



## UvA-DARE (Digital Academic Repository)

### Position and Altitude of the Nao Camera Head from Two Points on the Soccer Field Plus the Gravitational Direction

Oomes, S.; Visser, A.

**DOI**

[10.1007/978-3-031-85859-8\\_14](https://doi.org/10.1007/978-3-031-85859-8_14)

**Publication date**

2025

**Document Version**

Final published version

**Published in**

RoboCup 2024: Robot World Cup XXVII

**License**

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment>)

[Link to publication](#)

**Citation for published version (APA):**

Oomes, S., & Visser, A. (2025). Position and Altitude of the Nao Camera Head from Two Points on the Soccer Field Plus the Gravitational Direction. In E. Barros, J. P. Hanna, H. Okada, & E. Torta (Eds.), *RoboCup 2024: Robot World Cup XXVII* (pp. 164-175). (Lecture Notes in Computer Science; Vol. 15570), (Lecture Notes in Artificial Intelligence). Springer. [https://doi.org/10.1007/978-3-031-85859-8\\_14](https://doi.org/10.1007/978-3-031-85859-8_14)

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*



# Position and Altitude of the Nao Camera Head from Two Points on the Soccer Field Plus the Gravitational Direction

Stijn Oomes<sup>1</sup> and Arnoud Visser<sup>2</sup>(✉) 

<sup>1</sup> Oomes Vision Systems, Amsterdam, The Netherlands

<sup>2</sup> Intelligent Robotics Lab, Universiteit van Amsterdam,  
Amsterdam, The Netherlands

a.visser@uva.nl

<https://www.intelligentroboticslab.nl/>

**Abstract.** To be able to play soccer, a robot needs a good estimate of its current position on the field. Ideally, multiple features are visible that have known locations. By applying trigonometry we can estimate the viewpoint from where this observation was actually made. Given that the Nao robots of the Standard Platform League have quite a limited field of view, a given camera frame typically only allows for one or two points to be recognized.

In this paper we propose a method for determining the  $(x, y)$  coordinates on the field and the height  $h$  of the camera from the geometry of a simplified tetrahedron. This configuration is formed by two observed points on the ground plane plus the gravitational direction. When the distance between the two points is known, and the directions to the points plus the gravitational direction are measured, all dimensions of the tetrahedron can be determined.

By performing these calculations with rational trigonometry instead of classical trigonometry, the computations turn out to be 28.7% faster, with equal numerical accuracy.

The position of the head of the Nao can also be externally measured with the OptiTrack system. The difference between externally measured and internally predicted position from sensor data gives us mean absolute errors in the 3–6 cm range, when we estimated the gravitational direction from the vanishing point of the outer edges of the goal posts.

## 1 Introduction

RoboCup is an event where robot capabilities are tested by playing a game of soccer between one team against another [12]. Already from the beginning of the competitions, the importance of visual localization was recognized [3, 11]. Since that start methods like Visual Odometry have been developed, but those methods rely on enough dense and distinctive textures [1]. In the RoboCup often not much more than the green turf is visible, with an occasional white line (if not obstructed by other robots nearby). When a L-, T- or X-intersection on a

white line-crossing can be observed [16], a distinctive point on the field has been recognized (although due to the symmetry of the field unfortunately not unique). Two of those distinctive points are enough to estimate the location where this observation is made, as demonstrated in this paper.

## 2 Rational Trigonometry

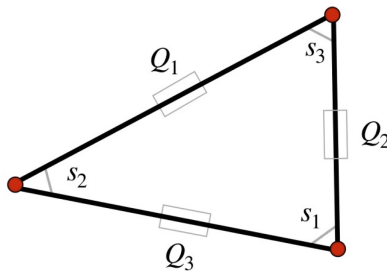
To understand the relations between observable variables (measurements with one or more cameras) and not directly observable variables (3D properties of objects and their configurations) transformations have to be defined. Rational Trigonometry [23] uses quadrances and spreads instead of distances and angles. The quadrance is defined as the quadratic sum of differences in the coordinate values of two points. The spread is defined as a proportion of quadrances; for a right triangle it is the opposite quadrance divided by the hypotenuse quadrance. Both quantities can be easily converted back and forth to the well-known concepts of distance and angle, although they are defined *independently*. A **quadrance**  $Q$  is the square of the distance  $d$ . This makes it in essence a measure of area in Euclidean space. Quadrances have dimension  $m^2$ .

