algorithm corrects the corners to estimate the corners of the full marker. This is done by rotating each of the marker points 45° and scaling it out by the marker border ratio, so it fits the full marker. These rotations and scalings are applied to the corners of the marker. After this, the homography is recalculated using the new corners.

The addition of the white rhombus adds a new detection type for every marker detection. For every marker detection, we now have 2 types:

1. **black square** detections: The marker detections achieved from the black square of the marker. These detections are equivalent to the normal STag detection process, but using the adjusted marker design. If detected, these are stored at the back of the list of marker detections.

2. **white rhombus** detections: The marker detections achieved from the white rhombus of the marker. These detections include the Rhombus correction phase. If detected, these are stored at the front of the list of marker detections.

In addition to the four detections generated by STag (see Sec. 4.2.1), the white rhombus detection method incorporates an additional four detections. This extension allows for the handling of occlusion scenarios involving the corners of both the white rhombus and the black square, further enhancing the algorithm's robustness and ability to handle various occlusion conditions.

6 Experimental setup

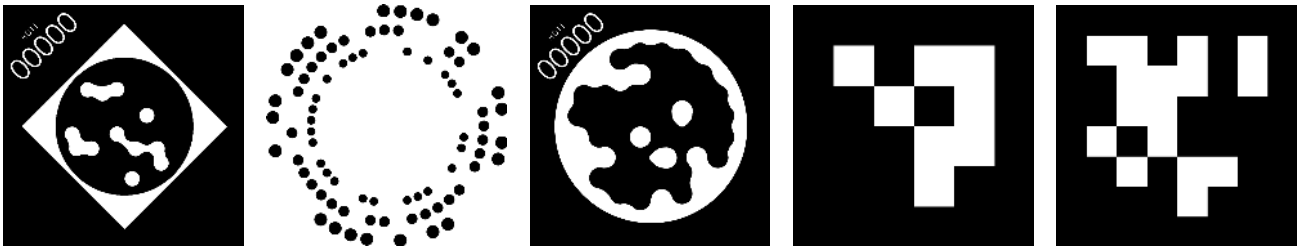


Figure 28: Markers to be compared: Stag2 (id: 0), RUNE-Tag (id: 8) [BART11], STag (id: 0) [BTA17], ArUco (4x4, id: 0) [GJMSMCMJ14], AprilTag (36h11, id: 0) [Ols11]

To assess the performance of the proposed STag2 marker, we compared it to the state-of-the-art for different aspects (see Fig. 28):

- **ArUco**: The ArUco marker system is one of the most popular fiducial markers, since it is included in the OpenCV library and therefore easily accessible.
- **AprilTag**: The aprilTag marker system is one of the most feature-rich marker systems and has been used by big companies such as NASA [MBS+20].
- **RUNE-Tag**: The RUNE-Tag marker system is currrently the best fiducial marker system in terms of occlusion resilience, but it underperforms when it comes to efficiency and challenging lighting scenarios.
- **STag**: The STag marker system forms the baseline for STag2.

The investigations in this paper can be split into 2 types:

1. **Baseline experiments**, where we compare the general performance of the proposed marker to the state-of-the-art.
2. **Occlusion experiments**, where we compare the occlusion-based markers at different occlusion levels.

In both cases, we investigate single marker configurations based on a marker comparison paper by Michail Kalaitzakis et al [KCA+20]. Like [KCA+20], we compare the detection rate and accuracy of each marker at different distances, but we add an extra axis, the X-axis, to this comparison. We also investigate as many rotation possibilities as our experimental setup allows us to.
Although there are more use cases for the proposed marker, the focus of this paper is testing the occlusion-resilience and general performance (compared to the state-of-the-art marker systems) of the proposed marker for mainly AR purposes. For most other use cases, it is comparable to STag, which has already been assessed in [KCA+20].

We employed a simple xy-table equipped with a CNC shield v3 (see Fig. 31c) to assess the performance (with and without occlusion) of the proposed marker STag2 along with the other state-of-the-art fiducial markers. The markers were of sizes approximately 5.89cm, and the camera

used for detection was a Logitech Webcam C505e HD 720p, boasting a FOV of 60° and 30 fps. Due to the use of a tripod without proper fixation, our setup was prone to slight camera movements between experiments, although the camera remained stable during each individual experiment. These slight inconsistencies may have an impact on marker comparisons.
To ensure adequate illumination, we set up a generic study lamp behind the camera, the height of which was approximately 40.5 cm. The lamp was angled at an angle of 45 degrees from the side and had a length of about 40.5 cm. To control the xy-table, we used a CNC GRBL 1.1h software with two stepper motors and two limit switches, which included soft limits.

The distance from the camera to the marker (on the xy-table) is 60cm. We moved the markers 1 cm at a time on the xy-table, using a zigzag pattern to cover an area of 31 x 37 cm. We tested each marker for different rotations, exploring every combination of the rotation angles of 0°, 30°, 45°, and 60° along the axis perpendicular to the marker (roll of the marker holder). Furthermore, each of these rotations was tested at every combination of the rotation along the horizontal center axis (pitch of the marker holder) at 0°, 30°, 45°, 60°, and 75° (see Fig. 29). This results in some different yaw, pitch, roll rotations of the marker. We did not rotate the yaw of the marker holder, because this would hide the marker from the camera in some locations.

See Fig. 31 for the experimental setup and Fig. 30 for visual representation of the experimental setup.

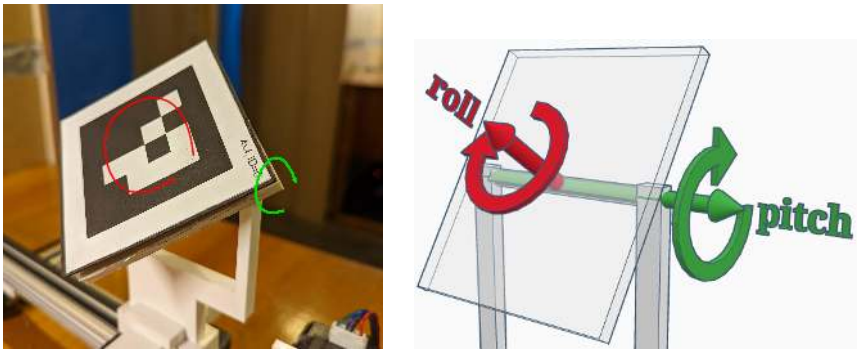


Figure 29: Marker holder rotations with holder roll (red arrow) and holder pitch (green arrow)

In recording the data, we made sure to note the marker ID, detected and real positions (x, y, z) in cm in relation to the camera, detected and real rotation (yaw, pitch, roll) in degrees, and detection time. This amounts to 1147 single frame detections per experiment for every rotation combination (20). This information was carefully stored in a CSV.

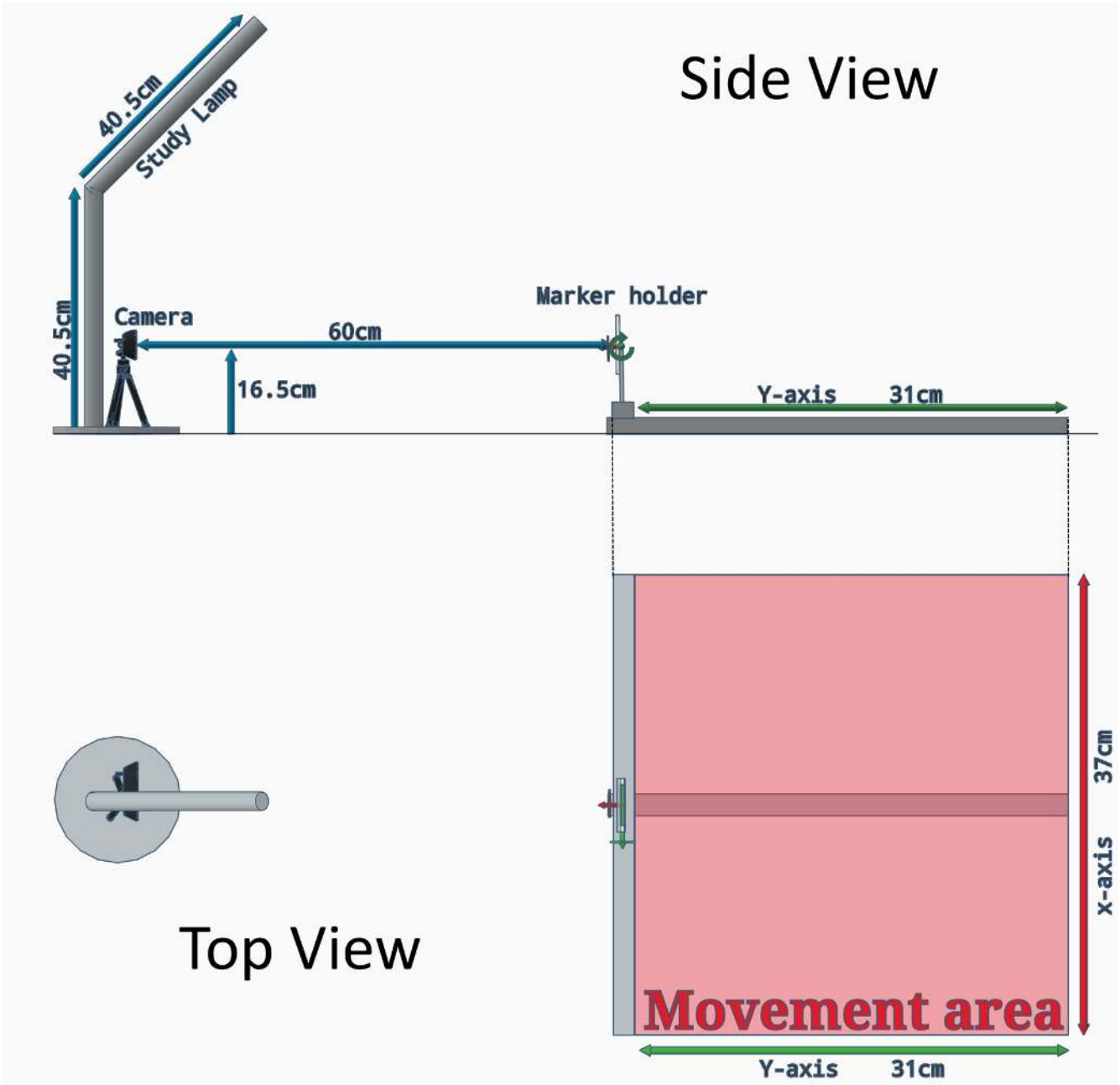
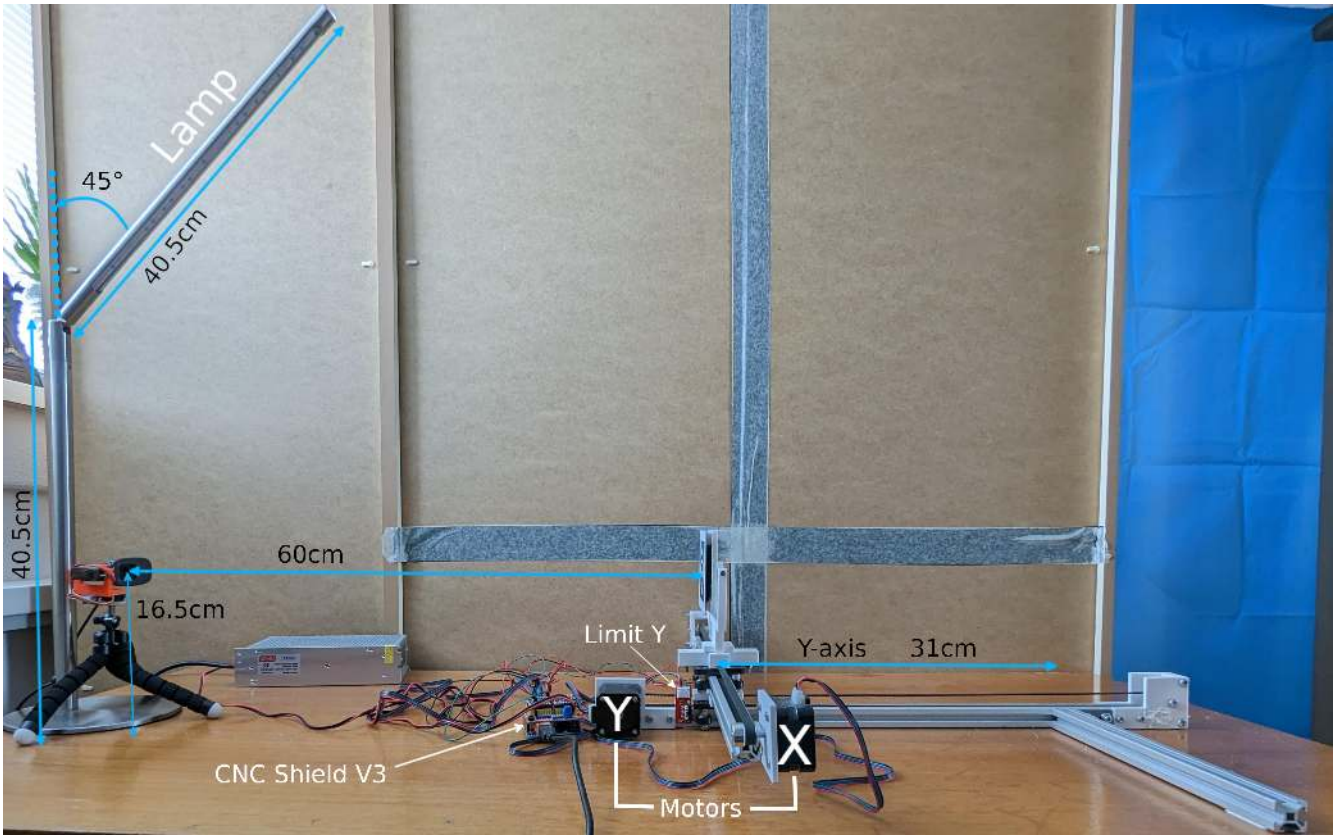
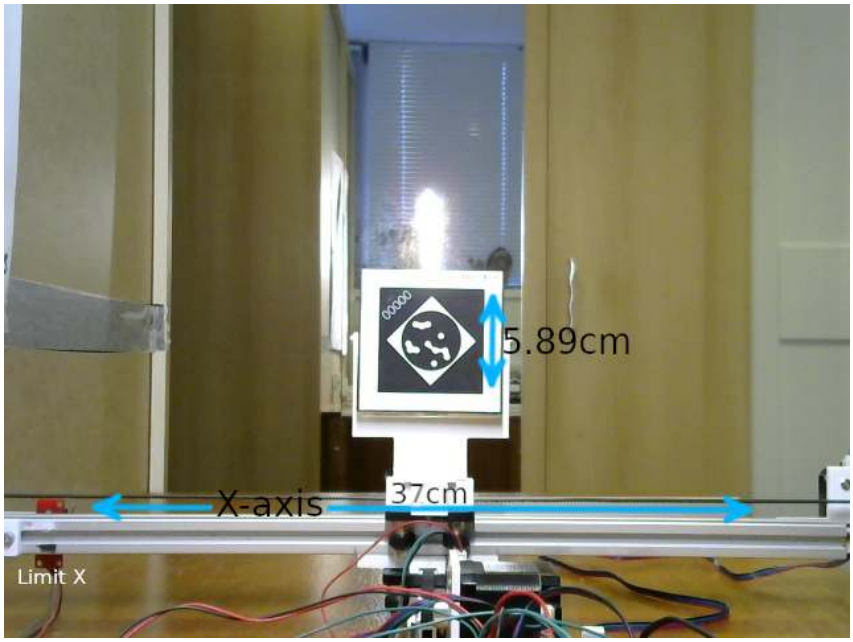


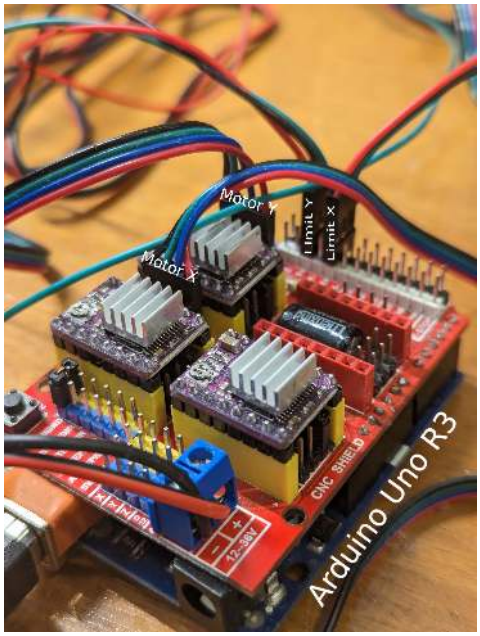
Figure 30: Visual representation of experimental setup.



(a)
Side view



(b)
Front view



(c)
CNC setup

Figure 31: Experimental setup.

6.1 Occlusion experiments

We performed occlusion experiments for two types of occlusions:

1. **Single corner occlusions with the top right corner occluded**: We occluded only the top right corner of the markers as we assumed that the results would be comparable for each corner. Specifically, we occluded approximately 21.1% of the marker encoding circle in both cases. For STag, this resulted in approximately 25.4% of the marker being occluded, while for STag2, approximately 39.1% of the marker was occluded. The difference in occlusion percentage can be attributed to the variation in encoding circle size between the two marker types. We only conducted this experiment for the proposed STag2, as it was not possible for STag.

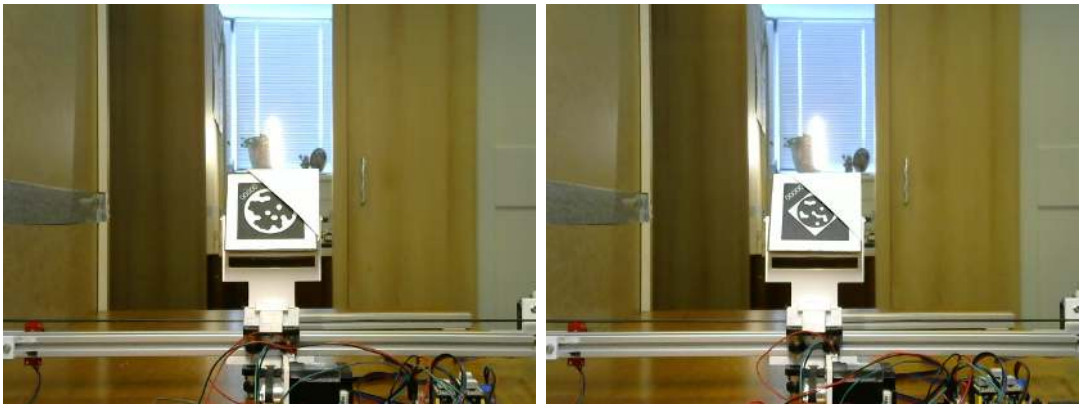


Figure 32: Single corner occlusion setup for STag and STag2.

2. **Occlusions with approximately half of the left side of the marker (vertically) occluded**: We occluded approximately half of the right side of the marker (vertically) as we assumed that the results would be comparable to all sides, both horizontally and vertically. We covered the marker as much as we could to still generate results. The occlusion accounted for approximately 22.1% of the marker encoding circle. Specifically, for STag2, about 36.4% of the marker was occluded. We only conducted this experiment for the proposed STag2, as it was not possible for STag.

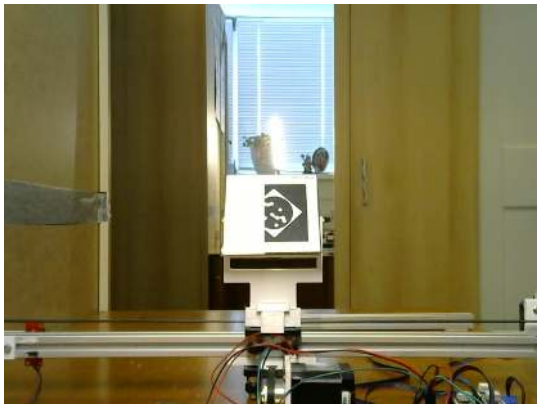


Figure 33: Half occlusion setup for STag2.

The occlusion experiments were performed to evaluate the performance of STag and STag2 in challenging scenarios where the markers might be partially occluded. The experimental setup aimed to assess the robustness of the markers and provide insights into their usability in real-world applications.

7 Experimental results

All experimental results for every rotation combination were plotted as heatmaps (which can be found in Appendix A, B and C). This includes both the baseline and occlusion results. The heatmap represents the similarity scores from 0 (completely dissimilar) to 1 (exactly the same) in terms of location (x, y, z) and rotation (yaw, pitch, roll) between the real and estimated detections. The specific ranges of location and rotation values used in the experiments are discussed in Section 6. Note that the location data in our study is expressed in centimeters (cm), while the rotation data is presented in degrees. The similarity score is calculated per detection using the following formula:

$$similarity = \frac{(similarity_{location} + similarity_{rotation})}{2} \tag{3}$$

, where the location and rotation similarity scores are calculated as follows:

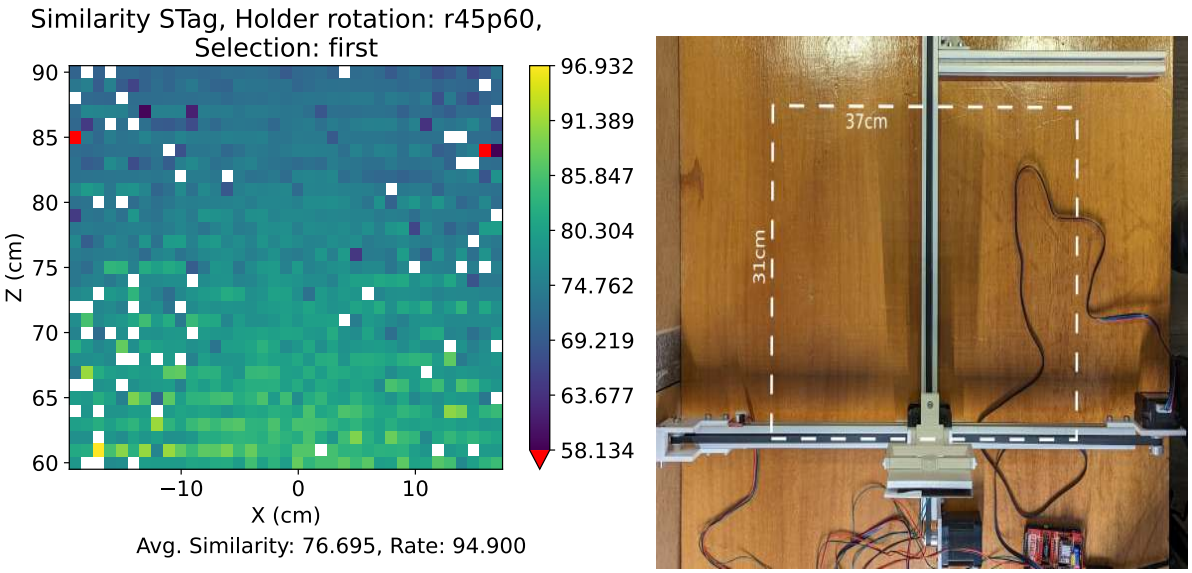
$$similarity_{location} = \frac{1}{1 + distance} \tag{4}$$

distance represents the euclidian distance in cm between the real and estimated location (x,y,z) of the marker. The location similarity is 1 if *distance* is 0 (meaning the estimated and real locations are identical) and approaches 0, as *distance* increases.

$$similarity_{rotation} = \cos(angle\_difference) \tag{5}$$

$angle\_difference$ represents the (absolute) angle difference between the real and estimated rotation of the marker (yaw, pitch, roll) in degrees. It ranges from 0 to 1, where 1 represents the 2 detections being exactly the same, 0.5 represents the 2 detections being orthogonal (angle difference is 90°) and 0 represents the 2 detections being inverted (see Fig. 35)

The heatmap is plotted based on the top down view of the setup (see Fig. 34).



(a)
Heatmap, where Holder rotation represents the roll and pitch of the marker holder, white spots represent incorrect or no detections and red spots represent inverse detections (see Fig. 35).

(b)
Top view corresponding to heatmap

Figure 34: Heatmap (a) with corresponding view (b)

7.1 Baseline results

Since STag and STag2 can have multiple detections of a single marker to handle occlusions, we initially compared the first and last detections of these marker systems. In the case of STag, the first and last detections are obtained from different corners but represent the same marker (as described in Sec. 4.2.1). For STag2, the first detection corresponds to the **white rhombus (WR)** detection, and the last detection represents the **black square (BS)** detection, which utilizes the same detection algorithm as STag (as explained in Sec. 5.2.1). These initial comparisons helped us determine the optimal detection to be used in the subsequent experiments.

7.1.1 STag2 vs STag

In order to evaluate the first and last values of the STags, we compared the averages of certain metrics (distance in cm, angle difference in degrees and similarity score) and represented these as tables (see Table 2, 3 and 4), similar to [KCA+20].

		STag				STag2			
		first (BS)		last (BS)		first (WR)		last (BS)	
Holder roll	Holder pitch	mean (cm)	std. (cm)	mean (cm)	std. (cm)	mean (cm)	std. (cm)	mean (cm)	std. (cm)
0°	0°	0.983	0.332	**0.981**	**0.324**	1.02	0.848	**0.763**	**0.267**
	30°	**1.18**	**0.327**	**1.18**	0.328	1.25	1.1	**0.898**	**0.308**
	45°	**1.14**	**0.392**	1.15	0.396	1.56	1.83	**1**	**0.272**
	60°	1.08	0.44	**1.07**	**0.436**	2.34	2.55	**1.01**	**0.196**
	75°	0.897	0.312	**0.89**	**0.298**	0.631	**0.407**	**0.628**	0.545
30°	0°	**0.892**	0.205	0.894	**0.204**	2.11	1.62	**0.723**	**0.165**
	30°	**1.19**	**0.303**	**1.19**	0.307	1.31	1.42	**0.591**	**0.225**
	45°	**1.2**	**0.267**	**1.2**	0.278	1.42	2.02	**0.59**	**0.257**
	60°	1.22	**0.342**	**1.21**	0.352	1.16	1.43	**0.638**	**0.256**
	75°	**1.02**	**0.353**	1.2	1.64	**0.782**	**0.315**	0.792	0.329
45°	0°	**0.857**	0.183	0.859	**0.182**	1.27	1.13	**0.652**	**0.198**
	30°	0.985	0.271	**0.981**	**0.265**	1.09	1.32	**0.571**	**0.193**
	45°	**0.971**	0.289	0.973	**0.282**	1.26	1.87	**0.59**	**0.329**
	60°	**0.922**	0.372	**0.922**	**0.342**	0.772	0.972	**0.642**	**0.271**
	75°	**0.959**	**0.361**	1.11	0.581	**0.747**	**0.32**	0.814	0.469
60°	0°	**0.715**	**0.159**	0.72	0.163	1.74	1.45	**0.687**	**0.162**
	30°	**0.911**	0.268	0.913	**0.266**	1.65	1.76	**0.636**	**0.162**
	45°	**0.876**	**0.266**	0.88	0.275	1.88	2.2	**0.618**	**0.326**
	60°	**0.81**	**0.283**	0.816	0.306	1.24	1.33	**0.615**	**0.356**
	75°	**0.851**	**0.308**	0.858	0.34	0.883	0.446	**0.868**	**0.415**

Table 2: Average distance (in cm) between estimated and real pose locations for STags per rotation combination. The optimal value of each metric has been highlighted for each rotation combination (row) and tag.

Looking at the distance results for STag and STag2, as listed in Table 2, we can see that the last values of both STags perform the best, since for STag2, the last value clearly performs better, and for STag, the difference is very small since it is technically the exact same detection. Looking at the STag2 plots in Appendix A, we can see that the first values (white rhombus) perform worse due to the fact that the distance results become more inconsistent the farther the marker gets from the camera, since the corners of the white rhombus get harder to detect accurately.

