EMERGING TECH CONFERENCE – Edge Intelligence

Volume 01, 2022, Page 311 – 314

Proceedings of Emerging Tech Conference: Edge Intelligence 2022

An embedded vision warehouse management system architecture for pallet identification and tracking

Christos Theocharatos*, Vassilis Tsagaris, Ioannis Papakonstantinopoulos, Dimitris Kastaniotis, Nikos Fragoulis

Irida Labs, Patras InnoHub – Kastritsiou 4, Magoula Patras 26504, Greece

htheohar@iridalabs.com

Abstract

An embedded vision architecture for the detection, identification, tracking and pose estimation in the 3D space of different types of pallets in warehouse management and logistics is presented here. The proposed hardware and software architecture was implemented in the context of the EV-Lib funded project and is able to provide a fully working system for monitoring and assessing the types, 3D positions and locations of pallets and boxes in real-time in a warehouse environment or distribution center, based on a vision system infrastructure and edge artificial intelligence methodology.

1. Introduction

Visual recognition and target tracking is one of the most challenging research areas in computer vision. With the advancements and wide adaption of deep learning techniques, visual object detection, classification and tracking has gained significant attention in the scientific community but also in practical industrial application fields. In warehouses and distribution centers, specifically, a robust management system can potentially grow productivity and inventory accuracy during real-time operations. Warehouses tend to store maximum inventory with ideal material usage and storage. This includes the monitoring of processes like stack height and occupancy in different warehouse areas, pallet/material storage and tracking across the warehouse (for safety reasons), inspection of track loading and uploading operations, and efficient inventory control. Commercial systems rely on the use of bar code scanning technology to recognize and track pallets, materials and goods. Visual technology, on the other hand, can help progress warehouse management systems to the next level, providing efficient solutions to inventory control, material identification and handling, storage location instructions and related safety awareness in real-time. With the advancements of deep learning techniques and the production of related HW infrastructure like high-end vision sensors and edge processing devices, warehouse information can be provided with high accuracy and incorporated in industrial enterprise systems.

2. Embedded pallet tracking system

In this section, all the components of the pallet tracking system for warehouse management are described. The solution incorporates two main components: (1) HW Equipment – The camera units, processing devices and related infrastructure for pallet detection, identification, tracking and pose estimation; (2)SW – The Machine Learning (ML) software, database, UserInterface (UI) and execution engine responsible for processing, storing, displaying and interfacing with the HW.

2.1 System architecture

The overall architecture of the vision-based pallet tracking system is illustrated in Fig. 1. In short, a set of (minimum) four IP cameras are integrated in the warehouse in positions able to cover the entire inspection area. The cameras are connected through a switch to an NVidia AGX Xavier for running the AI models for pallet identification, tracking and pose estimation. Containerized updates are made via docker files through secure VPN connection between the edge device and the accessible teams. Finally, a user interface (UI) running at the edge device is able to demonstrate the functionalities, visualize all visual sensors and provide access to the AI-generated metadata to the system operator.

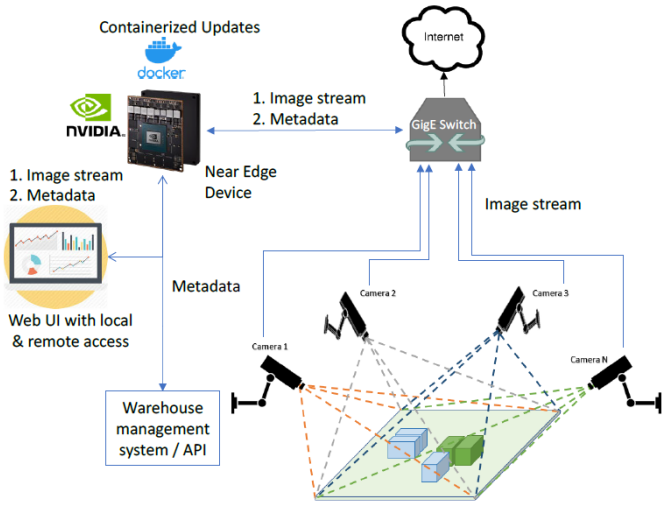

Figure 1: Vision system architecture of the Zeld-e approach.

2.2 HW components of the vision system

The vision system utilizes four industrial cameras in a configuration able to cover a warehouse of up to 500m2 and also support system flexibility. The cameras are a Basler acA5472-5gm ethernet rolling shutter model, a high-resolution camera with a Sony IMX183 CMOS sensor that can deliver 5 frames/sec at 20 MP resolution. A C-mount Kowa lens (LM6HC F1.8) having 6mm focal length is attached to each camera module, being able to cover the entire inspection area. The cameras transmit image data via the local network to a processing unit. The utilized edge processing device is an NVidia Jetson Xavier AGX Xavier, which is able to run machine learning algorithms (i.e. CNN models for detection, recognition and tracking) in real-time, and can also host the database and the UI server for visualization.