$$Q = d^2 \quad (1)$$

A spread measures the divergence between two lines. A **spread**  $s$  is the square of the sine of an angle  $\alpha$ .

$$s = \sin^2 \alpha \quad (2)$$

Spreads have a range between  $[0 - 1]$  and are dimensionless, but is actually the ratio of quadrances (so dimension  $\frac{m^2}{m^2}$ ). This characteristic has important consequences, because this range is better balanced than the ranges of *circs* ( $[0 - 360^{circ}]$ ) and radians ( $[0 - 2\pi]$ ) once translations and rotations have to be combined. It also means that the spread has no sign, and is actually the same value for adjacent angles. This prevents many ambiguity checks in the calculations.



**Fig. 1.** The quadrances  $Q_i$  and spreads  $s_i$  of a triangle.

The relation between quadrances and spreads of triangle, as illustrated in Fig. 1 is given by the **Spread law** of rational trigonometry:

$$\frac{s_1}{Q_1} = \frac{s_2}{Q_2} = \frac{s_3}{Q_3} \quad (3)$$

Yet, for the method applied in this paper mostly the **Cross law** of rational trigonometry is used:

$$(Q_1 + Q_2 - Q_3)^2 = 4Q_1Q_2(1 - s_3) \quad (4)$$

The Cross law (Eq. 4) can be used to calculate an unknown quadrance  $Q_3$  from two known quadrances  $Q_1$  and  $Q_2$  and the spread  $s_3$  between those two quadrances (and opposite the unknown quadrance  $Q_3$ ). Once  $s_3$  and all three quadrances  $Q_{1\dots 3}$  of a triangle are known, the other two spreads  $s_1$  and  $s_2$  can be calculated from the Spread law (Eq. 3).

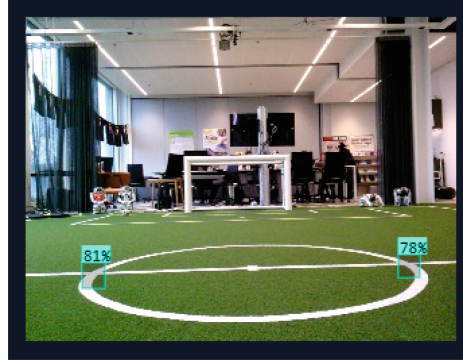
### 3 Method

The first phase is the detection of the relevant objects in the world, that is, on the soccer field. Those are the white lines on the green field, especially the X-, T- and L-intersections. Of course there are also objects like the goals, the ball and the other robots; team-members and opponents are distinguishable by their different colored jerseys [18]. The distance between two of those points detected on the field can be used to estimate the position where those points are observed from the camera position. The position of the X-, T- and L-intersections are specified in the Standard Platform League rule book [21], so their distances are known.

For efficiency reasons in RoboCup typical scan-line techniques are used to detect lines, searching for regions of white pixels surrounded with green pixels [16]. Yet, many other white objects are present on the field (such as goal-posts, the ball and other robots), so additional filters are needed to reduce the number of false positives [4]. Instead of hand-crafting such filters for the lines and their crossings, we used a state-of-the-art detector.

The X-, T- and L-intersections in this paper are detected with the detector based on YOLO v8, a detector which is slightly easier to fine-tune on new types of objects than YOLO v9 [22]. The detector is based on the YOLO v8 XL model pretrained on the COCO dataset, further finetuned on 1309 L-intersections, 1229 T-intersections and 806 X-intersections. These intersections were already annotated in the TORISO-21 dataset [2], but further extended with annotated images recorded by the Dutch Nao Team [10]. This dataset, including the YOLO v8 XL model trained on this data, is publicly available<sup>1</sup>. This model achieved a precision of 86.1% and a recall of 84.2%. An example of such detection of two X-intersections can be seen in Fig. 2. The intersections further away are also detected, but with a too low confidence to be displayed (a confidence below 50%). From the rule book it is known that the distance between the two X-intersections is 1.5 m.

<sup>1</sup> <https://universe.roboflow.com/fieldmarks/splfieldmark>, courtesy Gijs de Jong.



