Enhancing the accuracy of aerial imagery-based object position estimation

Zsombor Páncsics[1,2][0009-0002-9170-6908], Csaba Marosi[2][0009-0002-1309-0845], Zsombor Széplaki[2], Zoltán Berényi[2], Gergely Treplán[2][0009-0003-2123-5503], Szende Sándor[2], Levente Hajder[1][0000-0001-9716-9176], and Imre Benedek Juhász[2][0000-0002-6881-4604]

[1] Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary
[2] Robert Bosch Kft. Budapest, Hungary
Zsombor.Pancsics@hu.bosch.com

Abstract. This paper introduces compensation and robustification methods to enhance the accuracy of aerial imagery-based object position estimation, serving as an enabler technology for higher level aerial surveillance tasks such as real-time urban monitoring and traffic management in smart cities. We address perspective distortion through mathematical derivations and a dedicated simulator-based compensation approach. Furthermore, to mitigate camera instabilities, like tilt and shaking, and to correct coordinate system inaccuracies, we introduce and evaluate multi reference marker-based setups. To quantify the accuracy enhancement of these proposed methods, real-world measurement scenarios were designed and evaluated under controlled conditions by incorporating a reference sensor system. Our results demonstrate that the simulator-based approach reduces perspective distortion error by up to 15 cm in distant areas. Moreover, our case study highlights the critical sensitivity of simpler single reference marker-based setup, where even a slight deviation from a direct downward view can cause errors up to 90 cm in coordinate system. In contrast, our multi reference marker-based approach robustly handles this common error source, maintaining a root mean square error (RMSE) of less than 2 cm.

Keywords: aerial imagery · autonomous driving · bird's-eye view · drone · error analysis · error compensation · perspective distortion · reference

1 Introduction

The rise of smart cities has driven a growing demand for in-depth traffic analysis, making aerial imagery an important tool for efficient urban management and real-time monitoring. It has gained increasing attention over traditional methods, hence a single downward-looking camera can provide a cost-effective solution with a broad view of traffic. It has an advantage over single cameras installed on vehicles/walls by avoiding occlusion issues; furthermore, its simplicity is a considerable benefit compared to multi-sensor systems, which would introduce calibration, synchronization, and data fusion challenges.

Aerial imagery based object pose estimation requires video processing, object detection, tracking, and spatiotemporal transformation steps. The main challenges, rooting from the measurement constellation are perspective distortion (during the object detection) and coordinate system inaccuracies (during the spatial transformation), often leading to errors in resulting metric positions. Although prior research has addressed some of these challenges [10, 11, 17, 20], a comprehensive study that systematically integrates and evaluates the effectiveness of compensation methods is still lacking.

This paper aims to fill that gap by introducing and validating a set of compensation and robustification methods through structured real-world scenarios. We present a simulator-based compensation approach for perspective distortion and introduce multi reference marker-based setups to mitigate camera and coordinate system inaccuracies. The accuracy enhancement of our proposed strategies is quantified by comparing the results against a high-precision reference system.

1.1 Contributions

- By improving the fundamental accuracy of object localization, our work serves as an enabler technology for higher level aerial surveillance tasks.
- We introduce and compare mathematical and simulation-based approaches for the compensation of perspective distortion.
- We demonstrate that using a multi reference marker-based setup substantially improves the scaling and orientation accuracy of the coordinate system, by relying on the precise localization of board centers over a large baseline, instead of detecting small pattern features on a single board.
- We present how a four marker-based setup, leveraging homography, enhances system robustness under dynamic drone movements and camera tilts.
- Through a series of real-world experiments, we quantify the impact of key error sources and validate the effectiveness of our compensation methods.

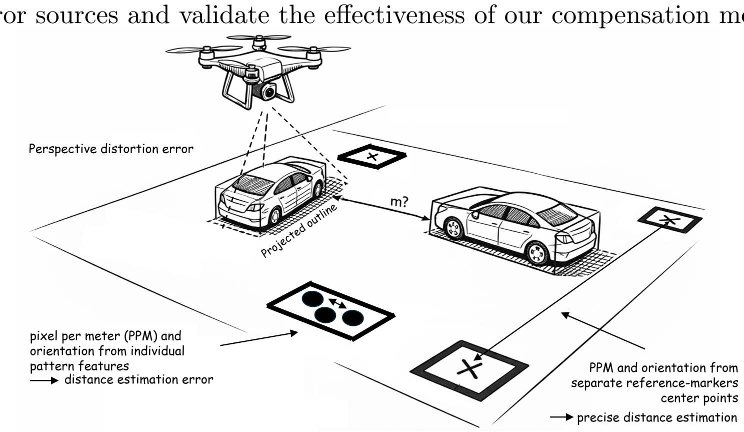

Fig. 1. Our core contributions: (1) compensation of perspective distortion with a mathematical model and a dedicated simulator; and (2) applying robust multi reference marker setups for camera instability mitigation and coordinate system stabilization.