Holder roll	Holder pitch	STag				STag2			
		first (BS)		last (BS)		first (WR)		last (BS)	
		mean (°)	std. (°)	mean (°)	std. (°)	mean (°)	std. (°)	mean (°)	std. (°)
0°	0°	**8.21**	**3.62**	8.37	3.65	9.17	8.3	**5.06**	**4.65**
	30°	**1.66**	0.575	1.69	**0.57**	11.6	22.2	**0.916**	**0.449**
	45°	1.78	0.424	**1.76**	**0.415**	33.4	43	**1.35**	**2.76**
	60°	1.69	**0.415**	**1.68**	0.419	33.3	53.3	**1.08**	**7.25**
	75°	**2.47**	**6.36**	2.49	**6.36**	2.67	12	**2.34**	**9.79**
30°	0°	**5.67**	**2.37**	**5.67**	2.39	12.3	8.74	**2.82**	**2.39**
	30°	**3.34**	0.381	3.35	**0.369**	16.4	24.3	**1.69**	**3.88**
	45°	**2.24**	0.407	2.29	**0.39**	21	35.9	**1.64**	**4.75**
	60°	2.65	**0.422**	**2.64**	0.432	12.9	35.4	**1.12**	**7.37**
	75°	**3.3**	**0.271**	4.06	10.8	3.29	**5.02**	**3.27**	**5.02**
45°	0°	**6.92**	4.15	6.96	**4.13**	11.2	8.28	**4.49**	**4.2**
	30°	**5.71**	**0.747**	**5.71**	0.75	17.2	22.8	**7.14**	**2.59**
	45°	3	0.391	**2.99**	**0.382**	16.5	31.2	**4.3**	**7.25**
	60°	**4.3**	**5.06**	4.31	**5.06**	5.78	18.3	**3.61**	**9.35**
	75°	**6.39**	**9.19**	8.19	18.3	2.74	0.329	**2.72**	**0.285**
60°	0°	**5.01**	**2.46**	5.04	2.53	10.7	8.26	**3.05**	**2.29**
	30°	**3.1**	**0.53**	3.13	0.539	13.7	22.9	**1.62**	**1.84**
	45°	3.47	2.59	**3.4**	**0.473**	19.2	34.7	**1.07**	**4.68**
	60°	**4.63**	0.34	4.65	**0.33**	14.4	37	**1.45**	**7.19**
	75°	**5.96**	**0.218**	6.25	6.23	2.49	11.5	**1.77**	**4.7**

Table 3: Average difference between estimated and real pose rotation angles (in degrees) for STags per rotation combination. The optimal value of each metric has been highlighted for each rotation combination (row) and tag.

Looking at the rotation results for the STags and STag2, as listed in Table 3, we observe a similar performance behavior as for the distance results. The white rhombus (first) values of STag2 perform worse than the black square (last) values of STag2, because the farther away the marker gets from the camera, the less accurately the white rhombus can be detected. This also makes it more prone to inverse detections (see Fig. 35), causing it to perform even worse.

Similar to the distance and angle difference, the results in Table 4 show that the similarity score is in general also the highest for the last value.

		STag				STag2			
		first (BS		last (BS)		first (WR		last (BS)	
Holder roll	Holder pitch	mean	std.	mean	std.	mean	std.	mean	std.
0°	0°	**0.756**	0.0442	**0.756**	**0.044**	0.764	0.0646	**0.788**	**0.0448**
	30°	**0.734**	**0.0345**	**0.734**	0.0352	0.727	0.0884	**0.769**	**0.0385**
	45°	**0.741**	**0.0399**	0.74	0.0401	0.653	0.165	**0.753**	**0.0298**
	60°	0.75	**0.0491**	**0.752**	0.051	0.601	0.193	**0.75**	**0.0336**
	75°	0.769	0.0481	**0.77**	**0.0479**	0.812	0.0574	**0.814**	**0.0504**
30°	0°	**0.766**	0.0283	**0.766**	**0.0282**	0.693	0.0921	**0.792**	**0.0268**
	30°	**0.732**	0.0321	**0.732**	**0.0319**	0.727	0.101	**0.818**	**0.0393**
	45°	**0.731**	**0.0283**	**0.731**	0.0286	0.703	0.144	**0.82**	**0.0446**
	60°	0.731	**0.0386**	**0.732**	0.0393	0.734	0.147	**0.809**	**0.0474**
	75°	**0.755**	**0.0444**	0.741	0.0665	**0.786**	**0.0444**	0.785	0.045
45°	0°	**0.769**	**0.0269**	**0.769**	**0.0269**	0.748	0.08	**0.805**	**0.0341**
	30°	0.755	0.0336	**0.756**	**0.0332**	0.747	0.109	**0.821**	**0.0396**
	45°	**0.758**	0.0358	**0.758**	**0.0345**	0.735	0.144	**0.819**	**0.0467**
	60°	**0.767**	0.0488	0.766	**0.0479**	0.791	0.0848	**0.807**	**0.0499**
	75°	**0.76**	**0.0557**	0.74	0.0862	**0.792**	**0.0388**	0.786	0.0467
60°	0°	**0.793**	**0.0267**	0.792	0.0271	0.714	0.0881	**0.798**	**0.0273**
	30°	**0.766**	0.0373	**0.766**	**0.0372**	0.713	0.117	**0.808**	**0.0297**
	45°	**0.771**	**0.0392**	**0.771**	0.0394	0.677	0.15	**0.814**	**0.0393**
	60°	**0.782**	**0.0457**	**0.782**	0.0471	0.725	0.158	**0.816**	**0.0479**
	75°	**0.776**	**0.0468**	0.775	0.055	0.77	0.0578	**0.774**	**0.0429**

Table 4: Average similarity (ranging from 0 to 1), as defined in Equation 3, between the estimated and real pose of STag and STag2 for each rotation combination. The optimal value of each metric has been highlighted for each rotation combination (row) and tag.

Since we can observed that the last values of both STag and STag2 represent the optimal performance of the marker system, we use these last values in our comparisons with the other marker systems.

7.1.2 STag2 (black square) vs the state-of-the-art

To compare the performance of all the marker systems, we used the same metrics as for the comparisons of STag and STag2, but we also included the detection rate (see Table 5, 6 and 7), similar to [KCA+20], i.e., the percentage that the marker's ID, location and orientation is detected.

Holder roll	Holder pitch	AprilTag mean (cm)	AprilTag std. (cm)	AprilTag det. rate (%)	ArUco mean (cm)	ArUco std. (cm)	ArUco det. rate (%)	RUNETag mean (cm)	RUNETag std. (cm)	RUNETag det. rate (%)	STag mean (cm)	STag std. (cm)	STag det. rate (%)	STag2 mean (cm)	STag2 std. (cm)	STag2 det. rate (%)
0°	0°	**0.738**	**0.224**	**100**	1.3	0.419	99.4	3.51	0.168	2.96	0.981	0.324	99.6	0.763	0.267	99
	30°	**0.867**	0.319	**100**	1.18	0.547	99.4	ND	ND	ND	1.18	0.328	99.9	0.898	**0.308**	99.1
	45°	**0.839**	0.322	**100**	1.34	0.401	97.1	ND	ND	ND	1.15	0.396	99.7	1	**0.272**	98.7
	60°	**0.828**	0.321	**100**	1.51	0.49	99.7	ND	ND	ND	1.07	0.436	99	1.01	**0.196**	98.1
	75°	0.779	0.361	56.2	1.73	0.575	**92.7**	ND	ND	ND	0.89	**0.298**	69.7	**0.628**	0.545	64.6
30°	0°	0.858	0.193	**100**	1.81	0.474	99.7	3.01	0.188	1.57	0.894	0.204	99.3	**0.723**	**0.165**	99.4
	30°	0.724	0.29	**100**	1.53	0.537	99.5	ND	ND	ND	1.19	0.307	98.7	**0.591**	**0.225**	99.2
	45°	0.832	0.296	**100**	1.15	0.478	99	ND	ND	ND	1.2	0.278	98.5	**0.59**	**0.257**	99.1
	60°	0.811	0.304	**100**	1.45	0.551	99.7	ND	ND	ND	1.22	0.352	97.5	**0.638**	**0.256**	96.2
	75°	0.806	0.362	31.6	1.61	0.582	**99.4**	ND	ND	ND	1.2	1.64	81.7	**0.792**	**0.329**	82.8
45°	0°	0.885	**0.179**	**100**	1.87	0.564	99.5	3.02	0.183	3.49	0.859	0.182	98.5	**0.652**	0.198	98.6
	30°	0.73	0.286	**100**	1.57	0.626	99	ND	ND	ND	0.981	0.265	99.2	**0.571**	**0.193**	98.4
	45°	0.724	0.302	**100**	1.6	0.763	99.7	ND	ND	ND	0.973	**0.282**	98.7	**0.59**	0.329	98
	60°	0.768	0.291	**100**	1.91	0.597	98.6	ND	ND	ND	0.922	0.342	97.7	**0.642**	**0.271**	92.7
	75°	**0.657**	**0.418**	45.8	1.87	0.649	**99.7**	ND	ND	ND	1.11	0.581	90.8	0.814	0.469	88.1
60°	0°	0.935	0.165	**99.9**	1.63	0.468	99.4	3.03	0.118	3.92	0.72	0.163	99.3	**0.687**	**0.162**	98.8
	30°	0.685	0.277	**100**	1.69	0.633	**100**	ND	ND	ND	0.913	0.266	99.2	**0.636**	**0.162**	98.8
	45°	0.693	0.301	**100**	1.79	0.58	99.6	ND	ND	ND	0.88	**0.275**	99	**0.618**	0.326	98.7
	60°	0.725	**0.3**	**100**	1.82	0.606	99.2	ND	ND	ND	0.816	0.306	98.8	**0.615**	0.356	97.5
	75°	**0.671**	0.515	25.9	2.19	0.649	**99.4**	ND	ND	ND	0.858	**0.34**	94.8	0.868	0.415	91.5

Table 5: Average distance (in cm) between estimated and real pose locations of marker systems for each rotation combinations. The optimal value of each metric has been highlighted for each rotation combination (row) and tag. "ND" indicates that the marker was not detected or was incorrectly detected in all frames.

In terms of accuracies over different distances, the experimental results presented in Table 5 reveal a somewhat unexpected finding: STag2 outperforms STag even in the absence of occlusions. This outcome is intriguing considering that STag2 features a smaller encoding circle, which would generally imply a potential decrease in detection accuracy, especially of its ID. A further analysis of this observation is given in Section 7.1.3.
Otherwise, the most obvious observation is that RUNETag has the least amount of detections due to the detection range of RUNETag not being high enough. Therefore, we will exclude it from all comparisons. Furthermore, we can see that in terms of detection rate AprilTag performs the best and is the most reliable of the marker systems in terms of distance performance and detection rate, even if the detection rate drops significantly at holder pitch of 75°.

Holder roll	Holder pitch	AprilTag mean (°)	AprilTag std. (°)	AprilTag det. rate (%)	ArUco mean (°)	ArUco std. (°)	ArUco det. rate (%)	RUNETag mean (°)	RUNETag std. (°)	RUNETag det. rate (%)	STag mean (°)	STag std. (°)	STag det. rate (%)	STag2 mean (°)	STag2 std. (°)	STag2 det. rate (%)
0°	0°	**3.1**	**1.07**	**100**	8.35	3.94	99.4	5.03	1.02	2.96	8.37	3.65	99.6	5.06	4.65	99
	30°	1.23	**0.232**	**100**	1.55	1.11	99.4	ND	ND	ND	1.69	0.57	99.9	**0.916**	0.449	99.1
	45°	**0.96**	**0.252**	**100**	1.08	0.5	97.1	ND	ND	ND	1.76	0.415	99.7	1.34	2.76	98.7
	60°	1.68	**0.171**	**100**	1.13	0.426	99.7	ND	ND	ND	1.68	0.419	99	**1.08**	7.25	98.1
	75°	**1.57**	**0.187**	56.2	1.62	6.55	**92.7**	ND	ND	ND	2.49	6.36	69.7	2.34	9.79	64.6
30°	0°	3.07	**1.56**	**100**	8.14	5.22	99.7	4.24	2.23	1.57	5.67	2.39	99.3	**2.82**	2.39	99.4
	30°	**1.66**	**0.198**	**100**	2.78	0.737	99.5	ND	ND	ND	3.35	0.369	98.7	1.69	3.88	99.2
	45°	**1.3**	**0.224**	**100**	2.75	0.476	99	ND	ND	ND	2.29	0.39	98.5	1.64	4.75	99.1
	60°	1.83	**0.15**	**100**	5.29	0.354	99.7	ND	ND	ND	2.64	0.432	97.5	**1.12**	7.37	96.2
	75°	**1.79**	**0.153**	31.6	4.67	10.8	**99.4**	ND	ND	ND	4.06	10.8	81.7	3.27	5.02	82.8
45°	0°	**2.68**	**1.33**	**100**	8.93	4.99	99.5	4.14	2.09	3.49	6.96	4.13	98.5	4.49	4.2	98.6
	30°	**1.88**	**0.218**	**100**	2.95	1.88	99	ND	ND	ND	5.71	0.75	99.2	7.14	2.59	98.4
	45°	**1.29**	**0.233**	**100**	2.5	1.5	99.7	ND	ND	ND	2.99	0.382	98.7	4.3	7.25	98
	60°	**1.87**	**0.185**	**100**	3.5	0.482	98.6	ND	ND	ND	4.31	5.06	97.7	3.61	9.35	92.7
	75°	2.79	9.09	45.8	11.5	9.47	**99.7**	ND	ND	ND	8.19	18.3	90.8	**2.72**	**0.285**	88.1
60°	0°	**2.56**	**2.05**	**99.9**	8.52	5.49	99.4	3.49	1.61	3.92	5.04	2.53	99.3	3.05	2.29	98.8
	30°	2.27	**0.216**	**100**	2.63	0.867	**100**	ND	ND	ND	3.13	0.539	99.2	**1.62**	1.84	98.8
	45°	1.63	**0.188**	**100**	2.12	0.586	99.6	ND	ND	ND	3.4	0.473	99	**1.07**	4.68	98.7
	60°	**1.31**	**0.177**	**100**	5.04	0.531	99.2	ND	ND	ND	4.65	0.33	98.8	1.45	7.2	97.5
	75°	1.9	8.69	25.9	2.37	**0.399**	**99.4**	ND	ND	ND	6.25	6.23	94.8	**1.77**	4.7	91.5

Table 6: Average difference between estimated and real pose rotation angles (in degrees) of marker systems for all rotation combinations.The optimal value of each metric has been highlighted for each rotation combination (row) and tag. "ND" indicates that the marker was not detected or was incorrectly detected in all frames.

In terms of angle difference in degrees, as listed in Table 6, we can clearly see that AprilTag outperforms all the other marker systems. We can also see that most of the standard deviation(std) values where the holder pitch is 75° are pretty high. This is due to the possibility that certain detection get inverted (see Fig. 35), causing a big reported difference in the angle. This could be due to the camera not being calibrated well enough for those 75° angles.

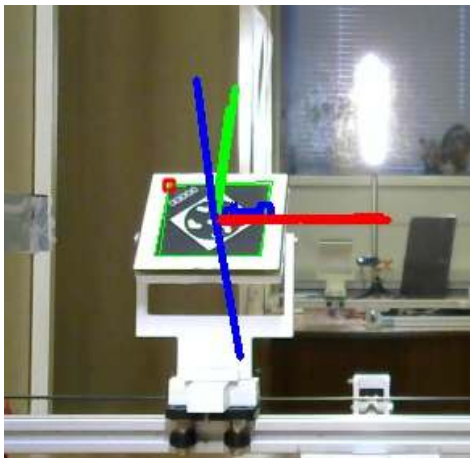


Figure 35: Inverted STag2 detection. [BTA17]

Holder roll	Holder pitch	AprilTag			ArUco			RUNETag			STag			STag2		
		mean	std.	det. rate (%)	mean	std.	det. rate (%)	mean	std.	det. rate (%)	mean	std.	det. rate (%)	mean	std.	det. rate (%)
0°	0°	**0.792**	**0.037**	**100**	0.721	0.04	99.4	0.61	0.004	2.96	0.756	0.044	99.6	0.788	0.045	99
	30°	**0.775**	0.045	**100**	0.739	0.045	99.4	ND	ND	ND	0.734	**0.035**	99.9	0.769	0.038	99.1
	45°	**0.78**	0.048	**100**	0.72	0.037	97.1	ND	ND	ND	0.74	0.04	99.7	0.753	**0.03**	98.7
	60°	**0.782**	0.05	**100**	0.706	0.039	99.7	ND	ND	ND	0.752	0.051	99	0.75	**0.034**	98.1
	75°	0.793	0.059	56.2	0.689	**0.043**	**92.7**	ND	ND	ND	0.77	0.048	69.7	**0.814**	0.05	64.6
30°	0°	0.772	0.028	**100**	0.68	0.031	99.7	0.624	0.005	1.57	0.766	0.028	99.3	**0.792**	**0.027**	99.4
	30°	0.798	0.049	**100**	0.707	0.046	99.5	ND	ND	ND	0.732	**0.032**	98.7	**0.818**	0.039	99.2
	45°	0.78	0.046	**100**	0.744	0.056	99	ND	ND	ND	0.731	**0.029**	98.5	**0.82**	0.045	99.1
	60°	0.784	0.049	**100**	0.713	0.049	99.7	ND	ND	ND	0.732	**0.039**	97.5	**0.809**	0.047	96.2
	75°	**0.788**	0.056	31.6	0.698	0.057	**99.4**	ND	ND	ND	0.741	0.066	81.7	0.785	**0.045**	82.8
45°	0°	0.767	**0.026**	**100**	0.677	0.036	99.5	0.624	0.005	3.49	0.769	0.027	98.5	**0.805**	0.034	98.6
	30°	0.797	0.047	**100**	0.706	0.051	99	ND	ND	ND	0.756	**0.033**	99.2	**0.821**	0.04	98.4
	45°	0.799	0.053	**100**	0.704	0.05	99.7	ND	ND	ND	0.758	**0.035**	98.7	**0.819**	0.047	98
	60°	0.791	0.051	**100**	0.679	**0.036**	98.6	ND	ND	ND	0.766	0.048	97.7	**0.807**	0.05	92.7
	75°	**0.813**	0.069	45.8	0.676	0.052	**99.7**	ND	ND	ND	0.74	0.086	90.8	0.786	**0.047**	88.1
60°	0°	0.76	**0.023**	**99.9**	0.692	0.036	99.4	0.624	0.004	3.92	0.792	0.027	99.3	**0.798**	0.027	98.8
	30°	0.805	0.05	**100**	0.696	0.047	**100**	ND	ND	ND	0.766	0.037	99.2	**0.808**	**0.03**	98.8
	45°	0.805	0.056	**100**	0.687	0.04	99.6	ND	ND	ND	0.771	**0.039**	99	**0.814**	**0.039**	98.7
	60°	0.799	0.056	**100**	0.684	**0.04**	99.2	ND	ND	ND	0.782	0.047	98.8	**0.816**	0.048	97.5
	75°	**0.818**	0.079	25.9	0.663	**0.032**	**99.4**	ND	ND	ND	0.775	0.055	94.8	0.774	0.043	91.5

Table 7: Average similarity (ranging from 0 to 1), as defined in Equation 3, between the estimated and real pose of all marker systems for each rotation combination. The optimal value of each metric has been highlighted for each rotation combination (row) and tag. "ND" indicates that the marker was not detected or was incorrectly detected in all frames.

Considering the observed similarity in optimal performance between STag and STag2 in the absence of occlusions, we can deduce that the occlusion-resilience changes of STag2 do not seem to impact the general performance.
Looking at the holder angles, we can see that the results with holder roll at 30° and 60° are comparable. For this reason, we will exclude them during the occlusion experiments, since they are expected to perform roughly the same.

7.1.3 Further analysis of STag vs STag2 results

As highlighted in Section 7.1.2, the analysis of Table 5 reveals unexpected outcomes regarding the performance of STag2 in comparison to STag. Despite having a smaller encoding circle, STag2 outperforms STag, which is consistent with the observations presented in Table 6 and Table 7. To investigate the unexpected performance results of STag and STag2, we employed a visualization technique to analyze the errors throughout their entire trajectory. Specifically, we focused on the

scenario where the holder pitch and holder roll were both set to 0. The visualization of these errors is depicted in Figure 36.

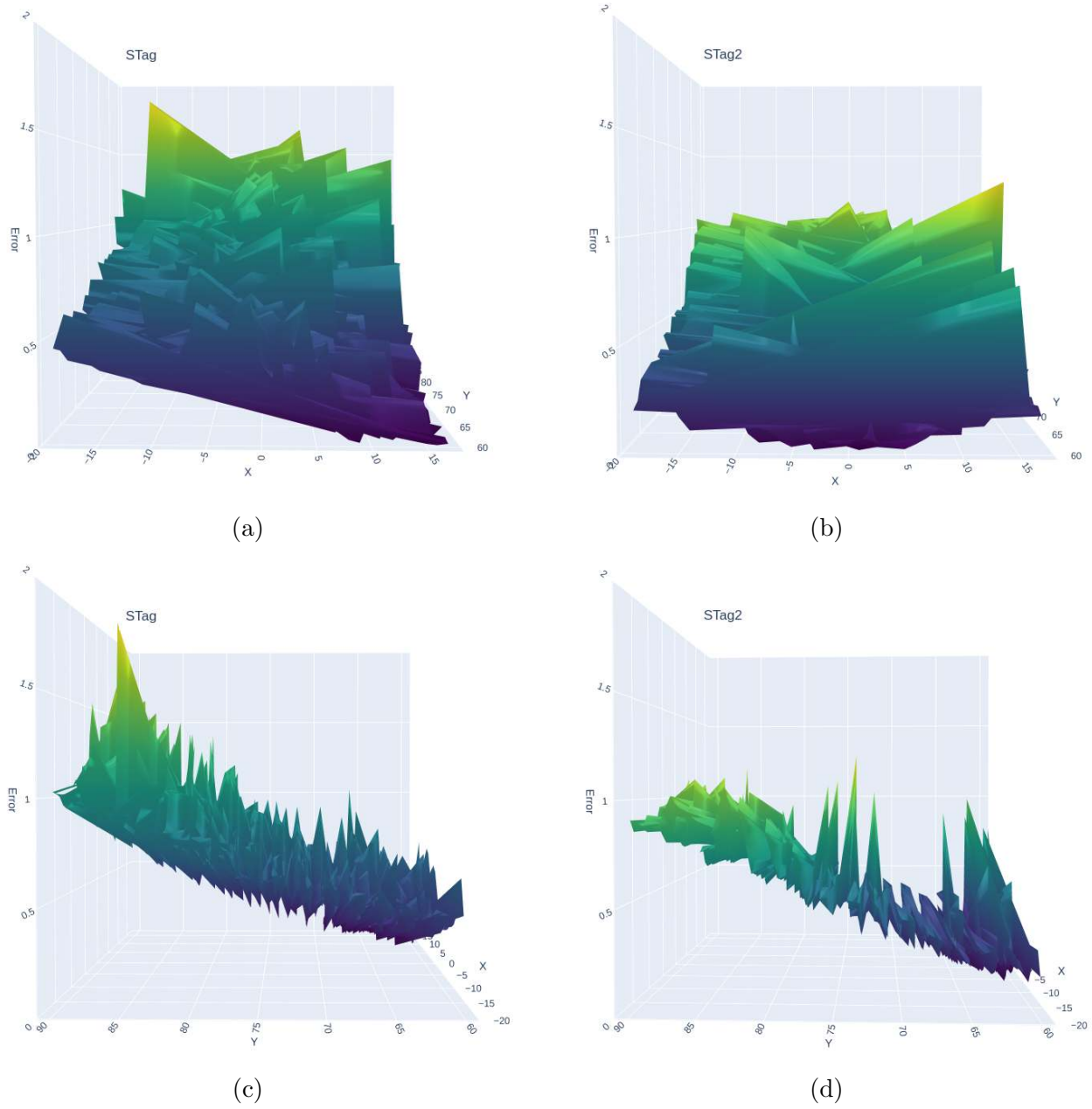


Figure 36: STag error plot from the front (a) and the side(c) and STag2 error plots from the front (b) and the side (d).

Upon examining these two plots, the results initially suggest that STag2 outperforms STag. However, upon closer inspection, we observe that the error patterns for STag exhibit relatively small systematic errors, ranging from approximately 0 to ± 0.2 cm over a 10 cm range. These errors may be attributed to minor variations in camera position or orientation during the experiments. These systematic errors are present to varying (mostly low) degrees in all of the experiment. It is important to

note that these errors have an impact on the significance of the observed performance differences. Addressing and mitigating these systematic errors in the performance measurements could be an avenue for future research to explore.

7.2 STag2: Occlusion results

In order to compare the occlusion results, we represented the results for the occlusion experiments in the same way as in Section 7.1.1.

7.2.1 STag2 (black square): Single corner occlusion

The single corner experiments work the same across both marker systems, meaning that, for STag2, the white rhombus detections do not get used.

		STag						STag2					
		first (BS)			last (BS)			first (WR)			last (BS)		
Holder roll	Holder pitch	mean (cm)	std. (cm)	det. rate (%)	mean (cm)	std. (cm)	det. rate (%)	mean (cm)	std. (cm)	det. rate (%)	mean (cm)	std. (cm)	det. rate (%)
0°	0°	**1.25**	**0.643**	**64.2**	**1.25**	**0.643**	**64.2**	2.23	1.17	45.3	2.23	1.17	45.3
	30°	**1.06**	**1**	**95.4**	**1.06**	**1**	**95.4**	1.31	1.22	66.2	1.31	1.21	66.2
	45°	**0.951**	**0.541**	**95.7**	**0.951**	**0.541**	**95.7**	1.04	1.22	52.8	1	1.11	52.8
	60°	**0.877**	**1.27**	**85.4**	**0.877**	**1.27**	**85.4**	0.972	1.54	10.7	0.972	1.54	10.7
	75°	**0.855**	**0.382**	**7.14**	**0.855**	**0.382**	**7.14**	10.3	1.38	1.03	10.3	1.37	1.03
30°	0°	**1.1**	**0.697**	**50.6**	**1.1**	**0.697**	**50.6**	1.43	1.02	46.5	1.43	1.02	46.5
	30°	**0.995**	**0.831**	**68.3**	**0.995**	**0.831**	**68.3**	1.68	1.52	44.3	1.68	1.52	44.3
	45°	**1.2**	**1.07**	**67.2**	**1.2**	**1.07**	**67.2**	1.42	1.13	42.1	1.42	1.13	42.1
	60°	**1.42**	**0.959**	**60.7**	**1.42**	**0.959**	**60.7**	2.11	1.58	6.51	2.11	1.58	6.51
	75°	**1.61**	**1.01**	**30.9**	**1.61**	**1.01**	**30.9**	3.08	2.23	0.434	3.08	2.23	0.434
45°	0°	**1.05**	**0.821**	**54.1**	**1.05**	**0.821**	**54.1**	1.38	1.21	32.1	1.38	1.21	32.1
	30°	**1.67**	1.68	**42**	**1.67**	1.68	**42**	1.8	**1.39**	20.9	1.8	**1.39**	20.9
	45°	2.86	3.5	**36.2**	2.86	3.5	**36.2**	**1.59**	**1.16**	18	**1.59**	**1.16**	18
	60°	**1.9**	**1.1**	**30.7**	**1.9**	**1.1**	**30.7**	1.97	1.43	9.95	1.97	1.43	9.95
	75°	**1.56**	**1.16**	**22.9**	**1.56**	**1.16**	**22.9**	2.2	1.96	2.67	2.2	1.96	2.67

Table 8: Average distance (in cm) between estimated and real pose locations for STags with top right corner occluded per rotation combination.The optimal value of each metric has been highlighted for each rotation combination (row) and tag.

