Functionalized Soft robotic gripper for delicate produce harvesting powered by imitation learning-based control

# D5.1 Annotated datasets of human expert demonstrations

| | |
|---|---|
| **Deliverable Number** | 5.1 |
| **Work package Number and Title** | WP5 - Robotic Skill Acquisition algorithm development |
| **Lead Beneficiary** | University of Essex |
| **Version - Status** | 1.0 - Final |
| **Due Date** | 31/3/2022 |
| **Deliverable Type**[1] | DEM |
| **Dissemination Level**[2] | PU |
| **Internal Reviewer** | TEAGASC |
| **Filename** | D5.1_Annotated datasets of human expert demonstrations |

[1] **Please indicate the type of the deliverable using one of the following codes**:
R = Document, report
DEM = Demonstrator, pilot, prototype
DEC = Websites, patent fillings, videos, etc.
ETHICS = Ethics requirement
ORDP = Open Research Data Pilot
DATA = data sets, microdata, etc.
[2] **Please indicate the dissemination level using one of the following codes**:
PU = Public
CO = Confidential, only for members of the consortium (including the Commission Services)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 101017054

# Document info

### Authors


| Surname | First name | Organization | E-mail |
|---|---|---|---|
| Porichis | Antonis | University of Essex | a.porichis@essex.ai |
| Vasios | Konstantinos | University of Essex | k.vasios@essex.ai |
| Mavridis | Panagiotis | TWI Hellas | panagiotis.mavridis@twi.gr |
| Mastrogeorgiou | Thanasis | TWI Hellas | thanasis.mastrogeorgiou@twi.gr |


### Internal Reviewers

| Surname | First name | Organization | E-mail |
|---|---|---|---|
| Uccello | Andrea | Teagasc | andrea.uccello@teagasc.ie |
| Grogan | Helen | Teagasc | Helen.grogan@teagasc.ie |

### Document History

| Date | Version | Editor | Change | Status |
|---|---|---|---|---|
| 15/03/2022 | 0.1 | Antonis Porichis | Initial Draft | Draft |
| 17/03/2022 | 0.2 | Thanasis Mastrogeorgiou, Panagiotis Mavridis | Data collection setup | Draft |
| 18/03/2022 | 0.3 | Antonis Porichis | Pressure maps and force measurement visualisations | Draft |
| 21/03/2022 | 0.4 | Andrea Uccello | Identified a few minor typos and made several comments | Draft |
| 29/03/2022 | 0.5 | Helen Grogan | Inserted image of harvesting glove with mushroom residue; added Chemicals after Mitsui to give full name | Draft |
| 31/03/2022 | 1.0 | Antonis Porichis | Integrated internal reviewers' feedback | Final |

# TABLE OF CONTENTS

# 1 INTRODUCTION

The purpose of this report is to document the collection and post-processing of the datasets pertaining to mushroom picking demonstrations. These demonstrations took place in TEAGASC's facilities in Dublin, Ireland over a period of five days from Feb 7th to Feb 11th, 2022.

## 1.1 DATA STREAMS AND RECORDING EQUIPMENT

Expert demonstrations comprise two main data streams, the tactile sensing and the RGB-D video stream. These streams along with the respective pieces of equipment used to record them are described below.

### 1.1.1 RGB-D VIDEO STREAM

The RGB-D stream provides visual and depth information which will be crucial for obtaining high-level semantic-rich information. Four RGB-D sensors were installed on separate locations to obtain a comprehensive view through multiple viewpoints. The RGB-D sensors used in the expert demonstration recording sessions were four instances of the ZED 2 camera model[3]. The layout of the four sensors around the mushroom bed is illustrated in Figure 1 below while a picture of the setup is shown in Figure 2 in the next page.