2 Related Work

Aerial imagery has become an increasingly outstanding data source in traffic analysis [2, 5, 8, 9, 12, 13, 15, 16, 18, 21] due to its versatile applicability and cost-effectiveness, motivating a range of studies focused on analyzing and improving its accuracy and reliability. While early work primarily focused on object detection performance, recent researches placed growing attention on mitigating errors that arise from the aerial measurement setup. Our previous work [17] serves as a solid reference for systematically collecting the most critical error factors impacting aerial imagery-based position estimation. These include 3D projection error, reference marker misalignment, coordinate system and camera inaccuracies. Among these, perspective distortion is one of the most complex and critical, affecting all aerial imagery-based measurements, especially in areas farther from the nadir of the camera. Classical computer vision literature has already addressed this error [4,7,19]. Lu et al. [11] proposed a calibration framework specifically designed to reduce systematic biases caused by camera pose uncertainty and distortion parameters, demonstrating that even small orientation errors can lead to notable localization inaccuracies on the ground plane. Simulator-based solutions also appeared [17], where the complexity of the model stems from the large number of unknown parameters and multi-stage error propagation.

Camera inaccuracies, encompassing issues like camera motion or shaking, altitude variations, and lens distortions, pose another critical challenge. Recent work has addressed these challenges, for example, lane-based spatial calibration has been proposed for image stabilization and mitigating the effects of camera motion [20]. Other studies have examined errors caused by topography elevation, landmark positional inaccuracies, and distortions induced by the camera lenses [1]. The impact of sensor position inaccuracy was also analyzed by Nuske et al. [14], while Dille et al. [6] investigated the inaccuracy of motion prediction.

While these contributions offer valuable theoretical insights for specific error sources, there remains a gap in studies that propose and evaluate practical compensation strategies for the most impactful error types in aerial imagery-based position estimation.

3 Methodology

In this chapter we introduce the error-mitigation approaches and the measurement constellations targeting the evaluation of their effectivity. For the compensation of perspective distortion we present two solutions: a mathematical model-based one and a simulator-based one. To enhance the accuracy of the coordinate system and mitigate errors due to camera instability, we introduce various multi reference marker-based solutions. Last but not least we show the designs of measurement setups and the metrics for the evaluation.

3.1 Compensation of perspective distortion

The most important and complex error factor, affecting all aerial measurement: **the perspective distortion** occurs during 3D-2D projection. It can be compensated using a mathematical model or simulator-based approaches.

Mathematical model-based approach is able to calculate and compensate the effect of perspective distortion assuming preset vehicle types and corresponding vehicle dimensions, the distance of the traffic participant location from the camera nadir and the camera altitude.

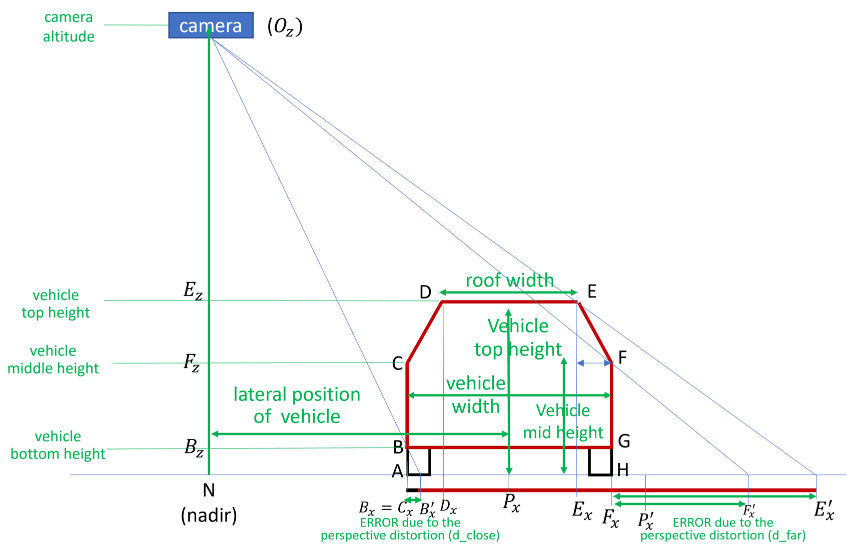

Fig. 2. Geometric model of perspective distortion

Figure 2 shows the mathematical model, derived by the following equations:

$$\overrightarrow{P_xP'_x} = \frac{\overrightarrow{\mathrm{d}_{\mathrm{close}}} + \overrightarrow{\mathrm{d}_{\mathrm{far}}}}{2}, \tag{1}$$

where $\overrightarrow{P_xP'_x}$ presents the center point deviation due to the perspective distort.

