THESIS BOOKLET

Klaudia Éva Zeleny

Computer Vision-Based Framework for Product Identification in Dynamic Environments

Supervisors:

Dr. Andrea Kő
Professor, Head of Institute

Dr. Tamás Ruppert
Associate Professor

BUDAPEST, 2026

Institute of Data Analytics and Information Systems

Thesis Booklet

Klaudia Éva Zeleny

Computer Vision-Based Framework for
Product Identification in Dynamic Environments

for the Ph.D. dissertation

Supervisors:

Dr. Andrea Kő, Professor, Head of Institute

Dr. Tamás Ruppert, Associate Professor

Table of Contents

1. Research Background and Rationale

This dissertation investigates the role of visual perception in supporting product identification and the generation of inventory-related evidence in indoor warehouse environments, focusing on high-mix low-volume (HMLV) intralogistics systems. In such systems, dynamic product placement, non-uniform packaging, and frequent product turnover limit the applicability of conventional, structured identification workflows.

Warehouse Management Systems typically rely on explicit machine-readable identifiers, such as barcodes, QR codes, and Data Matrix codes, to link physical objects to digital records. However, the accuracy and timeliness of this information depend on manual scanning and controlled acquisition processes, and product location and product identification remain related but distinct operational problems: reaching a recorded storage location does not by itself guarantee that the correct item is present, and an imprecise or outdated location record can still require manual search.

HMLV warehouse environments are characterized by a large number of distinct product types, low repetition rates, non-fixed or temporary storage locations, and changing product and identifier orientations. These conditions complicate the use of automated visual-identification strategies that assume stable layouts, standardized appearances, or predefined object models, and they introduce substantial visual variability, including changing scale, viewpoint, illumination, occlusion, and motion blur.

Although HMLV products already carry item-level, machine-readable Data Matrix identifiers, existing workflows typically treat code scanning as a discrete, manually triggered event rather than as part of a continuous perception process. The related literature reveals a fragmentation between methodological advances in visual detection and decoding on one hand, and the operational requirements of dynamic HMLV intralogistics on the other; no prior work was identified that systematically combines deep learning-based detection, dedicated Data Matrix decoding, and similarity-based validation into a single pipeline evaluated under realistic HMLV-inspired acquisition variability.

The primary objective of the dissertation is therefore to develop and evaluate a computer vision-based product identification pipeline that enables reliable Data Matrix recognition and supports product localization to inform inventory-related decision support in HMLV warehouse environments. The investigation is guided by three research questions, addressing, respectively, the system-level architectural design that integrates prior knowledge (RQ1), the implementation and empirical validation of the proposed pipeline under controlled environmental variability (RQ2), and the operational interpretation of temporally validated identification outcomes as evidence for inventory-related decision support (RQ3).

2. Methods Applied

The methodological framework of the research is based on CRISP-DM (Cross-Industry Standard Process for Data Mining), which provides a structured, six-phase approach for data-driven projects. Although the research involves the construction of a novel computer vision-based identification pipeline and a purpose-built dataset, and shares certain characteristics with design science research, CRISP-DM was adopted as the primary methodological framework because the study is positioned mainly as an empirical, data-driven investigation. The six phases were retained, but reinterpreted for a research context focused on methodological rigor, reproducibility, and the generation of new knowledge rather than business deployment.

In the Business Understanding phase, the research boundary was defined as visual Data Matrix identification under variable mobile-camera acquisition, excluding warehouse-management deployment or inventory-state updating. The systematic literature review, research questions, and objectives established modularity, controlled variability, reference validation, and the separation of frame-level perception from episode-level identification as guiding design principles.

The Data Understanding and Data Preparation phases identified the acquisition factors to be represented in the dataset, namely illumination, viewpoint, blur, apparent symbol size, partial visibility (occlusion), the presence of multiple visible symbols, and negative frames, and converted the recordings into annotated frame and episode manifests. Polygon-based object annotations were defined, together with normal, too-small, and occluded object-condition categories and an independent blur flag. A macro-scene-level train–validation–test separation was applied so that the same physical code layout and identifier set could not appear across partitions.