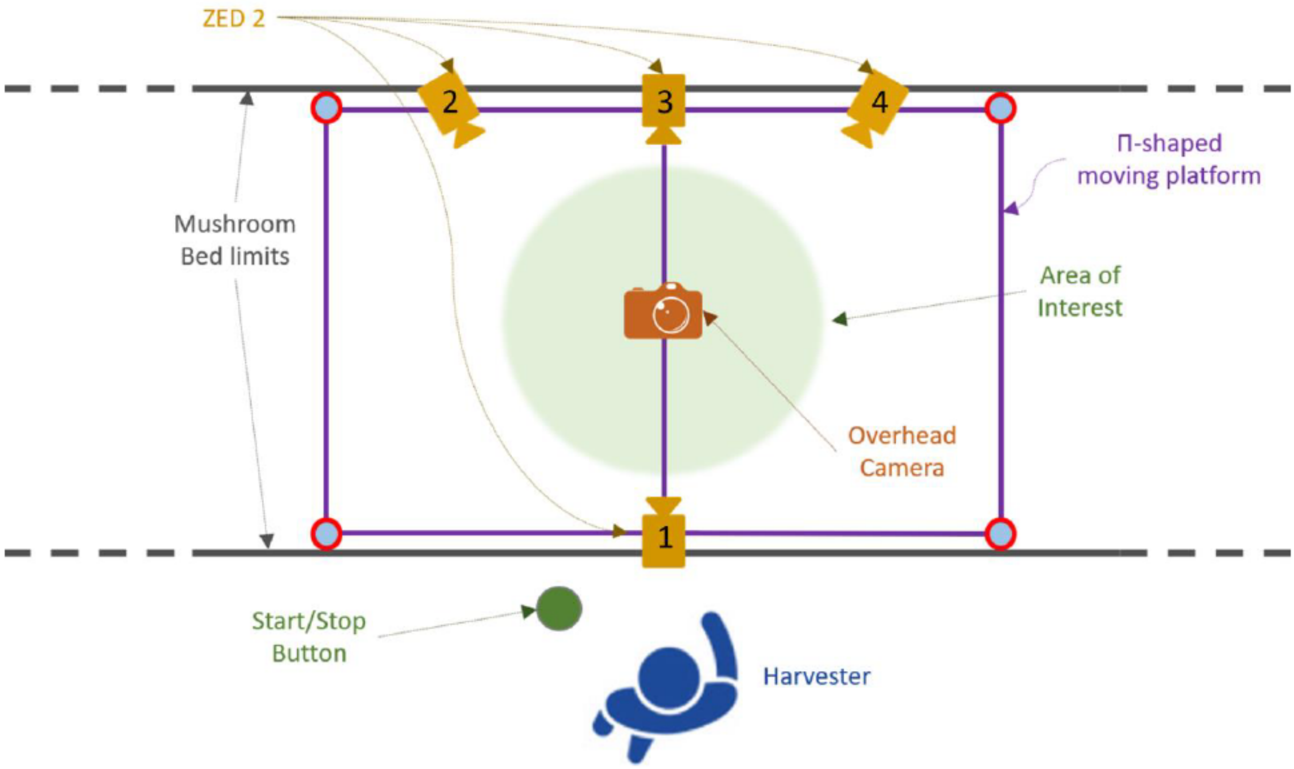



Figure 1 RGB-D camera positions

[3] ZED 2 Camera and SDK Overview available at https://www.stereolabs.com/zed-2/

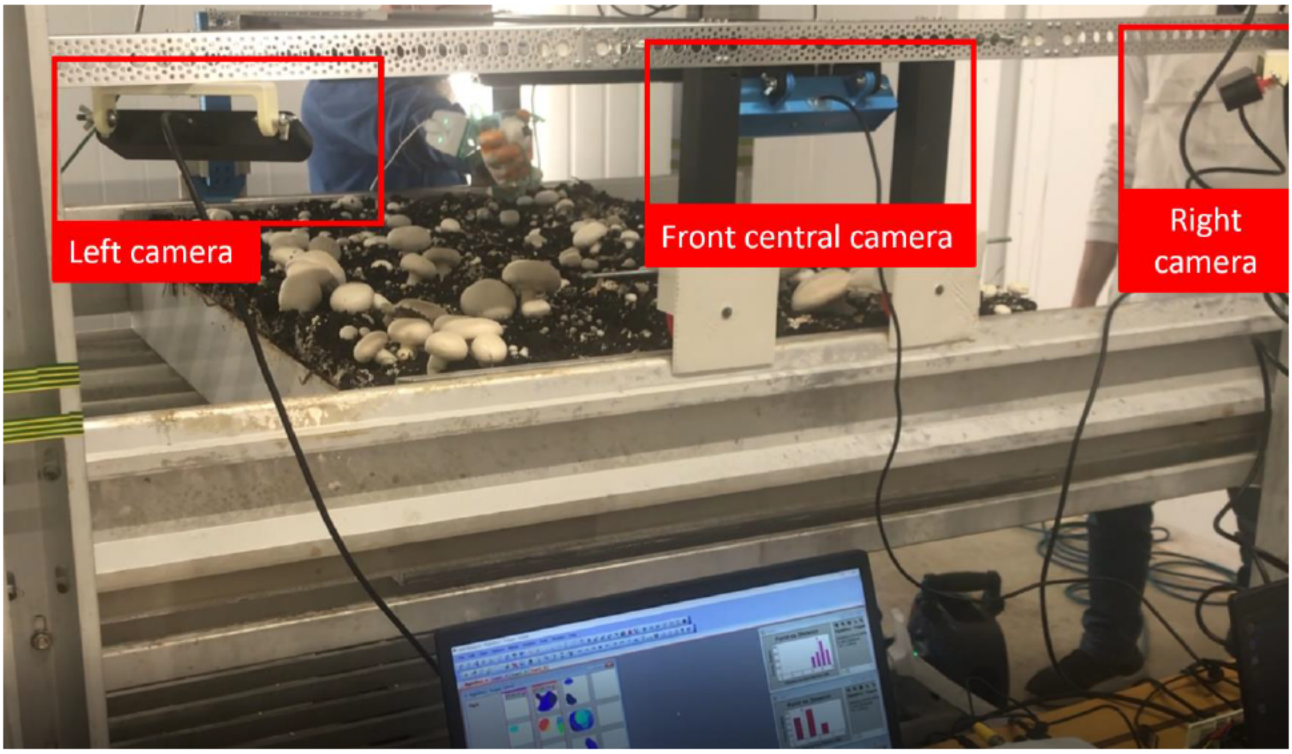



Figure 2 Expert demonstration collection setup

### 1.1.2 TACTILE SENSING STREAM

The main tactile sensing recording system used was the Tekscan Grip System[4]. The system provides >300 sensing elements distributed across the surface of the user's hand providing high-resolution pressure mapping information as seen in Figure 3. In the context of mushroom picking demonstration recording, we focused on the sensing elements placed on the Thumb, the Index and the Middle Finger.

Figure 3 Pressure map profiles tracked by Tekscan Grip System

[4] Datasheet available https://www.tekscan.com/products-solutions/systems/grip-system

The sensing elements were mounted on a standard mushroom pickers' glove as shown in Figure 4. This choice was made to obtain more reliable data, as the original set up shown in the Figure 3 did not allow a natural hand movement and the required sensitivity to apply the required force on mushrooms Figure 5 illustrates a mushroom picking experiment carried out by Andrea Uccello from TEAGASC.