$$\overrightarrow{\mathrm{d}_{\mathrm{close}}} = \overrightarrow{B_xB'_x} = \frac{||\overrightarrow{NB_z}||(||\overrightarrow{NP_x}|| - \frac{||\overrightarrow{F_xC_x}||}{2})}{||\overrightarrow{O_zB_z}||}\frac{NP_x}{||NP_x||}, \tag{2}$$

where the calculation is based on triangle similarity, $||\overrightarrow{F_xC_x}||$ is the vehicle$_{\mathrm{width}}$.

$$\overrightarrow{\mathrm{d}_{\mathrm{far}}} = \max \begin{cases} \overrightarrow{\mathrm{d}_{\mathrm{top}}} \\ \overrightarrow{\mathrm{d}_{\mathrm{middle}}} \end{cases} . \tag{3}$$

$$\overrightarrow{\mathrm{d_{top}}} = \overrightarrow{F_x E_x'} = \left(\frac{\overrightarrow{||NE_z||}(\overrightarrow{||NP_x||} + \frac{\overrightarrow{||E_x D_x||}}{2})}{\overrightarrow{||OzE_z||}} - \frac{\overrightarrow{||F_x C_x||} - \overrightarrow{||E_x D_x||}}{2} \right) \frac{NP_x}{||NP_x||}, \tag{4}$$

where the calculation is based on triangle similarity, $\overrightarrow{||E_x D_x||}$ is the roof$_{\mathrm{width}}$.

$$\overrightarrow{\mathrm{d_{mid}}} = \overrightarrow{F_x F_x'} = \frac{\overrightarrow{||NF_z||}(\overrightarrow{||NP_x||} + \frac{\overrightarrow{||F_x C_x||}}{2})}{\overrightarrow{||O_z F_z||}} \frac{NP_x}{||NP_x||}. \tag{5}$$

Based on the mathematical derivation, it can be seen that within a certain distance (at a small camera angle of incidence), the distortion is smaller, as it is determined by the mid-height of the car. Meanwhile, at a larger angle of incidence, the steepness of the distortion trend increases, as it is caused by the car roof point of the car beyond this distance. Substituting real-world parameters and a camera altitude of 100 m, it can be calculated that the inflection point appears at a distance of approximately 8-10 meters, depending on the passenger vehicle dimensions.

Simulator-based approach utilizes and validates our simulator presented in our earlier work [17] for perspective distortion compensation for real measurement scenarios. The simulator is designed to address all previously mentioned error factors, including the estimation (and implicitly the compensation) of the inaccuracies caused by the perspective distortion. Its operation requires the input of the exact dimensions, position, and orientation of the traffic participant, as well as the drone camera position and distortion parameters. In contrast to the article [11], relying on neural network-based compensation, our simulator uses physical models, granting a broader applicability, beyond aerial measurements to any other measurement task conducted with a perspective camera.

3.2 Enhancement of coordinate system accuracy

The enhancement of the accuracy of the coordinate system and the mitigation of the effect of camera instabilities is essential for two primary reasons: first, for the precise determination of the position and orientation of the relative coordinate system (and, if the reference marker absolute position is known, the absolute coordinate system); second, for accurate calculation of the pixel density or pixel per meter (PPM), which directly influences the precision of all distances within the resulting metric system.

When using a single reference marker, a primary limitation arises: its size will be relatively small from the high altitude. For instance, even an A1-sized reference board, when captured at an altitude of 120 meters with a 4K camera, may only occupy approximately 40x30 pixels in the image. Consequently, individual pattern features would occupy just 4x4 pixels, which extremely challenges their accurate localization, both in terms of position and orientation, resulting in notable errors in PPM and orientation estimation of the coordinate-system. To improve robustness against detection inaccuracies and camera instabilities, we propose a series of multi reference marker configurations. This approach takes advantage of the precise localization capability of reference marker centers, which can be determined with an accuracy of ± 0.5 cm based on our initial measurements. While four-board setup is the optimal (applying homography), practical constraints often prevent deploying all boards on a horizontal plane at the measurement location. Therefore, two and three-board setups were also analyzed.

Two reference markers approach offers two key advantages. Firstly, it allows for the robust localization of the reference boards centers, rather than relying on the less precise detection of individual, small pattern features of a single reference board. Secondly, by substantially increasing the baseline over which the PPM is calculated (from 23 cm to 100 meter), the impact of any remaining localization error is greatly reduced. This approach leads to a much more stable and accurate PPM, as the same detection error is distributed over a larger distance. Ideally the reference boards should be positioned close to the edges of the observed area, while avoiding image edges due to distortions.