In the Modeling phase, a modular rather than end-to-end system was implemented: a YOLOv8 instancesegmentation model for symbol detection, mask-based geometric normalization (cropping, warping, perspective correction), a dedicated Data Matrix decoder, reference validation against a predefined identifier list, failure-activated recovery based on cluster decoding and soft-Hamming similarity ranking, and frame- and episode-level evidence aggregation through a dominance-based temporal decision rule.

In the Experimental Design and Evaluation phase, the held-out test split, the ground-truth target, and the reporting units were fixed before the results were interpreted. Detection, decoding, frame-level identification, and episode-level identification were evaluated separately, complemented by component-level ablations (of the similarity measure and the accumulation rule) and condition-specific reporting across illumination, viewpoint, blur, object condition, and symbol-size categories, in order to localize where losses occur within the pipeline.

The Deployment phase was interpreted as research dissemination and conceptual operational interpretation rather than system rollout: the dissertation documents the dataset, pipeline, decision rules, results, and limitations, and discusses how accepted, ambiguous, and unresolved episode outcomes could serve as bounded evidence for later inventory-related decision support, without implementing a warehouse-management integration, inventory update, or production deployment.

3. New Scientific Results of the Dissertation

The main scientific results of the dissertation can be summarized in the following points.

3.1 A modular, uncertainty-aware identification architecture

The dissertation formulates an integrated identification architecture that explicitly separates frame-level perception from sequence-level (episode-level) identification. Detection, geometric normalization, dedicated Data Matrix decoding, reference validation, failure-activated recovery, and temporal evidence aggregation are combined within a single modular pipeline, in which frame-level outputs constitute evidence rather than final operational conclusions.

3.2 A structured dataset and a layered evaluation protocol under controlled variability

An evaluation-oriented dataset was constructed with explicitly controlled acquisition variability (illumination, viewpoint, blur, apparent symbol size, occlusion, and negative samples) and macro-scene-level, code-disjoint train–validation–test separation. A layered evaluation protocol spanning detection, decoding, frame-level, and episode-level units decomposes identification reliability into its component sources rather than reporting a single aggregate metric.

3.3 Empirically validated frame- and episode-level identification performance

On a held-out test split of 1,015 annotated objects and 34 episodes, detection localized 98.5% of objects; primary decoding alone identified 35.5% of objects, and the recovery branch raised the frame-level identification rate to 59.1%, with 98.7% precision among accepted identifiers. At episode level, the complete pipeline returned an identifier for 91.2% of episodes and was correct for 85.3%, with 93.5% precision, against a 73.5% retrospective single-frame decoding ceiling; temporal integration resolved four episodes for which no single frame produced a correct decode, and lost none.

3.4 Identification of the conditions that most strongly limit identification reliability

The stratified results show that limited symbol visibility is the strongest limiting factor: objects annotated as occluded or too small were identified in only 9.2% of cases, against

67.9% for normal objects. Viewpoint and blur produced differences of 30.8 and 25.4 percentage points respectively, and identification accuracy increased with apparent symbol area, from 37.6% in the smallest area category to 86.0% in the largest.

3.5 Localization of the residual error source within the recovery mechanism

No validated decode produced an incorrect identifier at either the frame or the episode level. All observed errors were traced to a single mechanism, dominance-based similarity acceptance from a single or weakly supported candidate, which indicates that the decision rule would benefit from an explicit minimum-evidence condition in addition to the relative-separation test already applied.

3.6 A conceptual framework for uncertainty-aware inventory-related decision support

The dissertation establishes a conceptual framework in which accepted, ambiguous, and unresolved episode-level identification outcomes are interpreted as uncertainty-aware evidence for inventory-related decision support. This framework positions validated visual identification as a candidate input to inventory-state estimation, while explicitly defining the boundary between such evidence and an actual warehouse-management integration or spatial localization, which remain outside the implemented system.

