1st German Robotics Conference
March 13-15, 2025. Nuremberg, Germany

An Inland Port Monitoring System using Aerial and Ground Imagery

Emre Gülsoylu[*,1], André Kelm[1], Lennart Bengtson[2], Christian Wilms[1], Janick Edinger[2] and Simone Frintrop[1]

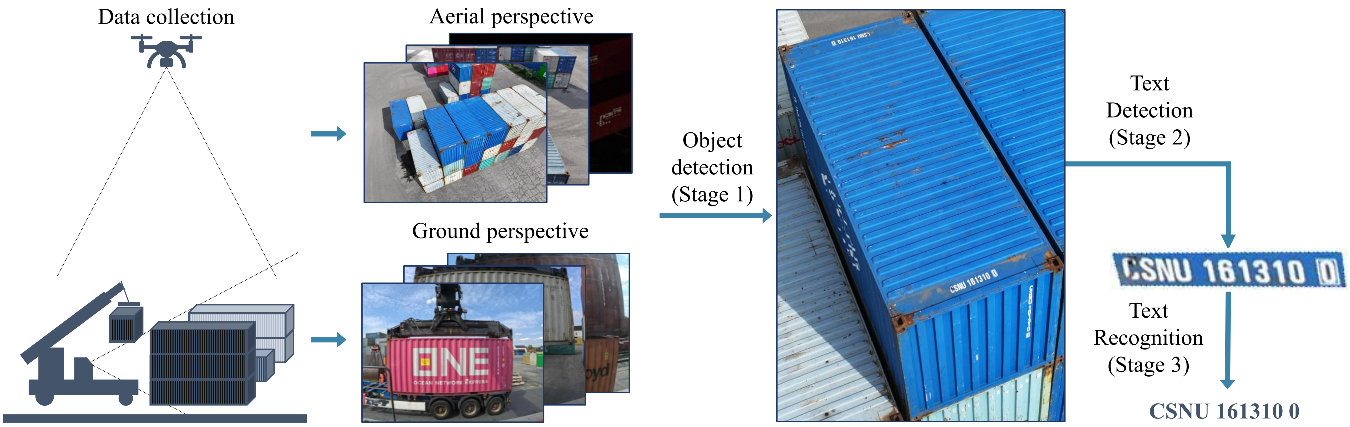

Fig. 1: This study presents a framework for inland ports that uses unmanned aerial vehicle (UAV)-based imaging, and reach stacker sensor data including a camera to monitor port operations. A three-stage identification pipeline detects transportation units (TUs) as the first stage, then markings of detected TUs such as ISO6436 compliant ID codes are detected as the second stage. Finally, a text recognition model extracts their IDs.

I. Introduction

Multi-purpose terminals in inland ports handle a variety of transportation units (TUs), such as containers and cranable semi-trailers, in variable volumes. This results in constantly evolving operational processes, changing storage configurations, manual effort for operations, and poor traceability. Increasing traceability for effective terminal monitoring requires identification of individual TUs, such as containers and semi-trailers. In accordance with ISO6346 all TUs feature standardised markings, which enable the unambiguous visual identification.

Current solutions attempt to solve TU identification tasks through various methods, including manual scanning, RFID tags [1], and fixed-camera networks [2]. However, these approaches either require additional hardware installation on each TU, which is impractical and costly, or depend on fixed infrastructure that lacks flexibility. This limited flexibility is not ideal for smaller ports that change their configuration based on seasonal contracts. Moreover, existing computer vision solutions primarily focus on two-stage TU identification approaches that combine text detection and text recognition [3]. These methods rely on the placement of TUs in front of a fixed-camera and can monitor only a limited area in terminals. Additionally, they prove insufficient especially for text detection in scenarios involving mobile cameras [4], in which objects appear in different scales and orientations due to the changes in perspective.

[1]University of Hamburg, Faculty of MIN, Department of Informatics, Computer Vision Group Hamburg, Germany [*]emre.guelsoylu@uni-hamburg.de

[2]University of Hamburg, Faculty of MIN, Department of Informatics, Distributed Operating Systems Group Hamburg, Germany

To overcome these limitations, we present a system (see Fig. 1) that uses images from unmanned aerial vehicles (UAVs), and from sensor-equipped reach stackers for creating a digital representation of Alberthafen, an inland port in Dresden. Its features represent not only the current state of the terminal but also enable analysis of the past and prediction of the future, providing valuable insights for process optimisation. Here, we will focus on the computer vision and robotics challenges of port monitoring by presenting two key innovations: First, we propose a three-stage TU identification pipeline comprising a stage for TU detection, followed by text field detection, and text recognition. Second, we implement a data collection strategy that uses aerial imagery from UAVs as well as sensor-equipped reach stackers, providing coverage of both stacked and single-layer TU arrangements.

