AprilTags 3D: Dynamic Fiducial Markers for Robust Pose Estimation in Highly Reflective Environments and Indirect Communication in Swarm Robotics

Luis A. Mateos
MIT

Abstract

Although fiducial markers give an accurate pose estimation in laboratory conditions, where the noisy factors are controlled, using them in field robotic applications remains a challenge. This is constrained to the fiducial maker systems, since they only work within the RGB image space. As a result, noises in the image produce large pose estimation errors. In robotic applications, fiducial markers have been mainly used in its original and simple form, as a plane in a printed paper sheet. This setup is sufficient for basic visual servoing and augmented reality applications, but not for complex swarm robotic applications in which the setup consists of multiple dynamic markers (tags displayed on LCD screen).

This paper describes a novel methodology, called AprilTags3D, that improves pose estimation accuracy of AprilTags in field robotics with only RGB sensor by adding a third dimension to the marker detector. Also, presents experimental results from applying the proposed methodology to swarm autonomous robotic boats for latching between them and for creating robotic formations.

1. Introduction

Fiducial markers are commonly use in computer vision (CV) and augmented reality (AR) applications for detection and identification. In robotic applications, fiducial markers have been of crucial importance for obtaining an accurate pose estimation of the marker. Since, estimating the tag’s 6 Degrees of Freedom (DOF) can lead a robotic arm to grab an item, or guide an autonomous robotic boat to a docking station or to latch another robot [12], see Figure 1.

There are different families of markers, with circular and squared shapes, and color base [4]. Circular tags such as Intersense [13] and Rune tags [3] provide accurate pose estimation in short distances with a high computational cost. While, squared fiducial tags such as ARTags [6], ARToolkit [10], ArUco [7], AprilTags [14] and AprilTag2 [16] have a low computational cost and can be detected from a further distance.

In robotics, the AprilTags framework has been preferred for having a lower false positive rate and higher detection rates, even in challenging viewing angles at further distances. However, AprilTags marker systems rely on RGB image space for detection and pose estimation, which is susceptible to the perspective ambiguity noise from sub-pixel detection in their corner locations. Resulting in rotational errors, making the pose estimation challenging without additional information [9].

Sensor fusion methods integrate additional sensors to cope with the RGB detection problems, such as RGBD, which correlates the markers pose detection with the plane...
they are positioned [2]. Other approaches in the Simulta-
neous Localization and Mapping (SLAM) community pro-
pose fusing RGB with inertial data using extended Kalman
filters to enhance the tracking algorithms accuracy [8].
However, for our robotic setup we can only rely on RGB
camera for pose estimation without additional sensor infor-
mation.

In laboratory conditions where the light is regulated, a
single tag is able to give accurate position and orienta-
tion. However, in field robotic applications with multiple
light sources and reflections, relying on one tag will result
in noisy and inaccurate estimations. In order to improve
the pose estimation in noisy environments, a collection of
AprilTags can be used simultaneously to extract a single
"bundle pose", rather than the poses of the individual tags,
see Figure 1b. This is helpful since bundle pose estimation
makes use of every detected tags' corners (making for 4
points if n tags are detected) rather than single tag pose esti-
mation, which only uses 4 points corresponding to the 4 tag
corners [17]. However, the tags are set on the same plane
and a reflection affecting one of them may affect all of them,
and still resulting in a noisy pose estimation.

In order to obtain a reliable pose estimation with April-
Tags using only RGB camera, we propose a simple rather
powerful method, called AprilTags3D, which integrates two
or more tags with the constraint that they are not laying in
the same plane, see Figure 1c. Instead, the tags are rotated
each other, so that in highly reflective environments at least
one tag can still be detected and a proper pose estimation
can be computed. The AprilTags3D can be described as a
tag bundle where each tag is set on its own unique plane.

