10.2352/EI.2022.34.10.IPAS-419

Accuracy Evaluation of Methods for Pose Estimation from Fiducial Markers

Ugurcan Budak, Olli Suominen, Atanas Gotchev; Tampere University; Tampere, Finland
Emilio Ruiz Morales; Fusion for Energy; Barcelona, Spain

Abstract

Estimating the pose from fiducial markers is a widely researched topic with practical importance for computer vision, robotics and photogrammetry. In this paper, we aim at quantifying the accuracy of pose estimation in real-world scenarios. More specifically, we investigate six different factors, which impact the accuracy of pose estimation, namely: number of points, depth offset, planar offset, manufacturing error, detection error, and constellation size. Their influence is quantified for four non-iterative pose estimation algorithms, employing direct linear transform, direct least squares, robust perspective-n-point, and infinitesimal planar pose estimation, respectively. We present empirical results which are instructive for selecting a well-performing pose estimation method and rectifying the factors causing errors and degrading the rotational and translational accuracy of pose estimation.

Introduction

Pose estimation is an extensively studied topic in computer vision, robotics and photogrammetry and there is notable number of findings in the literature. The problem of estimating an object pose with respect to a reference pose from 3D- and 2D-point correspondences is referred to as the Perspective-n-Point (PnP) problem. The used set of points can be redundant ($n \geq 5$) or non-redundant ($n = 4$) [8]. Both redundant and non-redundant point sets have been studied and various solutions have been presented. A general conclusion is that in the presence of noise, the higher the number of points is, the higher the accuracy [1].

A particular case of interest is when the corresponding points are representation of markers in real-world use cases [2], particularly circular fiducial markers in this study. To be detected by an image sensor, the markers, which are simple physical objects, are manufactured and placed in a 3D world scene. The number of markers and the area filled by markers; the manufacturing, placement and detection processes affect the accuracy of pose estimation. Thus far, in the literature, the overall detection error has been the most frequently considered error factor. Our motivation in this study is the need of quantification of the various factors which form the detection error alongside their impact on the accuracy. This should bring an informative insight into the selection of an appropriate method for pose estimation depending on the particular use case and how to optimize the factors.

Previous works have focused on the impact of the configuration of points on the accuracy [1] and [3], or have studied either the effect of the number of point sets [4], or the impact of noise levels on accuracy [5].

Pose estimation algorithms are generally categorized into two groups: iterative and non-iterative. Whereas iterative methods consider the PnP problem as a non-linear least squares problem; non-iterative methods convert it into a large system of equations [1]. The two groups have their own drawbacks; while iterative approaches have the instability of the cost function due to local minima and great cost of computation; non-iterative approaches are unstable in the presence of noise [8]. The pose estimation methods used in this study fall into the non-iterative group and are as follows: Direct Linear Transformation (DLT) [6], Direct Least-squares (DLS) [7], Robust Perspective-n-point (RPnP) [8] and Infinitesimal Planar Pose Estimation (IPPE) [9].

DLT has been considered one of the most accurate methods for pose estimation. DLT is based on homography estimation between model and image plane and minimizing the reprojection error utilizing the Levenberg-Marquardt algorithm. A variant, referred to as Normalized DLT aims at improving the condition number in the system of equations [6].

DLS has been proposed with the aim to find all possible solutions and select a solution which minimizes the cost function based on nonlinear least-squares. DLS defines the rotation matrix through so-called Cayley parameterization [7].

Unlike other non-iterative solutions, RPnP has been demonstrated to be stable for the case of non-redundant points, i.e., $n = 4$. It is considerably less time-consuming, having computational complexity of $O(n)$, while working for 3D, planar and quasi-singular cases [8].

IPPE is a recently proposed method, which utilizes the observation that true transformation between the world scene and image is better at certain areas on the scene than at other areas, when the homography is estimated through noisy point correspondences. The method is based on determining a point on the world scene in which the transformation is best estimated, next solving the pose with a local non-redundant first order partial differential equation [9].