II. System Overview

1) Three-stage TU Identification: The core of our system is a three-stage TU identification approach, consisting of lightweight models suitable for edge devices. We found that the integration of TU detection as a prior stage to text detection reduces the search space thus addresses the challenge of detecting small text instances. While conventional approaches circumvent this challenge by arranging their infrastructure accordingly [2], this is infeasible for cameras mounted UAVs or reach stackers [4].

The first stage employs the YOLOv5-S model [5] to detect TUs in images. The second stage uses DBNet++ [6] for text detection within the detected TU regions, addressing challenges such as varying orientations and environmental conditions including bad condition of TU markings. Finally, TrOCR [7] recognises text and generates ID information. To

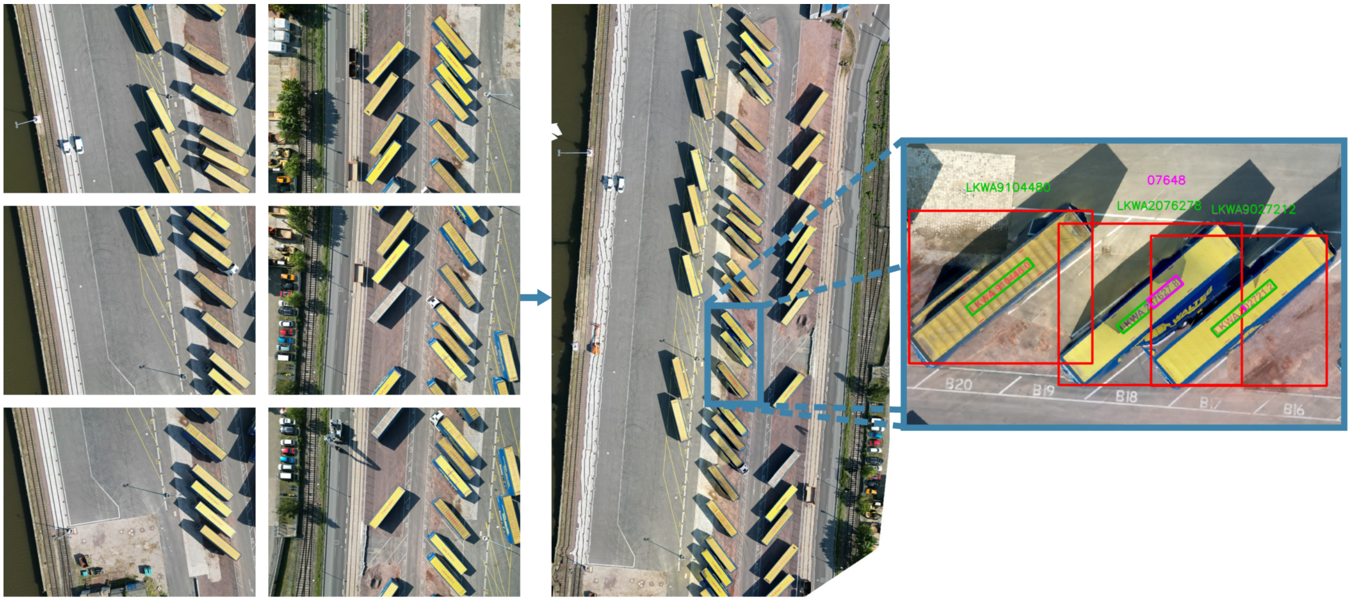

Fig. 2: Images captured during overview flights (left) are stitched to create an orthophoto (middle). Then the three stage TU identification is applied. Object detection model is able to detect the transportation units despite the artefacts created during orthophoto creation (right). Thanks to the georeferenced orthophoto, it is possible to extract locations of the detected TUs. After detection, markings are detected and recognised. Extracted text is assigned on the related TU if it is ISO6346 compliant (green text), if not (pink text), it is discarded.

support the development and evaluation of our system, we created the TRansportation Unit Detection and Identification (TRUDI) dataset, containing 25,941 labelled instances across five classes (container, trailer, tank-container, logo, ID text).

2) Aerial and Ground-level Data Collection: Our system employs UAVs such as DJI Mavic Pro 3 and DJI Mini 2 conducting systematic overview flights, capturing overlapping images to generate high-resolution orthophotos (geometrically corrected aerial photographs) [8] of terminals by using the pipeline of OpenDroneMap [9]. The creation of orthophotos facilitates the TU localisation by offering a georeferenced, large-scale representation of the terminal layout. These flights are scheduled after transloading operations, ensuring that the system maintains an accurate representation of the terminal's current state (see Fig. 2).

To address the limitations of orthophotos in identifying stacked containers, where only the top most container is visible (see Fig. 2), we developed a ground-level identification system. This system equips reach stackers with sensor boxes containing high-resolution cameras, real-time kinematic positioning (RTK) sensors, and distance sensors. The integration of distance sensors on the combined spreader enables detection of TU handling events, automatically triggering RTK measurements during pickup and placement operations. This approach ensures accurate tracking of container movements and stack configurations, complementing the aerial perspective with ground-level data.

Aerial images captured from bird's eye view or reach stacker's camera cover only one side of the TUs. Therefore, illegible IDs on the visible surface require checking other surfaces of the TU. We address this problem by conducting targeted missions, positioning the UAV iat angles such that other IDs become legible.

III. Preliminary Findings

Initial results show that the proposed three-stage approach for TU identification successfully tackles the problems with small text detection introduced by the mobile cameras. The text detection model benefits from detection of TUs under challenging conditions including small scale or occluded text, especially the aerial perspective.

All three stages were evaluated on the TRUDI dataset containing both UAV and ground level images. The object detection model achieved a mean Average Precision of 0.71 at a threshold of 50 % (mAP@50). The DBNet++ model performed balanced text detection with a recall of 0.87, precision of 0.84, and an F1-score of 0.85 at a threshold of 60 %. Finally, the TrOCR model achieved a low character error rate (CER) of 0.02. These results suggest robust capabilities in both detecting and recognizing text even from UAV view, compared to previous work [4] with 0.50 F1-score for detection and 0.16 CER for recognition.

IV. Future Directions and Conclusion

The integration of our three-stage identification approach marks a significant advancement in TU identification technology. Our system demonstrates robust performance across various operational conditions and provides a foundation for future developments in port automation and optimisation.

Future work will focus on enhancing the system's integration, while improving the performance of the individual models. Developing new methods for text detection and recognition will be prioritised to achieve more efficient TUs identification.

Acknowledgements

The project is supported by the Federal Ministry for Digital and Transport (BMDV) in the funding program Innovative Hafentechnologien II (IHATEC II).

References

[1] X. Shi, D. Tao, and S. Voß, "RFID technology and its application to port-based container logistics," *JOCEC*, vol. 21, no. 4, pp. 332–347, 2011.

[2] J. Zhao, N. Jia, X. Liu, G. Wang, and W. Zhao, "A practical unified network for localization and recognition of arbitrary-oriented container code and type," *IEEE Trans. on Instrumentation and Measurement*, 2024.

[3] M. Yu, S. Zhu, B. Lu, Q. Chen, and T. Wang, "A two-stage automatic container code recognition method considering environmental interference," *Applied Sciences*, vol. 14, no. 11, p. 4779, 2024.

[4] J. Teegen, A. Kelm, O. Grasse, M. Hillemann, E. Gülsoylu, and S. Frintrop, "Drone-based identification of containers and semi-trailers in inland ports," in *"12th Int'l Conf. on Logistics and Maritime Systems"*, 2024.

[5] G. J. et. al., "ultralytics/yolov5: v6.0 - YOLOv5n 'Nano' models, Roboflow integration, TensorFlow export, OpenCV DNN support," Oct. 2021. [Online]. Available: https://doi.org/10.5281/zenodo.5563715

[6] M. Liao, Z. Wan, C. Yao, K. Chen, and X. Bai, "Real-time scene text detection with differentiable binarization," in *Proc. of the AAAI Conf. on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 11 474–11 481.

[7] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, and F. Wei, "TrOCR: Transformer-based optical character recognition with pre-trained models," in *Proc. of the AAAI Conf. on Artificial Intelligence*, vol. 37, no. 11, 2023, pp. 13 094–13 102.

[8] G. Gao, M. Yuan, Z. Ma, J. Gu, W. Meng, S. Xu, and X. Zhang, "GeoROS: Georeferenced real-time orthophoto stitching with unmanned aerial vehicle," in *2022 IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 2250–2256.

[9] OpenDroneMap Authors, "ODM - a command line toolkit to generate maps, point clouds, 3D models and DEMs from drone, balloon or kite images." github.com/OpenDroneMap/ODM, 2020.