JARVIS

PROCESS AND HUMAN ACTION PERCEPTION AND HUMAN INTENTION PREDICTION METHODS

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement nº 101135708. The dissemination of results herein reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.

INTRODUCTION

Human action perception and intention prediction methods are foundational for human-robot collaboration since they enable robots to understand, anticipate and adapt to human behaviour in real-time. In addition, detection of fine defects and anomalies in process execution is essential for task performance and safety. Predicting human intention in collaborative industrial settings is particularly challenging because shop floors are dynamic, unstructured, and often crowded environments. Effective prediction depends on robust activity perception, temporal reasoning, and contextual awareness that can generalize across different workstations and process configurations.

The Human Intention Perception and Prediction (HIPAP) module address this need by combining AI-based methodologies with multimodal - multisensory data sources, to evaluate human primitives, including poses and gestures, and predict human actions and intention during task execution. It implements an awareness system for anticipating the human operator's next action and movement to support both dynamic operator instructions and robot behaviour adaptation, and aims to minimize misunderstanding in task execution, reduce coordination errors between human and robot and enable robots to provide proactive assistance. HIPAP interacts with various modules including multisensory perception and intelligent digital twins (IDT) to acquire multisensory data from the assembly process and is a feeder of important data to the robot control (RCM), human robot interaction (HRIM) and behaviour adaptation (ROBA), and task planning (TPM) modules.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement n° 101135708. The dissemination of results herein reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.

MODULES OVERVIEW

The HIPAP module consists of three key submodules which establish three key cognition capabilities: i) human detection and pose estimation, ii) task execution and anomaly detection, and iii) human intention prediction (see Figure 1). The first two submodules lay the foundation of HIPAP focusing on a set of cognition capabilities that are needed to estimate human poses around the workplace and classify the different states of task execution in real time. The third submodule aims to infer the operator's intention from observable behaviour, enabling the system to anticipate operator needs, adapt robot motion and task strategy proactively, but also provide contextualized instructions to operators.

HUMAN DETECTION AND POSE ESTIMATION

Accurate human detection and pose estimation are fundamental prerequisites for recognizing human actions and intentions in collaborative industrial environments. Using state of the art deep learning frameworks the submodule uses multisensory inputs from cameras, RGBD sensors, IR, lidar etc. for multi-person detection and reliable body posture estimation and 2D-3D skeletal joint reconstruction that enables the downstream modules to interpret operator activity, predict motion trajectories and ensure safe robot behaviour. The extracted posture and 2D-3D joint data can be directly consumed by the rest of the HIPAP submodules responsible for action recognition and intention prediction.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement n° 101135708. The dissemination of results herein reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.

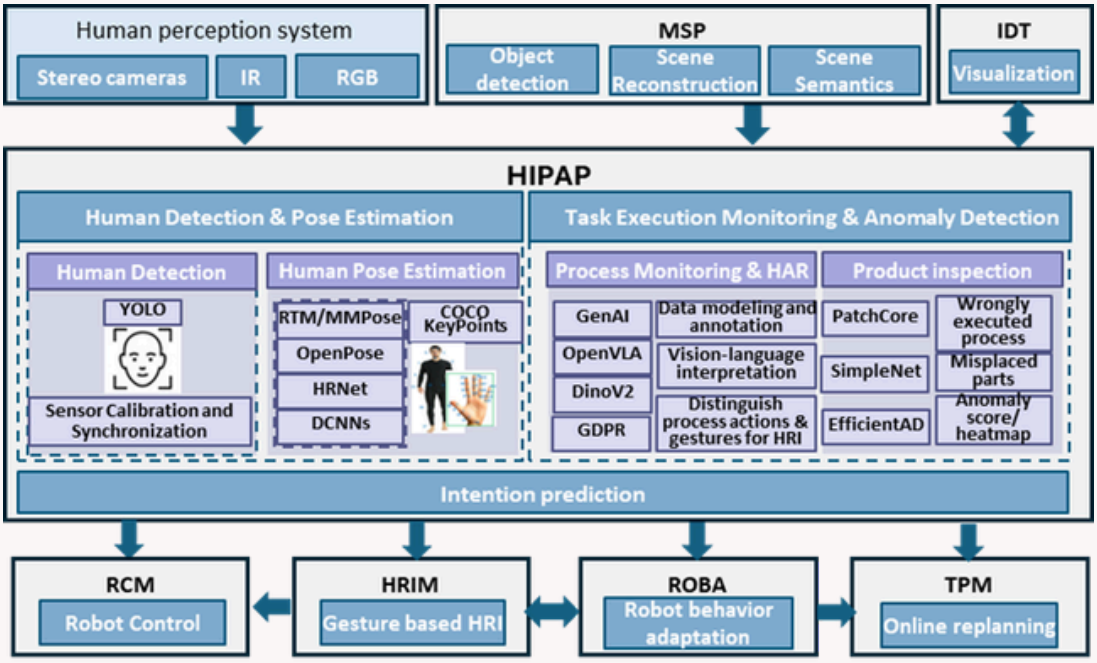

Figure 1: HIPAP architecture.