We present two main contributions in this paper. First,
we describe a novel methodology to improve the pose esti-
mation when only relying on RGB cameras. The method-
ology takes advantage of the third dimension by computing
the 6 DOF from a couple or more markers, shifted in differ-
ent connected planes and rotated a few degrees from each
other, creating a 3D object (similar to a "half-disco ball").
In this way, the arrangement of the rotated tags improves the
position estimation, by interpolating the position and orienta-
tion of the tags with a degree of reliability, mimicking a
Kalman filter for sensor fusion, resulting in a robust position
estimation in noisy conditions in real time.

Also, we propose a method for dynamic identification
and indirect communication in swarm robotic applications.
The method, instead of incorporating the classic paper
printed AprilTags on each robot, integrates LCD screens
that change the displayed tag depending on the state of the
robot without losing its position and orientation.

2. AprilTags3D

Among squared fiducial marker systems, AprilTags [14]
[16] perform the best when detecting smaller markers, in
high angle inclination and in different lighting levels [5].
This is the reason of their popularity in the robotics com-

Figure 2, shows the AprilTags framework. If the cam-
era is positioned in front of the tag, the lateral distance
d_{x} and angle ψ are zero. While the longitudinal distance
between the camera and tag is d_{c} in the X – axis. If the
camera changes its position with the same orientation, then
the lateral distance d_{x} reflects its change in position on the
Y – axis, d_{c} is the same longitudinal distance and ψ = 0.
If the camera in that position changes its orientation to face
the tag, then the lateral distance d_{y} = 0 and the angle reading
becomes the true angle ψ between the camera and tag.

In AprilTags the error in position and orientation varies
with the distance and rotation between the camera and the
marker. The error in distance is linear, since the error in-
creases the further away the marker is from the camera. The
work from [19] and [15] summarizes and quantifies the ac-
curacy of marker-based 3D visual localization.

2.1. AprilTags3D framework

A robotic setup integrating cameras and sensors may be
validated in laboratory conditions, however, this may fail
when used outdoors in real world conditions. This is due to
the noises from the environment that cannot be fully tested
in indoor setups, such as sunlight illumination and reflec-
tions of light from other objects.

The aim of AprilTags3D framework is to minimize the
error in markers detection and pose estimation in real case
scenarios. The framework consists of two or more tags that
are not laying on the same plane, as they are rotated but still
linked to each other, preventing that a single light source
distort or even cancel the tag detection and pose estimation.

The setup consists of at least a couple of markers, one
leader marker Tag_{leader} and one or more follower marker(s)
The markers are linked on one axis as if connected by a hinge and rotated certain degrees in a hyperbolic fashion. The degree of rotation is set from the distance and angle study to obtain the best detection rates [19] [15]. Further, the degree of rotation can be set upon the application and knowledge of the working environment of the robot. Hence, the tags are no longer forming a plane, as they become a 3D object. This novel setup enables us to keep the pose estimation even if the environment is extremely noisy. Moreover, we integrate LCD screens to display the tags dynamically, overcoming low light conditions and enabling indirect communication between robots. In this sense, a robot in a swarm can notify other robots about its state by changing the tag displayed on its LCD without affecting the tag detection, neither the pose estimation, see Figure 3.

2.1.1 AprilTags3D markers

The simplest configuration integrates a couple of markers, a Tag\textsubscript{leader} and a Tag\textsubscript{follower} connected in the $Z$-axis and rotated $g$ degrees. The marker nearer the camera will receive a higher weight, since the error is lower in short distances [19]. The marker with the best orientation will receive another weight [15]. The process is to combine the weights and take the higher confidence level or “heavier”.

In the optimal case, when the two tags are detected, the data is filtered with a nonlinear digital filtering technique (median) and then the joint pose estimation is computed from the couple of 6 DOF to refine the real pose estimation. For example, Figure 3 shows the physical setup consisting of a couple of LCD screens, Tag\textsubscript{leader}, linked to Tag\textsubscript{follower} rotated $10^\circ$ in the $Z$-axis. As long as the Tag\textsubscript{leader} is visible and nearer to the camera its pose distribution is more reliable than the Tag\textsubscript{follower}, since the follower is rotated and with larger distance to the camera. In the case the Tag\textsubscript{leader} is the only one detected, it is taken into account as a single marker with its corresponding confidence degree.