Experimental Methodology

The first step in the pipeline is to detect the circular fiducial markers represented in Figure 1. A set of markers, which are basically points, in the 3D world scene and their corresponding 2D points in the image are used for pose estimation algorithms to estimate the pose of the camera. The point set is projected onto the image plane and it gets imaged into discrete pixels. After projection, but before the point coordinates are readily available for pose estimation, the projected points are exposed to various image processing operations. Since the markers are circular, they are detected as blobs in the image, and due to perspective projection, the circles are formed as ellipses in the image. Thus, operations, such as thresholding, blob detection and ellipse fitting are required and can result in an error while detecting the projected points, which forms the factor called detection error in our model.

In addition to detection of the markers, the number of markers and how much space they fill in the scene are important factors to investigate, since the application region can have constraints on size. These factors are called number of markers and constellation size. To clarify, constellation is a term used to describe the set of markers and constellation size refers how much space markers invade in the scene. Another factor results from the positions of the world points in the scene in these experiments. As the world points are actually represented by physical items in any real-world use case, i.e., markers, the manufacturing method utilized to build them affects their positions. Those stochastic situations are inevitable, since all manufacturing methods have a manufacturing tolerance. For instance, we assume our markers are manufactured by a CNC machine, which is known to be a precise manufacturing method.

Markers are settled on a surface and besides the relative errors inside the constellation, the related world points are likely to have a planar misalignment on the plane as an entity. An imprecise measurement process can cause planar misalignment, referred to as planar offset in this study, which subsequently causes erroneous placement of the markers when they are mounted on the surface. The erroneous placement of the plane where markers are settled is called depth offset in this study. Both planar offset and depth offset are modelled as bias.

In this study, we investigate the impact of six different factors on the accuracy of pose estimation by quantifying four non-iterative methods for pose estimation from fiducial markers. Of these six factors, two are constant elements, which are referred to as *planar offset* and *depth offset*; the two are modelled as additive Gaussian noise, which are called *manufacturing error* and *detection error*; and the other two are the properties of the markers, *number of points* and *constellation size*.

Figure 2 depicts the end-to-end pose estimation and errors affecting the process. In the beginning, the designed constellation model is materialized into a physical object, i.e., the markers. Although the illustration of the error caused by discretization of the image is individual in Figure 2, discretization is encompassed by detection error in our model, as discretization is tightly coupled with detection error and it is difficult to separate them.

In Figure 3, error factors in question are visualised. Whereas depth offset Δz occurs only along one direction; planar offset, Δx and Δy, occur along two directions. Thus, total displacement vector, $\Delta\mathbf{t}$, as observed in Figure 3 is defined as

$$\Delta\mathbf{t} = \begin{bmatrix}\Delta x & \Delta y & \Delta z\end{bmatrix}^\mathsf{T}. \tag{1}$$

The image points obtained by the projection of the world points are calculated as follows

$$\begin{bmatrix}u'\\ v'\\ Z_w\end{bmatrix} = \begin{bmatrix}f_x & 0 & p_x\\ 0 & f_y & p_y\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\begin{bmatrix}X\\ Y\\ Z\\ 1\end{bmatrix}, \tag{2}$$

$$\begin{bmatrix}u\\ v\end{bmatrix} = \begin{bmatrix}u'\\ v'\end{bmatrix}\frac{1}{Z_w}. \tag{3}$$

In Equation 2, $\mathbf{R}$ is the orthogonal rotation matrix and $\mathbf{t}$ is the translation vector. While f_x,f_y represents the focal length, p_x,p_y represents the principal point of the camera, and that whole matrix including these components is called intrinsic matrix. Moreover, $\begin{bmatrix}u & v\end{bmatrix}^\mathsf{T}$ in Equation 3 and $\begin{bmatrix}X & Y & Z & 1\end{bmatrix}^\mathsf{T}$ in Equation 2 are image points and world points, respectively. Since the world points are co-planar, Z in Equation 2 is zero. Taking perspective projection of the world points into consideration, the image points $\begin{bmatrix}u & v\end{bmatrix}^\mathsf{T}$ in Equation 3 are in PnP terminology the 2D correspondences.

The components of the total displacement $\Delta\mathbf{t}$ from Equation 1 are added to X, Y and Z in Equation 2 respectively as constant values.

Stochastic processes with Gaussian distribution, $\mathcal{N}(\mu,\sigma^2)$ are assumed for modelling various imperfections

$$\mathcal{N}(\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \tag{4}$$

In our simulations, we assume zero mean $\mu = 0$, thus the noise component is essentially defined by the variance, σ^2.