Figure 4 Tekscan Grip System sensing elements mounted on the mushroom picker's gloves

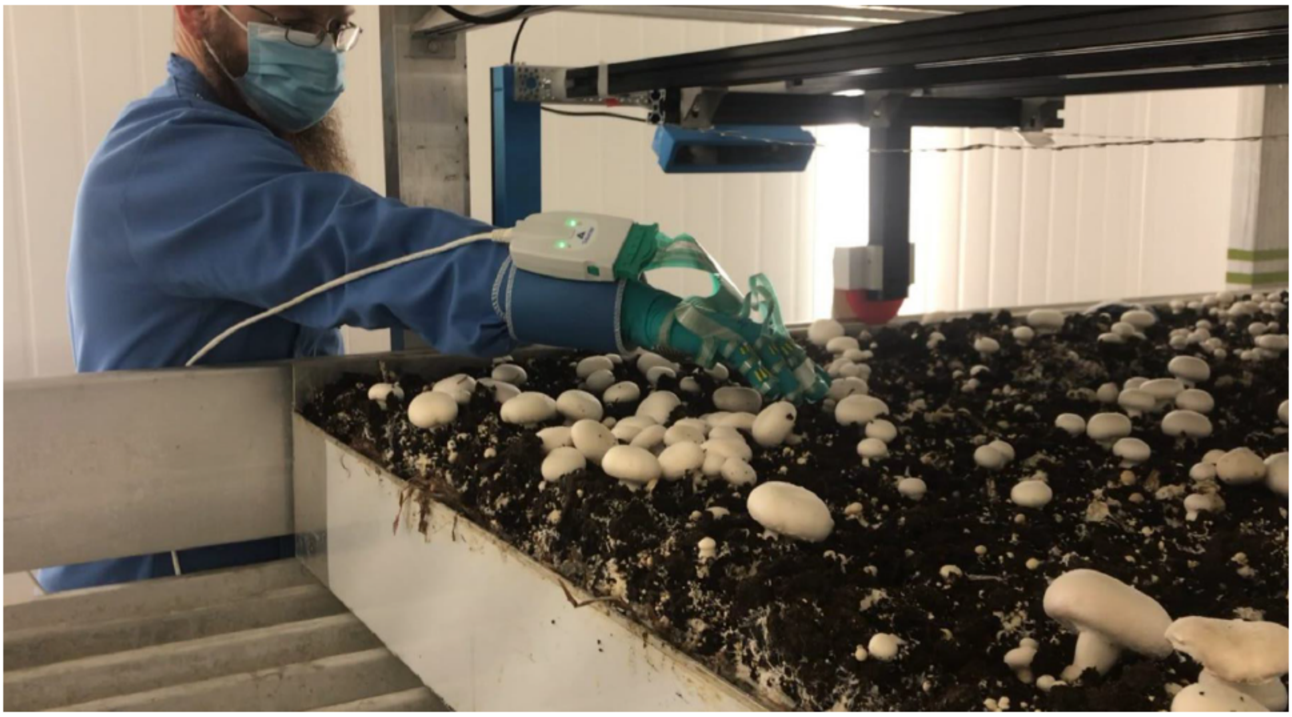


Figure 5 Mushroom Picking trial using the sensorised glove

Figure 6 MITSUI sensors mounted on the expert picker gloves

In order to investigate the domain shift between the measurements of the Tekscan Grip System and the tactile sensors envisaged for the soft robotic gripper developed within the project, a limited number of trials were conducted using the sensors developed by MITSUI Chemicals. Unlike the Tekscan Grip System sensing elements, which provide pressure maps, these are 6-axis tactile sensors measuring force and torque at a single point. Figure 6 above shows these sensors mounted on the mushroom picker's glove

It should be noted that both the force and the pressure sensors were applied in the areas of the fingers that most frequently get in contact with the mushrooms during the picking process. To identify those areas, a used glove was analysed, as the contact areas were clearly stained by mushroom residual tissues. This is shown in Figure 7 to the right

Figure 7 Used glove with stains indicating finger regions most frequently contacting the mushroom surface

## 1.2 DATA ACQUISITION HARDWARE AND MECHANICAL HARDWARE

In order for the camera and the glove recordings to be synchronized, the two streams should start simultaneously. To achieve that, a pulse signal was generated at the beginning of the recording procedure by the Tekscan hub output port. A GPIO pin on the Jetson devices was used to read this pulse, together with a pull-down resistor as shown in the Figure 8. The ZED2 cameras are connected to the Jetson via USB 3.0, are initialized and are waiting in standby mode for the pulse signal to initiate the recording. The output is an SVO file from every camera. The executable that interacts with the ZED cameras was developed in C++ using the ZED2 SDK. In addition, the gaosiy/JetsonXavierGPIO library was used to handle in the same app the GPIO signals sent from the Tekscan port. The mechanical structure that accommodates all devices can be seen in Figure 9. Lastly, it is important to mention that 4 cameras were used. The Jetson Xavier NX could handle 2 ZED2 cameras while one jetson nano per camera was needed for the last two cameras. Figure 10, illustrates how the Jetson devices and the cameras were connected during the data acquisition trials.

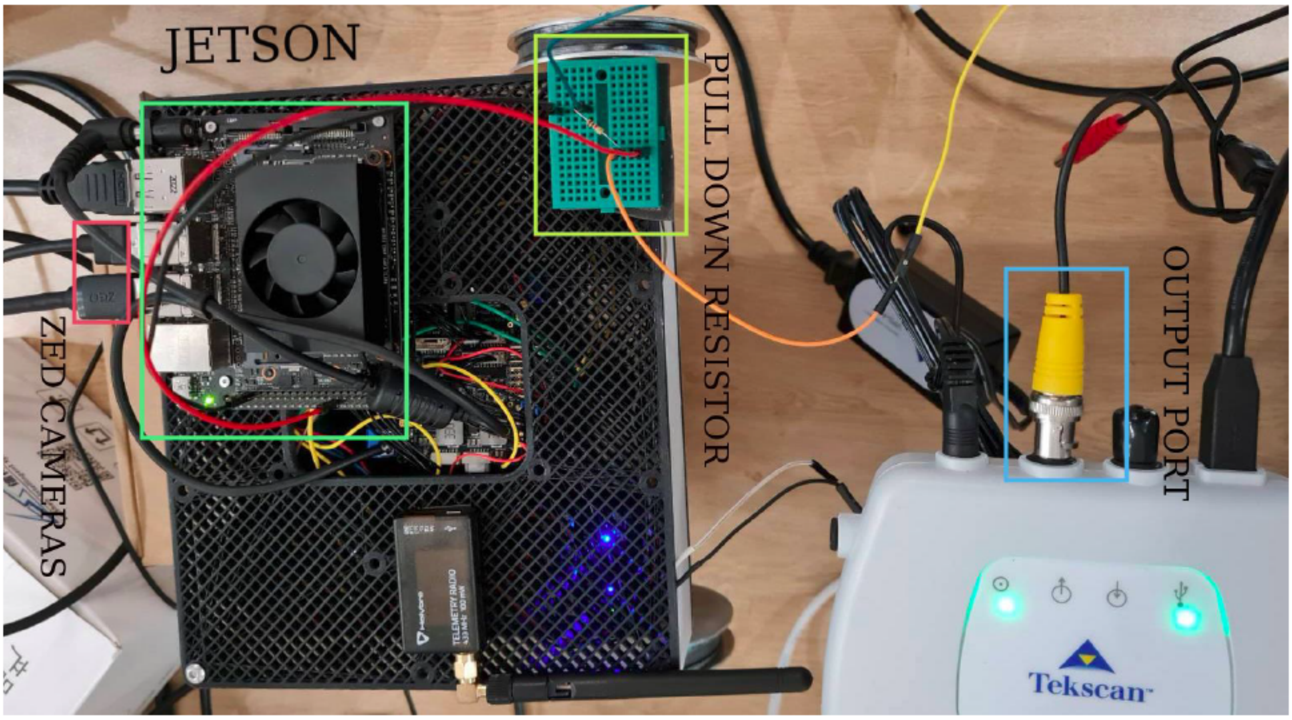
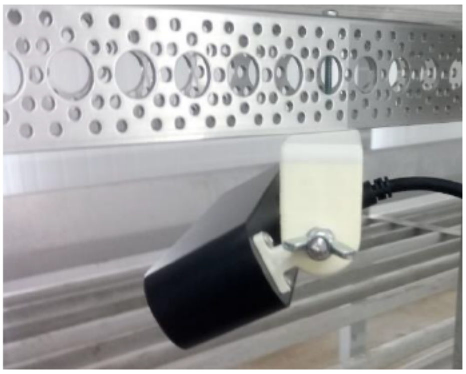



**Figure 8 The ZED camera is connected to the Jetson. A trigger signal initiates the recording procedure. This signal is sent from the Tekscan output port and received at a Jetson GPIO pin.**

(a)



(b)

**Figure 9 (a) The Passive carriage without the cameras. (b) A hinge joint was 3D-printed so that different camera positions and orientations could be tested. The 3D-printed hinge joint could be easily mounted on the perforated beam on the side of the carriage.**

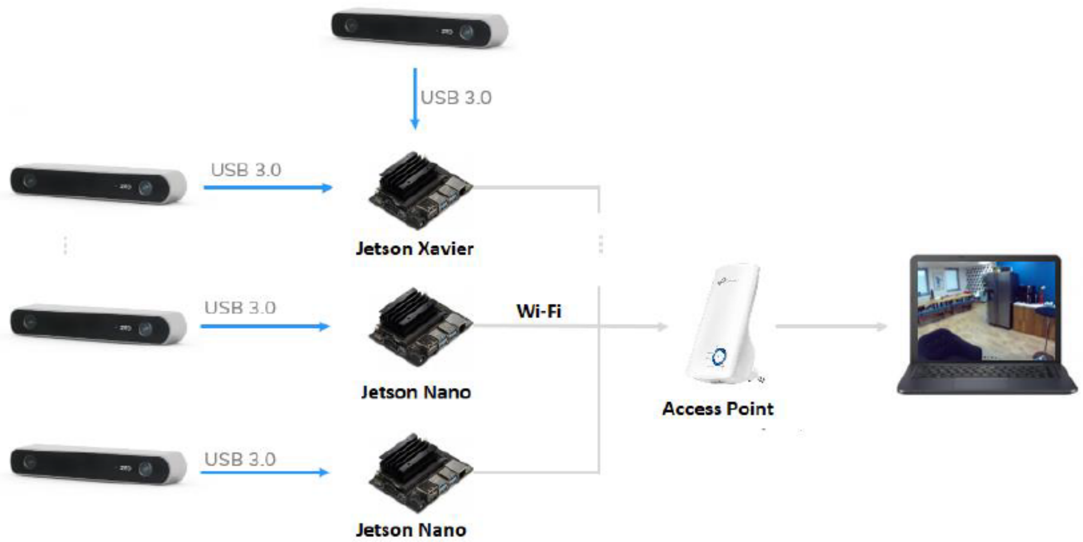



Figure 10 Topology of the Jetson devices, the ZED2 cameras and the PC that records the camera streams.

## 1.3 DATA FORMATS

The raw data recordings consist of sets of .svo and .csv files produced by the ZED 2 cameras and the Tekscan Grip System respectively. The SVO file format is a proprietary format that can only be read from the ZED SDK and its tools. It contains the unrectified images of the camera along with metadata information such as timestamps[5]. CSV stands for Comma-Separated Values, a file format which allows data to be saved in a tabular form.

In order to establish a common, unified basis for storing the data, we have written custom Python scripts for parsing both file formats and translate the data into Hierarchical Data Format version 5 (HDF5) format. This is an open source file format that supports large, complex, heterogeneous data. HDF5 uses a "file directory" like structure that allows users to organize data within the file in many different structured ways, in a fashion similar to how files are stored within folder structures[6]. The choice of HDF5 was made to satisfy three important criteria, namely ease-of-use, reliability, flexibility. This format is well-trusted and widely used by the research community for storing complex data and there are multiple open-source APIs for all major programming languages to parse HDF5 files.

In the context of expert data demonstrations, the structure of each file is illustrated in the following figure:

[5] SVO file parsing guide available at https://support.stereolabs.com/hc/en-us/articles/360009986754-How-do-I-convert-SVO-files-to-AVI-or-image-depth-sequences-

[6] Hierarchical Data Formats - What is HDF5?, available at: https://www.neonscience.org/resources/learning-hub/tutorials/about-hdf5

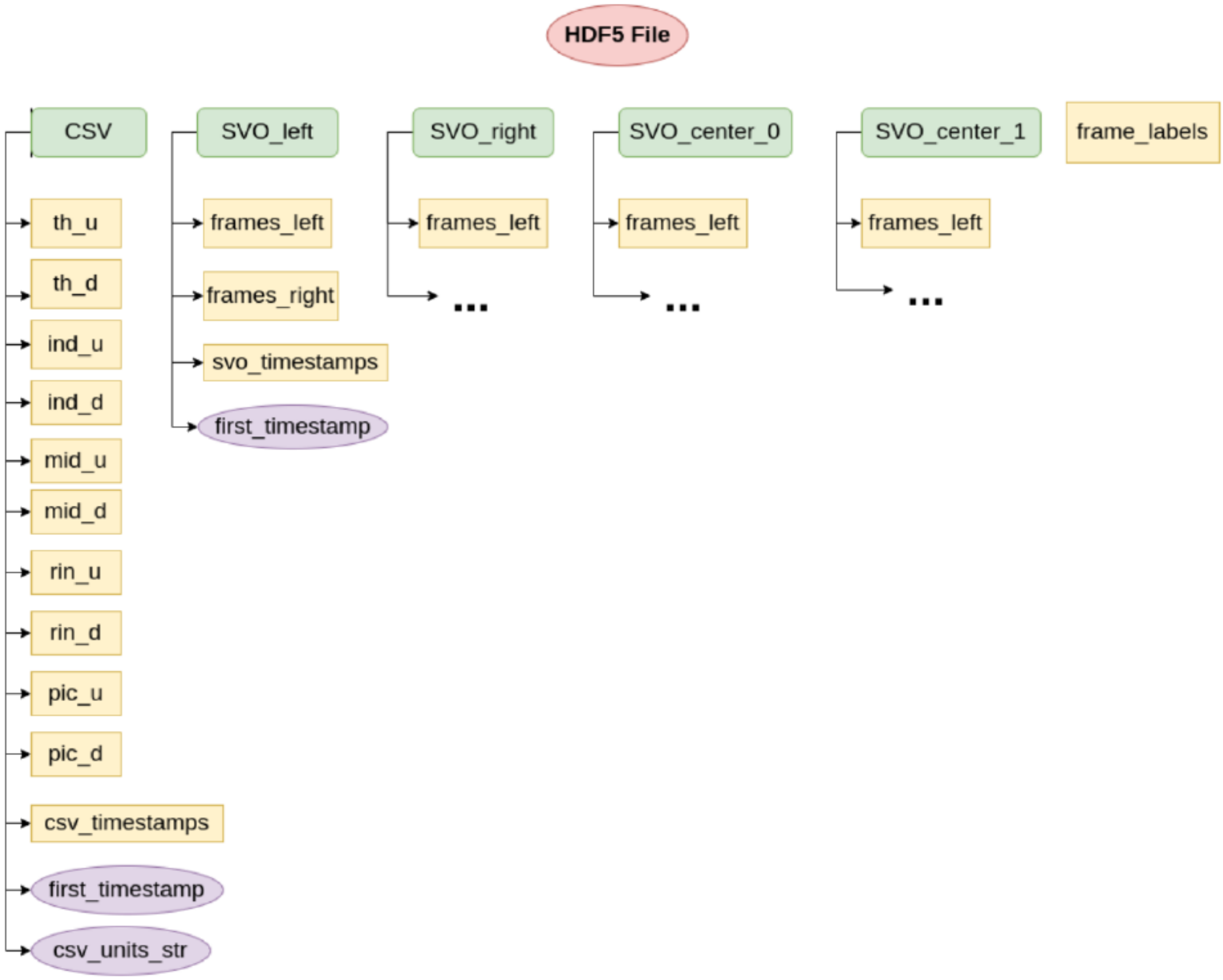



Figure 11 Expert demonstration .h5 file structure

The file consists of five groups and one dataset; one of the groups contains all the data extracted from the csv file of the trial, the other four the information from the corresponding cameras and the dataset the labels of the frames.

The CSV group contains ten 3-D arrays, two for each finger, which correspond to the sensors on the tip and the midsection of each finger. The correspondence of the array identifiers and the finger locations is illustrated in Figure 12. The timestamps of the glove are also saved, with starting time zero and the others following are calculated with the time distance of the frames. However, the first real timestamp is also stored in case of future calculations. Lastly, in the csv group the unit of pressure measurement is also stored.

The svo groups contain all the frames of the cameras in an array format from both camera lenses, an array of the corresponding timestamps, again starting from zero and the following calculated with the time distance of the frames, and the first actual timestamp.

The frame_labels is a matrix that keeps the labels of each frame in the trial. The states identified are the following:

- **Approaching (State 0)**: the picker's hand is moving towards the mushroom without yet touching it,
- **Touching (State 1)**: the picker's hand is touching the mushroom without yet twisting or tilting
- **Outrooting**: the picker's hand performs a combination of twisting and tilting motions while the mushroom is not yet outrooted

- **Lifting**: The mushroom is outrooted and is being lifted


Figure 12 Correspondence of CSV datasets with finger locations

## 2 RGB-D DATA

Indicative screenshots of the RGB-D streams are provided in the next figures. The camera numbering is provided in Figure 2, page 4.

Figure 13 RGB-D camera 1: Left Image, Right Image and Depth Array

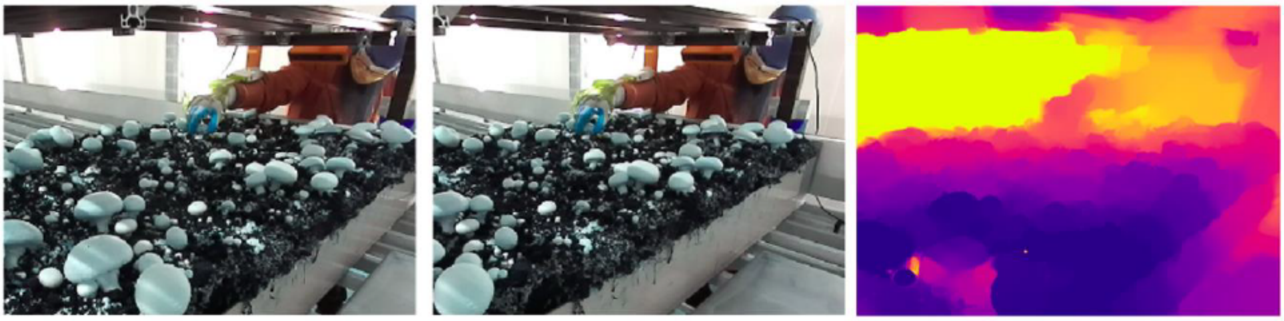


Figure 14 RGB-D camera 2: Left Image, Right Image and Depth Array

Figure 15 RGB-D camera 3: Left Image, Right Image and Depth Array

Figure 16 RGB-D camera 4: Left Image, Right Image and Depth Array

As seen in the figures above, the location of the cameras makes it difficult to obtain fine visual detail while depth information is very coarse. This is due to inherent limitations imposed by the geometry of the mushroom growing structures and ways to address these challenges should be further investigated.

# 3 TACTILE DATA

Graphical illustrations of the tactile data for each of the two sensing systems, namely the Tekscan Grip System and the sensors provided by MITSUI Chemicals are provided in the next paragraphs.

## 3.1 AGGREGATED PRESSURE AND FORCE MEASUREMENTS

The following figures illustrate the aggregated pressures retrieved from the Tekscan Grip System. The aggregation was performed by way of averaging the pressures recorded at each region of the finger. Figure 17 illustrates the time-series of the aggregated pressures on the tips of the three fingers (thumb, index and middle finger) while Figure 18 depicts the corresponding pressures at the finger mid-sections. These time-series are obtained for the same picking trial as the one illustrated in the series of figures from Figure 13 to Figure 16.

In this particular picking trial, it seems that the pressure from the thumb tip is dominant while the pressures from the index and middle fingers is relatively evenly distributed between the tip and the mid-section. The picker initially touches the mushroom around the 10th second of the recording and the twisting and tilting motion starts between the 11th second. The peak observed around the 12th second marks the successful breaking of the roots and from that point on the mushroom is being lifted.

To assess the domain shift between the pressure measurements retrieved by the Tekscan Grip System and the force measurements that will be available to the gripper, we also provide an indicative illustration of an expert demonstration recording using the sensors provided by MITSUI mounted on the on the experts' gloves as seen in Figure 6, page 7. The illustration of force measurements retrieved by the MITSUI sensors are provided in Figure 19. As observed in Figure 17 and Figure 19, pressure and force measurements by the two systems provide similar time-series profiles.

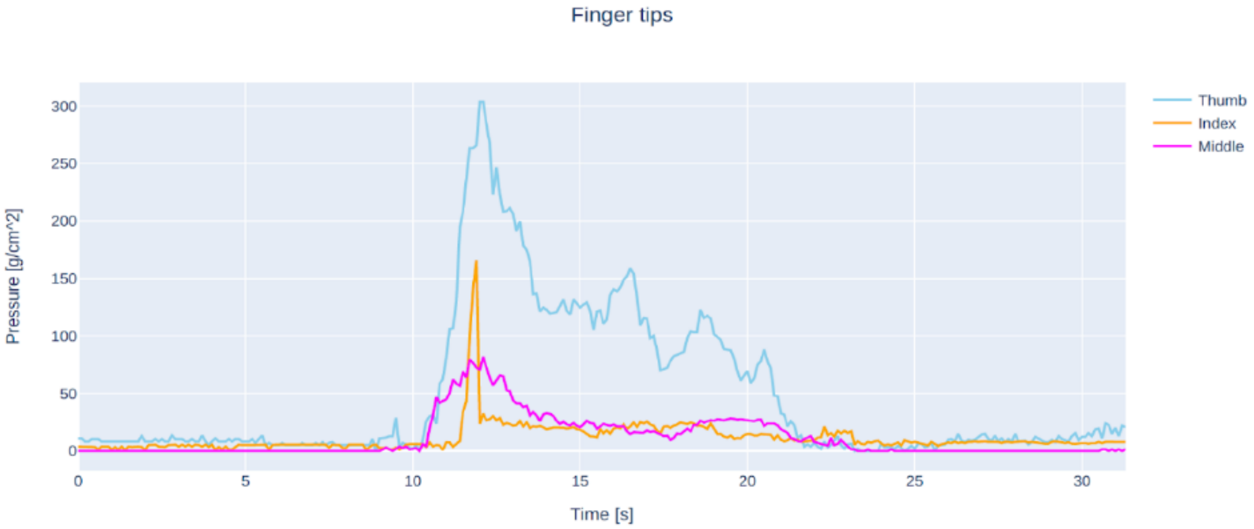



Figure 17 Aggregated pressures on finger tips (Tekscan Grip System)


Figure 18 Aggregated pressures on finger mid-sections (Tekscan Grip System)

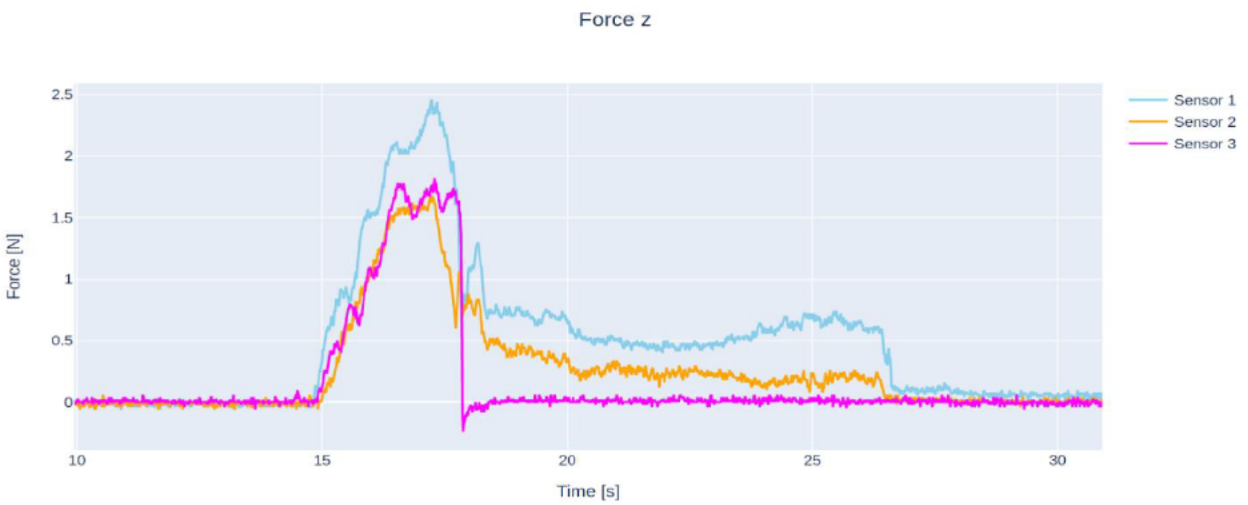



**Figure 19 Forces on finger tips (Mitsui sensors)**

## 3.2 PRESSURE RECORDING ANNOTATIONS

By examining the pressure time-series in combination with the visual streams we have performed an annotation of the recording based on four states, namely "Approaching", "Touching", "Outrooting" and "Lifting" as explained in paragraph 1.3 DATA FORMATS. The state annotation corresponding to the picking trial considered in Figure 17 and Figure 18 is plotted in Figure 20 below. Figure 21 illustrates the segmentation of the index tip pressure profile into the four states.