4. Main References

Almeida, T., Santos, V., Mozos, O. M., & Lourenco, B. (2021). Comparative analysis of deep neural networks for the detection and decoding of data matrix landmarks in cluttered indoor environments. Journal of Intelligent & Robotic Systems, 103(1), 13.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). The CRISP-DM user guide. In 4th CRISP-DM SIG Workshop in Brussels in March (Vol. 1999).

Edula, V., Ammisetty, K., Kotha, A., Ravipati, D., Davuluri, A., et al. (2023). A novel framework for QR code detection and decoding from obscure images using YOLO object detection and Real-ESRGAN image enhancement technique. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT) (pp. 1–6). IEEE.

Gan, Z. L., Musa, S. N., & Yap, H. J. (2023). A review of the high-mix, low-volume manufacturing industry. Applied Sciences, 13(3), 1687.

Karrach, L., & Pivarciová, E. (2020). Recognition of data matrix codes in images and their applications in production processes. Management Systems in Production Engineering, 28(3), 154–161. https://doi.org/10.2478/mspe-2020-0023

Karrach, L., & Pivarčiová, E. (2021). Comparative study of data matrix codes localization and recognition methods. Journal of Imaging, 7(9), 163.

Liao, L., Li, J., & Lu, C. (2022). Data extraction method for industrial data matrix codes based on local adjacent modules structure. Applied Sciences, 12(5).

Liu, Y., Song, Y., Gu, G., Luo, J., Wang, T., & Jiang, Q. (2024). A data matrix code recognition method based on L-shaped dashed edge localization using central prior. Sensors, 24(13), 4042.

Losada del Olmo, J. J., de Teruel, P. E. L., Ruiz, A., & García-Clemente, F. J. (2025). Computer vision on the edge: A scalable auto-ID solution for industrial logistics. Procedia Computer Science, 265, 276–284. https://doi.org/10.1016/j.procs.2025.07.182

Matuszczyk, D., & Weichert, F. (2023). Reading direct-part marking data matrix code in the context of polymer-based additive manufacturing. Sensors, 23(3), 1619.

Muallim, T., Kucuk, H., Bareket, M., & Kahraman, M. (2025). Lightweight deep learning model and novel dataset for restoring damaged barcodes and QR codes in logistics applications. Computer Modeling in Engineering & Sciences, 143(3), 3557.

Rybakova, E. O., Limonova, E. E., & Bezmaternykh, P. V. (2025). Improving data matrix mobile recognition via fast Hough transform and adaptive grid extractors. Computer Optics, 49(6), 1182–1190. https://doi.org/10.18287/COJ1804

Simeth, A., Kumar, A. A., & Plapper, P. (2025). Flexible and robust detection for assembly automation with YOLOv5: A case study on HMLV manufacturing line. Journal of Intelligent Manufacturing, 36(5), 3447–3463.

Tadjine, C., Ouafi, A., Benlamoudi, A., & Taleb-Ahmed, A. (2025). Computer vision in warehouse management automation: A survey on implemented methods with prototyping hardware. Engineering Applications of Artificial Intelligence, 160, 111886. https://doi.org/10.1016/j.engappai.2025.111886

Vaccari, L., Coruzzolo, A. M., Lolli, F., & Sellitto, M. A. (2024). Indoor positioning systems in logistics: A review. Logistics, 8(4), 126. https://doi.org/10.3390/logistics8040126

Winkelhaus, S., Grosse, E. H., & Morana, S. (2021). Towards a conceptualisation of order picking 4.0. Computers & Industrial Engineering, 159, 107511. https://doi.org/10.1016/j.cie.2021.107511

Xu, L., Kamat, V. R., & Menassa, C. C. (2017). Automatic barcode extraction for efficient large-scale inventory management. In Computing in Civil Engineering 2017 (pp. 340–348). ASCE.

5. Own Publications Related to the Topic

Zeleny, K., Ruppert, T., & Kő, A. (2026). A Proposed Framework for Robust Visual Product Identification in Dynamic Environments. Acta Polytechnica Hungarica, 23(8), 211–240. https://doi.org/10.12700/APH.23.8.2026.8.12