Manufacturing error is added to $\begin{bmatrix}X & Y & Z & 1\end{bmatrix}^\mathsf{T}$ in Equation 2 as additive Gaussian noise, which is defined as

$$\Delta m = \mathcal{N}(0,\sigma_m^2). \tag{5}$$

After adding the manufacturing error, Equation 2 takes the form

$$\begin{bmatrix}u'_m\\ v'_m\\ Z_w\end{bmatrix} = \begin{bmatrix}f_x & 0 & c_x\\ 0 & f_y & c_y\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}\mathbf{R} & \mathbf{t}\end{bmatrix}\begin{bmatrix}X+\Delta m_x\\ Y+\Delta m_y\\ Z\\ 1\end{bmatrix}. \tag{6}$$

Likewise, detection error is also modelled as additive Gaussian noise as

$$\Delta p = \mathcal{N}(0,\sigma_p^2). \tag{7}$$

The resulting noisy image points become

$$\begin{bmatrix}u_n\\ v_n\end{bmatrix} = \begin{bmatrix}u'_m\\ v'_m\end{bmatrix}\frac{1}{Z_w} + \begin{bmatrix}\Delta p_x\\ \Delta p_y\end{bmatrix}. \tag{8}$$

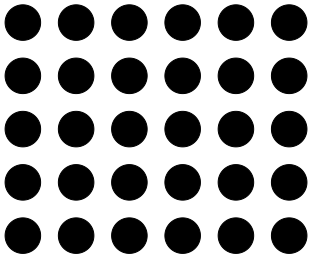

Figure 1: Circular pattern fiducial markers illustrated.

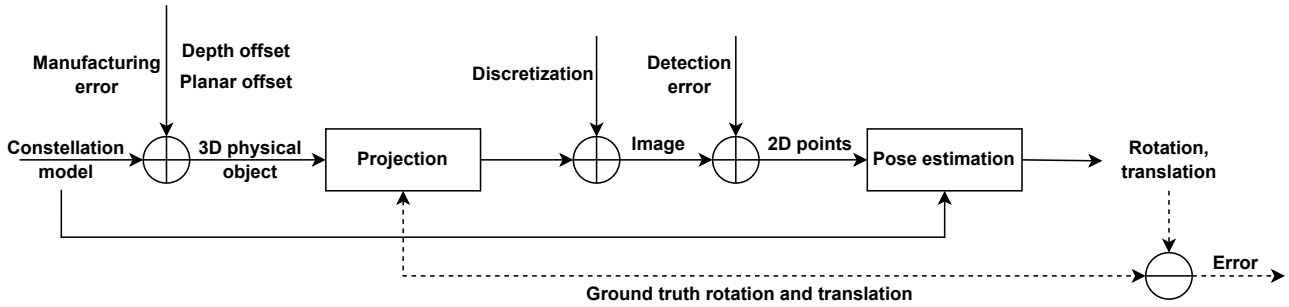

Figure 2: Modelling of the factors and their impact on the pose estimation.

Experimental Results

Experiments are conducted in MATLAB by simulating a scenario with parameters derived from a real-world use case [2]. The camera is set to a translation range of 300 to 500 millimetres through z-axis, and range of -100 to 100 millimetres through x- and y-axis. Camera resolution is 1920×1200, the focal length is set to 1090 to correspond to a 6 millimetres C-mount lens, and the principal point of the camera is assumed to be ideal, which makes it half of the camera resolution.

For each experiment, the algorithms are executed 500 times, and average errors of rotation and translation and their standard deviations are calculated. The absolute translational error is computed as the Euclidean distance between the ground truth and estimated translation from the algorithms. For the rotational error, rotation matrices are converted into axis-angle rotation form and the error is computed as the difference between estimated and ground truth rotation.

Four pose estimation algorithms explained in the Introduction are tested: DLT [6], DLS [7], RPnP [8] and IPPE [9]. The point set is formed by 8 points, of which both x- and y- coordinates are randomly selected within the range of $[-250\ 250] \times [-250\ 250]\ mm^2$.