**Fig. 2.** The X-intersection detected on the Standard Platform League field.

### 3.1 Position from 3 Known Points

When three points are known in the world, one can formulate a third *circ* polynomial which once solved gives 4 solutions. This is known as the **P3P** problem, which was first solved by Grunert [8] and further refined by Finsterwalder & Scheufele [5]. Grunert solution is nicely described by [9], which starts with the following three equations generated by the law of cosines:

$$\begin{aligned}
 x^2 + y^2 - 2xy \cos(\alpha) &= a^2 \\
 x^2 + z^2 - 2xz \cos(\beta) &= b^2 \\
 y^2 + z^2 - 2yz \cos(\gamma) &= c^2
 \end{aligned} \tag{5}$$

Finsterwalder has shown that this is equivalent with solving  $\lambda$  in this third *circ* polynomial for the P3P problem:

$$\begin{aligned}
 &\lambda^3 c^2 (c^2 \sin^2 \alpha - a^2 \sin^2 \gamma) + \\
 &\lambda^2 [a^2 (b^2 - a^2) \sin^2 \gamma - c^2 (2b^2 + c^2) \sin^2 \alpha + 2a^2 c^2 (1 - \cos \alpha \cos \beta \cos \gamma)] + \\
 &\lambda [a^2 (a^2 - c^2) \sin^2 \beta + b^2 (b^2 + 2c^2) \sin^2 \alpha - 2a^2 b^2 (1 - \cos \alpha \cos \beta \cos \gamma)] + \\
 &b^2 [a^2 \sin^2 \beta - b^2 \sin^2 \alpha] = 0
 \end{aligned} \tag{6}$$

After the overview of Haralick [9] several other researchers came with solutions for the P3P problems, such as [7] and [15].

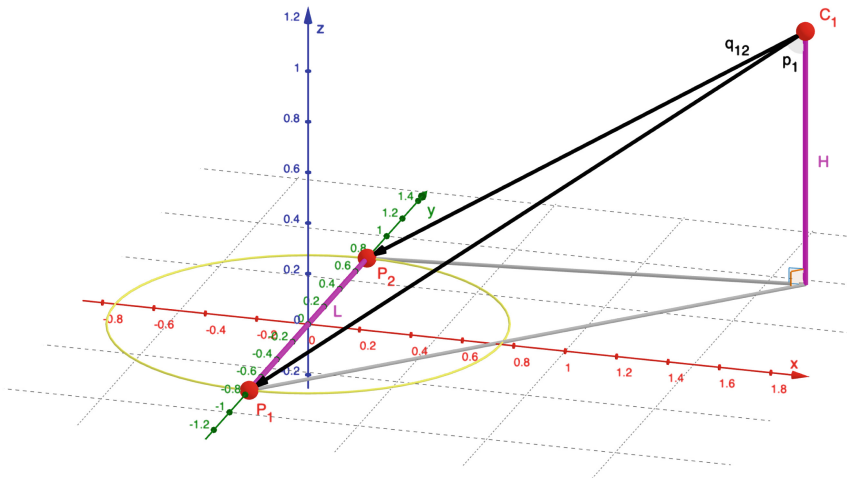
When a fourth point is known one can disambiguate between the four solutions, which makes P3P a special case of the PnP problem with  $n$  the number of known points. Yet, in our case only two points are known:  $n = 2$ .

### 3.2 Position from 2 Points on the Field and Gravity

The two point problem can be solved if we have a robot and a measure of the direction of gravity. Actually, there are two ways to get a measure on the direc-

tion of gravity. Initially we used the measurement of the an Inertial Measurement Unit (IMU) on-board of the robot. By also using the measurement values of the pitch and yaw angles of the head relative to the torso, this direction is transformed to the coordinate frame of the camera. Yet, together with timing errors small directional deviations resulted in positional errors in the range of 8–25 cm range. As a check, we also derived the gravitational direction from the perspective in the image, from the vanishing point of the outer edges of the goal posts. Here we found that the mean absolute errors are much better in the 3–6 cm range. So, for the rest of this paper we use the gravitational direction from the perspective, although this means that this direction can only be estimated when parallel verticals structures are visible. With better calibration of the gravitational direction from the IMU and the kinetic chain, we could skip this intermediate step.

