Open Access Database of Industry 4.0 Tasks for the Development of AI-based Classifier

Francesca Mongelli
Department of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy
francesca.mongelli@studenti.polito.it

Matteo Menolotto
Tyndall National Institute
University College Cork
Cork, Ireland
matteo.menolotto@tyndall.ie

Brendan O'Flynn
Tyndall National Institute
University College Cork
Cork, Ireland
brendan.oflynn@tyndall.ie

Danilo Demarchi
Department of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy
danilo.demarchi@polito.it

***Abstract*—Robots and humans coworkers are sharing more and more portions of the smart manufacturing globally, meeting the need for high flexibility and rapid changes in the production layout. To be fully effective, however, such transition from classic robotics to the so-called collaborative robotics has to address several open problems, mostly related with safety and task optimization. Promising answers are coming from the motion capture technology, where wearable and optoelectronic sensing devices are deployed to gather human centric data to provide the robots with some form of awareness respect with the human activity and position. Tracking the hand of the operator, in particular, offers many advantages as we use our hands to explore and interact with the surroundings and to communicate. This has been highlighted by the several works focusing on gesture hand configuration recognition. This work present HANDMI4, a new open access database of hand motion tracking data, which includes a wide range of static hand grasp configurations and some classic dynamic industry tasks. Such database was generated using two of the most mature technologies for motion capture: IMU-based data glove and camera-based triangulation. To test the capability of such dataset to foster AI-based task classifier, a set of machine learning techniques were implemented and tested. In particular, KNN weighted reached 94,4% and 100% of task classification accuracy for the data glove and the camera system, respectively. With this open access database we aim to boost the research around task classification through motion capture technology to enable the next revolution in smart manufacturing.**

***Index Terms*—Industry 4.0, IMU, motion capture, database, AI**

I. Introduction

Collaborative robots (cobots), robots able to safely share the same working space with their human co-workers [1], have intrinsic high task flexibility and scalability with reduced commissioning costs (e.g. minimal training, reduced impact on the existing layout) compared to classic automation. This is particularly attractive to small and medium scale enterprises, which can access smart automation within their investment capability. Moreover, collaborative robotics aims to bring the human operator back to the center of the industrial revolution, where smart systems support and relieve humans in a synergetic and holistic approach. If successful, this revolution will change how automation is employed in industry and in other fields, such as healthcare and households, with creation of new jobs and services as a result of the increased productivity and scalability, while advancing workers' safety, skillset and wellbeing.

To enable this scenario, human-robot collaboration (HRC) technology needs to go beyond the current state of the art. The term collaboration implies "direct interaction among participants by negotiation, discussions, sharing of knowledge and the accommodation of another's perspective" [2]. One of the main challenges for HRC is to create robots that allow a safe co-existence and a natural interaction with humans [3]. That implies that collaborative robots need to have at least a minimum form of autonomy and have the potential to show initiative [4]. However, the implementation of HRC introduces several issues related to safety, communication and task organization, as well as social-related aspects (e.g. trust, social acceptance, technology anxiety, perceived usefulness) [3], [5].

The communication between robot and operator can be attained physically and/or without direct contact [6]. In the first, the robot measures and compares external forces applied by the different agents (including the operator) and changes its speed and cinematic configuration accordingly. In contactless collaborations, there is no intentional physical interaction between robot and operator, and the communication is achieved via gesture, voice commands or human motion interpretation [7]. Several current contactless approaches involve motion tracking and interpretation of hand gesture, which have the double advantage of being a natural form of expression for humans and offering an effective detectable feature for pose detection [8].

This research was funded by Science Foundation Ireland under Grant number 16/RC/3918 (CONFIRM) which is co-funded under the European Regional Development Fund.