Three reference markers allow the computation of affine transformation that maps image-plane coordinates $(\mathbf{u}, \mathbf{v})$ to metric ground plane coordinates $(\mathbf{X}, \mathbf{Y})$ through scaling, rotation, and translation, under the assumption that the optical axis of the camera is perpendicular to the ground plane. Having the detected image positions of three reference boards and their corresponding metric ground coordinates, the parameters of the affine transformation matrix $\mathbf{A}$ can be determined. In our implementation, the affine parameters are estimated using *estimateAffinePartial2D* from OpenCV package [3]. The mapping is expressed in homogeneous form as:

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \\ 1 \end{pmatrix} \sim \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ 1 \end{pmatrix} . \tag{6}$$

Four reference markers enable the application of homography transformation. It is important in practice for addressing camera tilt, a common practical challenge, as drone cameras are rarely perfectly perpendicular to the ground. Assuming that the ground is flat, homography $\mathbf{H}$ transformation [7] maps the image points to the ground plane. In our implementation, the $\mathbf{H}$ matrix parameters are estimated using *findHomography* from OpenCV package [3]. Mathematically, point $(\mathbf{u}, \mathbf{v})$ is projected onto the ground plane $(\mathbf{X}, \mathbf{Y}, 1)$ as:

$$\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \\ 1 \end{pmatrix} \sim \mathbf{H} \begin{pmatrix} \mathbf{u} \\ \mathbf{v} \\ 1 \end{pmatrix} . \tag{7}$$

This method provides the most comprehensive geometric correction among the multi reference marker-based approaches, offering substantial robustness against variations in camera pose.

3.3 Measurement design and analysis method

In order to understand the impact of error factors and the effectiveness of their compensation approaches, real-world measurement constellations were designed.

Analysis of perspective distortion error compensation To quantify the effectiveness of our proposed compensation strategies for perspective distortion, we designed a real-world experiment. The evaluation compares the position of a vehicle estimated from aerial imagery against high-precision ground truth data, with and without the error correction applied. The primary metric for this comparison was the root mean square error (RMSE) calculated from the Euclidean distances between the estimated and actual vehicle center points.

$$\mathrm{RMSE}(y, \hat{y}) = \sqrt{\frac{\sum_{i=0}^{N-1} (y_i - \hat{y}_i)^2}{N}}, \tag{8}$$

where N is the number of data points, y_i represents the i-th reference measurement position, and $\hat{y}_i$ is the i-th estimated position.

The experiment was conducted under controlled conditions on a certified flat proving ground. A vehicle traversed a predefined trajectory (visualized on Fig. 4) at varying angular orientations and different radial distances from the nadir. For aerial recording, a DJI Mavic 3 Classic 2.0 drone hovered stationary at 30m altitude, using a single ground-placed reference marker to determine the coordinate system. A single marker was sufficient, as the low flight altitude makes the noise from pattern-feature mislocalization negligible. The footage was captured in 4K at 60 FPS (H.264), with automatic shutter speed and ISO settings.

Our evaluation process involved three main steps: data extraction, error compensation, and comparative analysis. For ground truth, an RTK-GNSS reference sensor system, installed in the host vehicle, provided precise positional data. It was extracted and processed using Foxglove software and stored in an MCAP modular container file. Concurrently, the position of the vehicle was also estimated via aerial imagery (according to [17]). Error compensation was then applied using both the mathematical model and the simulator-based approaches. Unlike the static mathematical model, the simulator provides a dynamic physical model of the measurement scenario, configured with the parameters of the real-world setup, including camera altitude, lens distortion, and the 3D dimensions of the vehicle. The simulator, utilizing the positions, orientations, and dimensions of the traffic participant, generated a frame-by-frame estimation of the perspective distortion error vector in a pandas DataFrame which was then used to correct the raw aerial estimates. Finally, the performance of the position estimation, both with and without error compensation, was compared against the RTK-GNSS ground truth. This comparative analysis was performed within a Python framework, utilizing pandas for data manipulation, SciPy for function approximation, and Matplotlib for visualization.

The coordinate system accuracy enhancement and camera instability compensation can be evaluated error-component wise and on system level.

For **error-component-wise analysis**, we isolated and observed the stability of an a-priori known ground-distance, focusing specifically on the error component related to the coordinate system determination, excluding the other error components, likethe perspective distortion. Therefore a multi reference marker and a two target (ArUco) marker-based (flat) measurement setup was designed, observing the stability of the estimated distance of the stationary target (ArUco) markers, using 1, 2, 3, and 4 reference board-based coordinate-system determination. The four-board measurement setup is shown on figure 3. The measurements were conducted on a certified flat surface. The reference markers were positioned with a cross line laser, achieving an accuracy of 0.5 cm. To analyze the effect of error factor at different altitudes, and use cases (stationary and dynamic), measurement videos were recorded from heights of 10, 60, 80, and 100 meters while the drone were performing hovering, descending, ascending, and rotating maneuvers, with different camera tilts, where 90° represents the direct downward bird's-eye view.