For two points on the soccer field we know the distance, so also the quadrance  $L$ . We can determine their directions in camera coordinates from the projection in the image. With the direction of gravity  $g$ , we can convert those two directions to the spreads  $p_1$  and  $p_2$  between the visual directions and the gravitational direction  $g$ . The three vectors form a tetrahedron with the camera on top, as illustrated in Fig. 3. The spread between the two visual directions is  $q_{12}$ . The quadrance  $H$  is along the gravitational direction  $g$  and that is the parameter to be determined.



**Fig. 3.** A line  $L$  between two known points seen from a camera. The red points on the ground plane are the intersections of middle line and circle of which the quadrance  $L$  is known. The red point on the top right is the camera point. Together with the foot of the altitude these points form a tetrahedron with two right angles. The altitude quadrance  $H$  can be determined from quadrance  $L$  and the spreads  $p_1$ ,  $p_2$ , and  $q_{12}$  (Color figure online).

The tetrahedron has 4 vertices, 6 edges, and 4 faces. So, in the tetrahedron 6 distances/quadrances can be distinguished (see the three colored pairs in Fig. 3) and 12 angles/spreads. One quadrance  $L$  and five spreads are known:  $(p_1, p_2, q_{12})$  and the two right angles formed by  $H$  and the ground plane. The other spreads and quadrances can be derived from those measurements.

From the three measured spreads  $p_1$ ,  $p_2$ , and  $q_{12}$  and the known quadrance  $L$  we can determine the vertical height, denoted as the quadrance  $H$ :

$$H = \frac{(1-p_1)(1-p_2)}{(1-p_1) + (1-p_2) - 2\sqrt{(1-p_1)(1-p_2)(1-q_{12})}} L \quad (7)$$

If we define P1 in Fig. 3 as the origin, and the Y-axis is the vector along line  $L$  in the direction of P2, the  $(X, Y)$  coordinate where quadrance  $H$  touches the ground plane can be calculated:

$$X = \frac{(1-p_1)(1-p_2)[q_{12} - (1-p_1) - (1-p_2) + 2\sqrt{(1-p_1)(1-p_2)(1-q_{12})}]}{[(1-p_1) + (1-p_2) - 2\sqrt{(1-p_1)(1-p_2)(1-q_{12})}]^2} L \quad (8)$$

$$Y = \frac{[(1-p_2) - \sqrt{(1-p_1)(1-p_2)(1-q_{12})}]^2}{[(1-p_1) + (1-p_2) - 2\sqrt{(1-p_1)(1-p_2)(1-q_{12})}]^2} L \quad (9)$$

Note that in both equations no transcendental trigonometric functions like  $\sin(\phi)$ ,  $\arccos(\phi)$  and  $\tan(\phi)$  are used.

### 3.3 Classical Trigonometry

The same method can also be formulated with distances and angles, which gives the following equations:

$$x = \frac{\cos(\alpha_1) \cos(\alpha_2) \sqrt{\sin^2(\beta_{12}) - \cos^2(\alpha_1) - \cos^2(\alpha_2) + 2\cos(\alpha_1) \cos(\alpha_2) \cos(\beta_{12})}}{\cos^2(\alpha_1) + \cos^2(\alpha_2) - 2\cos(\alpha_1) \cos(\alpha_2) \cos(\beta_{12})} l \quad (10)$$

$$y = \frac{\cos^2(\alpha_2) - \cos(\alpha_1) \cos(\alpha_2) \cos(\beta_{12})}{\cos^2(\alpha_1) + \cos^2(\alpha_2) - 2\cos(\alpha_1) \cos(\alpha_2) \cos(\beta_{12})} l \quad (11)$$

$$h = \frac{\cos(\alpha_1) \cos(\alpha_2)}{\sqrt{\cos^2(\alpha_1) + \cos^2(\alpha_2) - 2\cos(\alpha_1) \cos(\alpha_2) \cos(\beta_{12})}} l \quad (12)$$

When optimized (removing redundant calculations), this algorithm requires 17x a transcendental function-call to calculate  $\cos(\phi)$  and  $\sin(\phi)$ .

## 4 Computational Experiments

Wildberger claims that rational trigonometry increases both the computational accuracy and the computational speed when compared to classical trigonometry [23]. Let us check whether we have found any evidence for this claim.