The framework is similar to a Kalman filter for sensor fusion. In this case, instead having sensors there are markers, and the noise from the sensors in the filter is defined as the degree of confidence from the rotation and distance of the markers to the camera. In this scheme, the advantage is that although only one measurement is available, the pose estimation can be done. The more the measurement signals become available, the estimation will be more accurate, see Figure 4.
2.1.2 AprilTags3D for Indirect Communication in Swarm Robotics

Swarm robotics refer to robots controlled by themselves, without a centralized control coordinator. If individual robots integrate the AprilTags3D framework with LCD screens and cameras, then, these independent entities can indirectly communicate between them by being able to detect the tag on another robot and infer the meaning of its marker.

If a robot with a camera $R_c$ is following the tag on another robot $R_t$ and the tag is dynamically changing. The pose estimation computed by $R_c$ will not be affected by the changing tag. Instead, $R_c$ reads a different tagID that may communicate a changing state in $R_t$ or a direct order to $R_c$ to perform certain action. This change of tagID can be generalized for swarm robotics as one robot can communicate with all other robots or with only one is specific.

Moreover, the complexity of this indirect communication is minimal and can be extended to $N$ robots, as long as each robot can see the tag on another robot. Section 4.3 shows experimental results of AprilTags3D for robot formations.

3. Autonomous Robotic Boat

The robotic boat consists of a rectangular base (2:1 ratio) with four thrusters in the middle of the edges, see Figure 5a. In this relationship the robot is able to move forward, backward, sideways and able to rotate on its axis. The dimension of the robotic platform are $1000\text{mm} \times 500\text{mm} \times 150\text{mm}$. The dynamics and model of the robotic boat can be found in [18].

The robotic boat integrates a 360° lidar for navigation and obstacle avoidance for moving long distances $D > 10\text{m}$. While, for latching in short distances $d < 10\text{m}$ the camera - target principle is implemented.

3.1. Latching system

The robotic boats are able to latch to a docking station or to another boat with a couple of mechanisms: 1) an actuated funnel and 2) a pin with a bearing stud, see Figure 5c. In both cases the robot detects with its camera the tag on the other entity, computes the pose estimation and latch to the pin on the robot with the tag or to the docking station [12].

When the latching is achieved, the funnel and pin creates a spherical joint between the entities that enables them to overcome wave disturbances while connected.

3.1.1 Guiding system

The robot’s localization is based on lidar with NDT matching [1] and has an accuracy of $\pm 100\text{mm}$ in open spaces. However, this precision is not enough for putting together the boats on open water, since an accuracy of $\pm 40\text{mm}$ is required for performing the latching between the robots. The robotic boats simplify the 3D positioning to a 2D challenge.
Since, the entities are floating at similar levels above the water and the misalignment from the waves are compensated by the funnel, see Figure 7.

Therefore, we integrated the AprilTags frameworks in the autonomous robotic boats for guiding one robot to a target, which can be a docking station or to another robot.

The classic AprilTags framework is able to guide the robot in environments with low noise, while in noisy conditions, the yaw angle varies several degrees and the robot is not able to reach the target, see Section 4.1.1. On the other hand, the AprilTags3D framework is able to overcome noises from the environment and precisely position the robot within the tolerances and perform the latching, see Section 4.1.2.

4. Experimental Results

The challenge we want to solve with AprilTags3D is to have a reliable marker detection and high accuracy in pose estimation in highly reflective environments. In addition, we want to indirectly communicate the robot’s state with only the dynamic markers in a swarm robotics fashion.