Looking at the distance results for the single corner occlusion results, as listed in Table 8, the first thing we notice is that the first and last values are the same for both marker systems. This is because there is only one detection in the detection array as a result of the occlusion. Besides that, we can clearly see that STag outperforms STag2. In terms of detection rate, STag performs better due to having a larger encoding circle. Since STag2 has a smaller encoding circle, it can be harder

to detect after a certain distance.

		STag						STag2					
		first (BS)			last (BS)			first (WR)			last (BS)		
Holder roll	Holder pitch	mean (°)	std. (°)	det. rate (%)	mean (°)	std. (°)	det. rate (%)	mean (°)	std. (°)	det. rate (%)	mean (°)	std. (°)	det. rate (%)
0°	0°	**13.4**	**6.56**	**64.2**	**13.4**	**6.56**	**64.2**	16.2	8.08	45.3	16.2	8.08	45.3
	30°	**10.5**	**20.4**	**95.4**	**10.5**	**20.4**	**95.4**	14.9	23.4	66.2	14.9	23.4	66.2
	45°	**4.47**	**15.4**	**95.7**	**4.47**	**15.4**	**95.7**	9.4	24.4	52.8	9.31	24.4	52.8
	60°	**4.94**	**16**	**85.4**	**4.94**	**16**	**85.4**	7.49	23.1	10.7	7.49	23.1	10.7
	75°	**6.71**	**23.2**	**7.14**	**6.71**	**23.2**	**7.14**	15.7	45.1	1.03	15.7	45.1	1.03
30°	0°	**14**	**6.38**	**50.6**	**14**	**6.38**	**50.6**	14.6	7.71	46.5	14.6	7.71	46.5
	30°	**3.54**	10.4	**68.3**	**3.54**	10.4	**68.3**	5.32	**9.33**	44.3	5.32	**9.33**	44.3
	45°	3.69	12.6	**67.2**	3.69	12.6	**67.2**	**3.2**	**10.1**	42.1	**3.2**	**10.1**	42.1
	60°	2.11	10.5	**60.7**	2.11	10.5	**60.7**	2.08	1.02	6.51	**2.04**	**0.877**	6.51
	75°	7.06	20.5	**30.9**	7.06	20.5	**30.9**	**1.6**	**0.401**	0.434	**1.6**	**0.401**	0.434
45°	0°	**12.3**	**6.25**	**54.1**	**12.3**	**6.25**	**54.1**	15	7.07	32.1	15	7.07	32.1
	30°	8.4	16.2	**42**	8.4	16.2	**42**	**6.41**	**12.6**	20.9	**6.41**	**12.6**	20.9
	45°	10.4	23.2	**36.2**	10.4	23.2	**36.2**	**2**	**0.881**	18	**2**	**0.881**	18
	60°	8.16	19.6	**30.7**	8.16	19.6	**30.7**	**1.8**	**0.86**	9.95	**1.8**	**0.86**	9.95
	75°	9.12	22	**22.9**	9.12	22	**22.9**	**3.56**	**0.781**	2.67	**3.56**	**0.781**	2.67

Table 9: Average difference between estimated and real pose rotation angles (in degrees) for STags with top right corner occluded per rotation combination. The optimal value of each metric has been highlighted for each rotation combination (row) and tag.

In terms of angle difference, as listed in Table 9, both STag and STag2 perform somewhat similarly (keeping in mind that the results are in degrees) where STag2 has an acceptable detection rate. The limitations on the performance of STag2 here are similar to the distance results. The detection rate is low due to the smaller encoding circle. Looking at the last few values with Holder roll at 45°, we can see that STag2 performs better. However, these results are not comparable since the detection rate of STag2 is significantly lower than the detection rate of STag.

Holder roll	Holder pitch	STag						STag2					
		first (BS)			last (BS)			first (WR)			last (BS)		
		mean	std.	det. rate (%)	mean	std.	det. rate (%)	mean	std.	det. rate (%)	mean	std.	det. rate (%)
0°	0°	**0.727**	**0.053**	**64.2**	**0.727**	**0.053**	**64.2**	0.662	0.062	45.3	0.662	0.062	45.3
	30°	**0.744**	**0.077**	**95.4**	**0.744**	**0.077**	**95.4**	0.722	0.101	66.2	0.722	0.101	66.2
	45°	**0.756**	**0.065**	**95.7**	**0.756**	**0.065**	**95.7**	0.754	0.111	52.8	**0.756**	0.11	52.8
	60°	0.769	**0.068**	**85.4**	0.769	**0.068**	**85.4**	**0.785**	0.104	10.7	**0.785**	0.104	10.7
	75°	**0.765**	**0.094**	**7.14**	**0.765**	**0.094**	**7.14**	0.502	0.138	1.03	0.502	0.137	1.03
30°	0°	**0.748**	**0.065**	**50.6**	**0.748**	**0.065**	**50.6**	0.722	0.073	46.5	0.722	0.073	46.5
	30°	**0.771**	**0.077**	**68.3**	**0.771**	**0.077**	**68.3**	0.725	0.092	44.3	0.725	0.092	44.3
	45°	**0.753**	**0.091**	**67.2**	**0.753**	**0.091**	**67.2**	0.74	0.098	42.1	0.74	0.098	42.1
	60°	**0.731**	**0.084**	**60.7**	**0.731**	**0.084**	**60.7**	0.698	0.087	6.51	0.698	0.087	6.51
	75°	**0.707**	**0.102**	**30.9**	**0.707**	**0.102**	**30.9**	0.667	0.105	0.434	0.667	0.105	0.434
45°	0°	**0.758**	**0.061**	**54.1**	**0.758**	**0.061**	**54.1**	0.732	0.074	32.1	0.732	0.074	32.1
	30°	**0.721**	0.104	**42**	**0.721**	0.104	**42**	0.711	**0.098**	20.9	0.711	**0.098**	20.9
	45°	0.651	0.103	**36.2**	0.651	0.103	**36.2**	**0.729**	**0.088**	18	**0.729**	**0.088**	18
	60°	0.683	0.094	**30.7**	0.683	0.094	**30.7**	**0.707**	**0.092**	9.95	**0.707**	**0.092**	9.95
	75°	0.715	**0.104**	**22.9**	0.715	**0.104**	**22.9**	**0.718**	0.117	2.67	**0.718**	0.117	2.67

Table 10: Average similarity (ranging from 0 to 1), as defined in Equation 3, between the estimated and real pose of STag and STag2, with top right corner occluded, for each rotation combination. The optimal value of each metric has been highlighted for each rotation combination (row) and tag.

Combining the distance and rotation results together, as listed in Table 10, we can deduce that STag performs better than STag2 for single corner occlusion. The reason behind this is that, although (looking at appendix B) STag has a lower minimum performance, STag2 demonstrates a higher occurrence of subpar results and a lower detection rate. This is evident when comparing any of the Heatmaps between STag and STag2 in appendix B.
The lower detection rate can be attributed to the smaller size of the encoding circle of STag2, reducing the detection rate as the distance between the marker and camera increases. The high occurrence of subpar results is caused by the fact that more of the edges of the STag2 marker can be occluded. Although this may initially seem to improve occlusion resilience, the algorithm has a harder time correctly estimating the occluded corner, due to the lack of visible features to estimate the covered edges of the marker (see Fig. 37).

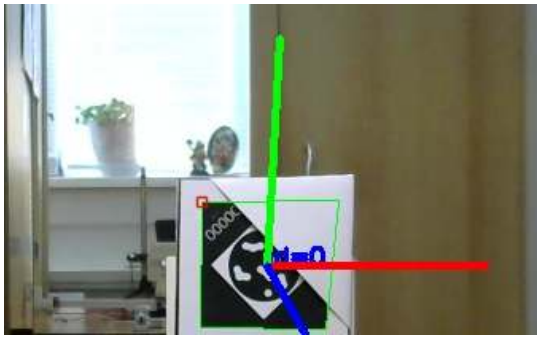


Figure 37: Incorrect estimation of occluded corner.

Using this incorrect detection to estimate the homography has a negative impact on the rotation estimation, leading to inaccurate pose estimation results. This and the low detection rate could be improved in future research by adjusting the STag2 marker design to have a bigger encoding circle.

7.2.2 STag2 (white rhombus): Half occlusion

Half occlusion is the main improvement STag2 tries to achieve and is not applicable to STag.

		STag2					
		first (WR)			last (BS)		
Holder roll	Holder pitch	mean (cm)	std. (cm)	det. rate (%)	mean (cm)	std. (cm)	det. rate (%)
0°	0°	3.41	2.28	15.2	3.41	2.28	15.2
	30°	4.54	2.14	22.9	4.54	2.14	22.9
	45°	5.76	2.2	21.8	5.76	2.2	21.8
	60°	8.9	1.34	0.741	8.9	1.34	0.741
30°	0°	2.17	1.58	40.5	2.17	1.58	40.5
	30°	2.63	1.59	34.5	2.63	1.59	34.5
	45°	3.32	2.5	30.1	3.32	2.5	30.1
	60°	4.26	1.64	1.05	4.26	1.64	1.05
45°	0°	3.19	2.23	24.5	3.19	2.23	24.5
	30°	2.21	1.76	30.8	2.21	1.76	30.8
	45°	2.06	2.04	22.2	2.06	2.04	22.2
	60°	1.68	1.54	1.63	1.68	1.54	1.63

Table 11: Average distance (in cm) between estimated and real pose locations per rotation combination for STags occluded about halfway.The optimal value of each metric has been highlighted for each rotation combination (row) and tag.

Looking at the distance results for half occlusion, as listed in Table 11, we can see that these results are not good. This has to do with the fact that the STag2 marker detections get less reliable at longer distances from the camera. This applies to the white rhombus, which in turn also applies to the encoding circle, due to their small size. The same idea of incorrect occluded corner estimation (see Fig. 37) also applies here. The more the corner of the white rhombus is occluded, the more inaccurate the detection.

		STag2					
		first (WR)			last (BS)		
Holder roll	Holder pitch	mean (°)	std. (°)	det. rate (%)	mean (°)	std. (°)	det. rate (%)
0°	0°	21.9	5.9	15.2	21.9	5.9	15.2
	30°	32.9	30.9	22.9	32.9	30.9	22.9
	45°	26.2	38.9	21.8	26.2	38.9	21.8
	60°	4.32	0.66	0.741	4.32	0.66	0.741
30°	0°	17.2	7.67	40.5	17.2	7.67	40.5
	30°	48	27.3	34.5	48	27.3	34.5
	45°	55.3	44.6	30.1	55.3	44.6	30.1
	60°	84.1	58	1.05	84.1	58	1.05
45°	0°	22	8.86	24.5	22	8.86	24.5
	30°	36.2	26.3	30.8	36.2	26.3	30.8
	45°	58.5	40.7	22.2	58.5	40.7	22.2
	60°	97.1	49.3	1.63	97.1	49.3	1.63

Table 12: Average difference between estimated and real pose rotation angles (in degrees) per rotation combination for STags occluded about halfway. The optimal value of each metric has been highlighted for each rotation combination (row) and tag.

In terms of the angle difference results for the half occluded STag2 experiments, as listed in Table 12, we see very poor performance. The main culprit for this is still the inaccuracy of the detections increasing at greater distances, as mentioned before, but worsening the performance even more is the inverted detections (see Fig. 35). Higher mean values imply more inverted detections. We can see that this gets worse for higher Holder pitch values. This shows that the impact of inverted detections is minimal on the results obtained at a Holder pitch of 0°. It is worth noting that future research endeavors could explore methods for correcting these inverted detections, potentially leading to improved results.

Holder roll	Holder pitch	STag2 first (WR) mean	STag2 first (WR) std.	STag2 first (WR) det. rate (%)	STag2 last (BS) mean	STag2 last (BS) std.	STag2 last (BS) det. rate (%)
0°	0°	0.626	0.078	15.2	0.626	0.078	15.2
	30°	0.536	0.093	22.9	0.536	0.093	22.9
	45°	0.516	0.12	21.8	0.516	0.12	21.8
	60°	0.551	0.007	0.741	0.551	0.007	0.741
30°	0°	0.678	0.077	40.5	0.678	0.077	40.5
	30°	0.562	0.108	34.5	0.562	0.108	34.5
	45°	0.497	0.173	30.1	0.497	0.173	30.1
	60°	0.345	0.177	1.05	0.345	0.177	1.05
45°	0°	0.63	0.076	24.5	0.63	0.076	24.5
	30°	0.624	0.104	30.8	0.624	0.104	30.8
	45°	0.545	0.155	22.2	0.545	0.155	22.2
	60°	0.417	0.165	1.63	0.417	0.165	1.63

Table 13: Average similarity (ranging from 0 to 1), as defined in Equation 3 between the estimated and real pose of STag2, occluded about halfway, for each rotation combination. The optimal value of each metric has been highlighted for each rotation combination (row) and tag.

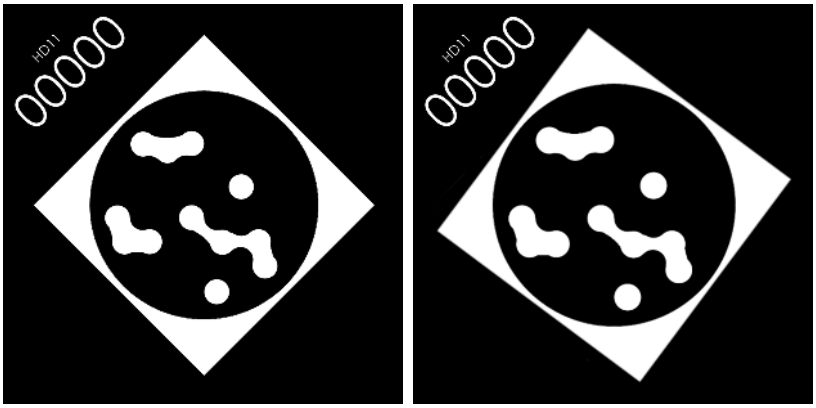
Looking at the distance and angle difference results separately, it is expected that the combination of both will not yield good results. We can see this in the similarity results listed in Table 13. We can see that with Holder pitch at 0°, we still have somewhat reasonable average results, due to the lack of inverted detections. Therefore, it makes sense that the results get increasingly worse, since the amount of inverted detections increases as the Holder pitch increases.

8 Conclusions and Further Research

The main purpose of this paper was to improve the occlusion-resilience of STag. Despite STag2 achieving satisfactory overall performance, we encountered challenges in enhancing the marker's resilience to occlusions. The primary factor contributing to this limitation is the small size of the rhombus, which subsequently results in a reduced encoding circle size. As the distance between the camera and marker increases, the accuracy of detection decreases, rendering the marker less reliabl.s.

Experimental Contributions Compared to the experiments done by Michail Kalaitzakis et al [KCA+20], the main improvement we made is adding another axis (X-axis). This allowed us to conduct our experiments with more degrees of freedom. However, this enhancement also posed challenges when comparing our results, particularly due to differences in distance ranges and methodologies.
Although the shortest distance value (75cm) used in [KCA+20] falls within the range of distances tested in our experiment, direct comparison is not feasible. This is because we analyze our distance results by considering the average distance per holder rotation combination, whereas the study by [KCA+20] presents results at various distances. In terms of the angle experiments, we can make a comparison between the results obtained in [KCA+20] and our own findings. However,these results do not exhibit a strong correlation. One plausible explanation for this might be that we conducted our angle experiments at longer distances from the camera, which could have lead to decreased accuracy.

Significant research opportunities exist to further advance the marker design and enhance the encoding circle, while carefully balancing the tradeoff between occlusion resilience and maximum detection distance, as discussed in Section 5.1 (see Fig. 38). Additionally, conducting future investigations on the performance of STag2 at shorter distances would be valuable, considering that our experiments commenced at a distance of 60 cm from the camera.
Moreover, there are several other promising areas for future research. One such area involves optimizing the minimunm occlusion of the white rhombus to accurately reflect the Bit Error Ratio (BER) of the marker encoding circle, as highlighted in Section 5.2.1. By finding an optimal balance between these factors, the overall performance and reliability of the marker system can be further enhanced.
Furthermore, addressing the issue of inverted detections, as discussed in Section 7.2.2, presents an interesting avenue for future research. Developing effective techniques for correcting inverted detections would contribute to improving the accuracy and robustness of marker detection.
Lastly, addressing and mitigating the systematic errors in the performance measurements discussed in Section 7.1.3 could be an area of future research. Exploring methods to minimize the impact of minor camera position or orientation variations, such as utilizing more stable camera fixation techniques, could help enhance the accuracy and reliability of the experimental setup and results.



(a) (b)

Figure 38: Current STag2 design (a) and possible STag2 design improvement (b)

References

[AT11] Cuneyt Akinlar and Cihan Topal. Edlines: Real-time line segment detection by edge drawing (ed). In *2011 18th IEEE International Conference on Image Processing*, pages 2837–2840, 2011.

[BART11] Filippo Bergamasco, Andrea Albarelli, Emanuele Rodolà, and Andrea Torsello. Rune-tag: A high accuracy fiducial marker with strong occlusion resilience. In *CVPR 2011*, pages 113–120, 2011.

[Braa] G. Bradski. Detection of aruco markers. https://docs.opencv.org/3.3.0/d5/dae/tutorial_aruco_detection.html. Accessed on 19.06.2023.

[Brab] G. Bradski. Detection of charuco boards. https://docs.opencv.org/4.x/df/d4a/tutorial_charuco_detection.html. Accessed on 19.06.2023.

[Brac] G. Bradski. Pose estimation. https://docs.opencv.org/3.4/d7/d53/tutorial_py_pose.html. Accessed on 19.06.2023.

[BTA17] Burak Benligiray, Cihan Topal, and Cuneyt Akinlar. Stag: A stable fiducial marker system. *CoRR*, 2017.

[CC12] Cuneytakinlar and Cihantopal. Edpf: A real-time parameter-free edge segment detector with a false detection control. *International Journal of Pattern Recognition and Artificial Intelligence*, 26, 06 2012.

[Fia05] M. Fiala. Artag, a fiducial marker system using digital techniques. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 590–596 vol. 2, 2005.

[GHS92] Lance B. Gatrell, William A. Hoff, and Cheryl W. Sklair. Robust image features: concentric contrasting circles and their image extraction. In William E. Stoney, editor, *Cooperative Intelligent Robotics in Space II*, volume 1612, pages 235 – 244. International Society for Optics and Photonics, SPIE, 1992.

[GJMSMCMJ14] 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.

[hel08] *The Helmholtz Principle*, pages 31–45. Springer New York, 2008.

[KB99a] 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, 1999.

[KB99b] 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, 1999.

[KCA+20] Michail Kalaitzakis, Sabrina Carroll, Anand Ambrosi, Camden Whitehead, and Nikolaos Vitzilaios. Experimental comparison of fiducial markers for pose estimation. In *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*, pages 781–789, 2020.

[Kum19] Teja Kummarikuntla. Camera calibration with opencv. https://medium.com/analytics-vidhya/camera-calibration-with-opencv-f324679c6eb7, Aug 2019. Accessed on 19.06.2023.

[MBS+20] Rudranarayan Mukherjee, Spencer Backus, Timothy Setterfield, Alexander Brinkman, Gregory Agnes, Eric Sunada, Junggon Kim, Blair Emanuel, Russell Smith, Jason Hyon, Laurie Chappell, John Lymer, and Alfred Tadros. A robotically assembled and serviced science station for earth observations. In *2020 IEEE Aerospace Conference*, pages 1–15, 2020.

[Ols11] Edwin Olson. Apriltag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*, pages 3400–3407, 2011.

[PKA+17] Janne Paavilainen, Hannu Korhonen, Kati Alha, Jaakko Stenros, Elina Koskinen, and Frans Mayra. The pokémon go experience: A location-based augmented reality mobile game goes mainstream. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, page 2493–2498, 2017.

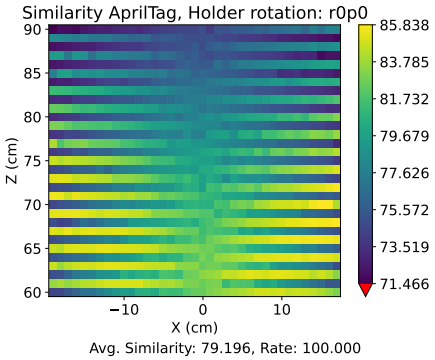
[Sha20] Munesh Kumar Sharma. Augmented reality navigation. *International Journal of Engineering Research and*, V9(06):670–675, June 2020.

[YCC+20] Paulo J. Younse, Chi Yeung Chiu, Jessica E. Cameron, Marco Dolci, Ethan Elliot, Alyssa Ishigo, Dima Kogan, Eloïse Marteau, John Mayo, Jason Munger, Preston Ohta, Jackson W. Strahle, Reg Willson, Eric Peter Olds, and Violet Malyan. Concept for an on-orbit capture and orient module for potential mars sample return. *2020 IEEE Aerospace Conference*, pages 1–22, 2020.

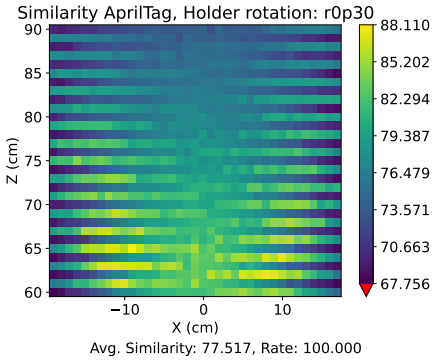
[ZPC21] Simon Zimmermann, Roi Poranne, and Stelian Coros. Go fetch! - dynamic grasps using boston dynamics spot with external robotic arm. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4488–4494, 2021.