If we look at the accuracy of both computations we do not find any significant differences. The deviations start showing up around the 12th decimal. That is way beyond the measurement accuracy and precision for our position measurements.

The code to calculate the camera positions from two points on the field with *quadrances* and *spreads* was executed multiple times on a Nao v6 robot, which indicates that one calculation takes on average  $2.47 \cdot 10^{-05}$  s. The implementation was in Python 2.7, the language supported by the standard Aldebaran NaoQi image, so could be further optimized in C++ or RUST. Yet, the Python implementation can already be used for a comparison with the same calculation with traditional method based on *angle* and *distance*, which needed for one calculation on average was  $3.18 \cdot 10^{-05}$  s (Table 1). This indicates that the rational algorithm on the Nao v6 robot is  $\approx 28.7\%$  faster, because it performs the calculations with only divisions and square-roots as computational heavy operations, in contrast with the trigonometric functions like  $\cos(\phi)$  and  $\sin(\phi)$  used in the classical method.

**Table 1.** Execution time of algorithm in seconds

# runs	50x	500x	5000x	50.000x	500.000x
classical algorithm	$1.58 \cdot 10^{-3}$	$1.59 \cdot 10^{-2}$	$1.61 \cdot 10^{-1}$	$1.59 \cdot 10^0$	$1.59 \cdot 10^{+1}$
rational algorithm	$1.24 \cdot 10^{-3}$	$1.24 \cdot 10^{-2}$	$1.23 \cdot 10^{-1}$	$1.23 \cdot 10^0$	$1.23 \cdot 10^{+1}$

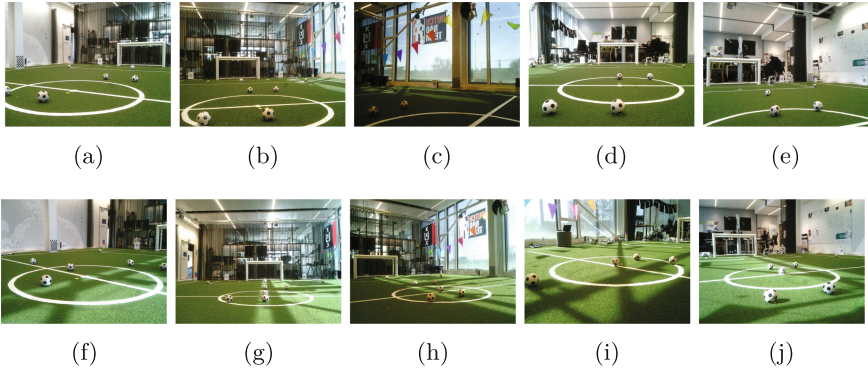
## 5 Metrical Experiments

To see the relative location  $(x, y, h)$  of the camera in the head of the Nao v6 robot relative to two points on the soccer field, we recreated two Simultaneous Localization and Mapping experiments, which were used as Technical Challenge in 2004 and 2005 [19, 20].

The 2004 SLAM challenge specified five points:  $[(160, 100), (180, -30), (50, -100), (-210, 0), (-100, 50)]$  - all cm from the origin at the center of the field. Those points were marked with red tape on the field. The 2005 SLAM challenge specified another five points:  $[(130, 120), (220, -150), (-160, -120), (-210, 90), (270, 0)]$ , they were marked with black tape on the field.

The soccer fields of 2004 and 2005 were smaller than the current field, so all points are sampled near the center (see Fig. 5a) The collected data is publicly available<sup>2</sup>. As can be seen from Fig. 4.(f-j), the black points from the 2005 SLAM challenge are slightly further from the center, so here the center circle is always fully visible. Also the weather conditions changed; the sun was casting shadows when the 2005 SLAM locations were recorded.

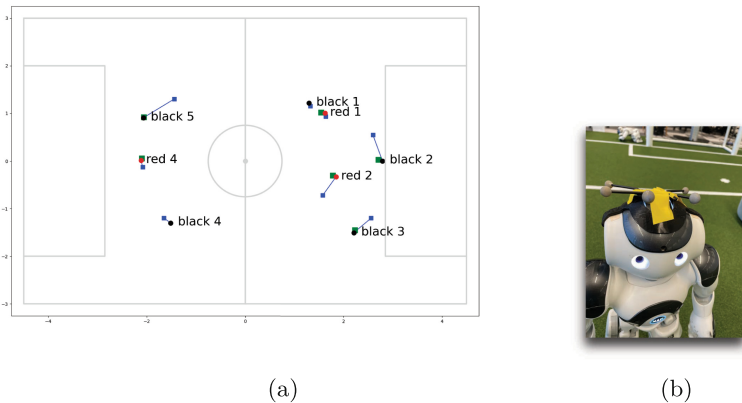
<sup>2</sup> [https://github.com/physar/landmark\\_based\\_slam\\_dataset](https://github.com/physar/landmark_based_slam_dataset).