TASK EXECUTION MONITORING AND ANOMALY DETECTION

Task execution monitoring and anomaly detection submodule identifies deviations from expected task sequences, whether caused by human error, robotic malfunction, or unforeseen environmental disturbances. To comprehensively perform this task, the submodule implements human action recognition (HAR), task execution monitoring and product inspection.

HAR detects and classifies operator action primitives such as walk, idle, reach, grasp, pick, align, insert, and tighten which can be further used to interpret higher-level tasks and for coupling the perception results to other modules. It uses the 2D-3D skeletal joint keypoints and object tracking in the human detection and pose estimation module as inputs, to capture patterns from human body motion and classify these action primitives.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement n° 101135708. The dissemination of results herein reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.

Task execution monitoring is using HAR and other multisensory perception data to identify the current state of the task execution and detect/classify anomalies. It learns representations of normality for each task and action during execution based on expert demonstrations and is able to classify an anomaly by detecting deviations from the normal operation.

Product inspection is deployed to identify if the results of human or robot task yield the expected results. Vision-based anomaly detection techniques trained on nominal video and images from the assembly or available public databases, are used to highlight deviations from potentially defect data to classify and determine the defects. Using visual features extracted from the training data, the product inspection produces an anomaly score which is used to highlight the segment of the image or video where the defect is.

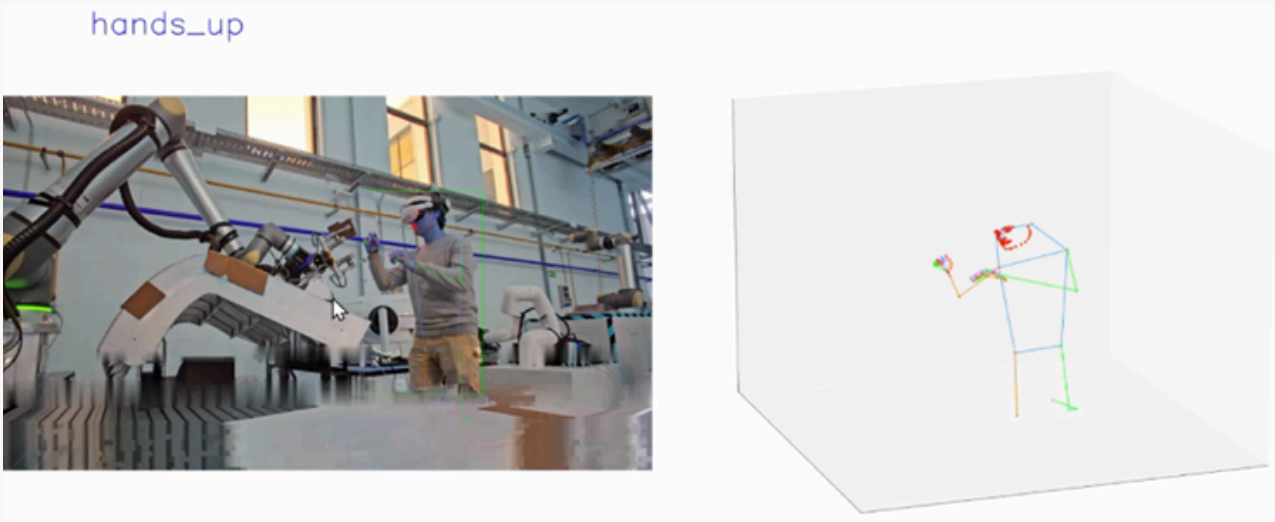

Figure 2: Human pose estimation and action recognition.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement n° 101135708. The dissemination of results herein reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.

HUMAN INTENTION PREDICTION

By human or operator intention in the context of the JARVIS project, we define the possible next actions of the assembly task and process execution and the location that the operator may be. The submodule uses inputs of the operator trajectory location and action classes, from the first two sub-modules, together with scene semantics from multisensory perception data (MSP) and process and task hierarchy from IDT to obtain probabilities of the intention (i.e. predicted operator actions and locations) of the operator. In addition, the submodule performs gestured-based intention recognition as a form of explicit human-robot interaction for the operator to convey intentions to the robot. Gesture potential as an interaction feature is further expanded with voice commands.