2.3 SW components for warehouse management

The SW components and algorithmic pipeline comprising the warehouse management system are illustrated in Fig.2. Initially, a convolutional neural network (CNN) detector processes the RGB camera stream to provide a region of interest (ROI) around the pallets. The CNN model is an optimized Python-based Yolov5 architecture [1] consisting of three parts: a modified CSPDarknet backbone model for feature extraction, followed by a PANet model for feature fusion and an optimized Yolo layer architecture for outputting the detection results (i.e. class, score, location, size). Pytorch framework was used for the training phase. During inference, the CNN receives an RGB image and returns a 2D bounding box that is provided to a CNN classification scheme for identifying the different types of pallets, followed by a tracking (by detection) algorithm, and a 3D pose estimation methodology for estimating the pose of the pallets. Pose estimation is comprised first by a landmark or key-point detection HRNet model [2], which is a universal architecture for visual recognition and object/human pose estimation. Pytorch framework is also used here for the training phase, including RGB images and their labels of 2D landmarks. During inference, the CNN returns a number of 2D landmark points, which are used in the next step for the estimation of the object's pose in the 3D space by solving the Perspective-n-Point (PnP) algorithm that determines the pose of a calibrated camera from n correspondences between 3D reference points and their 2D projection.

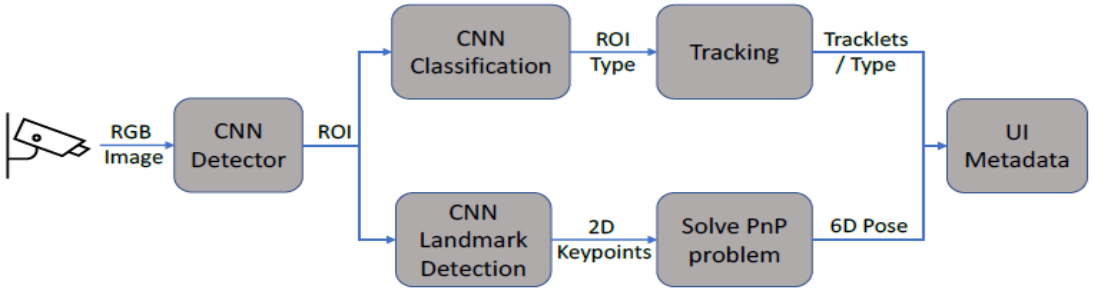

Figure 2: SW components' pipeline for the warehouse management system.

2.4 User Interface

A web-based UI has been developed for visualizing the camera streams and digesting the metadata in real-time, hosted on the Jetson Xavier AGX device and communicating with the sensors using the MQTT protocol. The interface provides access to the real-time display and recording of the multi-camera system (on a MongoDB database), provides search utilities through the recording sessions and displays the data statistics through the analytics screen (Fig. 3).

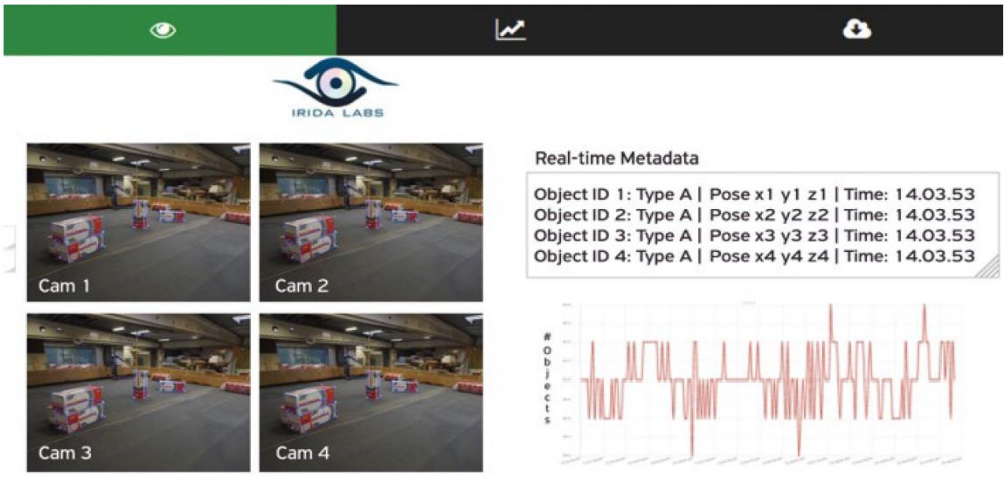

Figure 3: UI multi-camera visualization of the warehouse management system for pallet detection, identification, pose estimation and tracking.

Acknowledgements

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RE-SEARCH – CREATE – INNOVATE (project code: T1EDK – 00681, MIS 5067617).

References

[1] Redmon J, Divvala S, Girshick R, Farhadi A. You Only Look Once:Unified, Real-Time Object Detection. Computer Vision and PatternRecognition 2016.

[2] Wang J, Sun K, Cheng T, Jiang B, Deng C, Zhao Y, Liu D, Mu Y, Tan M,Wang X, Liu W, Xiao B. Deep High-Resolution Representation Learning for Visual Recognition, Computer Vision and Pattern Recognition 2020.