A. No occlusion

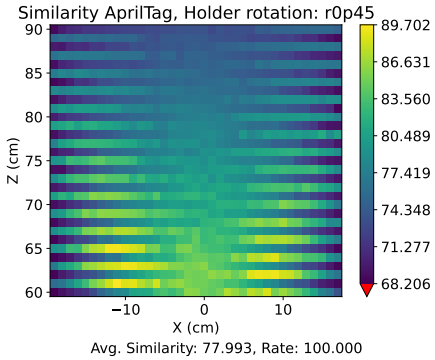
AprilTag



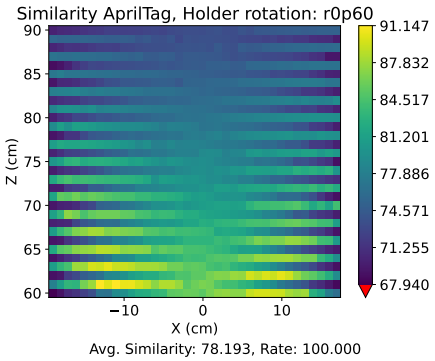
Avg. Similarity: 79.196, Rate: 100.000



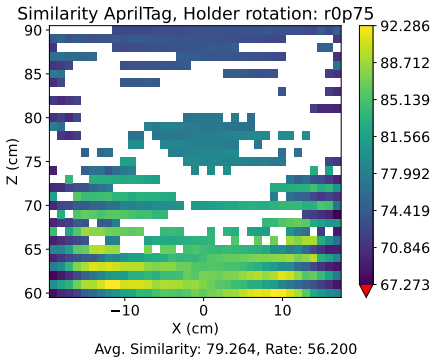
Avg. Similarity: 77.517, Rate: 100.000



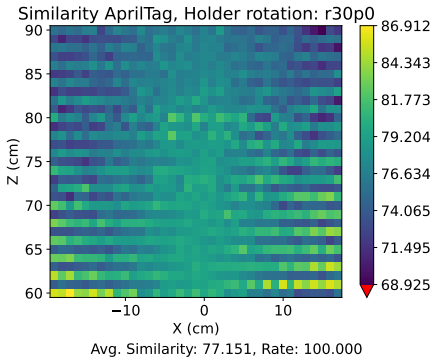
Avg. Similarity: 77.993, Rate: 100.000



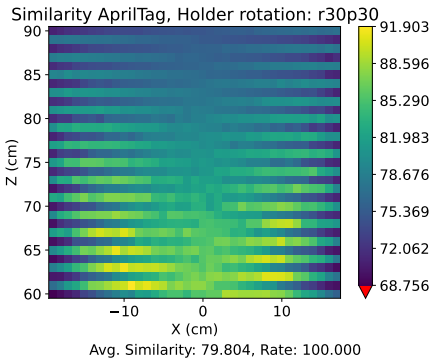
Avg. Similarity: 78.193, Rate: 100.000



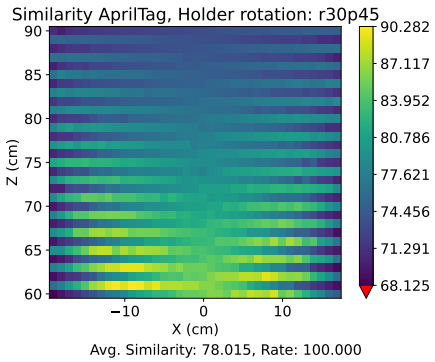
Avg. Similarity: 79.264, Rate: 56.200



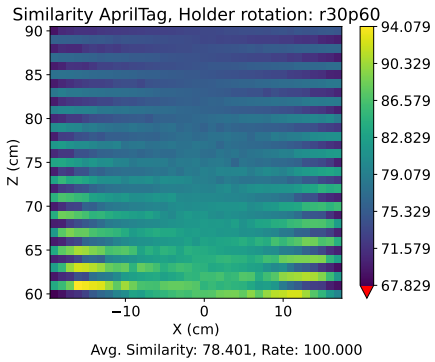
Avg. Similarity: 77.151, Rate: 100.000



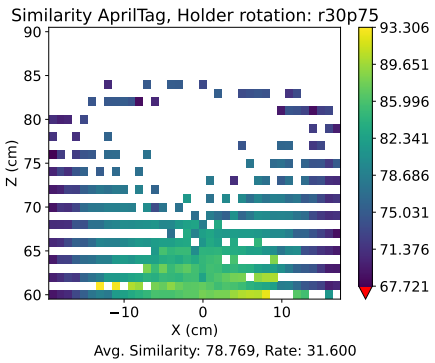
Avg. Similarity: 79.804, Rate: 100.000



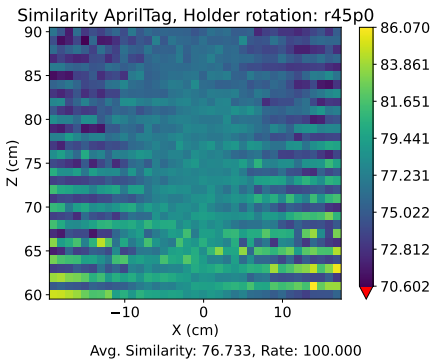
Avg. Similarity: 78.015, Rate: 100.000



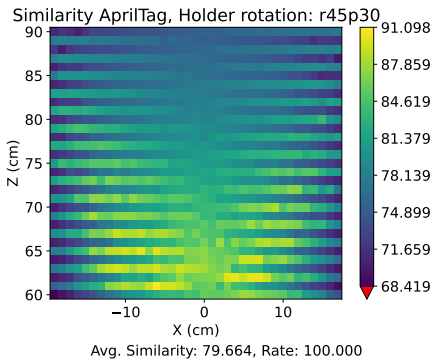
Avg. Similarity: 78.401, Rate: 100.000



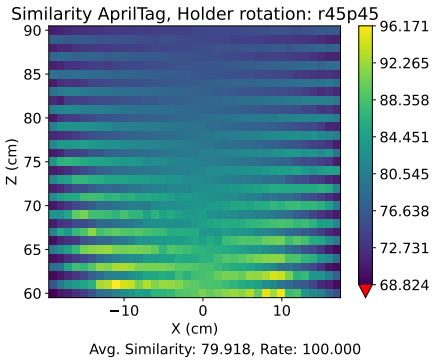
Avg. Similarity: 78.769, Rate: 31.600



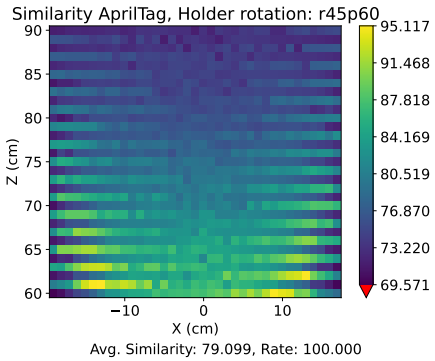
Avg. Similarity: 76.733, Rate: 100.000



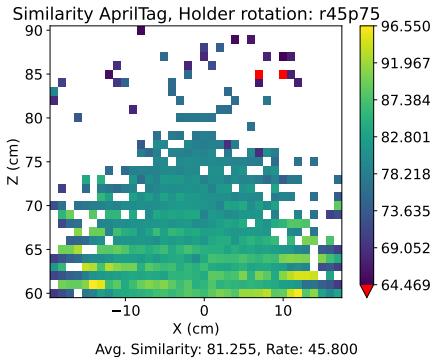
Avg. Similarity: 79.664, Rate: 100.000



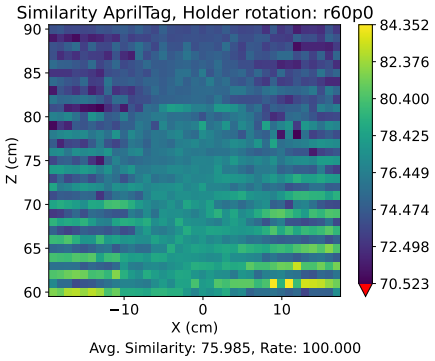
Avg. Similarity: 79.918, Rate: 100.000



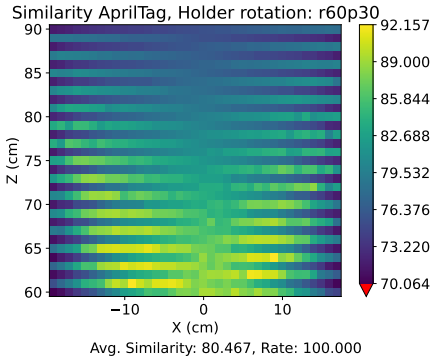
Avg. Similarity: 79.099, Rate: 100.000



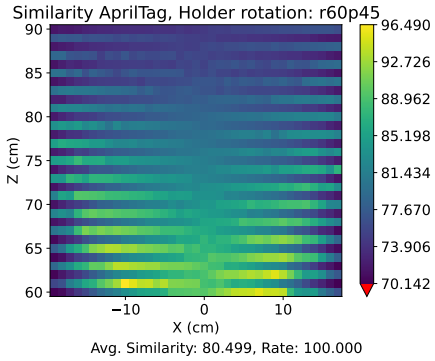
Avg. Similarity: 81.255, Rate: 45.800



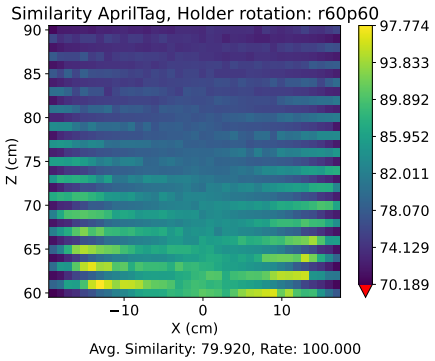
Avg. Similarity: 75.985, Rate: 100.000



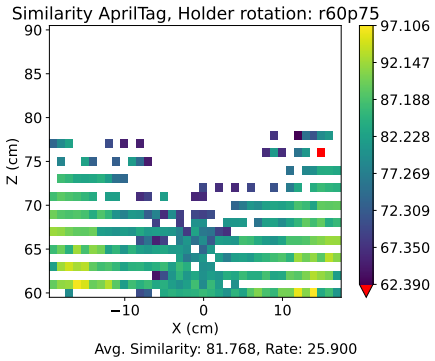
Avg. Similarity: 80.467, Rate: 100.000



Avg. Similarity: 80.499, Rate: 100.000

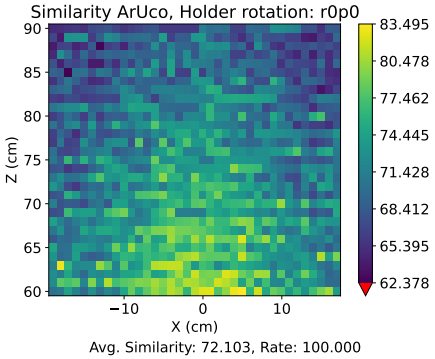


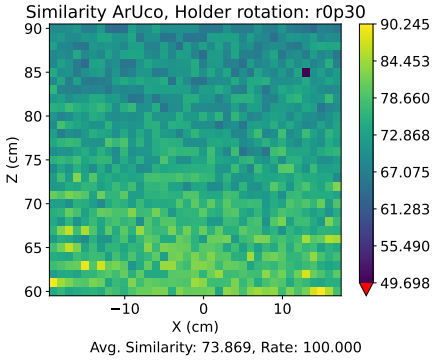
Avg. Similarity: 79.920, Rate: 100.000

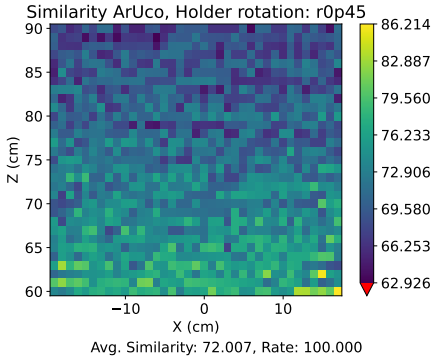


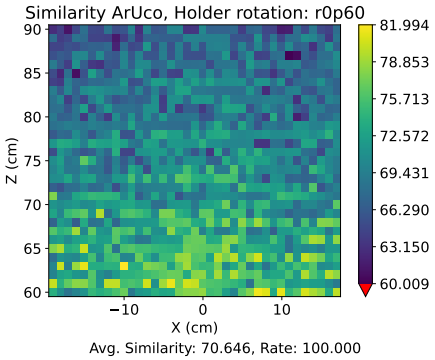
Avg. Similarity: 81.768, Rate: 25.900

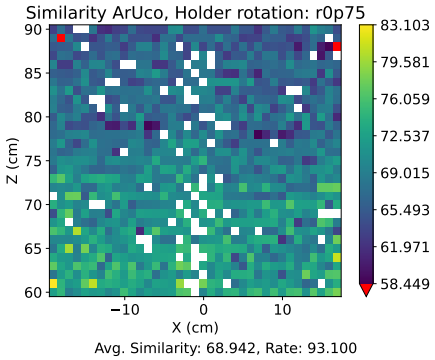
ArUco

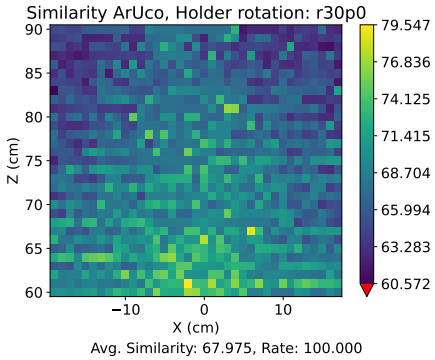

Similarity ArUco, Holder rotation: r0p0
Z (cm)
X (cm)
Avg. Similarity: 72.103, Rate: 100.000

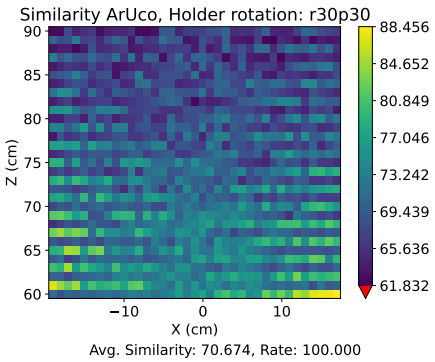

Similarity ArUco, Holder rotation: r0p30
Z (cm)
X (cm)
Avg. Similarity: 73.869, Rate: 100.000

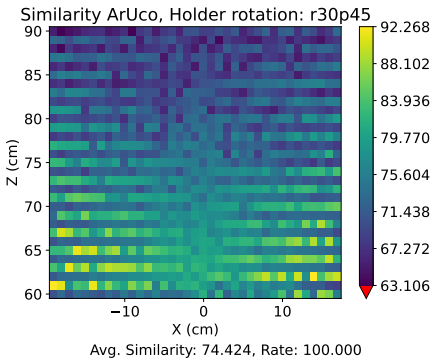

Similarity ArUco, Holder rotation: r0p45
Z (cm)
X (cm)
Avg. Similarity: 72.007, Rate: 100.000

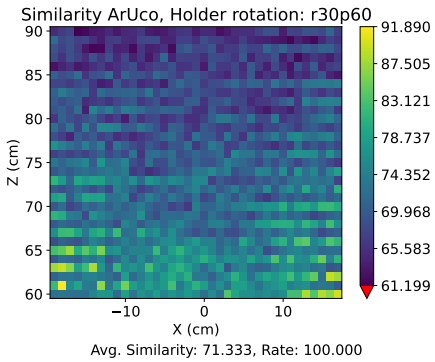

Similarity ArUco, Holder rotation: r0p60
Z (cm)
X (cm)
Avg. Similarity: 70.646, Rate: 100.000

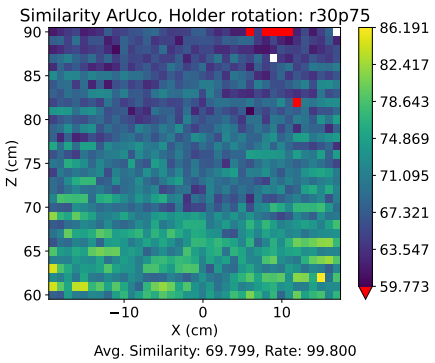

Similarity ArUco, Holder rotation: r0p75
Z (cm)
X (cm)
Avg. Similarity: 68.942, Rate: 93.100

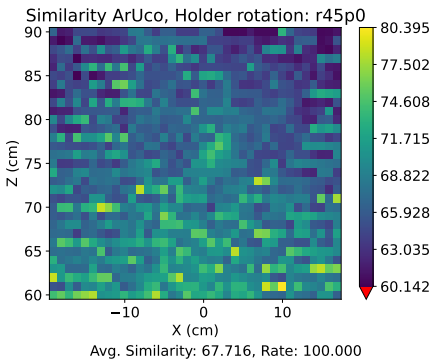

Similarity ArUco, Holder rotation: r30p0
Z (cm)
X (cm)
Avg. Similarity: 67.975, Rate: 100.000

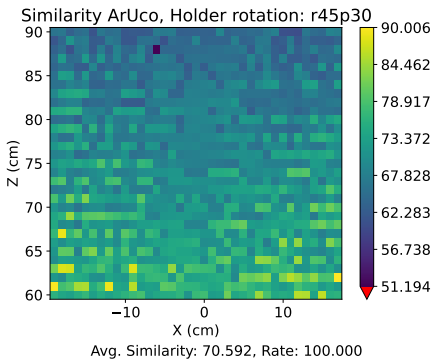

Similarity ArUco, Holder rotation: r30p30
Z (cm)
X (cm)
Avg. Similarity: 70.674, Rate: 100.000

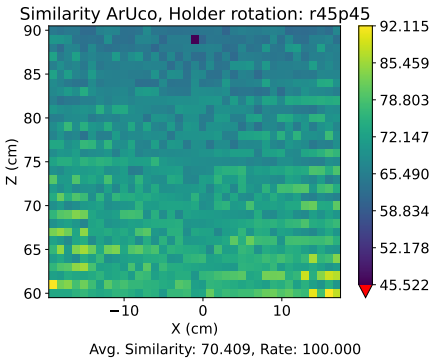

Similarity ArUco, Holder rotation: r30p45
Z (cm)
X (cm)
Avg. Similarity: 74.424, Rate: 100.000


Similarity ArUco, Holder rotation: r30p60
Z (cm)
X (cm)
Avg. Similarity: 71.333, Rate: 100.000

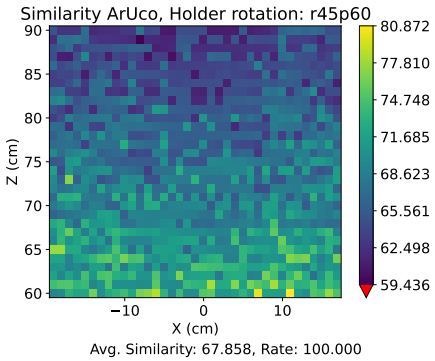

Similarity ArUco, Holder rotation: r30p75
Z (cm)
X (cm)
Avg. Similarity: 69.799, Rate: 99.800


Similarity ArUco, Holder rotation: r45p0
Z (cm)
X (cm)
Avg. Similarity: 67.716, Rate: 100.000

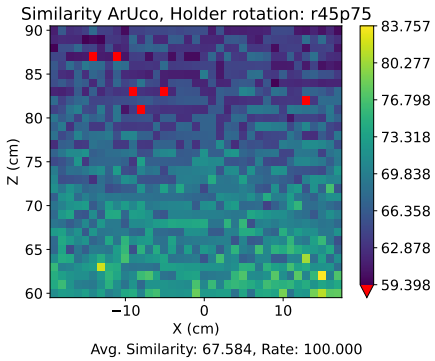

Similarity ArUco, Holder rotation: r45p30
Z (cm)
X (cm)
Avg. Similarity: 70.592, Rate: 100.000



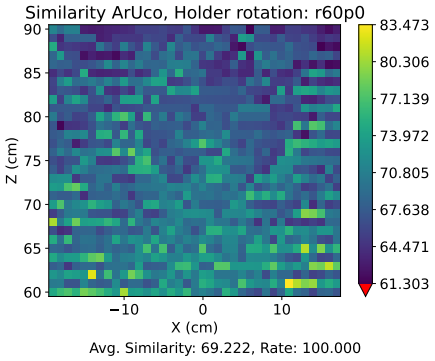
Avg. Similarity: 70.409, Rate: 100.000



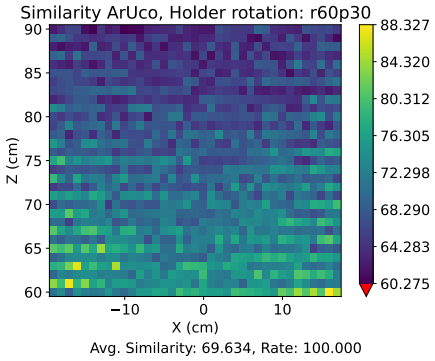
Avg. Similarity: 67.858, Rate: 100.000



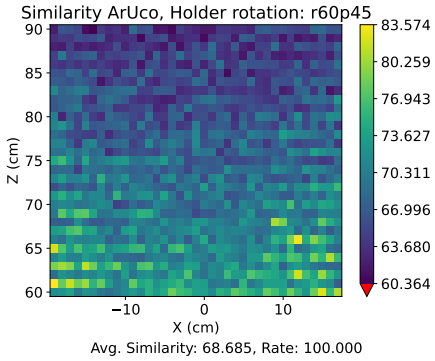
Avg. Similarity: 67.584, Rate: 100.000



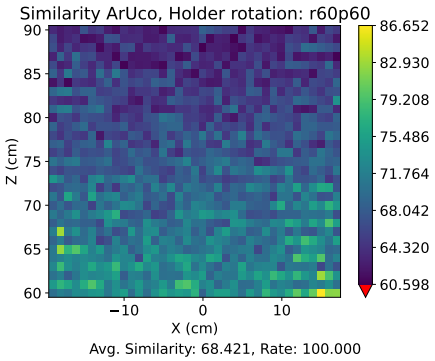
Avg. Similarity: 69.222, Rate: 100.000



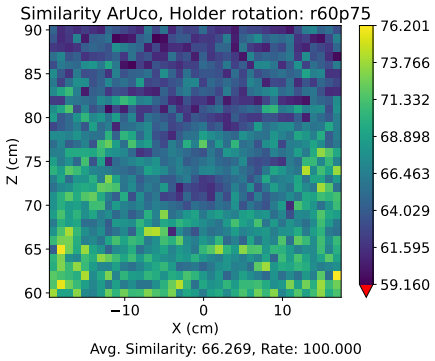
Avg. Similarity: 69.634, Rate: 100.000



Avg. Similarity: 68.685, Rate: 100.000

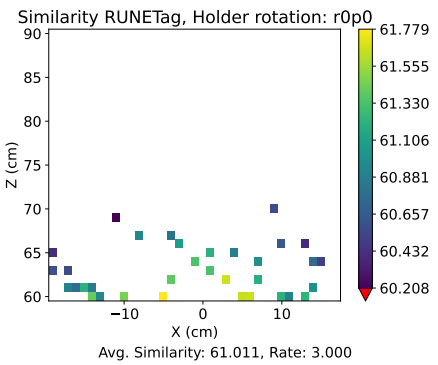


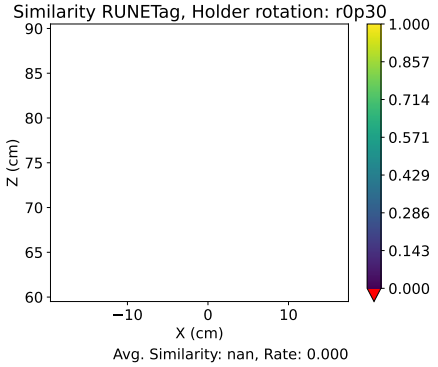
Avg. Similarity: 68.421, Rate: 100.000

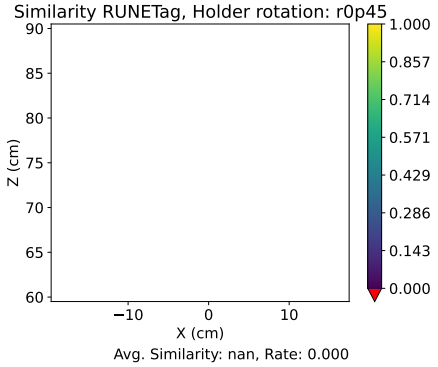


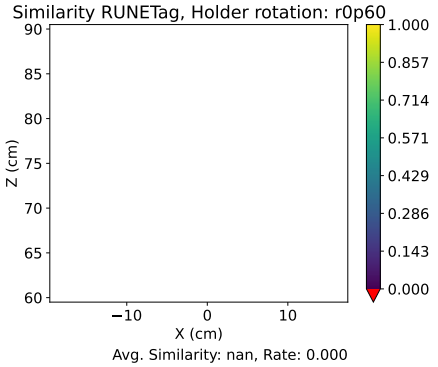
Avg. Similarity: 66.269, Rate: 100.000

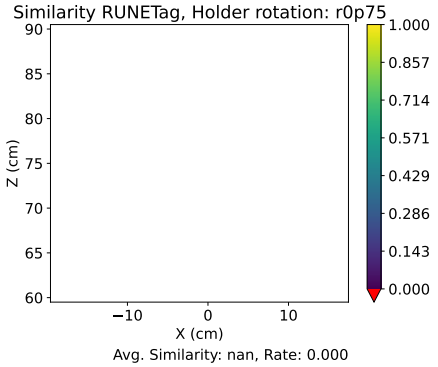
RUNETag

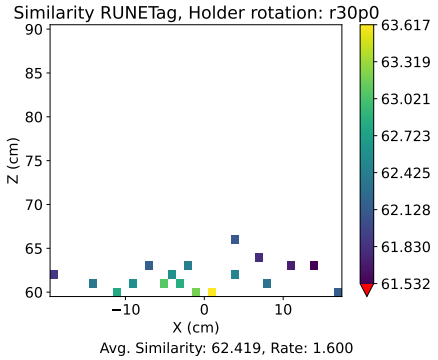

Similarity RUNETag, Holder rotation: r0p0
Z (cm)
X (cm)
Avg. Similarity: 61.011, Rate: 3.000

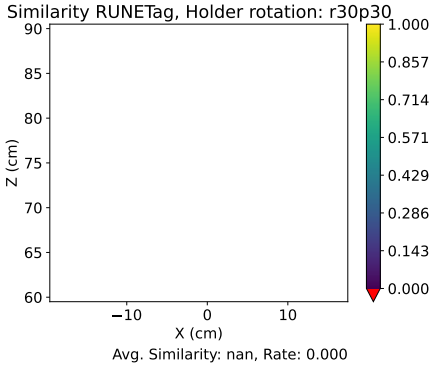

Similarity RUNETag, Holder rotation: r0p30
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

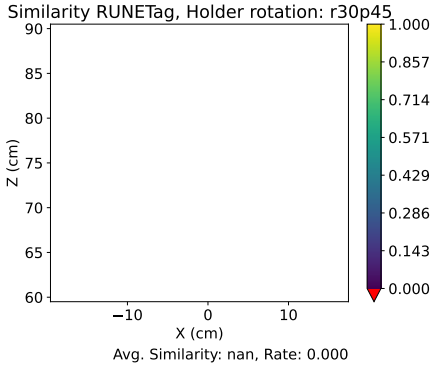

Similarity RUNETag, Holder rotation: r0p45
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

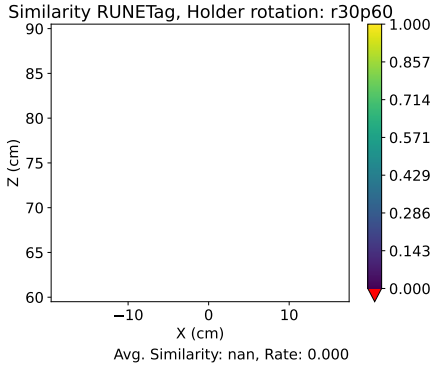

Similarity RUNETag, Holder rotation: r0p60
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

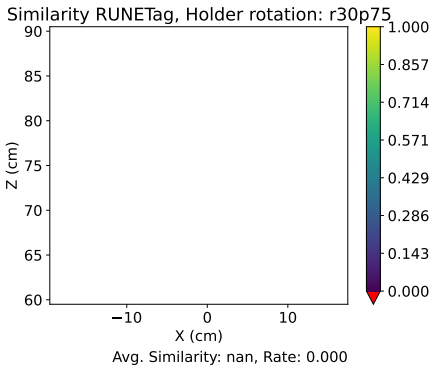

Similarity RUNETag, Holder rotation: r0p75
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

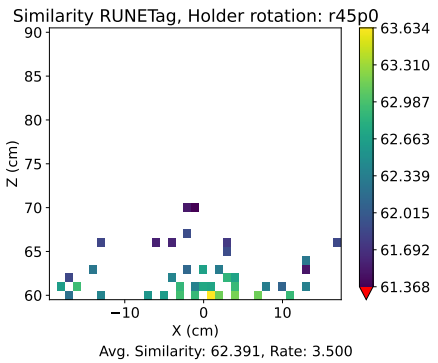

Similarity RUNETag, Holder rotation: r30p0
Z (cm)
X (cm)
Avg. Similarity: 62.419, Rate: 1.600

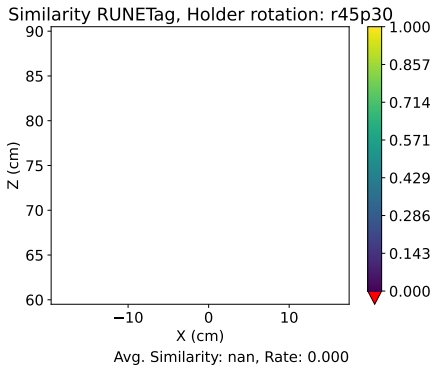

Similarity RUNETag, Holder rotation: r30p30
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

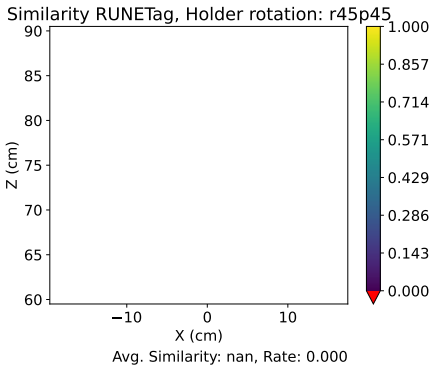

Similarity RUNETag, Holder rotation: r30p45
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

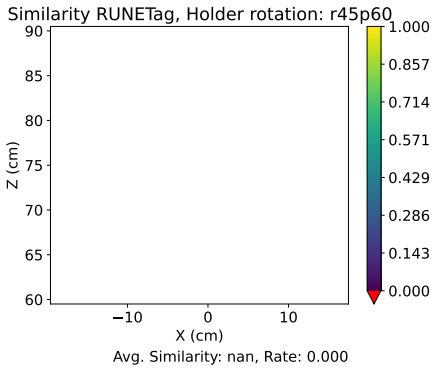

Similarity RUNETag, Holder rotation: r30p60
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

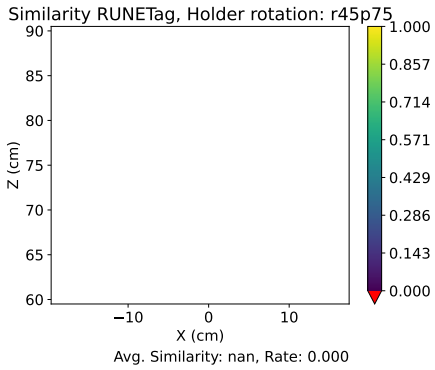

Similarity RUNETag, Holder rotation: r30p75
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

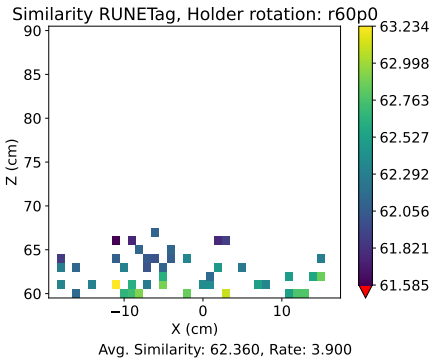

Similarity RUNETag, Holder rotation: r45p0
Z (cm)
X (cm)
Avg. Similarity: 62.391, Rate: 3.500

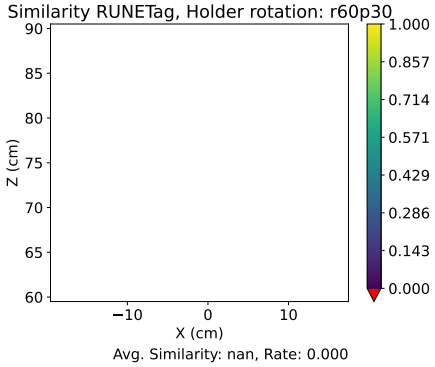

Similarity RUNETag, Holder rotation: r45p30
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

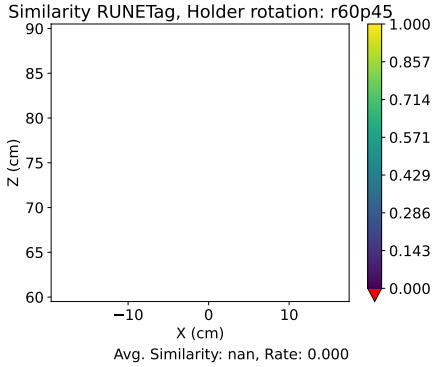

Similarity RUNETag, Holder rotation: r45p45
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

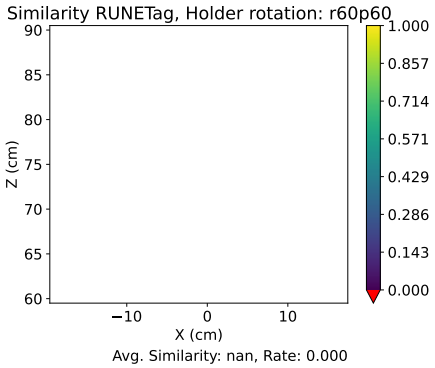

Similarity RUNETag, Holder rotation: r45p60
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

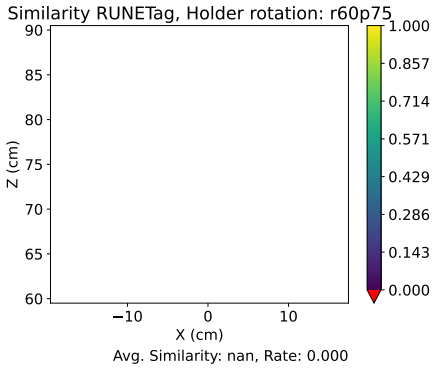

Similarity RUNETag, Holder rotation: r45p75
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000


Similarity RUNETag, Holder rotation: r60p0
Z (cm)
X (cm)
Avg. Similarity: 62.360, Rate: 3.900

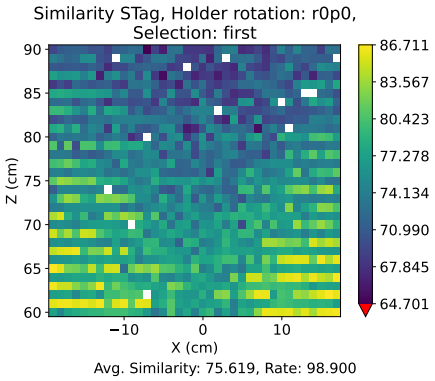

Similarity RUNETag, Holder rotation: r60p30
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000


Similarity RUNETag, Holder rotation: r60p45
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

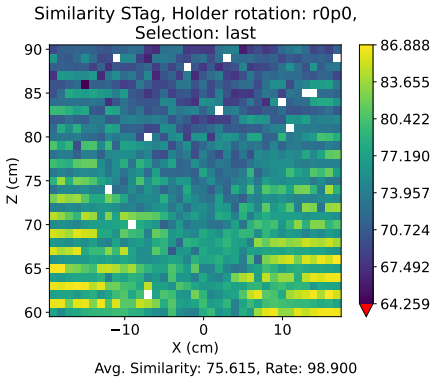

Similarity RUNETag, Holder rotation: r60p60
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000


Similarity RUNETag, Holder rotation: r60p75
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000

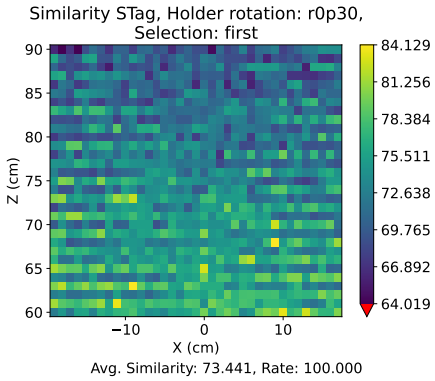
STag



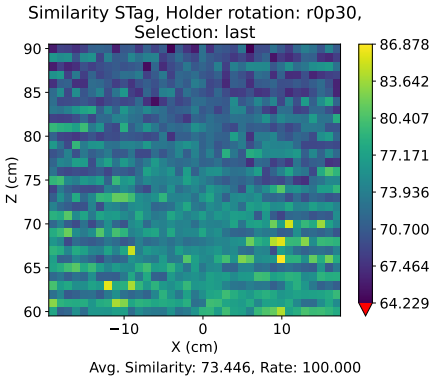
Avg. Similarity: 75.619, Rate: 98.900



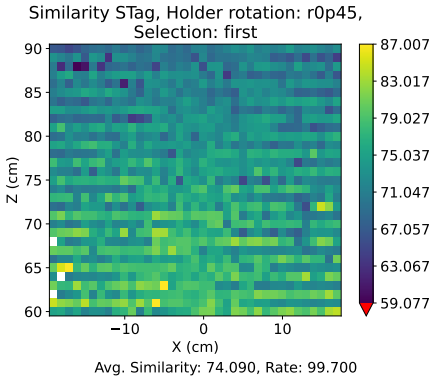
Avg. Similarity: 75.615, Rate: 98.900



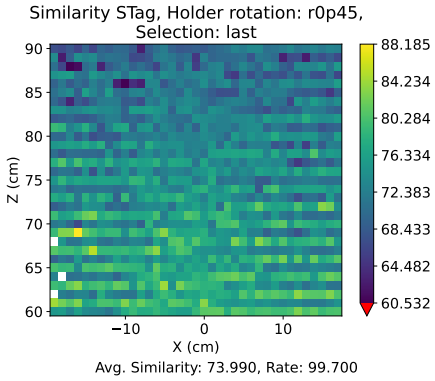
Avg. Similarity: 73.441, Rate: 100.000



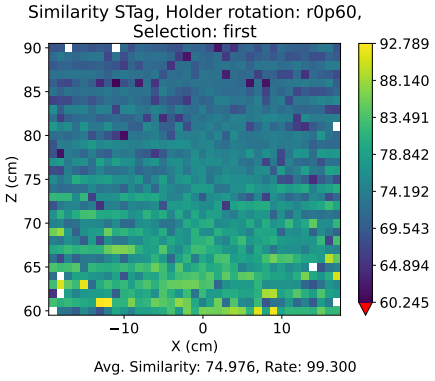
Avg. Similarity: 73.446, Rate: 100.000



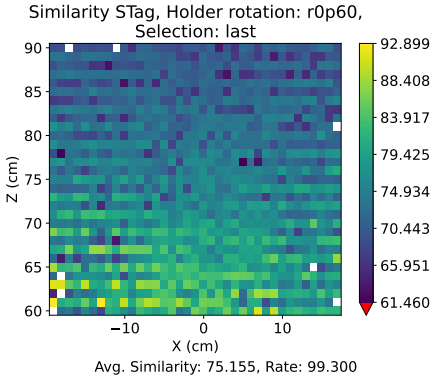
Avg. Similarity: 74.090, Rate: 99.700



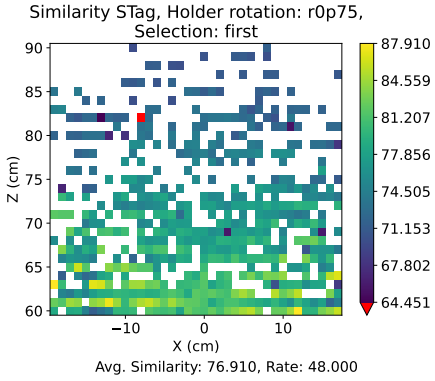
Avg. Similarity: 73.990, Rate: 99.700



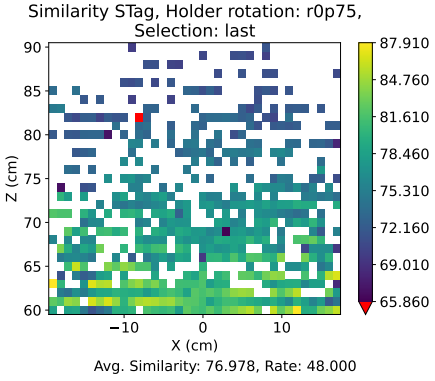
Avg. Similarity: 74.976, Rate: 99.300



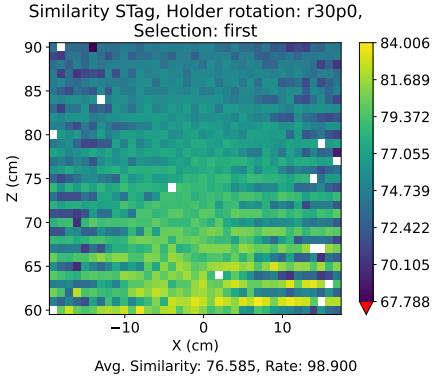
Avg. Similarity: 75.155, Rate: 99.300



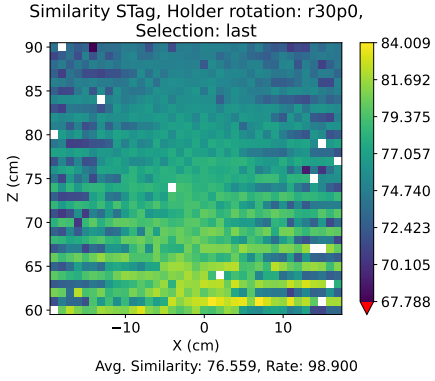
Avg. Similarity: 76.910, Rate: 48.000



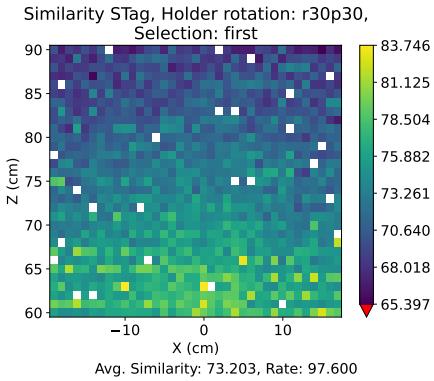
Avg. Similarity: 76.978, Rate: 48.000

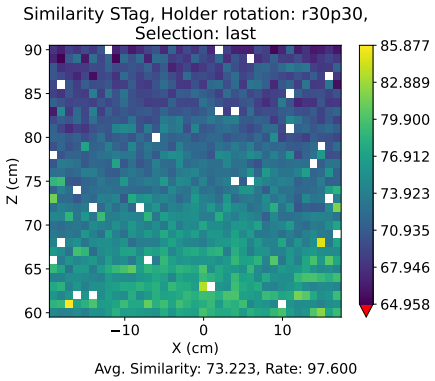


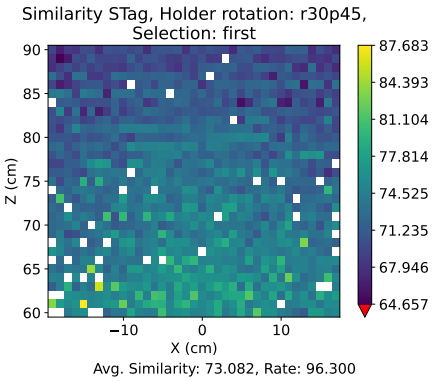
Avg. Similarity: 76.585, Rate: 98.900

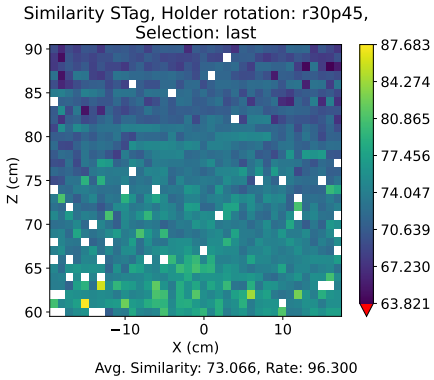


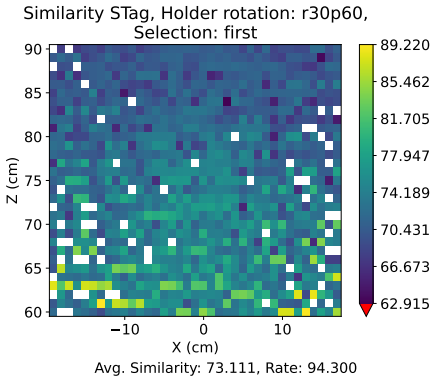
Avg. Similarity: 76.559, Rate: 98.900

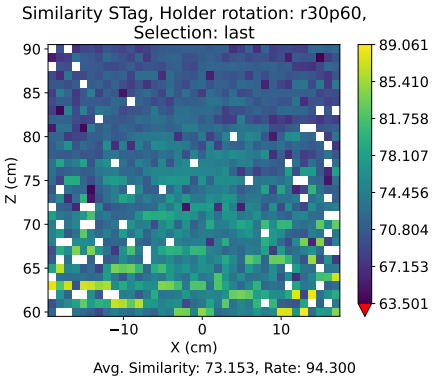

Similarity STag, Holder rotation: r30p30,
Selection: first
Z (cm)
X (cm)
83.746
81.125
78.504
75.882
73.261
70.640
68.018
65.397
Avg. Similarity: 73.203, Rate: 97.600

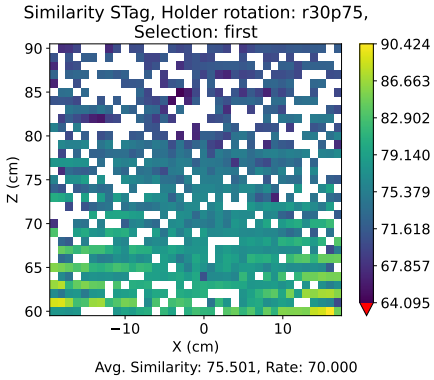

Similarity STag, Holder rotation: r30p30,
Selection: last
Z (cm)
X (cm)
85.877
82.889
79.900
76.912
73.923
70.935
67.946
64.958
Avg. Similarity: 73.223, Rate: 97.600

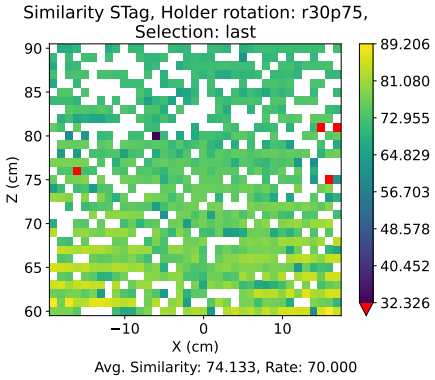

Similarity STag, Holder rotation: r30p45,
Selection: first
Z (cm)
X (cm)
87.683
84.393
81.104
77.814
74.525
71.235
67.946
64.657
Avg. Similarity: 73.082, Rate: 96.300

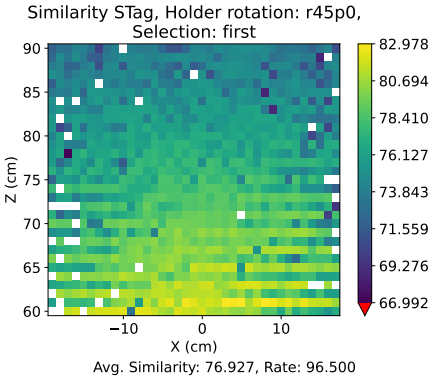

Similarity STag, Holder rotation: r30p45,
Selection: last
Z (cm)
X (cm)
87.683
84.274
80.865
77.456
74.047
70.639
67.230
63.821
Avg. Similarity: 73.066, Rate: 96.300

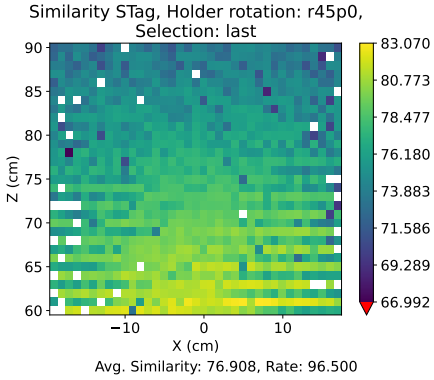

Similarity STag, Holder rotation: r30p60,
Selection: first
Z (cm)
X (cm)
89.220
85.462
81.705
77.947
74.189
70.431
66.673
62.915
Avg. Similarity: 73.111, Rate: 94.300

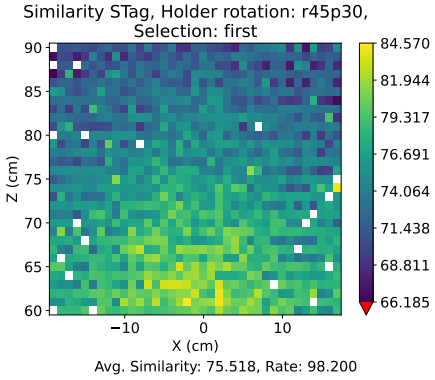

Similarity STag, Holder rotation: r30p60,
Selection: last
Z (cm)
X (cm)
89.061
85.410
81.758
78.107
74.456
70.804
67.153
63.501
Avg. Similarity: 73.153, Rate: 94.300

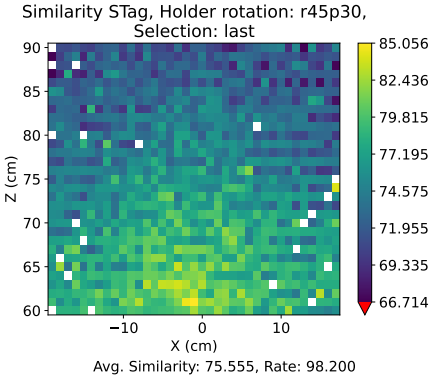

Similarity STag, Holder rotation: r30p75,
Selection: first
Z (cm)
X (cm)
90.424
86.663
82.902
79.140
75.379
71.618
67.857
64.095
Avg. Similarity: 75.501, Rate: 70.000

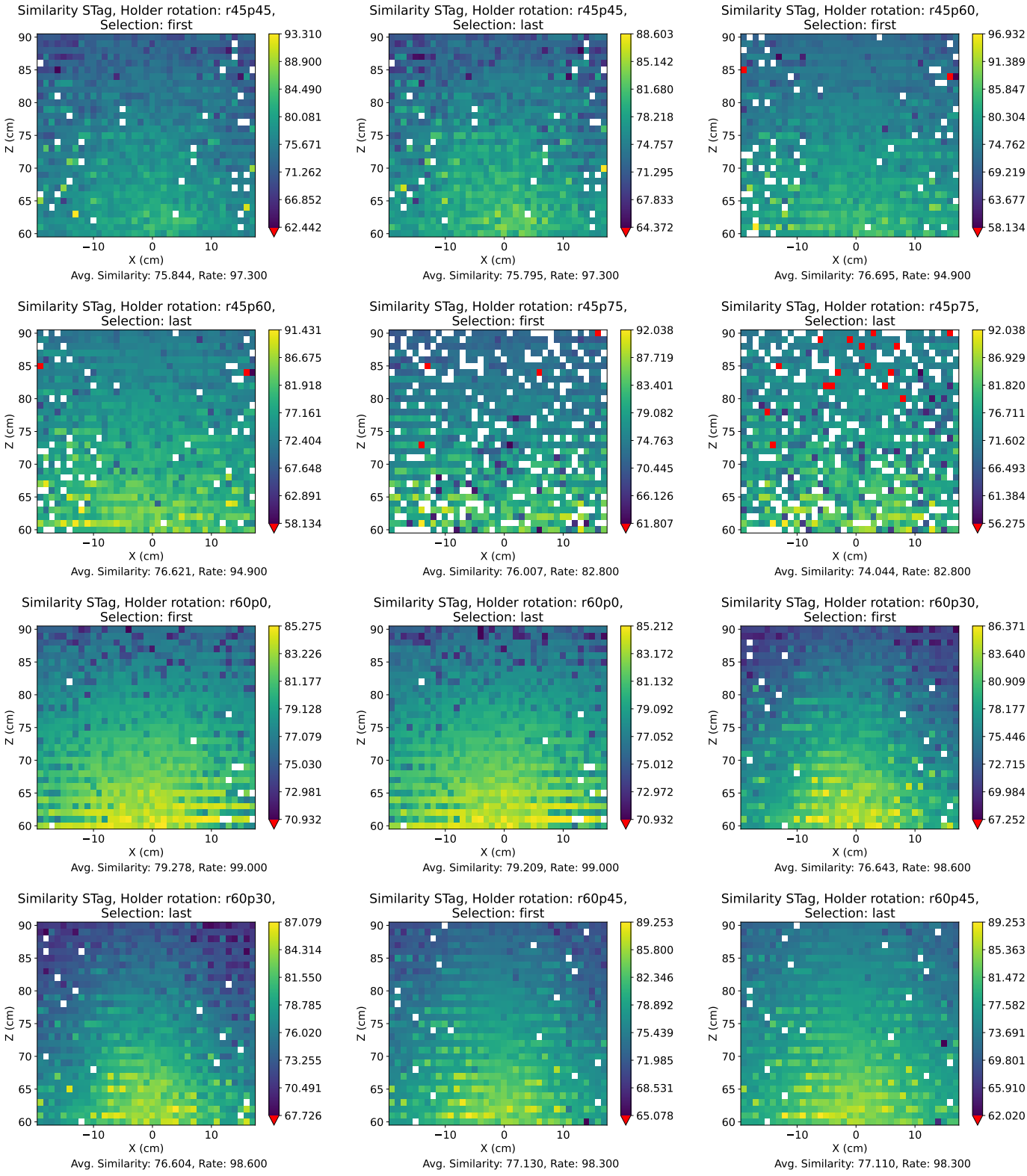

Similarity STag, Holder rotation: r30p75,
Selection: last
Z (cm)
X (cm)
89.206
81.080
72.955
64.829
56.703
48.578
40.452
32.326
Avg. Similarity: 74.133, Rate: 70.000

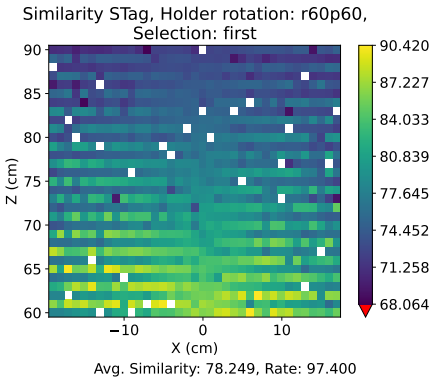

Similarity STag, Holder rotation: r45p0,
Selection: first
Z (cm)
X (cm)
82.978
80.694
78.410
76.127
73.843
71.559
69.276
66.992
Avg. Similarity: 76.927, Rate: 96.500

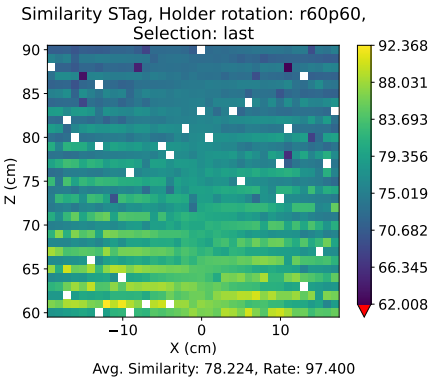

Similarity STag, Holder rotation: r45p0,
Selection: last
Z (cm)
X (cm)
83.070
80.773
78.477
76.180
73.883
71.586
69.289
66.992
Avg. Similarity: 76.908, Rate: 96.500

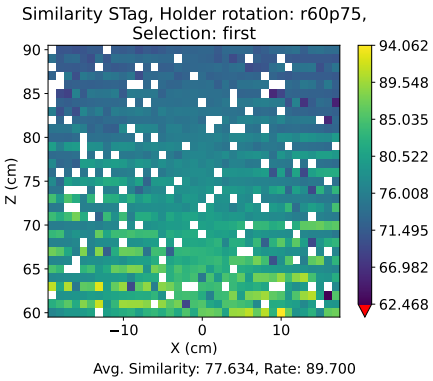

Similarity STag, Holder rotation: r45p30,
Selection: first
Z (cm)
X (cm)
84.570
81.944
79.317
76.691
74.064
71.438
68.811
66.185
Avg. Similarity: 75.518, Rate: 98.200

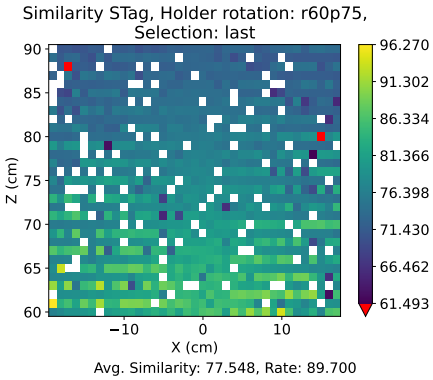

Similarity STag, Holder rotation: r45p30,
Selection: last
Z (cm)
X (cm)
85.056
82.436
79.815
77.195
74.575
71.955
69.335
66.714
Avg. Similarity: 75.555, Rate: 98.200


Similarity STag, Holder rotation: r45p45, Selection: first
Z (cm)
X (cm)
93.310 88.900 84.490 80.081 75.671 71.262 66.852 62.442
Avg. Similarity: 75.844, Rate: 97.300
Similarity STag, Holder rotation: r45p45, Selection: last
88.603 85.142 81.680 78.218 74.757 71.295 67.833 64.372
Avg. Similarity: 75.795, Rate: 97.300
Similarity STag, Holder rotation: r45p60, Selection: first
96.932 91.389 85.847 80.304 74.762 69.219 63.677 58.134
Avg. Similarity: 76.695, Rate: 94.900
Similarity STag, Holder rotation: r45p60, Selection: last
91.431 86.675 81.918 77.161 72.404 67.648 62.891 58.134
Avg. Similarity: 76.621, Rate: 94.900
Similarity STag, Holder rotation: r45p75, Selection: first
92.038 87.719 83.401 79.082 74.763 70.445 66.126 61.807
Avg. Similarity: 76.007, Rate: 82.800
Similarity STag, Holder rotation: r45p75, Selection: last
92.038 86.929 81.820 76.711 71.602 66.493 61.384 56.275
Avg. Similarity: 74.044, Rate: 82.800
Similarity STag, Holder rotation: r60p0, Selection: first
85.275 83.226 81.177 79.128 77.079 75.030 72.981 70.932
Avg. Similarity: 79.278, Rate: 99.000
Similarity STag, Holder rotation: r60p0, Selection: last
85.212 83.172 81.132 79.092 77.052 75.012 72.972 70.932
Avg. Similarity: 79.209, Rate: 99.000
Similarity STag, Holder rotation: r60p30, Selection: first
86.371 83.640 80.909 78.177 75.446 72.715 69.984 67.252
Avg. Similarity: 76.643, Rate: 98.600
Similarity STag, Holder rotation: r60p30, Selection: last
87.079 84.314 81.550 78.785 76.020 73.255 70.491 67.726
Avg. Similarity: 76.604, Rate: 98.600
Similarity STag, Holder rotation: r60p45, Selection: first
89.253 85.800 82.346 78.892 75.439 71.985 68.531 65.078
Avg. Similarity: 77.130, Rate: 98.300
Similarity STag, Holder rotation: r60p45, Selection: last
89.253 85.363 81.472 77.582 73.691 69.801 65.910 62.020
Avg. Similarity: 77.110, Rate: 98.300

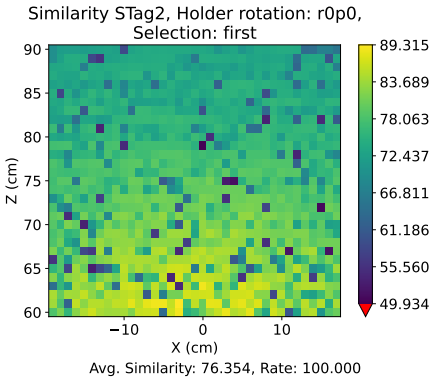

Similarity STag, Holder rotation: r60p60,
Selection: first
Z (cm)
X (cm)
Avg. Similarity: 78.249, Rate: 97.400


Similarity STag, Holder rotation: r60p60,
Selection: last
Z (cm)
X (cm)
Avg. Similarity: 78.224, Rate: 97.400

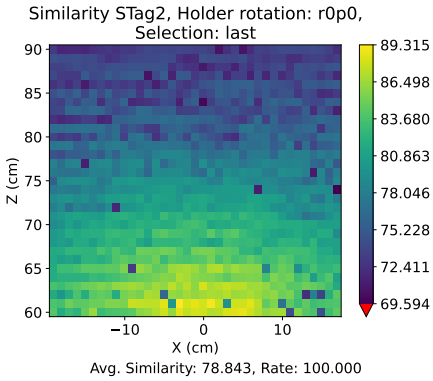

Similarity STag, Holder rotation: r60p75,
Selection: first
Z (cm)
X (cm)
Avg. Similarity: 77.634, Rate: 89.700


Similarity STag, Holder rotation: r60p75,
Selection: last
Z (cm)
X (cm)
Avg. Similarity: 77.548, Rate: 89.700

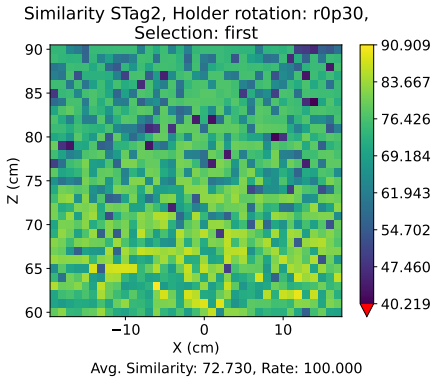
STag2



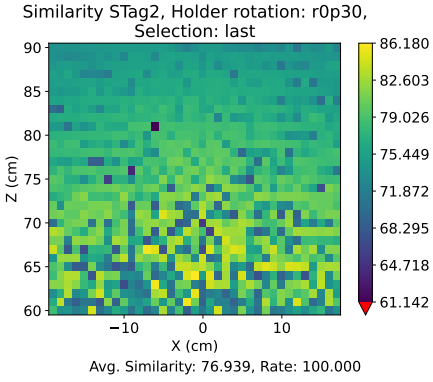
Avg. Similarity: 76.354, Rate: 100.000



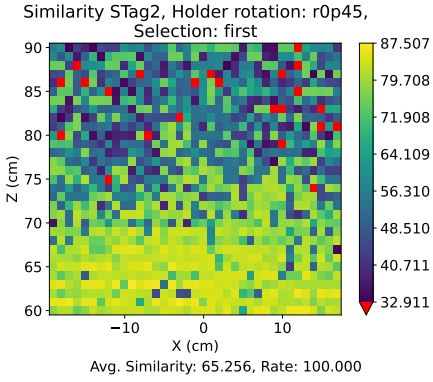
Avg. Similarity: 78.843, Rate: 100.000



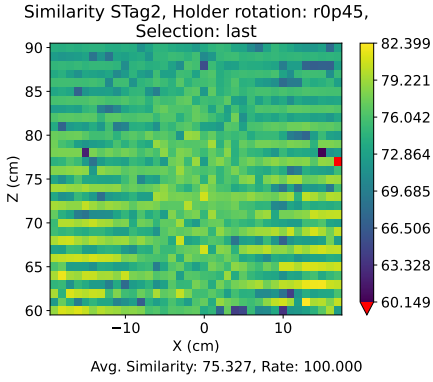
Avg. Similarity: 72.730, Rate: 100.000



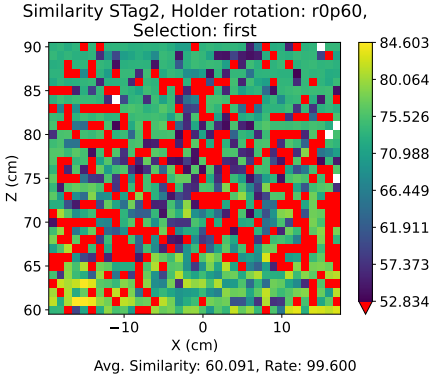
Avg. Similarity: 76.939, Rate: 100.000



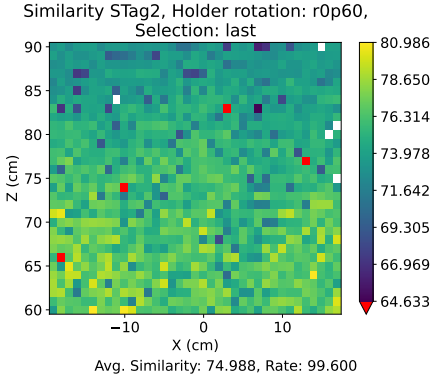
Avg. Similarity: 65.256, Rate: 100.000



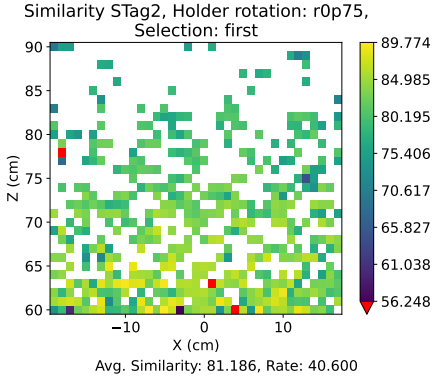
Avg. Similarity: 75.327, Rate: 100.000



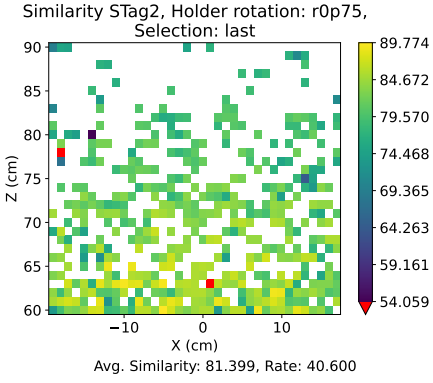
Avg. Similarity: 60.091, Rate: 99.600



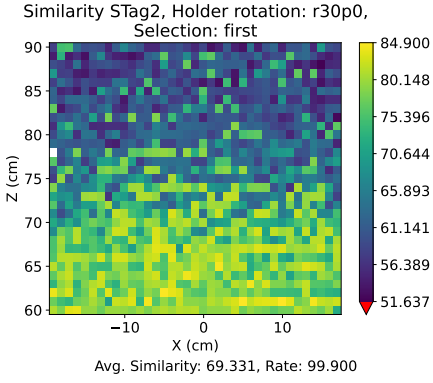
Avg. Similarity: 74.988, Rate: 99.600



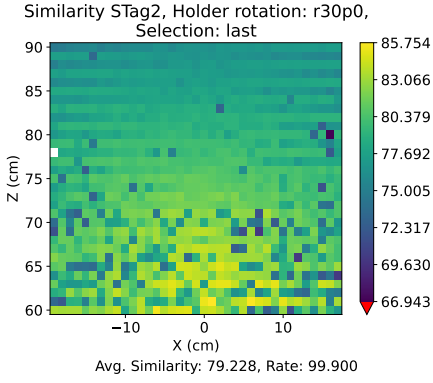
Avg. Similarity: 81.186, Rate: 40.600



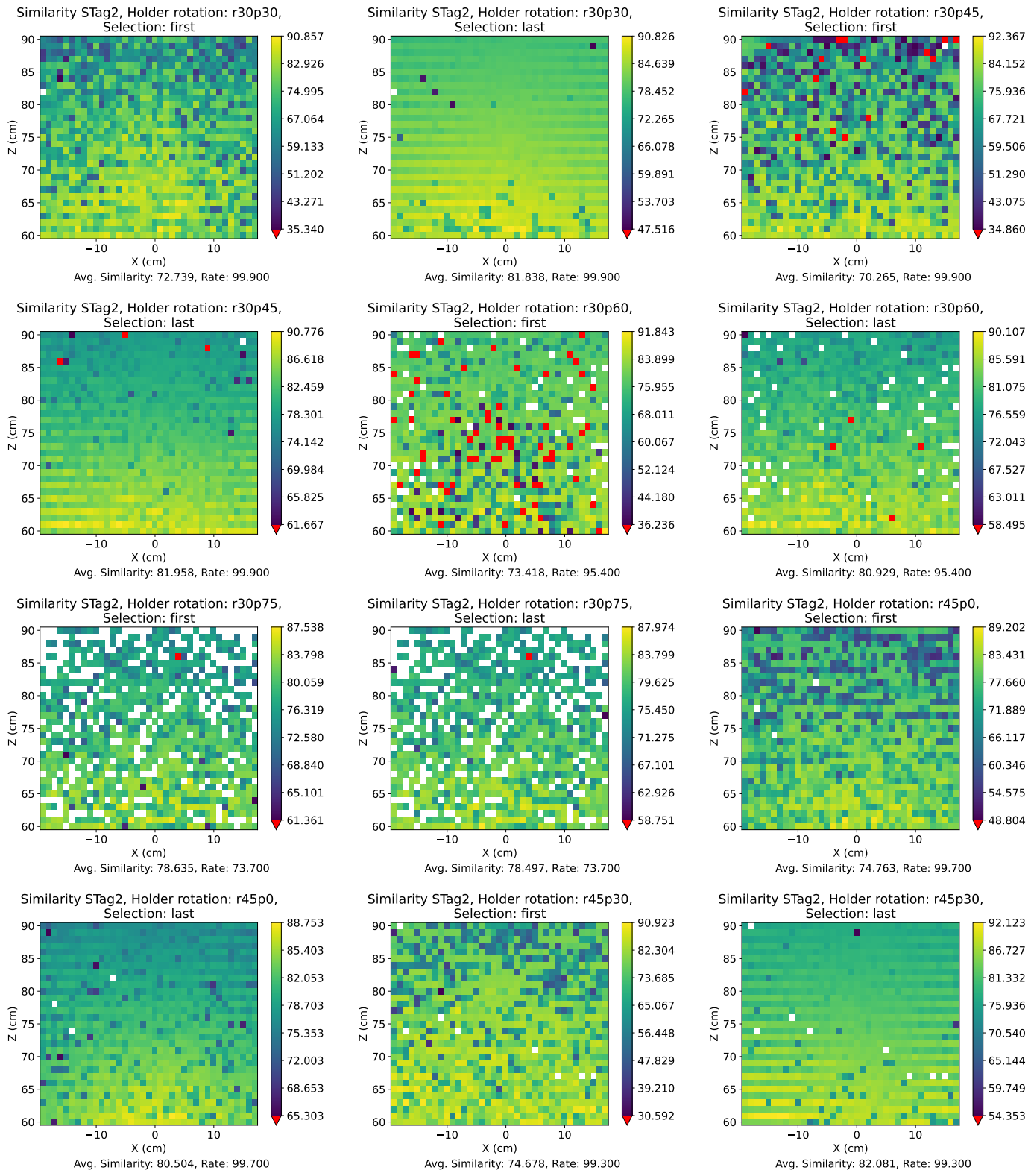
Avg. Similarity: 81.399, Rate: 40.600

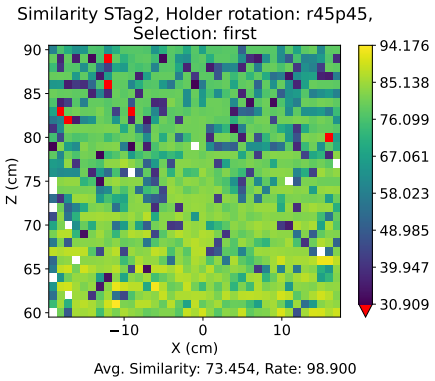


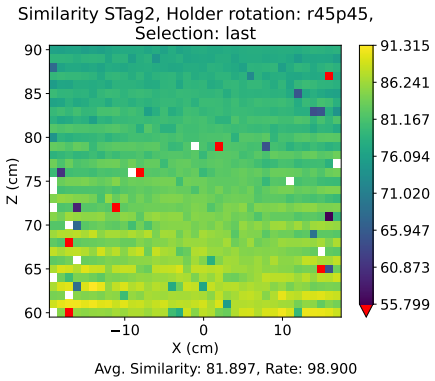
Avg. Similarity: 69.331, Rate: 99.900

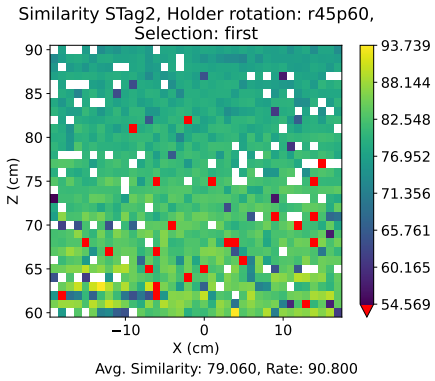


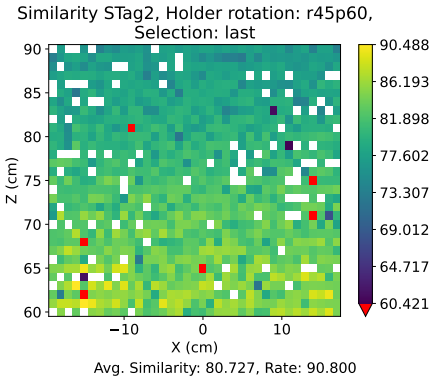
Avg. Similarity: 79.228, Rate: 99.900

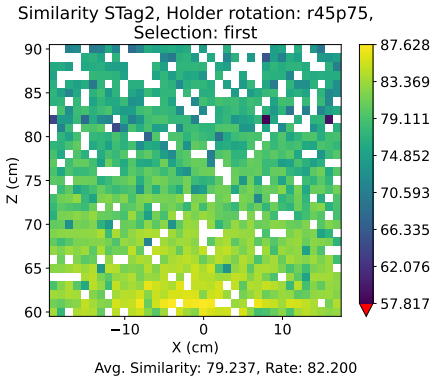

Similarity STag2, Holder rotation: r30p30, Selection: first
Avg. Similarity: 72.739, Rate: 99.900
Similarity STag2, Holder rotation: r30p30, Selection: last
Avg. Similarity: 81.838, Rate: 99.900
Similarity STag2, Holder rotation: r30p45, Selection: first
Avg. Similarity: 70.265, Rate: 99.900
Similarity STag2, Holder rotation: r30p45, Selection: last
Avg. Similarity: 81.958, Rate: 99.900
Similarity STag2, Holder rotation: r30p60, Selection: first
Avg. Similarity: 73.418, Rate: 95.400
Similarity STag2, Holder rotation: r30p60, Selection: last
Avg. Similarity: 80.929, Rate: 95.400
Similarity STag2, Holder rotation: r30p75, Selection: first
Avg. Similarity: 78.635, Rate: 73.700
Similarity STag2, Holder rotation: r30p75, Selection: last
Avg. Similarity: 78.497, Rate: 73.700
Similarity STag2, Holder rotation: r45p0, Selection: first
Avg. Similarity: 74.763, Rate: 99.700
Similarity STag2, Holder rotation: r45p0, Selection: last
Avg. Similarity: 80.504, Rate: 99.700
Similarity STag2, Holder rotation: r45p30, Selection: first
Avg. Similarity: 74.678, Rate: 99.300
Similarity STag2, Holder rotation: r45p30, Selection: last
Avg. Similarity: 82.081, Rate: 99.300
Z (cm)
X (cm)

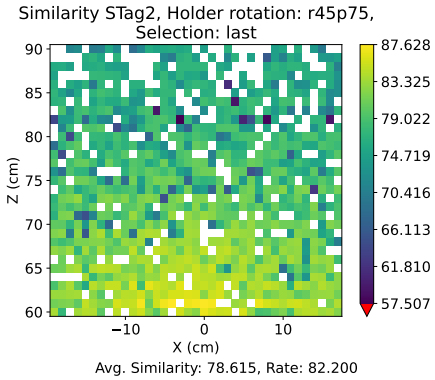

Similarity STag2, Holder rotation: r45p45,
Selection: first
Z (cm)
X (cm)
94.176
85.138
76.099
67.061
58.023
48.985
39.947
30.909
Avg. Similarity: 73.454, Rate: 98.900

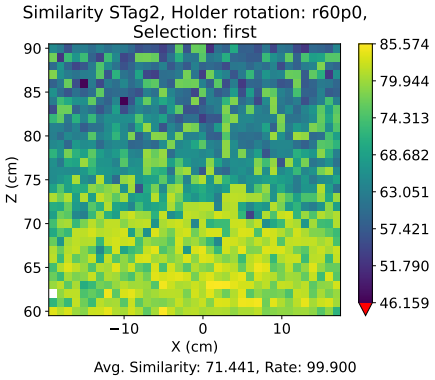

Similarity STag2, Holder rotation: r45p45,
Selection: last
Z (cm)
X (cm)
91.315
86.241
81.167
76.094
71.020
65.947
60.873
55.799
Avg. Similarity: 81.897, Rate: 98.900

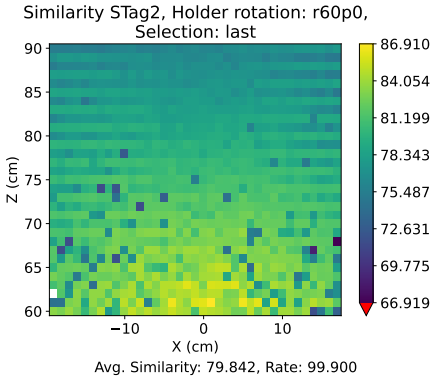

Similarity STag2, Holder rotation: r45p60,
Selection: first
Z (cm)
X (cm)
93.739
88.144
82.548
76.952
71.356
65.761
60.165
54.569
Avg. Similarity: 79.060, Rate: 90.800

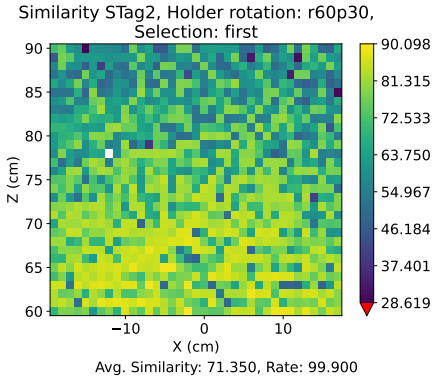

Similarity STag2, Holder rotation: r45p60,
Selection: last
Z (cm)
X (cm)
90.488
86.193
81.898
77.602
73.307
69.012
64.717
60.421
Avg. Similarity: 80.727, Rate: 90.800

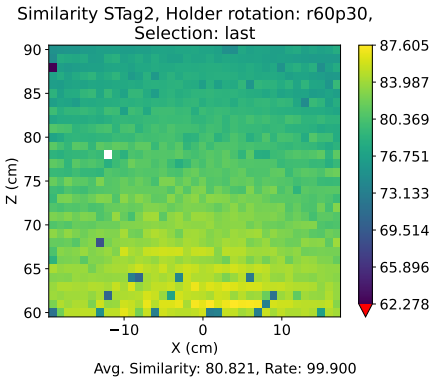

Similarity STag2, Holder rotation: r45p75,
Selection: first
Z (cm)
X (cm)
87.628
83.369
79.111
74.852
70.593
66.335
62.076
57.817
Avg. Similarity: 79.237, Rate: 82.200

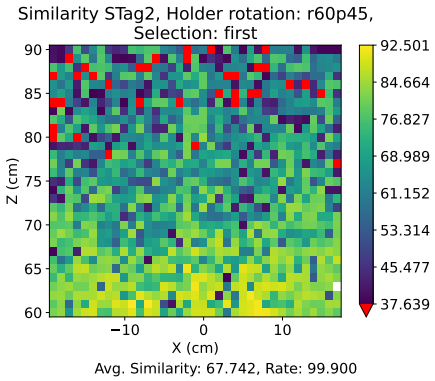

Similarity STag2, Holder rotation: r45p75,
Selection: last
Z (cm)
X (cm)
87.628
83.325
79.022
74.719
70.416
66.113
61.810
57.507
Avg. Similarity: 78.615, Rate: 82.200

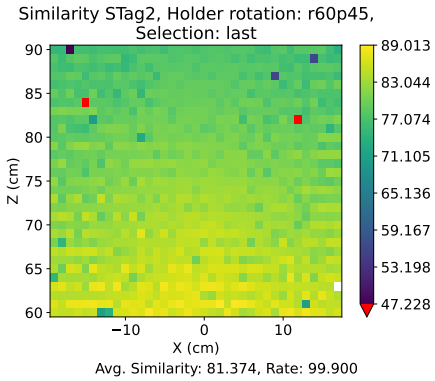

Similarity STag2, Holder rotation: r60p0,
Selection: first
Z (cm)
X (cm)
85.574
79.944
74.313
68.682
63.051
57.421
51.790
46.159
Avg. Similarity: 71.441, Rate: 99.900

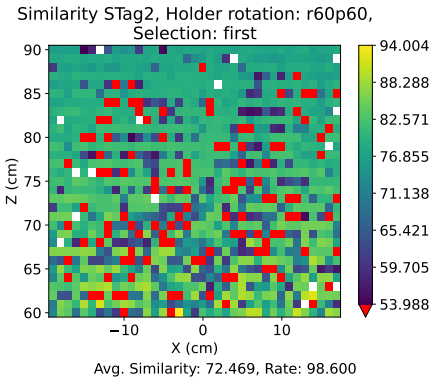

Similarity STag2, Holder rotation: r60p0,
Selection: last
Z (cm)
X (cm)
86.910
84.054
81.199
78.343
75.487
72.631
69.775
66.919
Avg. Similarity: 79.842, Rate: 99.900

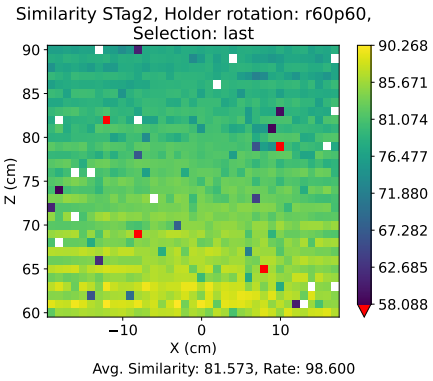

Similarity STag2, Holder rotation: r60p30,
Selection: first
Z (cm)
X (cm)
90.098
81.315
72.533
63.750
54.967
46.184
37.401
28.619
Avg. Similarity: 71.350, Rate: 99.900

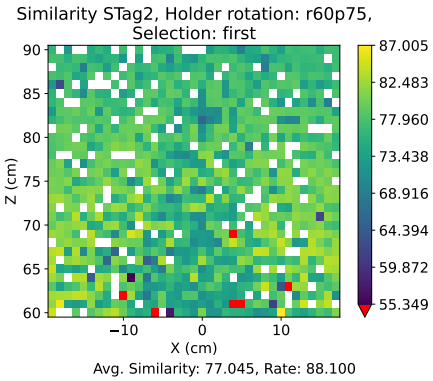

Similarity STag2, Holder rotation: r60p30,
Selection: last
Z (cm)
X (cm)
87.605
83.987
80.369
76.751
73.133
69.514
65.896
62.278
Avg. Similarity: 80.821, Rate: 99.900

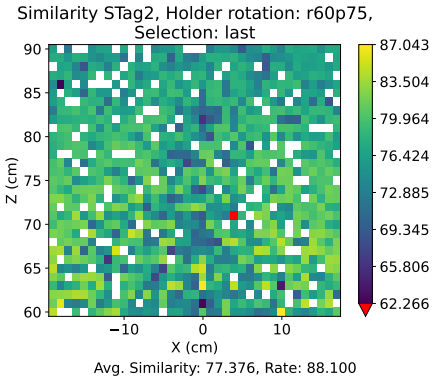

Similarity STag2, Holder rotation: r60p45,
Selection: first
Z (cm)
X (cm)
92.501
84.664
76.827
68.989
61.152
53.314
45.477
37.639
Avg. Similarity: 67.742, Rate: 99.900


Similarity STag2, Holder rotation: r60p45,
Selection: last
Z (cm)
X (cm)
89.013
83.044
77.074
71.105
65.136
59.167
53.198
47.228
Avg. Similarity: 81.374, Rate: 99.900


Similarity STag2, Holder rotation: r60p60,
Selection: first
Z (cm)
X (cm)
94.004
88.288
82.571
76.855
71.138
65.421
59.705
53.988
Avg. Similarity: 72.469, Rate: 98.600

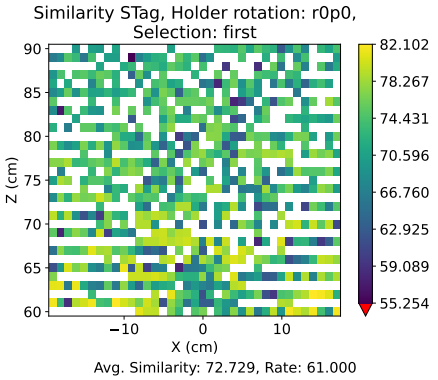

Similarity STag2, Holder rotation: r60p60,
Selection: last
Z (cm)
X (cm)
90.268
85.671
81.074
76.477
71.880
67.282
62.685
58.088
Avg. Similarity: 81.573, Rate: 98.600


Similarity STag2, Holder rotation: r60p75,
Selection: first
Z (cm)
X (cm)
87.005
82.483
77.960
73.438
68.916
64.394
59.872
55.349
Avg. Similarity: 77.045, Rate: 88.100


Similarity STag2, Holder rotation: r60p75,
Selection: last
Z (cm)
X (cm)
87.043
83.504
79.964
76.424
72.885
69.345
65.806
62.266
Avg. Similarity: 77.376, Rate: 88.100

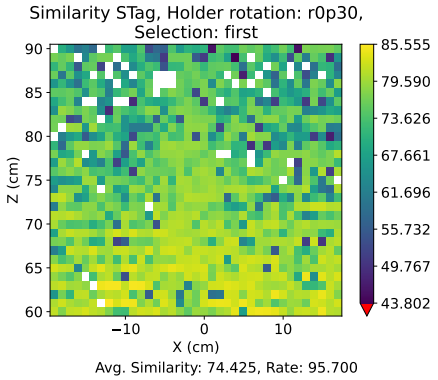
B. Top-Right Corner occlusion

STag

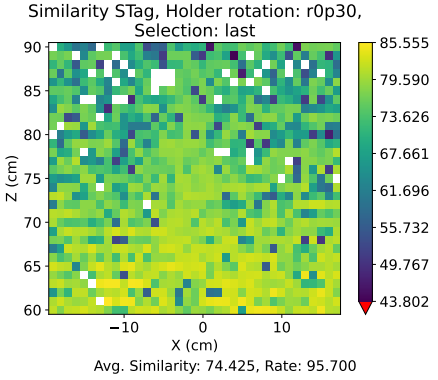


Avg. Similarity: 72.729, Rate: 61.000

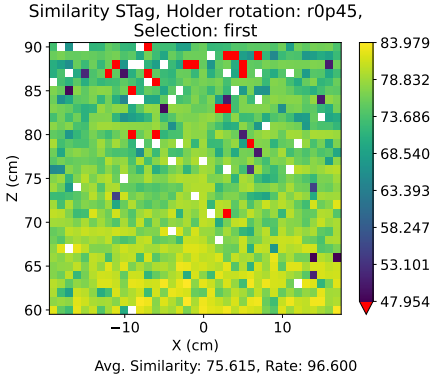
Avg. Similarity: 72.729, Rate: 61.000



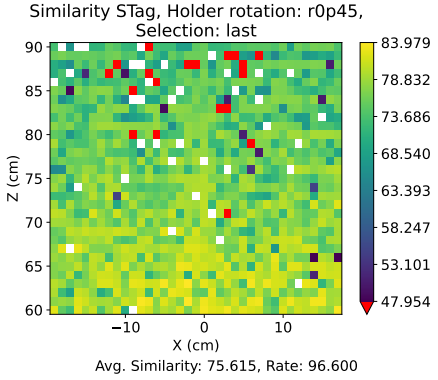
Avg. Similarity: 74.425, Rate: 95.700



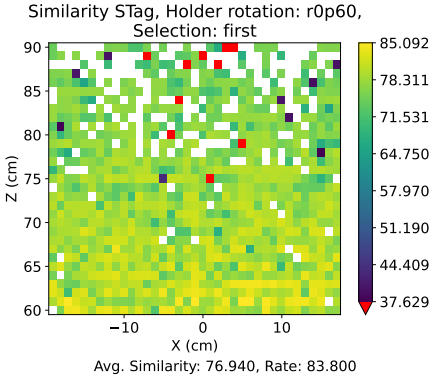
Avg. Similarity: 74.425, Rate: 95.700



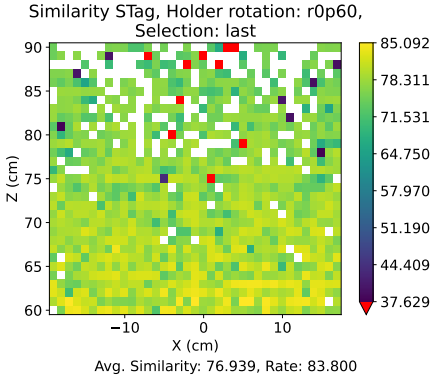
Avg. Similarity: 75.615, Rate: 96.600



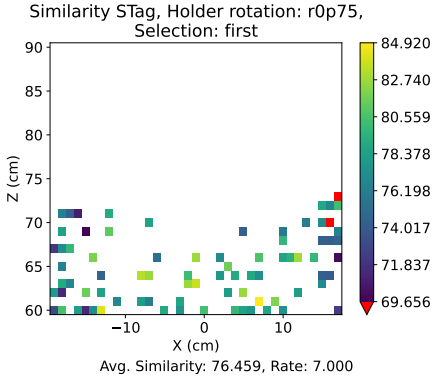
Avg. Similarity: 75.615, Rate: 96.600



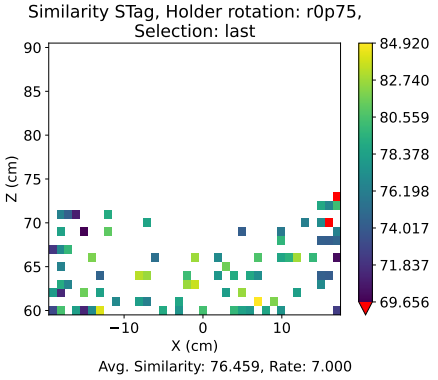
Avg. Similarity: 76.940, Rate: 83.800



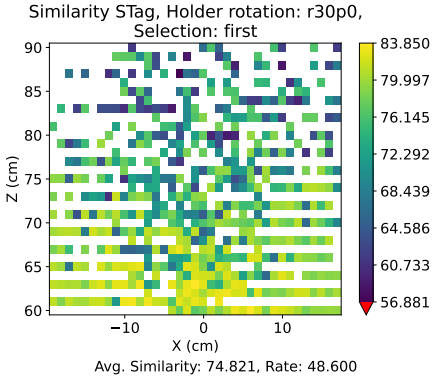
Avg. Similarity: 76.939, Rate: 83.800



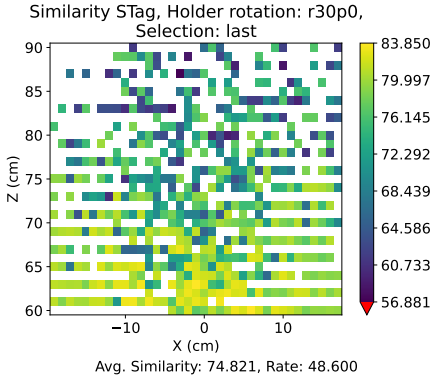
Avg. Similarity: 76.459, Rate: 7.000



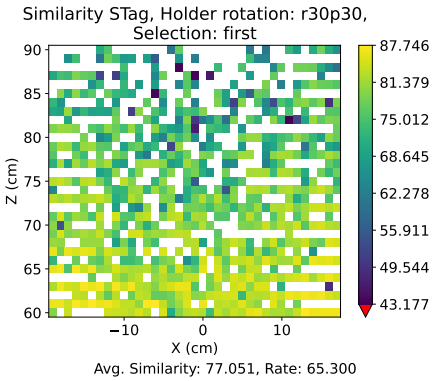
Avg. Similarity: 76.459, Rate: 7.000

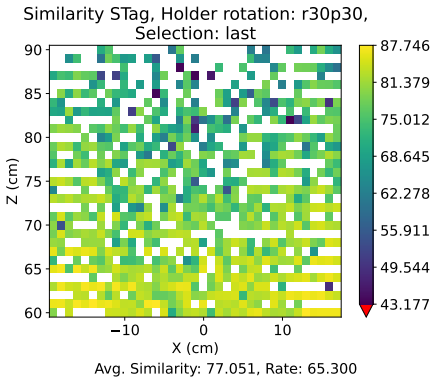


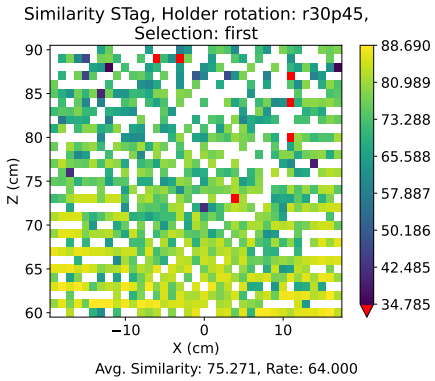
Avg. Similarity: 74.821, Rate: 48.600

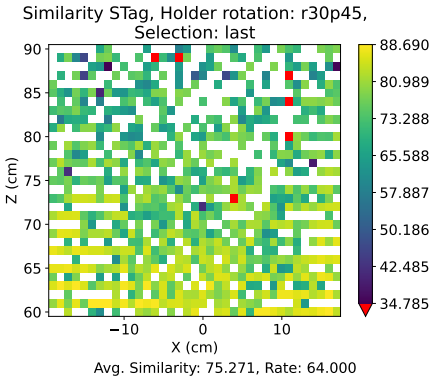


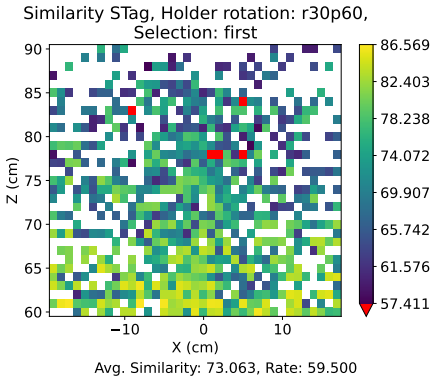
Avg. Similarity: 74.821, Rate: 48.600

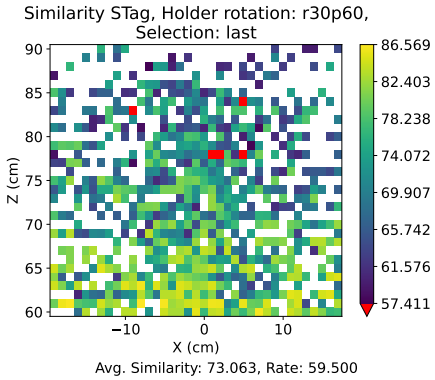

Similarity STag, Holder rotation: r30p30,
Selection: first
Z (cm)
X (cm)
87.746
81.379
75.012
68.645
62.278
55.911
49.544
43.177
Avg. Similarity: 77.051, Rate: 65.300

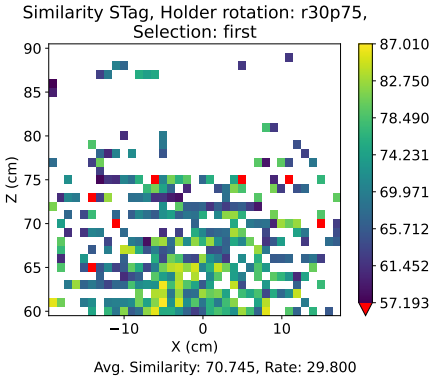

Similarity STag, Holder rotation: r30p30,
Selection: last
Z (cm)
X (cm)
87.746
81.379
75.012
68.645
62.278
55.911
49.544
43.177
Avg. Similarity: 77.051, Rate: 65.300

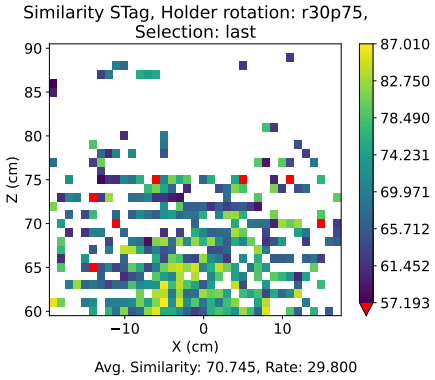

Similarity STag, Holder rotation: r30p45,
Selection: first
Z (cm)
X (cm)
88.690
80.989
73.288
65.588
57.887
50.186
42.485
34.785
Avg. Similarity: 75.271, Rate: 64.000

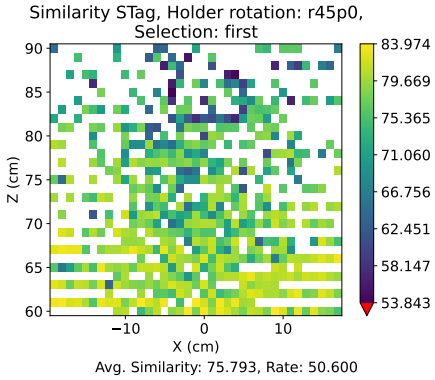

Similarity STag, Holder rotation: r30p45,
Selection: last
Z (cm)
X (cm)
88.690
80.989
73.288
65.588
57.887
50.186
42.485
34.785
Avg. Similarity: 75.271, Rate: 64.000

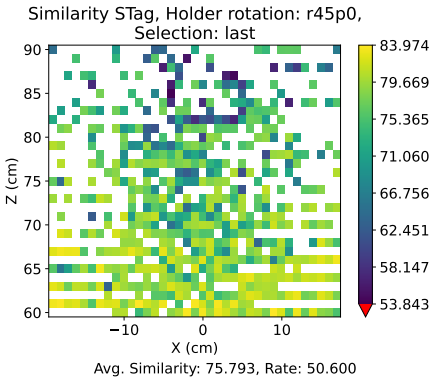

Similarity STag, Holder rotation: r30p60,
Selection: first
Z (cm)
X (cm)
86.569
82.403
78.238
74.072
69.907
65.742
61.576
57.411
Avg. Similarity: 73.063, Rate: 59.500

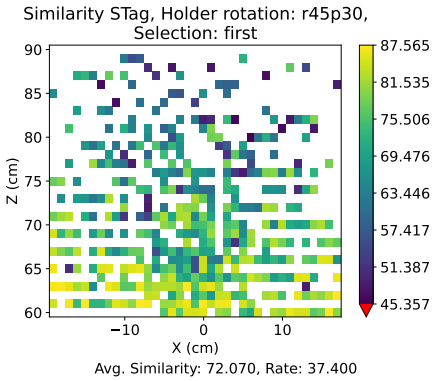

Similarity STag, Holder rotation: r30p60,
Selection: last
Z (cm)
X (cm)
86.569
82.403
78.238
74.072
69.907
65.742
61.576
57.411
Avg. Similarity: 73.063, Rate: 59.500

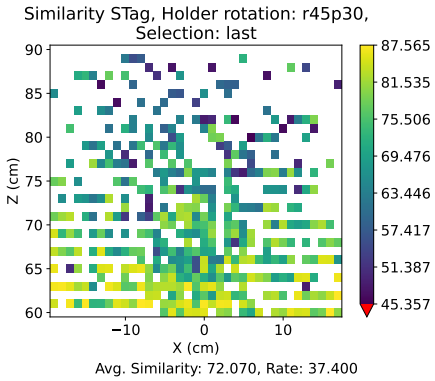

Similarity STag, Holder rotation: r30p75,
Selection: first
Z (cm)
X (cm)
87.010
82.750
78.490
74.231
69.971
65.712
61.452
57.193
Avg. Similarity: 70.745, Rate: 29.800

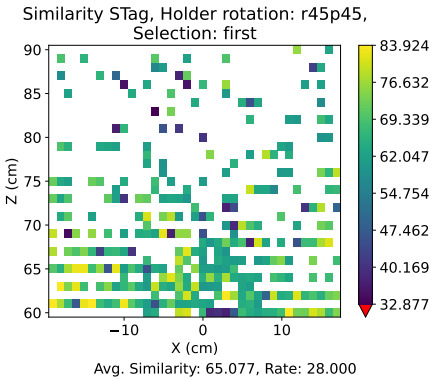

Similarity STag, Holder rotation: r30p75,
Selection: last
Z (cm)
X (cm)
87.010
82.750
78.490
74.231
69.971
65.712
61.452
57.193
Avg. Similarity: 70.745, Rate: 29.800

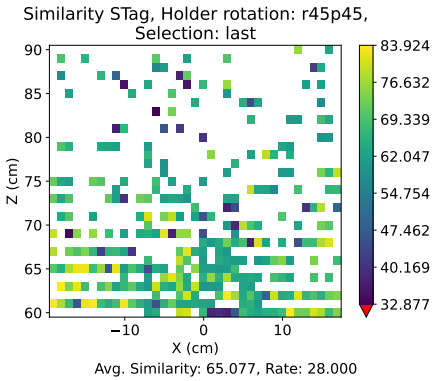

Similarity STag, Holder rotation: r45p0,
Selection: first
Z (cm)
X (cm)
83.974
79.669
75.365
71.060
66.756
62.451
58.147
53.843
Avg. Similarity: 75.793, Rate: 50.600

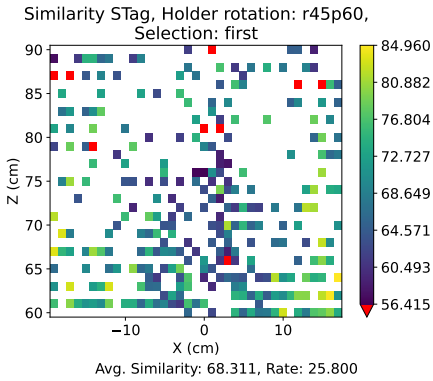

Similarity STag, Holder rotation: r45p0,
Selection: last
Z (cm)
X (cm)
83.974
79.669
75.365
71.060
66.756
62.451
58.147
53.843
Avg. Similarity: 75.793, Rate: 50.600

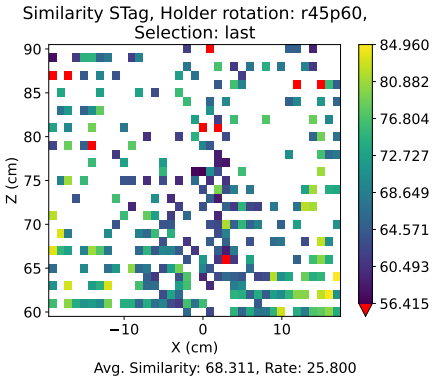

Similarity STag, Holder rotation: r45p30,
Selection: first
Z (cm)
X (cm)
87.565
81.535
75.506
69.476
63.446
57.417
51.387
45.357
Avg. Similarity: 72.070, Rate: 37.400

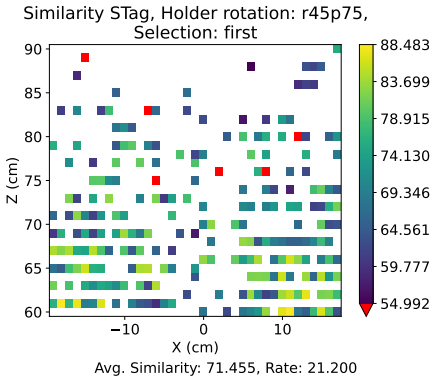

Similarity STag, Holder rotation: r45p30,
Selection: last
Z (cm)
X (cm)
87.565
81.535
75.506
69.476
63.446
57.417
51.387
45.357
Avg. Similarity: 72.070, Rate: 37.400

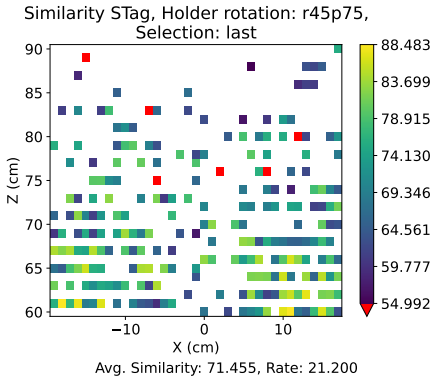

Similarity STag, Holder rotation: r45p45,
Selection: first
Z (cm)
X (cm)
83.924
76.632
69.339
62.047
54.754
47.462
40.169
32.877
Avg. Similarity: 65.077, Rate: 28.000


Similarity STag, Holder rotation: r45p45,
Selection: last
Z (cm)
X (cm)
83.924
76.632
69.339
62.047
54.754
47.462
40.169
32.877
Avg. Similarity: 65.077, Rate: 28.000

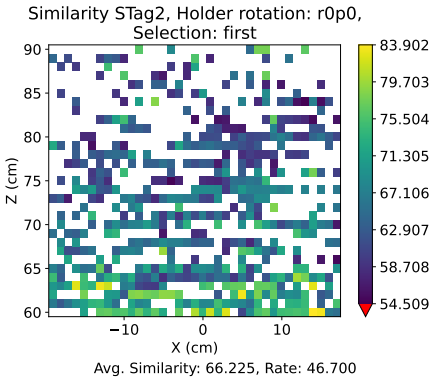

Similarity STag, Holder rotation: r45p60,
Selection: first
Z (cm)
X (cm)
84.960
80.882
76.804
72.727
68.649
64.571
60.493
56.415
Avg. Similarity: 68.311, Rate: 25.800


Similarity STag, Holder rotation: r45p60,
Selection: last
Z (cm)
X (cm)
84.960
80.882
76.804
72.727
68.649
64.571
60.493
56.415
Avg. Similarity: 68.311, Rate: 25.800

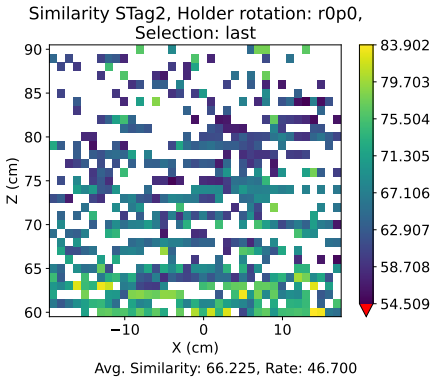

Similarity STag, Holder rotation: r45p75,
Selection: first
Z (cm)
X (cm)
88.483
83.699
78.915
74.130
69.346
64.561
59.777
54.992
Avg. Similarity: 71.455, Rate: 21.200


Similarity STag, Holder rotation: r45p75,
Selection: last
Z (cm)
X (cm)
88.483
83.699
78.915
74.130
69.346
64.561
59.777
54.992
Avg. Similarity: 71.455, Rate: 21.200

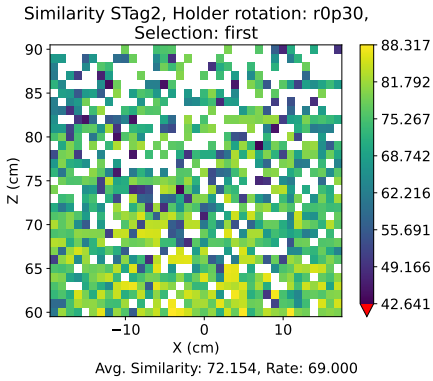
STag2



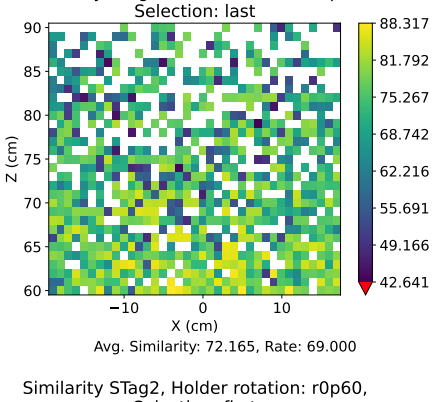
Avg. Similarity: 66.225, Rate: 46.700



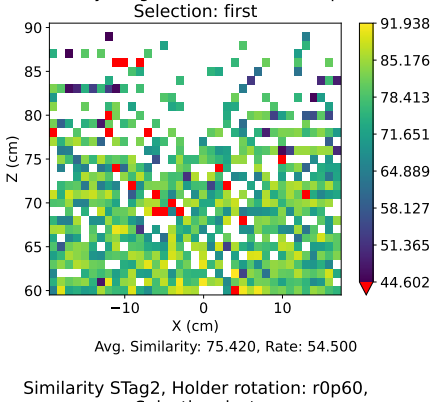
Avg. Similarity: 66.225, Rate: 46.700



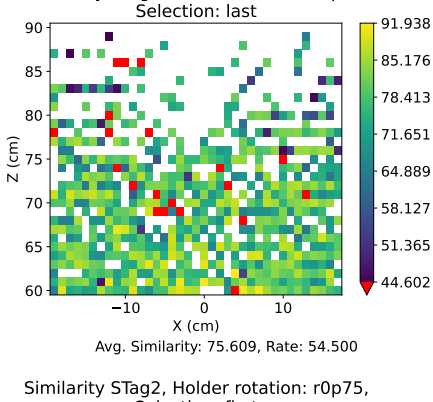
Avg. Similarity: 72.154, Rate: 69.000



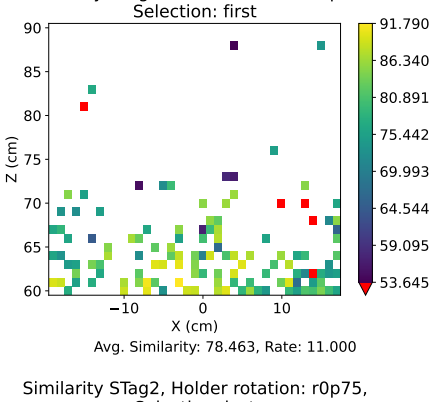
Avg. Similarity: 72.165, Rate: 69.000



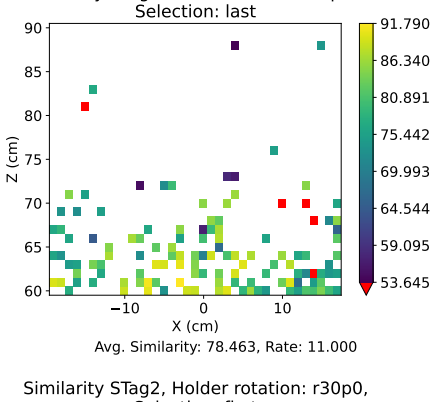
Avg. Similarity: 75.420, Rate: 54.500



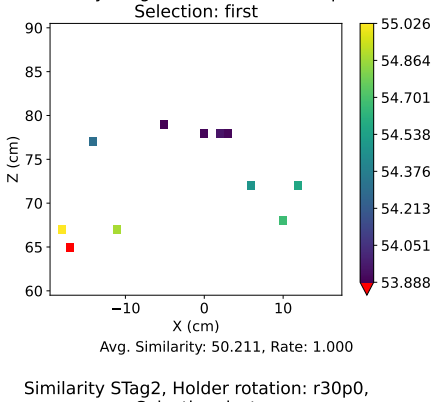
Avg. Similarity: 75.609, Rate: 54.500



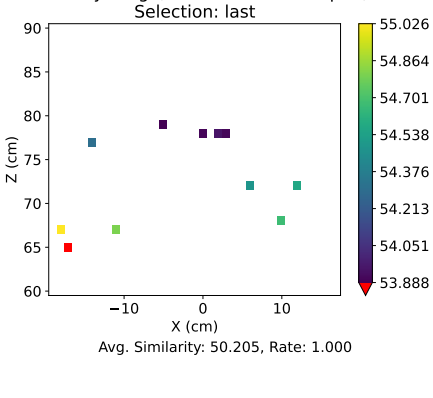
Avg. Similarity: 78.463, Rate: 11.000



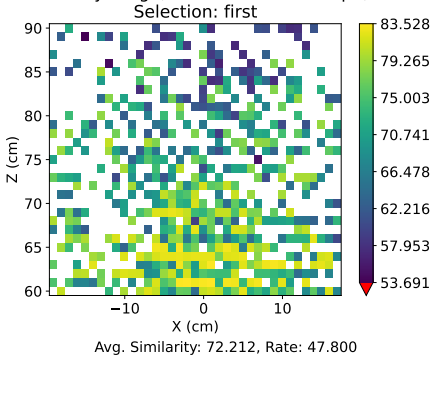
Avg. Similarity: 78.463, Rate: 11.000



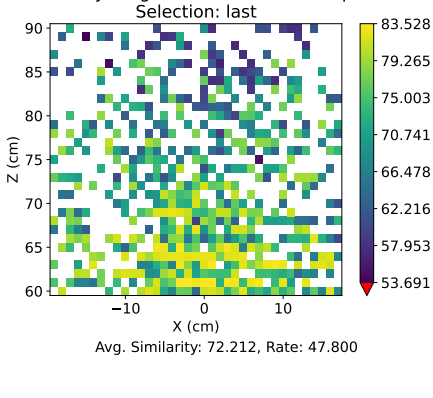
Avg. Similarity: 50.211, Rate: 1.000



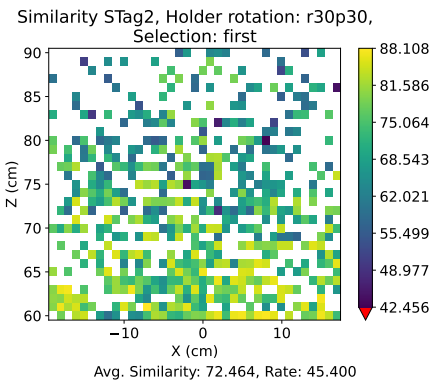
Avg. Similarity: 50.205, Rate: 1.000

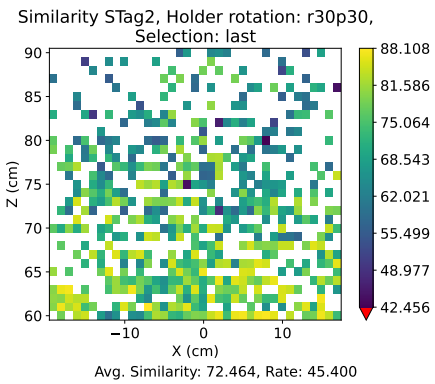


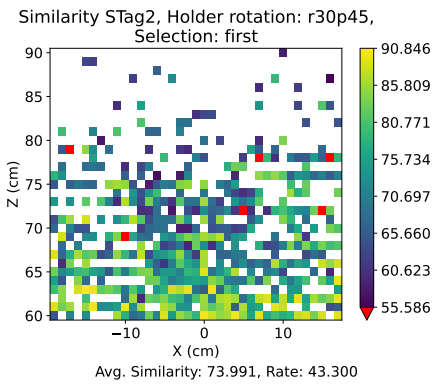
Avg. Similarity: 72.212, Rate: 47.800

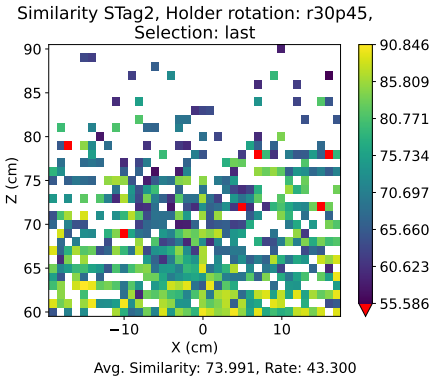


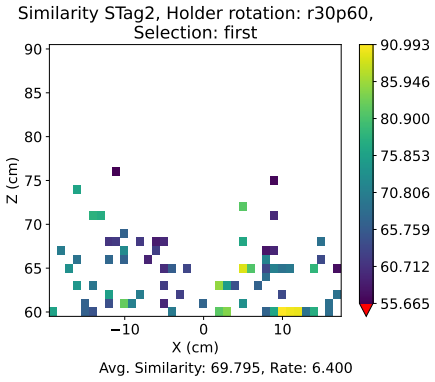
Avg. Similarity: 72.212, Rate: 47.800

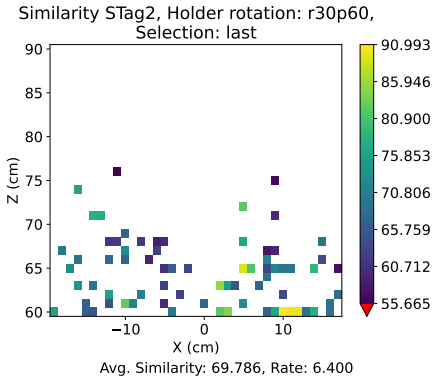

Similarity STag2, Holder rotation: r30p30,
Selection: first
Z (cm)
X (cm)
88.108
81.586
75.064
68.543
62.021
55.499
48.977
42.456
Avg. Similarity: 72.464, Rate: 45.400

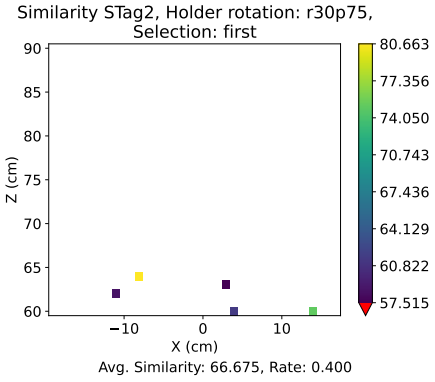

Similarity STag2, Holder rotation: r30p30,
Selection: last
Z (cm)
X (cm)
88.108
81.586
75.064
68.543
62.021
55.499
48.977
42.456
Avg. Similarity: 72.464, Rate: 45.400

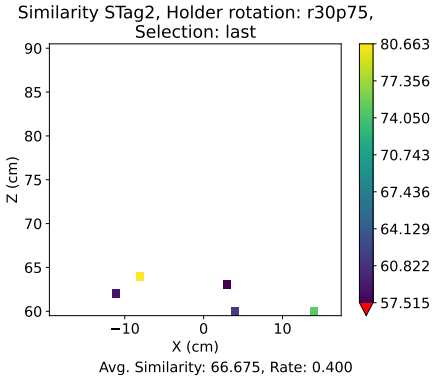

Similarity STag2, Holder rotation: r30p45,
Selection: first
Z (cm)
X (cm)
90.846
85.809
80.771
75.734
70.697
65.660
60.623
55.586
Avg. Similarity: 73.991, Rate: 43.300

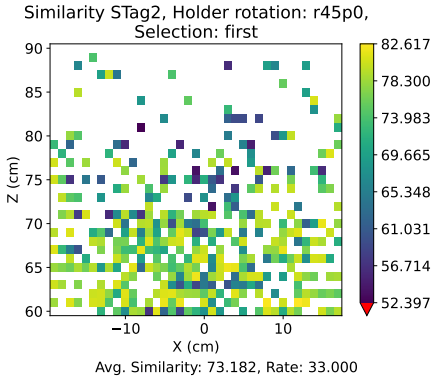

Similarity STag2, Holder rotation: r30p45,
Selection: last
Z (cm)
X (cm)
90.846
85.809
80.771
75.734
70.697
65.660
60.623
55.586
Avg. Similarity: 73.991, Rate: 43.300

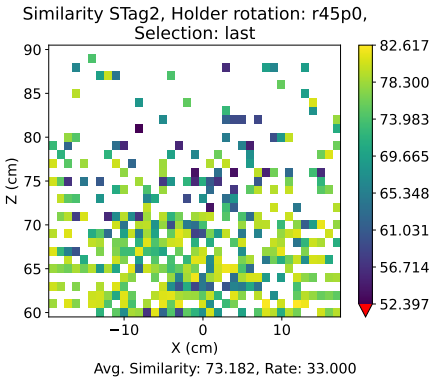

Similarity STag2, Holder rotation: r30p60,
Selection: first
Z (cm)
X (cm)
90.993
85.946
80.900
75.853
70.806
65.759
60.712
55.665
Avg. Similarity: 69.795, Rate: 6.400

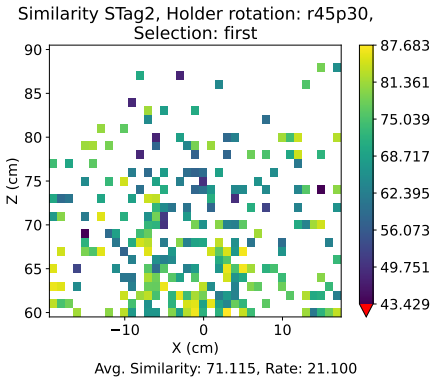

Similarity STag2, Holder rotation: r30p60,
Selection: last
Z (cm)
X (cm)
90.993
85.946
80.900
75.853
70.806
65.759
60.712
55.665
Avg. Similarity: 69.786, Rate: 6.400

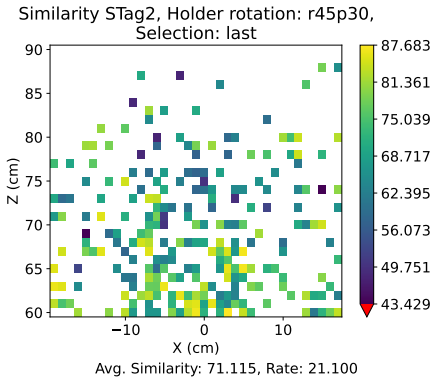

Similarity STag2, Holder rotation: r30p75,
Selection: first
Z (cm)
X (cm)
80.663
77.356
74.050
70.743
67.436
64.129
60.822
57.515
Avg. Similarity: 66.675, Rate: 0.400

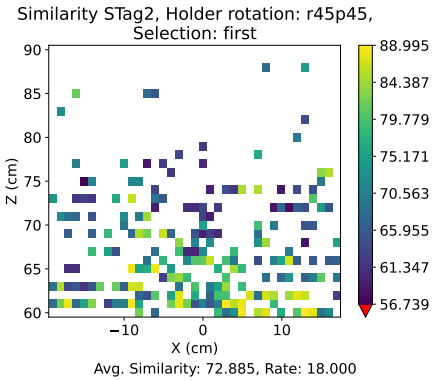

Similarity STag2, Holder rotation: r30p75,
Selection: last
Z (cm)
X (cm)
80.663
77.356
74.050
70.743
67.436
64.129
60.822
57.515
Avg. Similarity: 66.675, Rate: 0.400

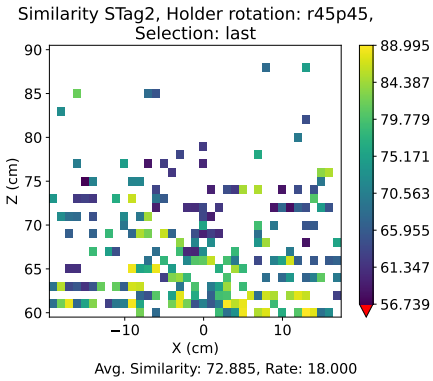

Similarity STag2, Holder rotation: r45p0,
Selection: first
Z (cm)
X (cm)
82.617
78.300
73.983
69.665
65.348
61.031
56.714
52.397
Avg. Similarity: 73.182, Rate: 33.000

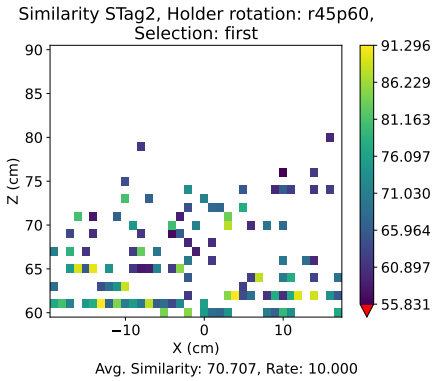

Similarity STag2, Holder rotation: r45p0,
Selection: last
Z (cm)
X (cm)
82.617
78.300
73.983
69.665
65.348
61.031
56.714
52.397
Avg. Similarity: 73.182, Rate: 33.000

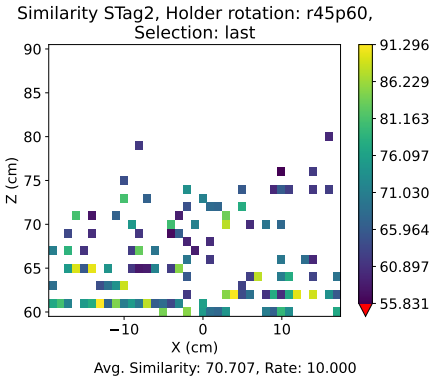

Similarity STag2, Holder rotation: r45p30,
Selection: first
Z (cm)
X (cm)
87.683
81.361
75.039
68.717
62.395
56.073
49.751
43.429
Avg. Similarity: 71.115, Rate: 21.100

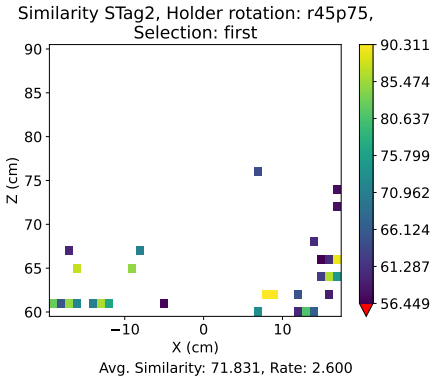

Similarity STag2, Holder rotation: r45p30,
Selection: last
Z (cm)
X (cm)
87.683
81.361
75.039
68.717
62.395
56.073
49.751
43.429
Avg. Similarity: 71.115, Rate: 21.100

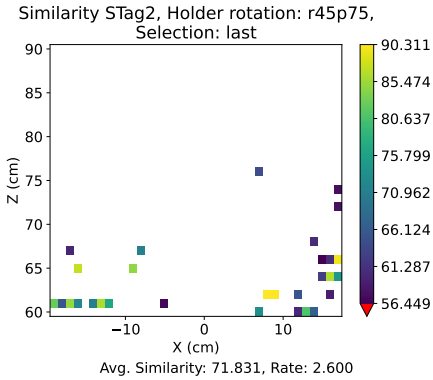

Similarity STag2, Holder rotation: r45p45,
Selection: first
Z (cm)
X (cm)
88.995
84.387
79.779
75.171
70.563
65.955
61.347
56.739
Avg. Similarity: 72.885, Rate: 18.000


Similarity STag2, Holder rotation: r45p45,
Selection: last
Z (cm)
X (cm)
88.995
84.387
79.779
75.171
70.563
65.955
61.347
56.739
Avg. Similarity: 72.885, Rate: 18.000


Similarity STag2, Holder rotation: r45p60,
Selection: first
Z (cm)
X (cm)
91.296
86.229
81.163
76.097
71.030
65.964
60.897
55.831
Avg. Similarity: 70.707, Rate: 10.000

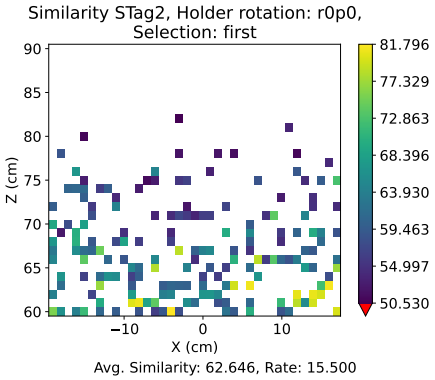

Similarity STag2, Holder rotation: r45p60,
Selection: last
Z (cm)
X (cm)
91.296
86.229
81.163
76.097
71.030
65.964
60.897
55.831
Avg. Similarity: 70.707, Rate: 10.000

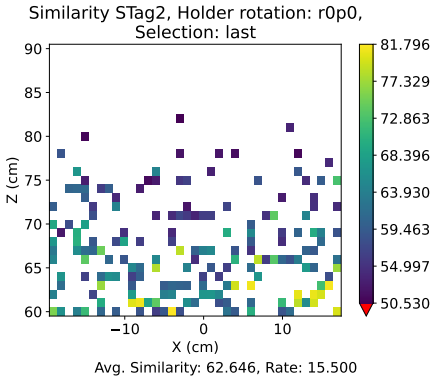

Similarity STag2, Holder rotation: r45p75,
Selection: first
Z (cm)
X (cm)
90.311
85.474
80.637
75.799
70.962
66.124
61.287
56.449
Avg. Similarity: 71.831, Rate: 2.600

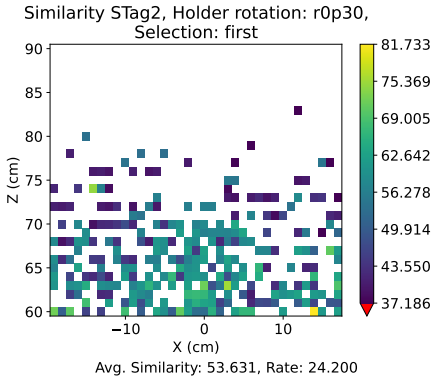

Similarity STag2, Holder rotation: r45p75,
Selection: last
Z (cm)
X (cm)
90.311
85.474
80.637
75.799
70.962
66.124
61.287
56.449
Avg. Similarity: 71.831, Rate: 2.600

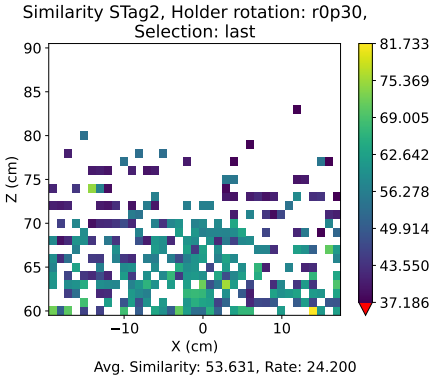
Half occlusion

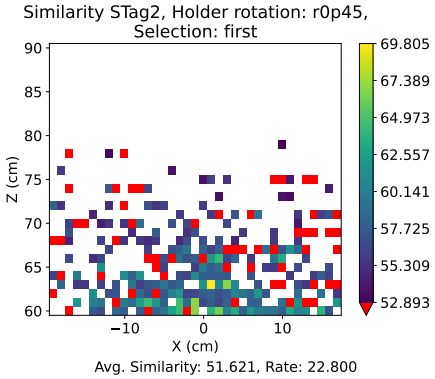
STag2

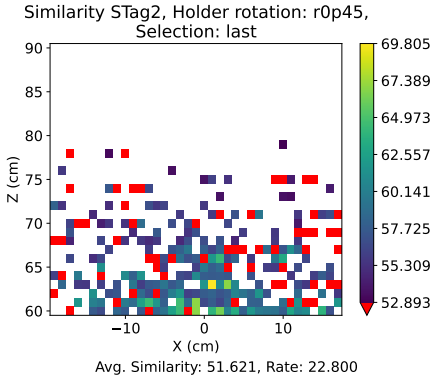


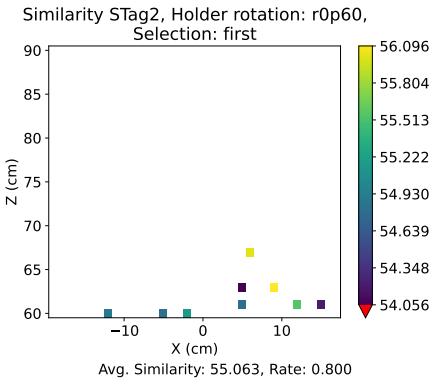


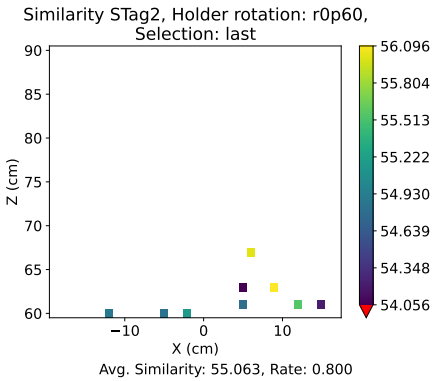


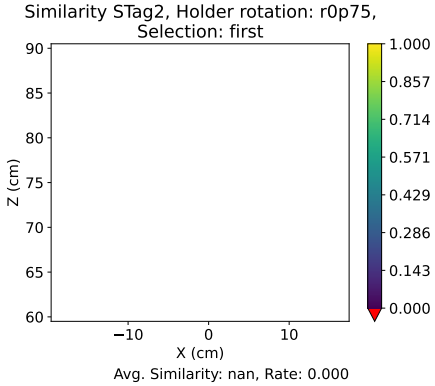


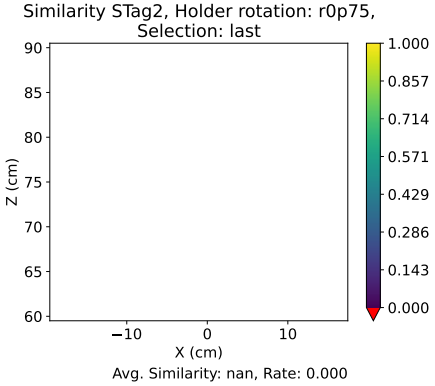


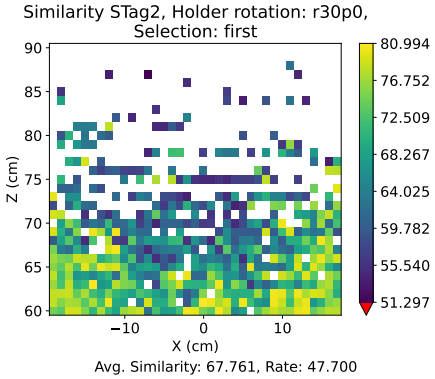


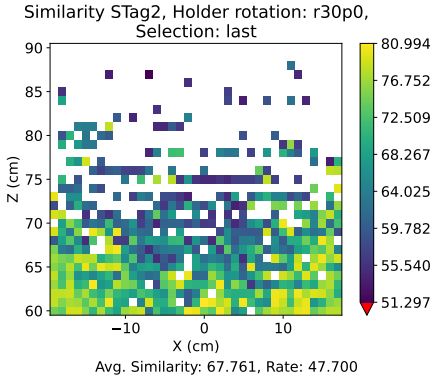


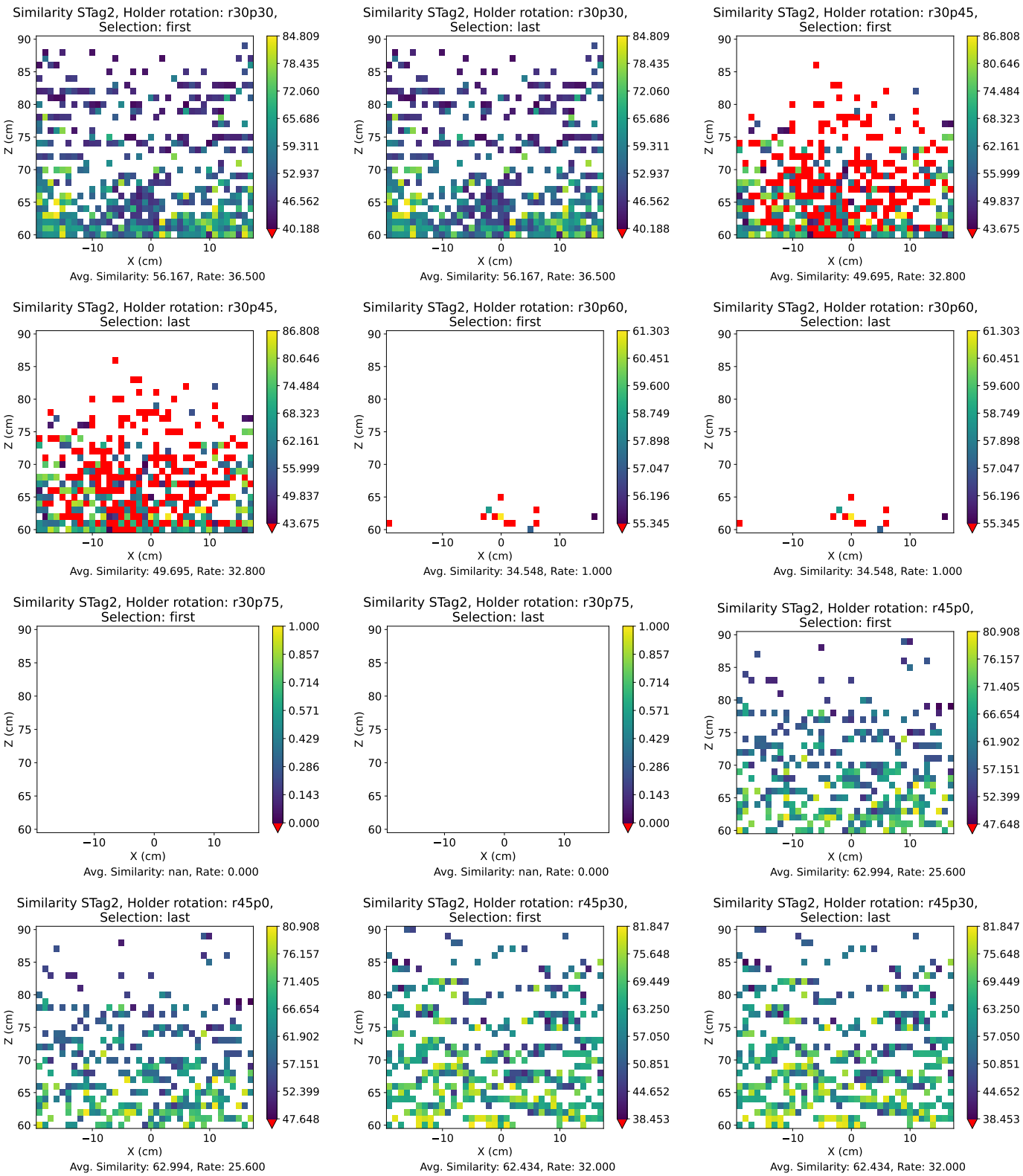


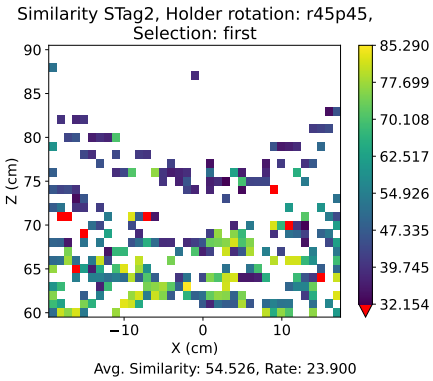


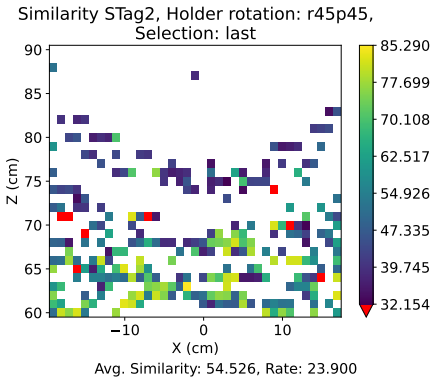


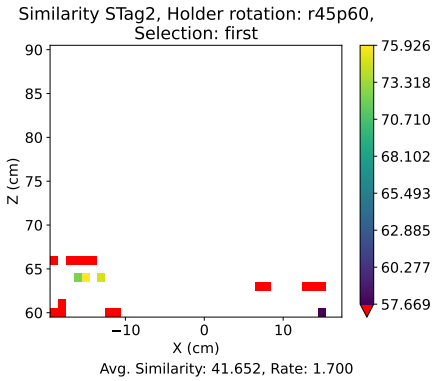


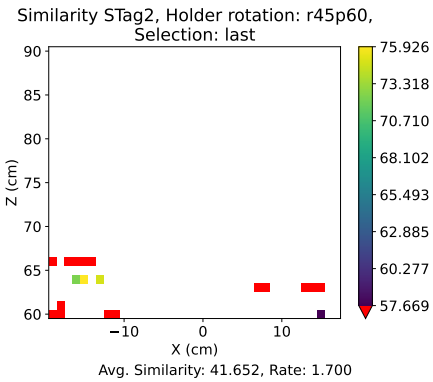


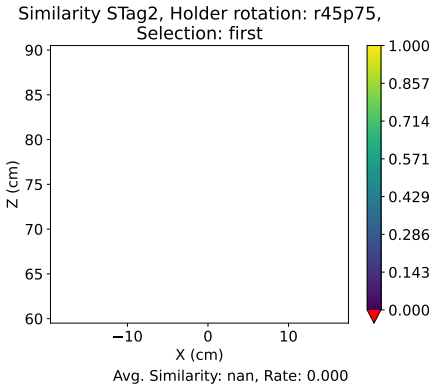

Similarity STag2, Holder rotation: r30p30, Selection: first
Avg. Similarity: 56.167, Rate: 36.500
Similarity STag2, Holder rotation: r30p30, Selection: last
Avg. Similarity: 56.167, Rate: 36.500
Similarity STag2, Holder rotation: r30p45, Selection: first
Avg. Similarity: 49.695, Rate: 32.800
Similarity STag2, Holder rotation: r30p45, Selection: last
Avg. Similarity: 49.695, Rate: 32.800
Similarity STag2, Holder rotation: r30p60, Selection: first
Avg. Similarity: 34.548, Rate: 1.000
Similarity STag2, Holder rotation: r30p60, Selection: last
Avg. Similarity: 34.548, Rate: 1.000
Similarity STag2, Holder rotation: r30p75, Selection: first
Avg. Similarity: nan, Rate: 0.000
Similarity STag2, Holder rotation: r30p75, Selection: last
Avg. Similarity: nan, Rate: 0.000
Similarity STag2, Holder rotation: r45p0, Selection: first
Avg. Similarity: 62.994, Rate: 25.600
Similarity STag2, Holder rotation: r45p0, Selection: last
Avg. Similarity: 62.994, Rate: 25.600
Similarity STag2, Holder rotation: r45p30, Selection: first
Avg. Similarity: 62.434, Rate: 32.000
Similarity STag2, Holder rotation: r45p30, Selection: last
Avg. Similarity: 62.434, Rate: 32.000
X (cm)
Z (cm)

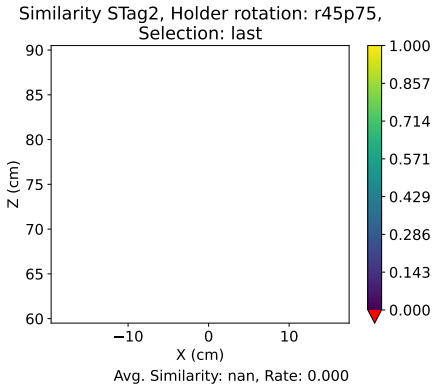

Similarity STag2, Holder rotation: r45p45,
Selection: first
Z (cm)
X (cm)
Avg. Similarity: 54.526, Rate: 23.900


Similarity STag2, Holder rotation: r45p45,
Selection: last
Z (cm)
X (cm)
Avg. Similarity: 54.526, Rate: 23.900


Similarity STag2, Holder rotation: r45p60,
Selection: first
Z (cm)
X (cm)
Avg. Similarity: 41.652, Rate: 1.700


Similarity STag2, Holder rotation: r45p60,
Selection: last
Z (cm)
X (cm)
Avg. Similarity: 41.652, Rate: 1.700


Similarity STag2, Holder rotation: r45p75,
Selection: first
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000


Similarity STag2, Holder rotation: r45p75,
Selection: last
Z (cm)
X (cm)
Avg. Similarity: nan, Rate: 0.000