The experimental results reveal the relation between the pose estimation accuracy and the number of points, depth offset, planar offset, manufacturing error, detection error and constellation size. The results are plotted as lines with vertical error bars representing the standard deviations of the average rotational and translational errors.

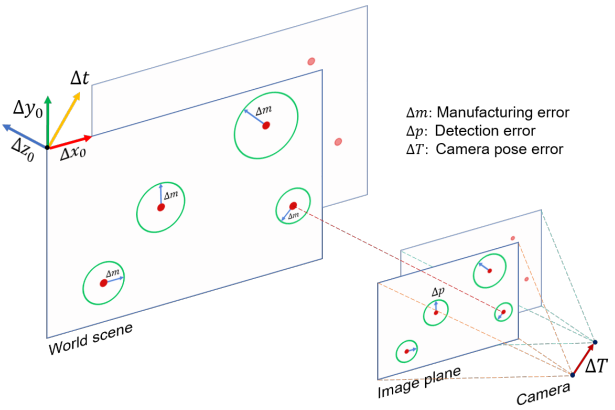

Figure 3: Manufacturing error, detection error and resulting camera pose error visualised.

Number of Points

The first factor quantified is the number of points, i.e., markers. As indicated in [1], escalating number of points raises the accuracy pose estimation, which is validated in Figure 4. For this experiment, the standard deviations of both manufacturing error and detection error are set to 0.2, which are feasible values for real-world use cases; and depth and planar offsets are set to 0.

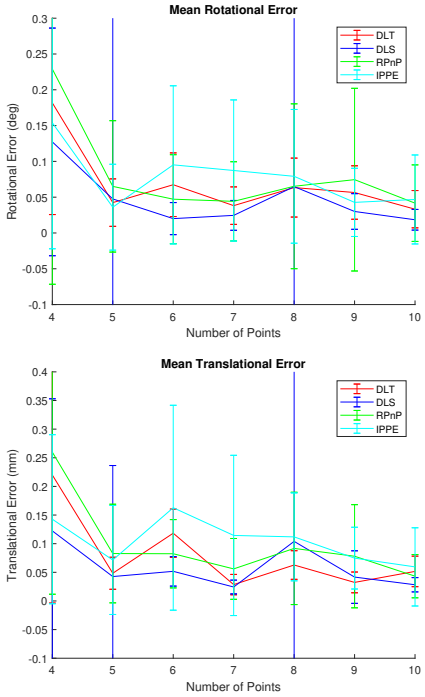

Figure 4: Mean rotational and translational error with respect to number of points.

The first outcome of Figure 4 is that increasing number of points brings about a decline in both rotational and translational error for all algorithms. Even though the algorithms appear to jump at certain number of points, in fact, general trend is following a gradual reduction in both error types. Unlike other algorithms, DLS depicts high standard deviations, especially at 5 and 8 points in the constellation. For all algorithms, particularly after 8 points in the world scene, accuracy is not notably affected by the increasing number of points.

Depth Offset

One error factor discussed above is the depth offset, which is the dislocation of the whole point set through the z-axis. Thus, depth offset becomes a bias added to point set in the world scene.

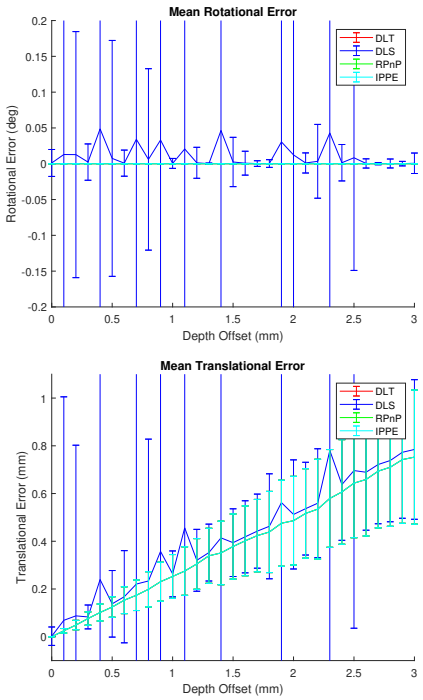

Figure 5: Mean rotational and translational error with respect to depth offset.