Operator trajectory prediction is formulated as a time-series forecasting problem, where the objective is to estimate the operator's future positions based on past observed trajectories and contextual task information. The system assumes a task-aware configuration, meaning that the upcoming task is known at any given moment, providing spatial context about where the operator is likely to move. Since different tasks correspond to specific workstation regions, this context significantly improves the prediction accuracy.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement n° 101135708. The dissemination of results herein reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.

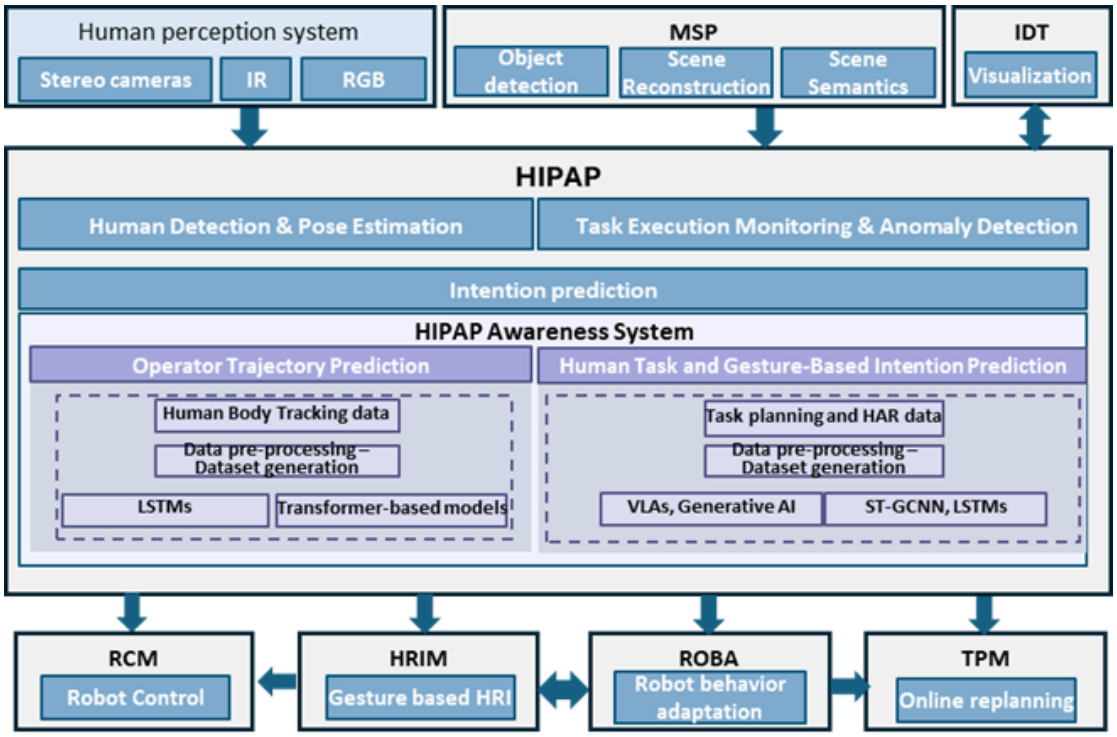

Figure 3: Human intention prediction – HIPAP Awareness System

The HIPAP module can be deployed through command-line, APIs and ROS2 endpoints on JARVIS's unified data exchange framework. These interfaces allow users, autonomous agents or other modules to interact with HIPAP to load batch or real-time data and retrieve in real-time human perception, action and intention outputs.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement n° 101135708. The dissemination of results herein reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.