We performed three different experiments, in the first two tests we used classic printed AprilTag and our AprilTags3D to compare the results: 1) One robotic boat latch to a docking station in an indoors swimming pool with light reflections from the windows and head lamps. 2) Instead of latching to a docking station, we set two robotic boats to latch outdoors on the open water. 3) We test the indirect communication for swarm robotics with three autonomous robotic boats performing the train link formation with the dynamic AprilTags3D. Videos from the experiments available in [11]

4.1. Indoor test in swimming pool

The setup consist of one autonomous robotic boat $R_1$ with a RGB camera and an actuated funnel, $R_1$ is located in a position $(R_{1x}, R_{1y})$ in the swimming pool. $R_1$ is dismissed, since the controllable space is defined as a 2D plane on the water $(W_x, W_y)$. A docking station with a squared tag dimension $l = 0.13m$ and a pin for latching is located at $(T_x, T_y, T_z)$.

The robot’s camera can detect and calculate the pose of the marker, which is located on the docking station a couple of meters apart. In case, the robot fails to latch due to noise in the pose estimation readings, it executes an auto-recovery algorithm to reposition itself to retry the docking action [12].

A successful latching is achieved when $d_x < 10mm$, $d_y < \pm 40mm$ and the yaw angle is $< \pm 27.5^\circ$.
The results show an improved detection and pose estimation, enabling the robotic boat to find and latch to the pin 99% of the time. Table 1 compares the detection rate and pose estimation error with both, the classic AprilTags and the AprilTags3D.

### 4.2. Outdoor test in River

The setup consist of two autonomous robotic boats performing a latching action on the river. The robot carrying the tag $R_1$ tries to maintain its position on $(x, y)$ from its lidar sensor, with an error in position in the range of $±100 mm$. While, robot $R_2$ tries to detect the tag, compute the pose estimation and latch to the "balancing $R_1$ robot". The robot $R_1$ integrates the tag and the pin, and the robot $R_2$ integrates the camera with the actuated funnel.

Initially, $R_2$ is set 2000 mm away from $R_1$ in a direct line, similar to the previous experiments in the swimming pool. In the same way, if the robot fails to latch at the first try, it will retry again autonomously until it latches.

In this experiments, a successful latch is achieved when $d_y < ±40 mm$, yaw angle is $< ±27.5^\circ$ and $d_x < 500 mm$.

Since, the tag is mounted inside the robot, see Figure 5b.

### 4.2.1 Outdoor test with classic printed AprilTag

In order to latch the robots, robot $R_2$ is guided by visual cues to a position $(x, y)$. This coordinate must be reach with an error $< ±40mm$ for a successful latching of robot $R_1$. In this challenging task, the robots are not stable on the ground, they are floating on the river, trying to maintain their positions, overcoming waves, wind and water currents.

From the camera point of view it is a big challenge to process the 6DOF from the tags located in another boat when it moves with the waves at a degree that the pitch and roll angles vary a few degrees. Further, we are only relying on RGB sensors to perform the latching on the boats.

On the river, the main problem with the classic markers was the detection. The marker was detected only 60% of the time, due to the sunlight reflections, see Figure 9. The reason for this low detection rate were the waves, since these created oscillations on both robots, changing their position.
4.2.2 Outdoor test with dynamic AprilTag3D

We integrated a couple of tags on the robot R₁. The Tag_{leader} replaced the printed tag, with position (R₁Tₓ, R₁Tᵧ, R₁Tₗ), and the Tag_{follower} was set similar to the indoor experiments, see Figure 5b.

The results revealed a consistent tag detection, detecting at least one tag 95% of time. The framework shows a minimization in the environmental noises and an improved detection rate, see Table 1.

4.3. Dynamic AprilTag3D for swarm robotics

In our framework for joining the robots, the latching state of the robot can be registered. When a robot latches, the actuated funnel closes a socket, trapping the bearing stud from another robot or from the docking station. This allows for a robot to communicate with another robot or from the docking station. In this sense, the robot performing the latching action can communicate the state of its actuated funnel by changing its tag dynamically, or on its LCD screen.