For **system-level evaluation**, the aerial imagery-based vehicle position estimation was compared against the actual vehicle position, by calculating their Euclidean distance. For the evaluation, a test measurement concept was designed - a grid, spanning 0 to 80 meters in the x-direction and -5 to 5 meters in the y-direction. The measurement points were selected at approximately 5-meter intervals in close areas, with this spacing gradually increasing further away. The coordinates of these grid points were determined using a cross line laser, achieving an accuracy of 0.5 cm. A reference vehicle traversed this designated area, stopping at these predefined grid points to collect data. The drone videos were recorded from the same altitudes as for error level analysis, with both stationary and dynamic maneuvers being performed.

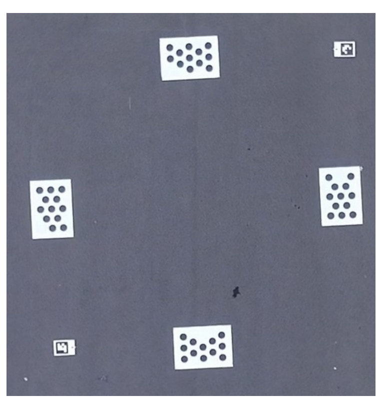

Fig. 3. Measurement design for error level analysis of coordinate system and camera instabilities mitigation with a four reference board-based setup.

4 Results

We structured our analysis into two main parts: first, we evaluate the methods for perspective distortion compensation; second, we analyze the effectiveness of different multi reference marker configurations for coordinate system accuracy enhancement and camera instability error mitigation, which is crucial for overall system precision, especially under dynamic conditions, where camera tilt and shaking is present.

4.1 Perspective distortion compensation

Figure 4 displays the actual (black) and the estimated trajectories based on aerial imagery with different error compensation methods. The measurement covered all four spatial quadrants of the scene and varying radial distances from the camera nadir which serves as the origin. The orange line shows the trajectory estimated with aerial imagery, without any compensation. The blue-orange dashed line includes compensation of perspective distortion via the mathematical model presented above, while the green-orange dashed line displays the simulator-based compensation.

The results clearly demonstrate the benefit of perspective distortion compensation, substantially improving the accuracy of aerial imagery-based position estimation. Based on figure 4 it can be stated that both the mathematical model and the simulator effectively compensate the error. It can also be observed that the error compensation of the simulator outperforms the mathematical model at greater distances from the camera nadir - where perspective distortion is most critical, while the mathematical model-based compensation leads to smaller errors closer to the camera nadir. The advantage of the simulator lies in its more comprehensive physical model, which not only utilizes a precise 3D vehicle geometry but also accounts for camera properties like lens distortion and tilt. The close-range capability of the simulator would certainly improve with an even more accurately modeled vehicle.

Table 1. RMSE of Euclidean error without and with error compensations for the whole dataset.

Compensation method	RMSE	Relative error reduction
Without error compensation	13.5 cm	
Mathematical model-based error compensation	9.5 cm	29,6 %
Simulator-based error compensation	10.0 cm	25,9 %

Table 2. RMSE of Euclidean error without and with error compensations at distances of more than 8 m from the nadir.

Compensation method	RMSE	Relative error reduction
Without error compensation	20.4 cm	
Mathematical model-based error compensation	11.3 cm	44,6 %
Simulator-based error compensation	8.3 cm	59,3 %