This work presents the creation of an open access database of hand motion tracking data generated by two prominent motion capture (MoCap) technologies (data glove and cameras) while performing classic industry 4.0 tasks. Preliminary results on the use of such database to train an AI-based classifier, and the comparison between the two datasets are also presented. The creation of such open access dataset aims to boost the development of real-time task classification strategy based on human motion interpretation. Moreover, has the ambition to foster the creation of new wearable sensing frameworks, whit features (sensor technology, position and number) able to boost wearability and usability while maintaining adequate information extraction performance. This could also better highlight the pros and cons of the use of a one MoCap technology against another, both in terms of data extraction and impact on the manufacturing infrastructure.

II. Materials and Methods

This work presents the creation of an open access database of hand motion tracking data collected from two different MoCap technologies: camera-based system and IMU-based data glove. Both technologies have in fact reached maturity and are currently being implemented in the smart industry ecosystem [9]. Participants have been recorded while performing different hand grasps, as per Cutkosky taxonomy [10], and four different manual handling tasks: pick and place, object inspection, tool assisted assembly and hand assembly. These tasks have been identify as more frequently performed in collaborative robotics operations [5]. Only from the manual handling dataset, motion-specific (e.g. acceleration, joint angles, and jerk) and task-specific (e.g. flexion-extension of the fingers over time) features were further extracted and labelled for AI training. The final database, called HANDMI4 (HAND Motion capture data for Industry 4.0), is openly accessible and usable under common attribution license (CC BY 4.0).

A. Participants and ethics statement

A total of 20 subjects (12 male and 8 female) were recruited for the data collection, with age ranges from 20-30 to 50-60 years. Before commencing the trial, a comprehensive briefing of the nature of the study was given, followed by the opportunity to confirm consent. We only retained participant information that was relevant to the interpretation of the data: age, sex (M/F) and dominant hand. No information that can directly identify participants was included in the data collection, in accordance with the General Data Protection Regulation 2018 of the EU. Ethics authorization from the University College Cork Ethics Committee was secured before starting the data acquisition, as per the law in force.

TABLE I
Demographic Information

	Demographic Information
Participants	20 Subjects
Age Range	30% in 20-29 65% in 30-39 5% in 50-59
Sex	60% Male 40% Female
Dominant Hand	90% Right-Handed and 10% Left-Handed

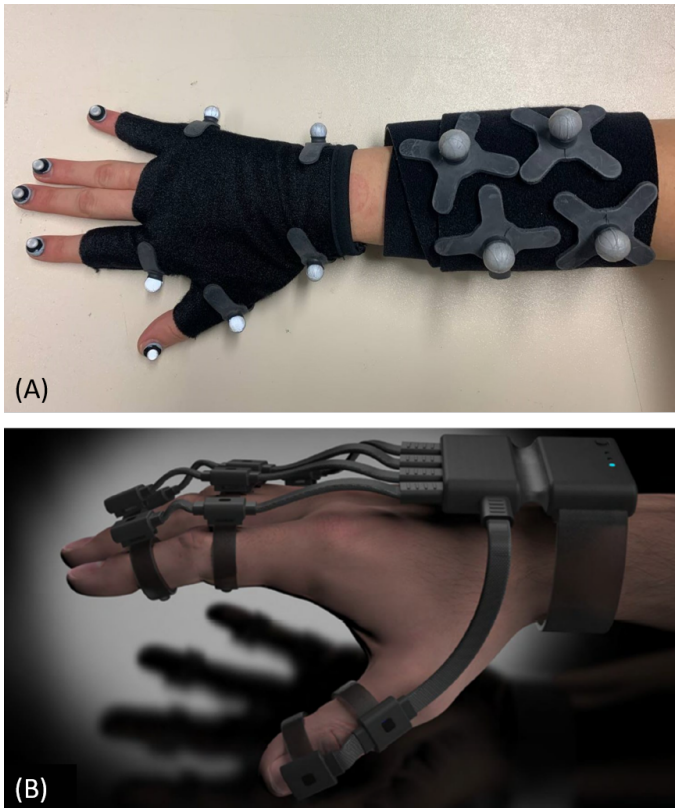

Fig. 1. (A) Hand with the 10 reflective markers and 4 markers for the forearm. (B) Tyndall Smart Glove (Tyndall National Institute, CO, IE).