Since each robot is an independent entity and each robot speaks the same language, meaning they understand a set of common tags. Then, they can communicate in a swarm robotic fashion with the constraint that the robots must be able to see and detect the tag on another robot.

This framework works for multiple case scenarios, Figure 10 shows the robotic boats performing a train link formation. Moreover, this case scenario can be scale to N robots with minimal complexity. The framework can be seen as the robots playing the telephone game. One robot passes a message to another and this one passes to the next ad so on. The passing of message can encode a robotic formation, a position and/or a change of state, see Figure 11.

5. Conclusion

We presented a novel methodology of squared fiducial markers called AprilTags3D which improves the detection and pose estimation of the classic AprilTags in field robotics in real environments.

We performed multiple tests indoors and outdoors with robotic boats that autonomously latch to a docking station and to another robot with the camera - target principle. The experimental results using both methodologies, revealed that our method improves the detection and pose estima-
tion in real case scenarios. Specially, in highly reflective environment, such as aquatic environments inside the city (rivers and lakes) with sunlight reflections from cars, building and other boats.

In the experiments we simplified the 3D space to a 2D plane on the water surface for guiding the robotic boats. Nevertheless, the proposed AprilTags3D framework can be integrated in drones, submarines or space robotics required to navigate in 3D space.

Also, we presented a novel concept to indirectly communicate robots in a swarm by only changing their dynamic tag depending on its state. This simple rather powerful framework can be scaled to \( N \) robots with minimal complexity for robot formations. We showed how three robotic boats "talk" to each other to perform robotic formations on the river.

References

[1] N. Akai, L. Y. Morales, E. Takeuchi, Y. Yoshihara, and Y. Ninomiya. Robust localization using 3d ndt scan matching with experimentally determined uncertainty and road marker matching. In 2017 IEEE Intelligent Vehicles Symposium (IV), pages 1356–1363, June 2017. 4

[2] A. Assa and F. Janabi-Sharifi. A robust vision-based sensor fusion approach for real-time pose estimation. IEEE Transactions on Cybernetics, 44(2):217–227, Feb 2014. 2

[3] F. Bergamasco, A. Albarelli, E. Rodol, and A. Torsello. Rune-tag: A high accuracy fiducial marker with strong occlusion resilience. In CVPR 2011, pages 113–120, June 2011. 1

[4] J. DeGol, T. Bretl, and D. Hoiem. Chromatag: A colored marker and fast detection algorithm. 2017 IEEE International Conference on Computer Vision (ICCV), pages 1481–1490, 2017. 1

[5] D. B. dos Santos Cesar, C. Gaudig, M. Fritsche, M. A. dos Reis, and F. Kirchner. An evaluation of artificial fiducial markers in underwater environments. In OCEANS 2015 - Genova, pages 1–6, May 2015. 2

[6] M. Fiala. Artag, an improved marker system based on artoolkit. In National Research Council Canada, volume 47419, page 2004, 2004. 1

[7] S. Garrido-Jurado, R. Muoz-Salinas, F. Madrid-Cuevas, and M. Marn-Jimnez. Automatic generation and detection of highly reliable fiducial markers under occlusion. Pattern Recognition, 47(6):2280 – 2292, 2014. 1

[8] O. S. Gedik and A. A. Alatan. Rgbd data based pose estimation: Why sensor fusion? In 2015 18th International Conference on Information Fusion (Fusion), pages 2129–2136, July 2015. 2

[9] P. Jin, P. Matikainen, and S. S. Srinivasa. Sensor fusion for fiducial tags: Highly robust pose estimation from single frame rgbd. In 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 5770–5776, Sept 2017. 1

[10] H. Kato. Artoolkit: library for vision-based augmented reality. In IEICE 2002, volume 6, pages 79–86, 2002. 1