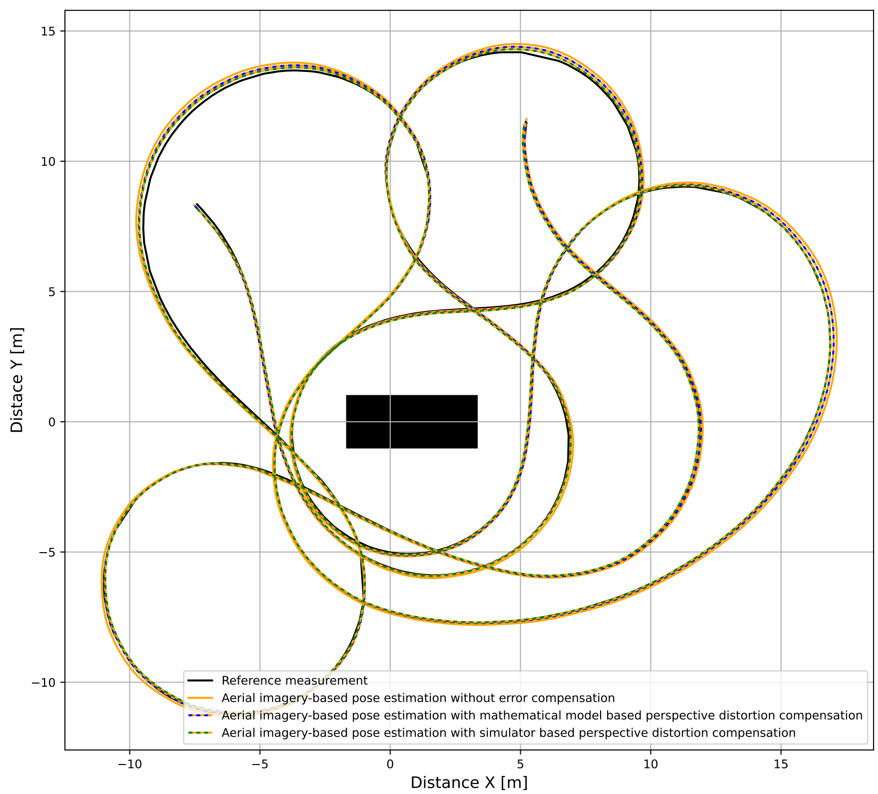

Fig. 4. The real (black) and the estimated trajectory of the traffic participant with mathematical model-based (marked with dashed blue), with simulator-based (marked with dashed green), and without (orange) error compensation.

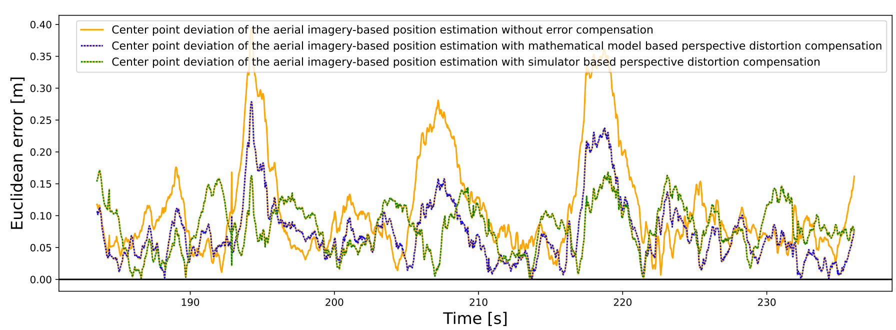

Fig. 5. The position estimation error as a function of time with mathematical model-based (marked with dashed blue), with simulator-based (marked with dashed green), and without (marked with orange) error compensation.

The temporal tendency of the Euclidean error can be seen in figure 5. It is evident that the simulator-based solution performs exceptionally well at remote areas, reducing the error by 15-20 cm. In figure 5 the orange curve illustrates the center point deviation of aerial imagery-based position estimation without error correction. The blue-orange dashed line shows the center point deviation of the same aerial imagery-based position estimation, compensated for perspective distortion using the mathematical model. The green-orange dashed line displays the trend when compensating for perspective distortion based on the simulator.

4.2 Coordinate system accuracy enhancement and camera tilt error mitigation

By error-component-wise evaluation the stability of the a-priori known target-marker distance were analyzed, calculated based on the different (1, 2, 3, and 4 reference marker-based) approaches. As the altitude was increased, we replaced the target-markers from ArUco markers to larger A0 boards for better detectability. Figure 6a illustrates the results when the drone was operated in a quasi-stationary state, hovering at various altitudes with a downward-facing camera. Figure 6b presents dynamic scenarios where the drone was performing horizontal movements, rotation, and in some cases additional camera tilts were applied.

Figure 6a reveals that with a quasi-stationary drone position, at an altitude of 10 meters, the error of individual reference pattern features mislocalization is negligible, as evidenced by the similarly noisy results across all board configurations. As the altitude increases, the error of individual reference pattern features mislocalization becomes critical and clear trend emerges: the stability of the distance-estimation improves with an increasing number of boards. The single board setup is the noisiest. The two-board setup outperformed it by relying on board center point detection instead of individual reference pattern features detection and by distributing the mislocalization error over a larger distance, thereby mitigating its noise. The four-board setup is the most stable one, as coordinate system transformations and camera tilts can be handled more effectively through homographic transformation. This trend can be consistently observed at different altitudes.

The true power of homography becomes evident in figure 6b, where the drone was operating under dynamic conditions, performing both translational (x-y) and rotational movements. Notably, in the last two scenarios the camera was intentionally tilted to 85° and 80° respectively, representing a 5° and a 10° deviation from the direct downward-looking bird's-eye view perspective. In these challenging scenarios, the importance of homography in handling non-parallel planes is clearly important. While 2 and 3 reference marker-based solutions introduce ~ 0.3 and ~ 0.6 meter RMSE respectively, the 4 board-based solution remains stable even in this scenario and has less than 0.02 meter RMSE.

By system-level evaluation we investigated how the different reference marker setups influence the resulting system error measured by the Euclidean distance of estimated and ground truth vehicle position. Figure 7 shows the Euclidean error progression for vehicle in near (marked with solid line), mid-range (marked with dashed line), and distant (marked with dotted line) positions from camera nadir, showcasing the performance of 1 (marked with blue line) and 2 (marked with orange line) board-based configurations from different altitudes. First, from figure 7, it can be concluded that the two-board setup outperforms the single board solution on system level as well, similarly to the error-component wise tests. Furthermore, it is also observable that the final system level distance prediction are smoother (less noisy) compared to the error-component-wise evaluation. This noise reduction is a direct result of applying a 99-frame smoothing window.

Fig. 6. The error-component-wise evaluation of 1 (marked with blue), 2 (marked with orange), 3 (marked with green), and 4 (marked with red) reference marker-based coordinate system accuracy enhancement and camera instability error mitigation in (a) stationary and in (b) dynamic use cases, from different camera positions and tilts, by ground truth target-marker distance (purple) comparison.

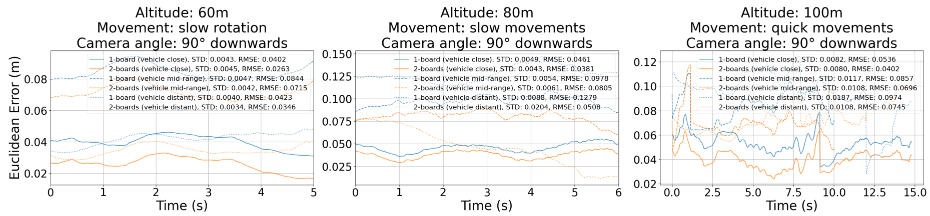

Fig. 7. The system-level evaluation of 1 (marked with blue) and 2 (marked with orange) reference marker-based setup, from different camera altitudes, with diverse vehicle positions (near - marked with solid line, mid-range - marked with dashed line, and distant - marked with dotted line), by Euclidean distance comparison of estimated and ground truth vehicle positions.

5 Conclusion

This paper provides a validated framework of methods that enhance the accuracy of aerial imagery-based object position estimation, acting as an enabler technology for advanced aerial surveillance tasks. First, we addressed perspective distortion with both mathematical model and simulator-based approaches. Our evaluation demonstrates that the simulator is capable of reducing this source of error by up to 15 cm (by 60%) at distant areas, offering valuable, quantitative insights for practitioners.

Second, we introduced and analyzed multi reference marker setups to mitigate the coordinate system inaccuracies (metric and orientation) and camera instabilities (tilt and shaking). Through a series of real-world measurement scenarios, we evaluated 1, 2, 3, and 4 board setups and found that the four-board setup is optimal, providing robustness against quick, dynamic drone movements, camera tilts, and rotations. This superiority stems from leveraging homography and from its ability to use a robust, board center localization-based approach, instead of relying on the less precise detection of individual pattern features of a single board. This was clearly demonstrated in our case study: while a minor 10° camera tilt (an 80° viewing angle) resulted in substantial inaccuracies in coordinate system determination, an error of ~ 90 and ~ 60 cm RMSE in the 1 and 2-board-based setups respectively, the 4-board-based solution remained robust with less than 2 cm RMSE. The four-board based setup performed the best in every scenario, compensating the error effectively in both stationary and dynamic drone operations, including scenarios with unavoidable camera tilts - a common problem in practice.

However, due to practical constraints, a planar four-board setup may not always be feasible at the measurement location. In such cases, height differences must be compensated for. If this is not possible, according to our results we recommend a minimum of a two-board configuration to maintain robustness against the individual pattern features detection inaccuracies inherent in a single board approach.

Future Work In the future, we would like to measure and test how perspective distortion error factor and error compensation perform in case of larger vehicles (such as trucks and vans) - where the problem is even more amplified. Furthermore we will focus on extending the framework to non-planar environments.

Acknowledgements Project no. C2286690 has been implemented with the support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation fund, financed under the KDP-2023. Furthermore, we would like to express our gratitude to our company: Robert Bosch Kft., Hungary, for supporting our project.

References

1. Babinec, A., Apeltauer, J.: On accuracy of position estimation from aerial imagery captured by low-flying uavs. International Journal of Transportation Science and Technology **5**(3), 152–166 (2016). https://doi.org/10.1016/j.ijtst.2017.02.002
2. Bock, J., Krajewski, R., Moers, T., Runde, S., Vater, L., Eckstein, L.: The ind dataset: A drone dataset of naturalistic road user trajectories at german intersections. In: 2020 IEEE Intelligent Vehicles Symposium (IV). pp. 1929–1934 (2020). https://doi.org/10.1109/IV47402.2020.9304839
3. Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000)
4. Collins, R.: Projective reconstruction of approximately planar scenes. Proceedings of SPIE - The International Society for Optical Engineering **1838**, 174–185 (1992). https://doi.org/10.1117/12.142793
5. Dhaouadi, O., Meier, J., Wahl, L., Kaiser, J., Scalerandi, L., Wandelburg, N., Zhou, Z., Berinpanathan, N., Banzhaf, H., Cremers, D.: Highly accurate and diverse traffic data: The deepscenario open 3d dataset. In: 2025 IEEE Intelligent Vehicles Symposium (IV). pp. 377–384 (2025). https://doi.org/10.1109/IV64158.2025.11097484
6. Dille, M., Grocholsky, B., Nuske, S.: Persistent visual tracking and accurate geolocation of moving ground targets by small air vehicles. In: Infotech@Aerospace 2011. American Institute of Aeronautics and Astronautics, Reston, Virigina (Mar 2011)
7. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, New York, NY, USA, 2 edn. (2003)
8. Krajewski, R., Bock, J., Kloeker, L., Eckstein, L.: The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp. 2118–2125 (2018). https://doi.org/10.1109/ITSC.2018.8569552
9. Krajewski, R., Moers, T., Bock, J., Vater, L., Eckstein, L.: The round dataset: A drone dataset of road user trajectories at roundabouts in germany. In: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). pp. 1–6 (2020). https://doi.org/10.1109/ITSC45102.2020.9294728
10. Kruber, F., Morales, E.S., Chakraborty, S., Botsch, M.: Vehicle position estimation with aerial imagery from unmanned aerial vehicles. In: 2020 IEEE Intelligent Vehicles Symposium (IV). pp. 2089–2096 (2020). https://doi.org/10.1109/IV47402.2020.9304794

11. Lu, L., Dai, F.: Accurate road user localization in aerial images captured by unmanned aerial vehicles. Automation in Construction **158**, 105257 (2024). https://doi.org//10.1016/j.autcon.2023.105257
12. Meier, J., Scalerandi, L., Dhaouadi, O., Kaiser, J., Araslanov, N., Cremers, D.: Carla drone: Monocular 3d object detection from a different perspective. In: Pattern Recognition. pp. 137–152. Springer Nature Switzerland (2025)
13. Nieuwenhuisen, M., Droeschel, D., Beul, M., Behnke, S.: Autonomous navigation for micro aerial vehicles in complex gnss-denied environments. Journal of Intelligent & Robotic Systems **84**(1-4), 199–216 (Dec 2016). https://doi.org/10.1007/s10846-015-0274-3
14. Nuske, S.T., Dille, M., Grocholsky, B., Singh, S.: Representing substantial heading uncertainty for accurate geolocation by small uavs. In: AIAA Guidance, Navigation, and Control Conference. American Institute of Aeronautics and Astronautics, Reston, Virigina (Aug 2010). https://doi.org/10.2514/6.2010-7886
15. Organisciak, D., Poyser, M., Alsehaim, A., Hu, S., Isaac-Medina, B., Breckon, T., Shum, H.: Uav-reid: A benchmark on unmanned aerial vehicle re-identification in video imagery. In: Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. pp. 136–146. SCITEPRESS - Science and Technology Publications (2022). https://doi.org/10.5220/0010836600003124
16. Patoliya, J., Mewada, H., Hassaballah, M., Khan, M.A., Kadry, S.: A robust autonomous navigation and mapping system based on gps and lidar data for unconstraint environment. Earth Science Informatics **15**(4), 2703–2715 (Dec 2022). https://doi.org/10.1007/s12145-022-00791-x
17. Páncsics, Z., Nyisztor, N., Tóth, T., Juhász, I., Treplán, G., Hajder, L.: Error analysis of aerial image-based relative object position estimation. In: Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: VISAPP. pp. 656–663. INSTICC, SciTePress (2024). https://doi.org/10.5220/0012433300003660
18. Spannaus, P., Zechel, P., Lenz, K.: Automatum data: Drone-based highway dataset for the development and validation of automated driving software for research and commercial applications. In: 2021 IEEE Intelligent Vehicles Symposium (IV). pp. 1372–1377 (2021). https://doi.org/10.1109/IV48863.2021.9575442
19. Szeliski, R.: Computer vision algorithms and applications (2011)
20. Zhang, C., Tang, Z., Zhang, M., Wang, B., Hou, L.: Developing a more reliable aerial photography-based method for acquiring freeway traffic data. Remote Sensing **14**(9) (2022). https://doi.org/10.3390/rs14092202
21. Zhang, Y., Wang, C., Yu, R., Wang, L., Quan, W., Gao, Y., Li, P.: The ad4che dataset and its application in typical congestion scenarios of traffic jam pilot systems. IEEE Transactions on Intelligent Vehicles **8**(5), 3312–3323 (2023). https://doi.org/10.1109/TIV.2023.3260902