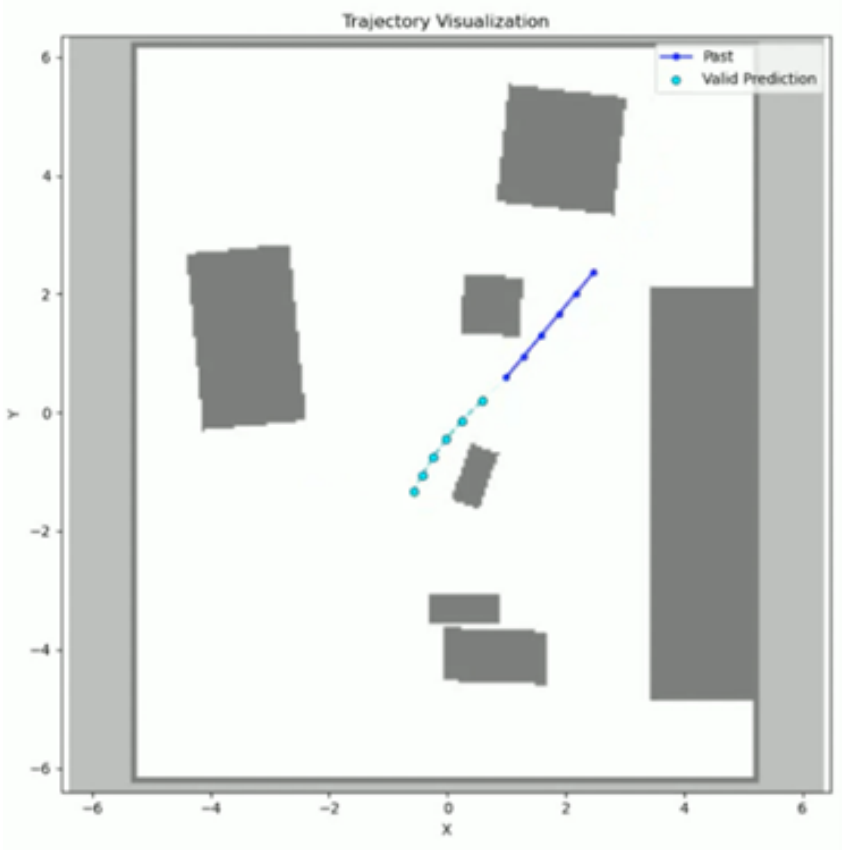

Figure 4: Operator trajectory prediction visualization.

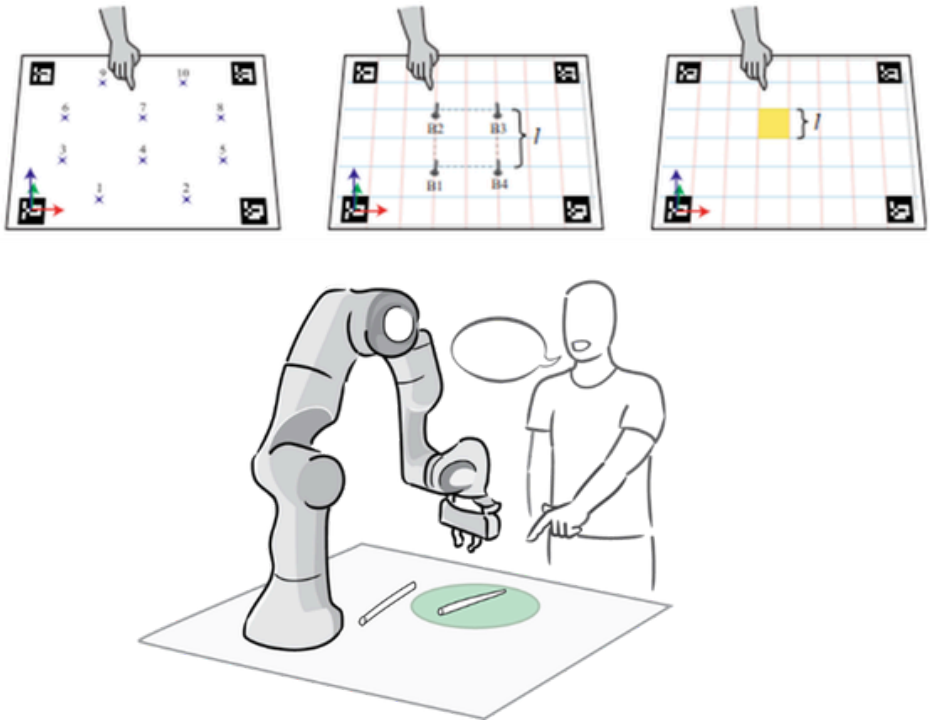

Figure 5: Gesture-based intention recognition with voice commands.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement n° 101135708. The dissemination of results herein reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.

SUMMARY

The HIPAP module implements a human action perception and intention prediction system in human-robot collaboration in assembly and manufacturing systems. It aims to improve task execution performance, safety and robot adaptability in a dynamic, unstructured and crowded collaborative process environments.

Based on AI approaches and fusing multisensory data it can accurately and efficiently:

- Detect and estimate human pose in relation to the workplace
- Recognise, categorise and predict operator actions during task execution
- Predict operator action and trajectory on the shopfloor
- Estimate the state of task execution and detect anomalies in them and therefore, provide a strong foundation capability for a robust, safe and efficient human-robot collaboration.

Funded by the European Union

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement n° 101135708. The dissemination of results herein reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains.