Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Emre Gülsoylu[1], Niklas Fiedler[2], Simone Frintrop[1]
[1] University of Hamburg, Faculty of MIN, Department of Informatics, Computer Vision Group
[2] University of Hamburg, Faculty of MIN, Department of Informatics, Technical Aspects of Multimodal Systems

German Robotics Conference

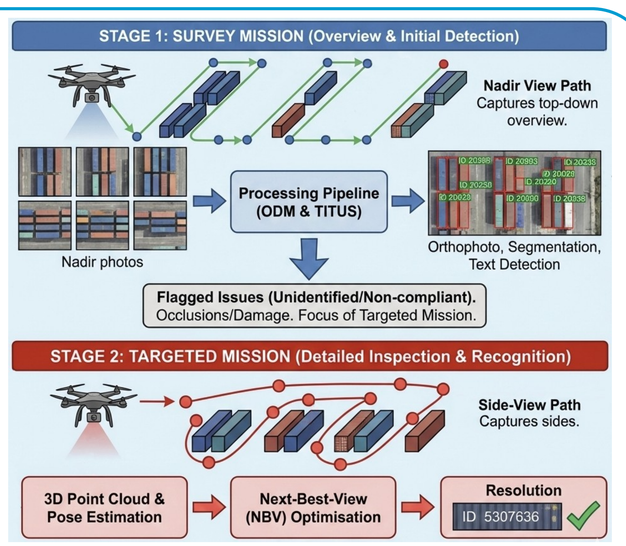
Next Best View for Text Detection and Recognition in Port Monitoring Unmanned Aerial Vehicles

Motivation

Task: Optimise intermodal loading unit (ILU) identification in port environments using unmanned aerial vehicles (UAVs). Specifically, the focus is on detecting and recognising textual identifiers on ILUs, which are occluded, degraded, or poorly visible from standard survey perspectives.

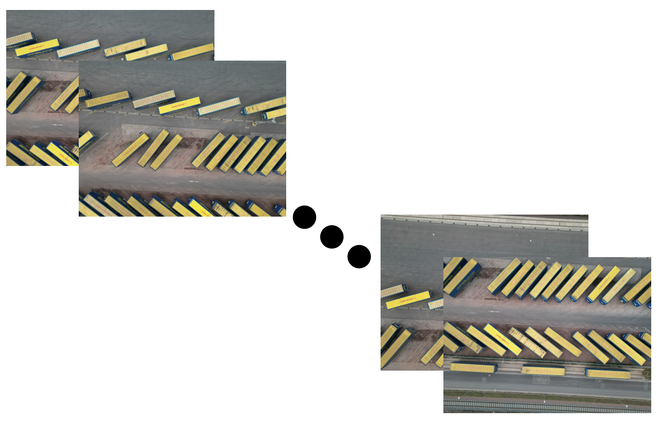
Problem:
- Existing methods rely on static viewpoints, which are insufficient for complex, dynamic port environments.
- Nadir camera perspectives often fail to capture legible ID codes due to stacking, damage or suboptimal lighting.
- Missed or incorrect text detection/recognition creates information gaps, leading to inefficiencies in port logistics.

Idea:
A two-stage UAV mission framework:
- Survey Mission: Capture nadir-view images, generate orthophotos, and identify ILUs with a three stage pipeline.
- Targeted Mission: Use Next-Best-View (NBV) optimisation to plan side viewpoints for unidentified ILUs.



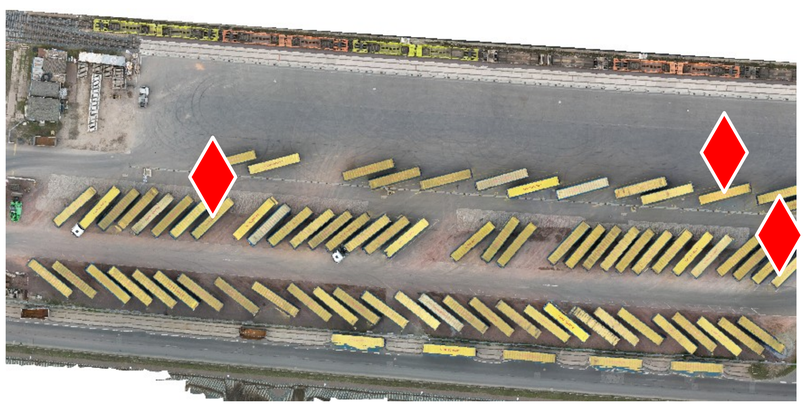
Survey Mission

Planning with DARP [1, 2]

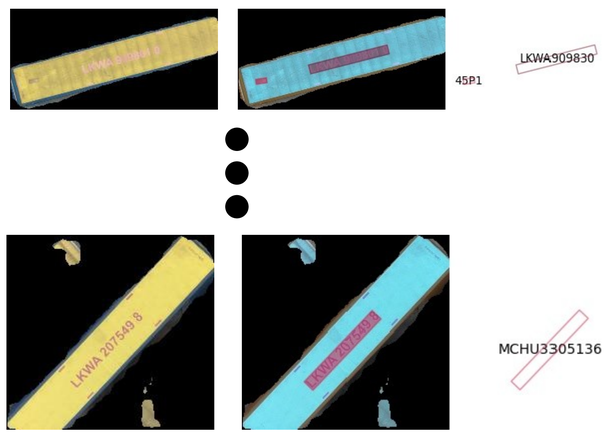


Executing missing for image collection

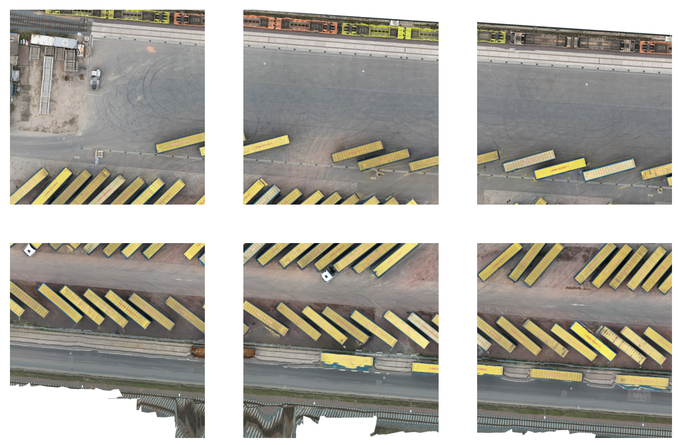
Orthophoto creation with ODM [3]



Defining target points based on unidentified ILUs

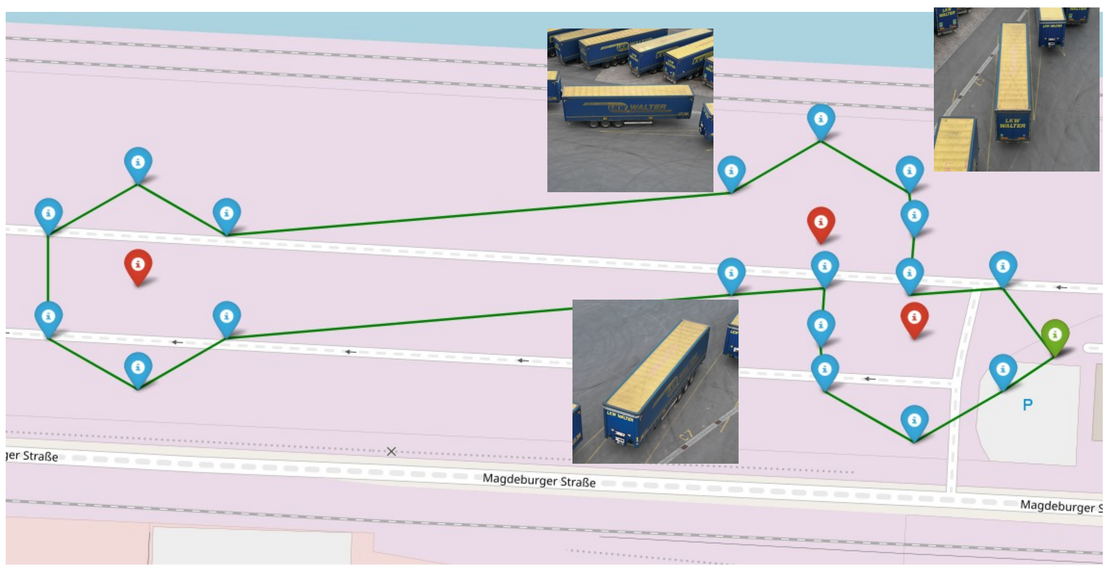


ILU identification with TITUS [4]



Tiling

Targeted Mission



Targeted mission template with Ant Colony Optimisation [5]

Legibility score (LS)

$$LS(w, f_i) = \underbrace{\left(\frac{1 - \hat{n}_i \cdot \hat{v}_w}{2}\right)}_{\text{Angle Term}} \cdot \underbrace{\left(\frac{1}{d(w, f_i) + 1}\right)}_{\text{Distance Term}},$$

Optimum waypoint selection

$$w^* = \arg\max_{w \in \mathcal{V}_f} \left(LS(w, T_i) \cdot V(w, f)\right)$$

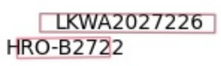
Pheremon update rule

$$\text{Cost}(w_i, w_j) = \alpha \cdot d(w_i, w_j) + \beta \cdot \frac{1}{LS(w_j, f)}$$

Instance Segmentation

Text Detection

Text Recognition



Acknowledgment: The project is supported by the German Federal Ministry for Digital and Transport in the funding program Innovative Hafentechnologien II.

inte great drones

Computer Vision Group

References

[1] Kapoutsis, et al. “DARP: Divide areas algorithm for optimal multi-robot coverage path planning,” Journal of Intelligent & Robotic Systems, vol. 86, no. 3, pp. 663–680, 2017.

[2] Apostolidis, et al., “Cooperative multi-UAV coverage mission planning platform for remote sensing applications,” Autonomous Robots, vol. 46, no. 2, pp. 373–400, 2022.

[3] OpenDroneMap Authors, “ODM - a command line toolkit to generate maps, point clouds, 3D models and DEMs from drone, balloon or kite images.”

[4] Gülsoylu, et al., “TRUDI and TITUS: A Multi-Perspective and A Three-Stage Recognition System for Transportation Unit Identification,” in In proceedings of the 36th British Machine Vision Conference (BMVC), 2025.

[5] M. Dorigo and T. St¨utzle, “Ant colony optimization: overview and recent advances,” Handbook of metaheuristics, pp. 311–351, 2018.