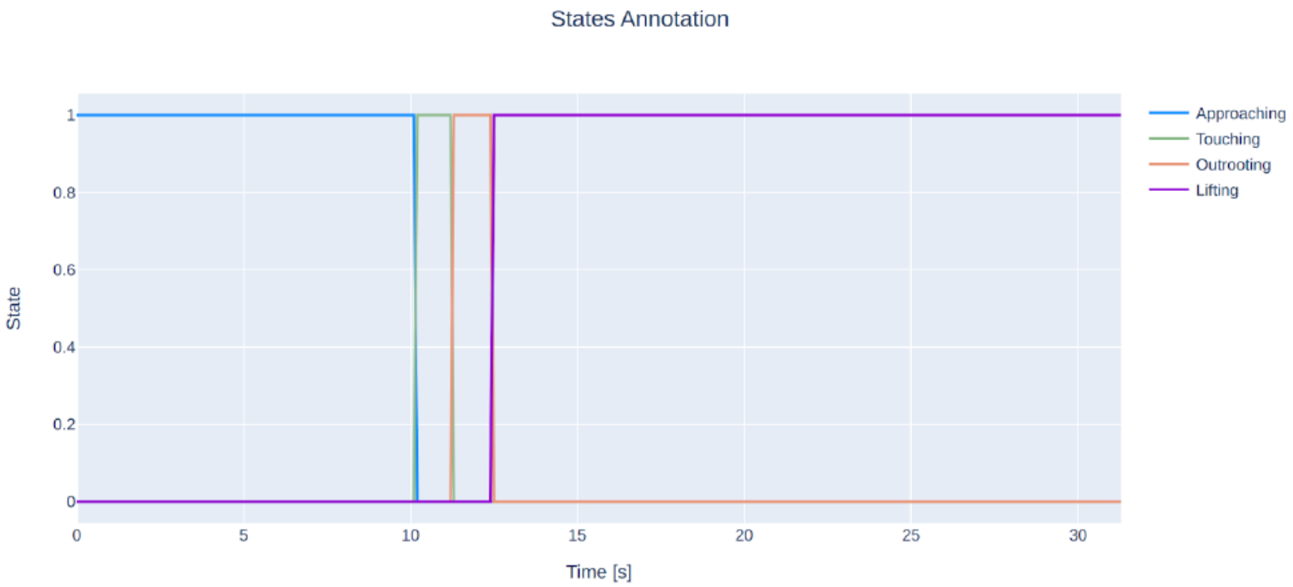



**Figure 20 State sequence for recording presented in Figure 17 and Figure 18**

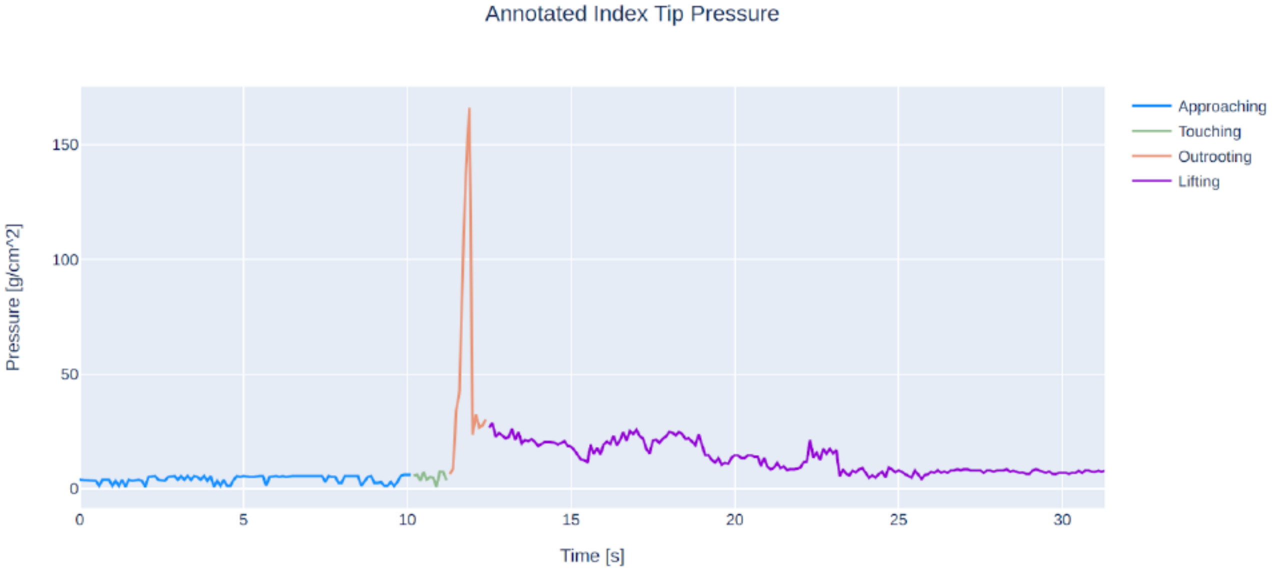



Figure 21 Segmentation of the index tip pressure profile

# 4 SYNCHRONISED SEQUENCES

For a comprehensive overview of the picking process both in terms of the visual and the pressure maps streams, Figure 23 illustrates four screenshots, each belonging to a different state, of the picking attempt considered throughout Figure 17, Figure 18, Figure 20 and Figure 21 with the pressure maps reading overlaid in the bottom right corner of each image. The correspondence of the each overlayed patch with the finger regions is provided in Figure 22. As seen in Figure 23, there is a clear distinction between the pressure patterns observed during the different states, with measurements being increasingly apparent during the "Touching" and "Outrooting" states and gradually receding towards the lifting phase.

Figure 22 Overlay - finger region correspondence

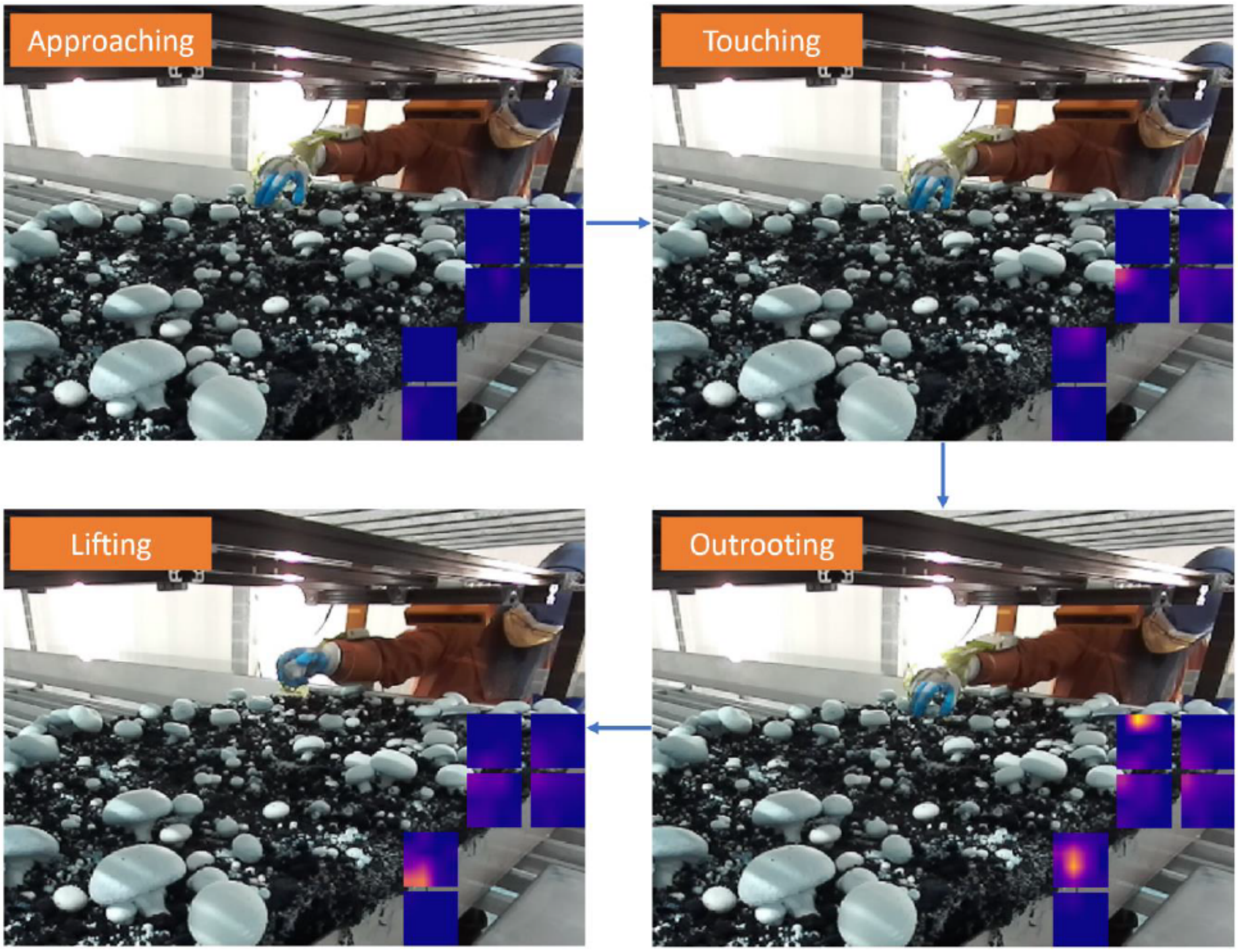



Figure 23 Mushroom picking sequence with pressure maps overlay

## 5 CONCLUSIONS

The expert data collection methodology and indicative results have been thoroughly presented. Available datasets are accessible upon request in form of:

- Datasets: contains the .h5 files storing the datasets as explained in paragraph 1.3 DATA FORMATS
- Videos: contains indicative videos of picking attempts with overlayed pressure maps in MP4 format