**Fig. 4.** Viewpoints from the 2004 and 2005 SLAM challenge locations

To provide ground truth, the Nao robot was equipped with a rigid body marker of the OptiTrack system (see Fig. 5b), consisting of eight Flex 13 cameras, which can record movements with a 3D accuracy of  $\pm 0.20$  mm in a tracking area of  $\pm 8$  m  $\times$  4 m [6].

The estimates of location of the observations made on the 10 marked positions from the 2004 and 2005 SLAM challenge are summarized in Table 2 and 4. As can be seen from Table 2, no predictions were made for red point 3 and point 5, because the center circle was not in full view. Also can be seen from the two tables combined that in half of the cases (4 of the 8 marked positions) a reasonable good estimate is given, while for the other half the estimate is clearly off (indicated with a red italic font). Yet, this clearly wrong location-estimates can be filtered out by setting a threshold on the predicted height, for instance by the constraint that the height as to be in the range (+0.440, +0.550).



**Fig. 5.** Top view of the 2004 and 2005 SLAM challenge locations and a Nao robot with an OptiTrack rigid body marker

**Table 2.** Predictions of the observations at the 2004 SLAM challenge locations

	$x$	$y$	$h$	
<b>2004 SLAM challenge</b>				
<b>red point 1</b>	+1.600	+1.000		m
OptiTrack reference	+1.626	+1.004	+0.449	m
prediction with $g$ -vector from IMU	+1.637	+0.939	+0.483	m
prediction with visual $g$ -vector	+1.545	+1.023	<i>+0.369</i>	m
<b>red point 2</b>	+1.800	-0.300		m
OptiTrack reference	+1.849	-0.330	+0.452	m
prediction with $g$ -vector from IMU	<i>+1.576</i>	<i>-0.721</i>	<i>+0.364</i>	m
prediction with visual $g$ -vector	+1.784	-0.306	+0.468	m
<b>red point 3</b>	+0.500	-1.000		m
<b>red point 4</b>	-2.100	0.000		m
OptiTrack reference	-2.115	+0.014	+0.452	m
prediction with $g$ -vector from IMU	-2.083	-0.125	+0.448	m
prediction with visual $g$ -vector	-2.100	+0.056	<i>+0.390</i>	m
<b>red point 5</b>	-1.000	+0.500		m

Yet, a reliable observation prediction in only half of the cases is not good enough to be useful as input for a localization algorithm. To test whether the origin of this failure comes from the IMU measurement of the  $g$ -vector, we also estimated the  $g$ -vector from the goal-posts, when they were visible in the field of view. This improved the prediction of the observation locations considerably. With a visual  $g$ -vector we found the following errors between the prediction and the reference measurement from the OptiTrack system as indicated in Table 3.

**Table 3.** Difference between predictions with visual  $g$ -vector and the OptiTrack reference system

	$x$	$y$	$h$	
mean absolute error	0.044	0.030	0.076	m
standard deviation	0.037	0.017	0.069	m

Note that for black point 1 and point 3 of the 2005 SLAM challenge (Table 4) no goal-posts were in the field-of-view, so visual  $g$ -vector could be estimated. Also note for red point 1 and point 4 of the 2004 SLAM challenge (Table 2) the  $(x, y)$  prediction with the visual  $g$ -vector is quite good, but that the constraint on the height-estimate would indicate that this observation could better be skipped.

In general, a small error in the viewing direction can give large errors in position estimates. The underlying origin of this small error in estimating the  $g$ -vector by the IMU can be drift of the accelerometer itself, or a buildup of an error the kinematic chain of several joints between orientation of the head relative to the torso (where the IMU is located).

**Table 4.** Predictions of the observations at the 2005 SLAM challenge locations

	$x$	$y$	$h$	
<b>2005 SLAM challenge</b>				
<b>black point 1</b>	+1.300	+1.200		m
OptiTrack reference	+1.296	+1.217	+0.460	m
prediction with $g$ -vector from IMU	+1.296	+1.157	+0.538	m
<b>black point 2</b>	+2.700	0.000		m
OptiTrack reference	+2.791	+0.003	+0.472	m
prediction with $g$ -vector from IMU	<i>+2.596</i>	<i>+0.549</i>	<i>+0.606</i>	m
prediction with visual $g$ -vector	+2.706	+0.027	+0.597	m
<b>black point 3</b>	+2.200	-1.500		m
OptiTrack reference	+2.207	-1.505	+0.474	m
prediction with $g$ -vector from IMU	<i>+2.560</i>	<i>-1.199</i>	<i>+0.577</i>	m
prediction with visual $g$ -vector	+2.230	-1.449	+0.502	m
<b>black point 4</b>	-1.600	-1.200		m
OptiTrack reference	-1.513	-1.298	+0.454	m
prediction with $g$ -vector from IMU	-1.654	-1.298	+0.499	m
<b>black point 5</b>	-2.100	+0.900		m
OptiTrack reference	-2.070	+0.907	+0.425	m
prediction with $g$ -vector from IMU	<i>-1.439</i>	<i>+1.302</i>	<i>+0.301</i>	m
prediction with visual $g$ -vector	-2.058	+0.919	+0.449	m

## 6 Discussion

Finding the viewpoint from a limited number of ‘known’ points depends strongly on the accuracy of the ‘known’ part. The Inertial Measurement Unit (IMU) should in principal give an accurate estimate of the  $g$ -vector. Yet, this  $g$ -vector is measured inside the torso of the robot, which requires a coordinate transformation from the torso to the camera in the head. Kneip [13] assumes that this transformation is fixed and accurately established with a calibration procedure [17], which is unfortunately not possible on a moving robot. When the  $g$ -vector was estimated from the perspective lines in the image, instead of the IMU measurement, the results actually improved. This indicates that our measurement strongly depends on a good estimate of the  $g$ -vector.

A better estimate could also be accomplished by combining the constraints from multiple known points, as demonstrated by several solutions available in the OpenGV library [14]. In this paper only the X-intersections of middle line and circle were used. Yet, the Standard Platform field has many other useful points, such as the L-corners of the penalty area or the T-intersections at the edge of the field. All these type of intersections can be recognized with the YOLO v8 detector. Yet, the field of the view of the Nao robot is limited, so even in most cases not more than two known points are visible. Including the

goal-posts as ‘known’ points would increase the number of candidates, but it remains a challenge to have at least three or four of those points visible. When playing a soccer match with the occlusion of the opponents nearby, the situation will be even worse. Having an algorithm that can give a reasonable estimate on just two ‘known’ points remains valuable.

## 7 Conclusion

In this paper the claim that the use of rational trigonometry can reduce the time to calculate a triangulation is reproduced in a robotics application. This is beneficial for a robotics platform like the Nao robot, where computation time is scarce.

In addition, this paper demonstrates that for a known orientation relative to the  $g$ -vector, observing two ‘known’ points in the environment is enough to get an estimate on the pose of the camera. The accuracy depends on an accurate measurement of the  $g$ -vector by either an Inertial Measurement Unit or the vanishing point from two vertical lines. Estimates that are completely off can be filtered out by setting a boundary of the height estimate; the height of the robot is known, measurements above that height are only possible when the robot is lifted from the ground.

The estimates of the relative position to two ‘known’ points can be the input of a full-scale localization or a landmark based simultaneous localization and mapping algorithm, which keeps track of those observations over time, combined a model of the movement of the robot.

**Acknowledgement.** We like to thank Madelon Bernardy for her support in collecting a dataset, including the reference measurements with the OptiTrack system. We like to thank Gijs de Jong for extending the YOLO v8 XL detector with annotated training data for the X- T- and L-intersections.

## References

1. Alkendi, Y., Seneviratne, L., Zweiri, Y.: State of the art in vision-based localization techniques for autonomous navigation systems. *IEEE Access* **9**, 76847–76874 (2021)
2. Bestmann, M., et al.: Torso-21 dataset: typical objects in RoboCup soccer 2021. bit-bots’ github repository (April 2021)
3. Enderle, S., Ritter, M., Fox, D., Sablatnög, S., Kraetzschmar, G., Palm, G.: Vision-based localization in RoboCup environments. In: *RoboCup 2000: Robot Soccer World Cup IV*, pp. 291–296. Springer, Berlin, Heidelberg (January 2001)
4. Farazi, H., Allgeuer, P., Behnke, S.: A monocular vision system for playing soccer in low color information environments. In: *Proceedings of the 10th Workshop on Humanoid Soccer Robots, International Conference on Humanoid Robots* (November 2015)
5. Finsterwalder, S., Scheufele, W.: Das rückwärtseinschneiden im raum. *Sitzungsberichte der Mathematisch-Physikalischen Klasse der K.B. Akademie der Wissenschaften zu München*, pp. 591–614 (November 1903). band XXXIII

6. Furtado, J.S., Liu, H.H.T., Lai, G., Lacheray, H., Desouza-Coelho, J.: Comparative analysis of optitrack motion capture systems. In: *Advances in Motion Sensing and Control for Robotic Applications*, pp. 15–31. Springer (June 2019)
7. Gao, X.S., Hou, X.R., Tang, J., Cheng, H.F.: Complete solution classification for the perspective-three-point problem. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(8), 930–943 (2003)
8. Grunert, J.A.: Das pothenot'sche problem, in erweiterter gestalt; nebst bemerkungen über seine anwendungen in der geodäsie. *Grunerts Archiv für Mathematik und Physik*, pp. 238–248 (1841)
9. Haralick, R.M., Lee, C.N., Ottenburg, K., Nölle, M.: Analysis and solutions of the three point perspective pose estimation problem. In: *CVPR*, vol. 91, pp. 592–598 (June 1991)
10. Hidde, G.J., Lekanne gezegd Deprez, H., Visser, A.: The Dutch Nao Team synthetic dataset. Universiteit van Amsterdam (June 2020). <https://doi.org/10.21942/uva.12570902.v1>
11. Iocchi, L., Nardi, D.: Self-localization in the RoboCup environment. In: *RoboCup-99: Robot Soccer World Cup III*, pp. 318–330. Springer (June 2000)
12. Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I., Osawa, E.: RoboCup: the robot world cup initiative. In: *Proceedings of the First International Conference on Autonomous Agents*, pp. 340–347 (February 1997)
13. Kneip, L., Chli, M., Siegwart, R.: Robust real-time visual odometry with a single camera and an IMU. In: *Proceedings of the British Machine Vision Conference 2011. British Machine Vision Association* (September 2011)
14. Kneip, L., Furgale, P.: Opengv: a unified and generalized approach to real-time calibrated geometric vision. In: *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–8 (September 2014)
15. Kneip, L., Scaramuzza, D., Siegwart, R.: A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. In: *CVPR 2011*, pp. 2969–2976 (August 2011)
16. Laue, T., de Haas, T.J., Burchardt, A., Graf, C., Röfer, T., Härtl, A., Rieskamp, A.: Efficient and reliable sensor models for humanoid soccer robot self-localization. In: *Proceedings of the 4th Workshop on Humanoid Soccer Robots*, pp. 22–29 (December 2009)
17. Lobo, J., Dias, J.: Relative pose calibration between visual and inertial sensors. *Int. J. Robot. Res.* **26**(6), 561–575 (2007)
18. Owusu, P.P.D.: Recognizing all opponents: a real-time object detection approach for robotic soccer jerseys. Bachelor thesis, Universiteit van Amsterdam (February 2024)
19. RoboCup Technical Committee: Technical challenges for the RoboCup 2004 legged league competition (June 2004)
20. RoboCup Technical Committee: Technical challenges for the RoboCup 2005 legged league competition (June 2005)
21. RoboCup Technical Committee: RoboCup Standard Platform league (NAO) rule book (March 2024)
22. Wang, C.Y., Yeh, I.H., Liao, H.Y.M.: Yolov9: learning what you want to learn using programmable gradient information. preprint arXiv 2402.13616 (February 2024)
23. Wildberger, N.J.: *Divine Proportions: Rational Trigonometry to Universal Geometry*. Wild Egg Pty Ltd., Sydney, Australia (August 2005)