Translational error of all algorithms in Figure 5, excluding DLS, appears to noticeably and indistinguishably enlarge with respect to growing depth offset. The behaviour of DLT, RPnP and IPPE are utterly similar or same at certain values; therefore, they look like a single turquoise line. DLS follows a linear trend regardless of the high standard deviation at certain depth offset values. The linear escalation results from the depth offset being a constant value added to world points. At the same time, the rotational error against increasing depth is practically zero.

Planar Offset

As discussed above, the planar offset characterizes the misalignment of the whole point set in the world scene on the $x-y$ plane. Figure 6 reveals the response of each algorithm to increasing planar offset values.

The rotational error is not affected by the misalignment in question. Similarly to Figure 5, DLT, RPnP and IPPE yield practically the same results for the translational error, which is expectantly linear, since planar offset values are only constant values added to world points. Excluding DLS, as in the depth offset experiment, all algorithms identically respond to increasing planar offset for translation error. The tendency is similar to the case of depth offset. To illustrate, displacing the plane, in which the point set is placed, by 1.5 mm leads to approximately 2 mm translational error for all algorithms, as shown in Figure 6.

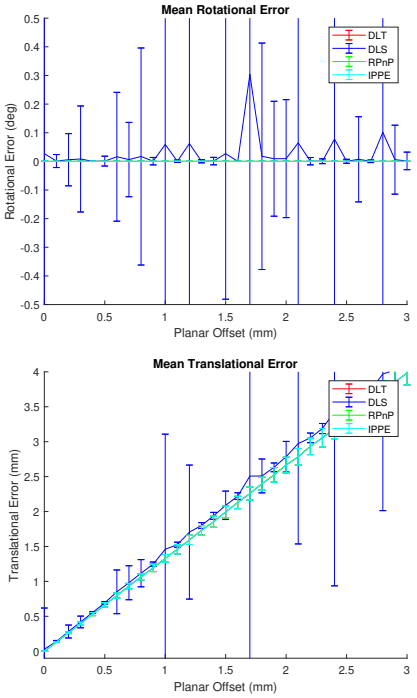

Figure 6: Mean rotational and translational error with respect to planar offset.

Manufacturing Error

Manufacturing error accounts for inevitable defects during the manufacturing process of the markers and is modelled as zero-mean additive Gaussian noise, which impacts the pose estimation accuracy.

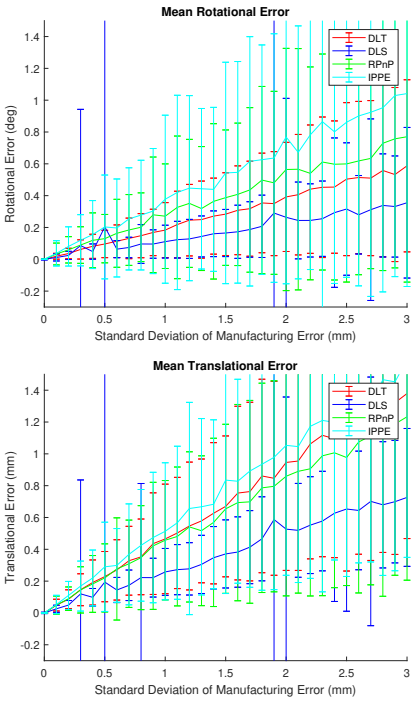

Figure 7: Mean rotational and translational error with respect to manufacturing error.

As shown in Figure 7, rising manufacturing error causes higher rotational and translational errors . The most accurate performance in terms of mean error is depicted by DLS, albeit high standard deviations. Additionally, all methods provide consistency with respect to rising manufacturing error values.

Detection Error

Detection error is frequently considered in the literature for pose estimation applications and is the type of error formed after the points are imaged by a sensor.

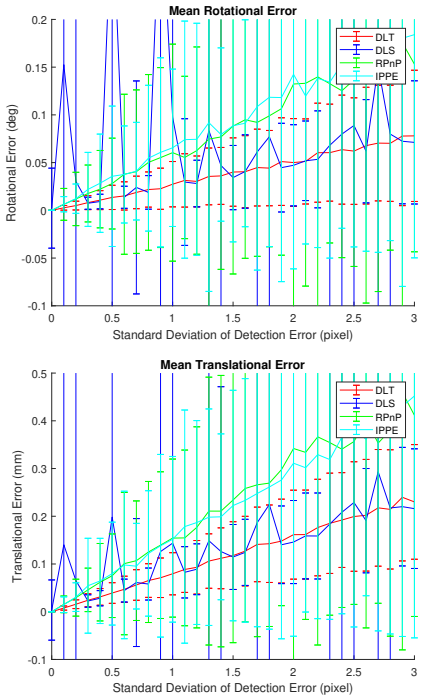

Figure 8: Mean rotational and translational error with respect to detection error.

Figure 8 depicts the impact of the detection error on rotational and translational errors. The trends of translational error graphs in Figure 7 and Figure 8 are similar, due to the way how the influencing factors are similarly modelled by additive Gaussian noise (manufacturing error expressed in millimetres, detection error expressed in pixels). In contrast, the rotational error behaviour is different. the detection error has relatively lower impact on the rotational accuracy compared to that of manufacturing error. DLT demonstrates the best performance in both rotational and translational accuracy with respect to the detection error.

Constellation Size

The final factor evaluated in this study is the area occupied by the constellation. In this experiment, the effect of the proportion of the area occupied by the constellation on the rotational and translational accuracy is investigated.

For this experiment, maximum area (100%) is set to $1000 \times 1000\ mm^2$; the standard deviations of detection error and manufacturing error are both set to 0.3 pixels and mm, respectively; the planar and depth offset are both set to 0. Figure 9 illustrates the accuracy of algorithms with respect to the frame size the constellation fills. We configure the markers to spread in $500x500\ mm^2$ area, which corresponds to 50% of constellation size in Figure 9. With 8 markers, and 0.3-standard deviation manufacturing and detection error, less than 0.05-degree error in rotation is acquired. In a real scenario, owing to lower standard deviation values, better accuracy can be achieved. Figure 9 demonstrates that raising constellation size causes a slight decrease in rotational error and slight increase in translational error.

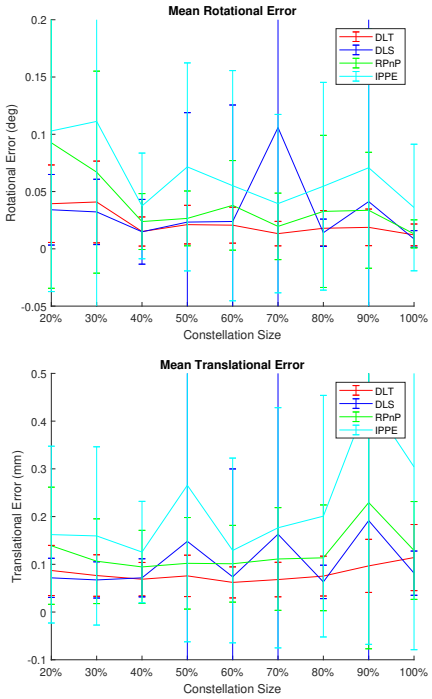

Figure 9: Mean rotational and translational error with respect to constellation size.

Conclusion

We have presented an empirical study of pose estimation accuracy of a group of methods with respect to a set of factors. We have specifically considered the case of fiducial markers placed on a plane. If the markers are manufactured by high-tolerance methods, DLT or RPnP are the best to work with. DLT is also the most advantageous option if a low-resolution camera is involved. If the settlement process of the markers to scene is not precisely performed, i.e. causing bias, DLT provides the highest accuracy.

DLS has demonstrated high accuracy in several experiments, however, it is also the method with highest number of outliers. The reason behind these outliers is believed to be the degeneracy caused by the Cayley parameterization. Cayley parameterization is known to degenerate when the rotation around x-, y- or z-axis is 180 degrees and the accuracy of pose estimation worsens if it gets closer to these singularities [11–13]. DLS is furthermore reported to be unstable when the points are co-planar [14].

Acknowledgments

This article reflects the views of the authors. F4E and Tampere University cannot be held responsible for any use which may be made of the information contained herein. The work presented in this paper was funded by Fusion for Energy (F4E) and Tampere University under the F4E grant contract F4E-GRT-0901.

References

[1] Raul Acuna, Volker Willert, Insights into the robustness of control point configurations for homography and planar pose estimation (2019).

[2] Laura Goncalves Ribeiro, Olli J. Suominen, Ahmed Durmush, Sari Peltonen, Emilio Ruiz Morales, Atanas Gotchev, Retro-Reflective-Marker-Aided Target Pose Estimation in a Safety-Critical Environment, Applied Sciences, 11, 1 (2021).

[3] Vincent Lepetit, Francesc Moreno-Noguer, Pascal Fua, EPnP: An Accurate O(n) Solution to the PnP Problem, International Journal of Computer Vision, 81 (2009).

[4] Adnan Ansar, Kostas Daniilidis, Linear pose estimation from points or lines, IEEE Transactions on Pattern Analysis and Machine Intelligence, 25, 5 (2003).

[5] Denis Oberkampf, Daniel DeMenthon, Larry Davis, Iterative pose estimation using coplanar points, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pg. 626-627. (1993).

[6] Richard Hartley, Andrew Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, Cambridge, 2004.

[7] Joel A. Hesch, Stergios I. Roumeliotis, A Direct Least-Squares (DLS) method for PnP, 2011 International Conference on Computer Vision, pg. 383-390. (2011).

[8] Shiqi Li, Chi Xu, Ming Xie, A Robust O(n) Solution to the Perspective-n-Point Problem, IEEE Transactions on Pattern Analysis and Machine Intelligence, 34, 7 (2012).

[9] Toby Collins, Adrien Bartoli, Infinitesimal Plane-Based Pose Estimation, International Journal of Computer Vision, 109 (2014).

[10] Peter Sturm, Algorithms for plane-based pose estimation, Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pg. 706-711. (2000).

[11] Yinqiang Zheng, Yubin Kuang, Shigeki Sugimoto, Kalle Åström, Masatoshi Okutomi, Revisiting the PnP Problem: A Fast, General and Optimal Solution, 2013 IEEE International Conference on Computer Vision, pg. 2344-2351. (2013).

[12] Lipu Zhou, Michael Kaess, An Efficient and Accurate Algorithm for the Perspective-n-Point Problem, 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pg. 6245-6252. (2019).

[13] Alexander Vakhitov, Luis Ferraz Colomina, Antonio Agudo, Francesc Moreno-Noguer, Uncertainty-Aware Camera Pose Estimation from Points and Lines, 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pg. 4657-4666. (2021).

[14] Gaku Nakano, Globally Optimal DLS Method for PnP Problem with Cayley parameterization, Proceedings of the British Machine Vision Conference (BMVC), pg. 78.1-78.11. (2015).

Author Biography

Ugurcan Budak received his B.Sc. in mechanical engineering from Istanbul Technical University (2018) and his M.Sc. in automation engineering with major of robotics from Tampere University (2021). He has worked in 3D Media Research Group in Faculty of Information Technology and Communications at the Tampere University. His research interest includes pose estimation, camera calibration and machine vision applications.

Olli J. Suominen received the B.Sc. and M.Sc.(Tech.) degrees in information technology, with a major in signal processing, from the Tampere University of Technology (TUT), in 2011 and 2012, respectively, where he is currently pursuing the Ph.D. degree in the 3D Media Group. His research interests are in applying visual technologies to improve situational awareness and spatial perception of operators in industrial contexts such as heavy mobile work machines and remote maintenance robotics. He is leading several industry driven research projects in these areas and developing relations with the industry.

Atanas Gotchev received the M.Sc. degrees in radio and television engineering (1990) and applied mathematics (1992), the Ph.D. degree in telecommunications from the Technical University of Sofia (1996), and the D.Sc.(Tech.) degree in information technologies from the Tampere University of Technology (2003). He is currently Professor of Signal Processing with the Tampere University. His recent work concentrates on developing methods for multi-sensor 3D scene capture, reconstruction, and display.

Emilio Ruiz Morales received the M.Sc. degree in electromechanical engineering and telecommunications from the École Polytechnique, Université Libre de Bruxelles, in 1990. He is currently the Project Manager for remote handling control systems of the ITER Project at the EU Fusion for Energy Agency. He has dedicated his career to the design and development of robotics control systems and advanced robotics applications in the fields of remote handling, nuclear, and surgical robotics.