B. Motion Capture Technologies

A 12-camera system from Optitrack (NaturalPoint, Inc., OR, US), composed by 10 PrimeX13 and 2 PrimeX41, was selected for the camera-based MoCap technology. The hand model was captured using a 10-markers model provided by the Optitrack motion capture software Motive. In such model, 5 passive infrared (IR) reflective markers of $9\,\mathrm{mm}$ are placed on the carpal section of the hand, mounted on a black non-reflective glove to help wearability and contrast, and 5 passive IR reflective markers of $6\,\mathrm{mm}$ placed on top of each nail (Figure 1,A). Smaller size markers on the finger's top help to prevent markers overlap during the recording, especially for fine manipulation tasks. The sampling rate for the data capture was set to $240\,\mathrm{Hz}$ and exposure time to $20\,\mu\mathrm{m}$. The forearm was included in the data capture as rigid body, by placing 4 markers of $14\,\mathrm{mm}$ on an elastic Velcro strap and wrapped around the subject's forearm.

As for the data glove, the Tyndall smart glove, mounting twelve 9-axis IMUs (inertial measurement units), each with a 3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer, was chosen [11]. The glove has an open structure with one IMU located at each middle and proximal phalangeal segment (distal and proximal for the thumb) and one at the palm

and wrist section. Such configuration allows the evaluation of the range of motion of the wrist, the metacarpophalangeal joints and the proximal interphalangeal joints (Figure 1, B). The IMU (MPU-9250) sensitivity of the tri-axis angular rate sensor (gyroscope) was set to a scale range of ±2000dps, the tri-axis accelerometer to ±4 g, and the tri-axis compass to ±4800 µT. In this work, we decided to use only three fingers (thumb, index, and middle) out of the five available from the glove. This to improve the device wearability while performing the grasps and fine manipulations.

C. Cutkosky Grasp Taxonomy Task

The Cutkosky Grasp Taxonomy was proposed by Mark Cutkosky in the 1990s and has became a widely used reference in the field of robotics and HRC [10]. It is used to describe the different ways in which objects can be held by the hand, either human or robotic. The categorization of grasps, shown in Fig.2, is based on the number of fingers used and their position relative to the object being grasped.

D. Industry 4.0 Tasks

The selected four manual handling tasks that simulate different activities involved in the Industry 4.0 are:

1) *Pick and place*: two boxes are placed in front of the subject on the table, one on the left and the other on the right. The subject is asked to move 4 objects one by one from the left box to the other box.
2) *Object inspection*: a cubic object is placed on the table in front of the subject. The cube presents different shapes engraved on each side. The subject is asked to take the cube inspect it and count how many different shapes are engraved on the cube.
3) *Hand assembly*: a seven pieces puzzle is placed on the table in front of the subject. The subject is asked to put together the puzzle.
4) *Tool assisted assembly*: the subject is asked to bolt together two rectangular objects using bolts and nuts, and to taking it apart again. The task is performed with the aid of a screwdriver and a wrench.

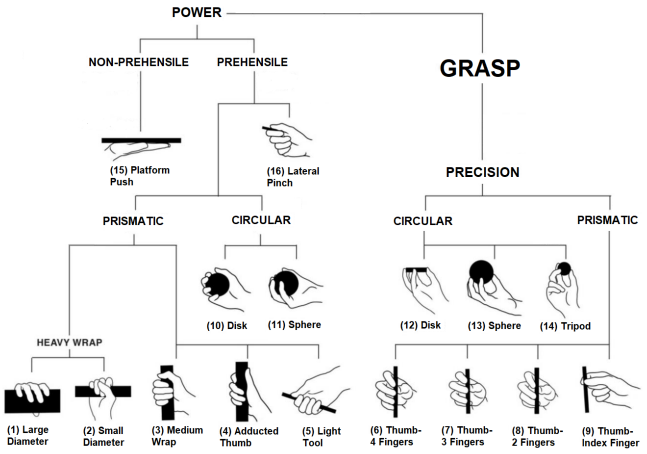

Fig. 2. Cutkosky Grasp Taxonomy Task.

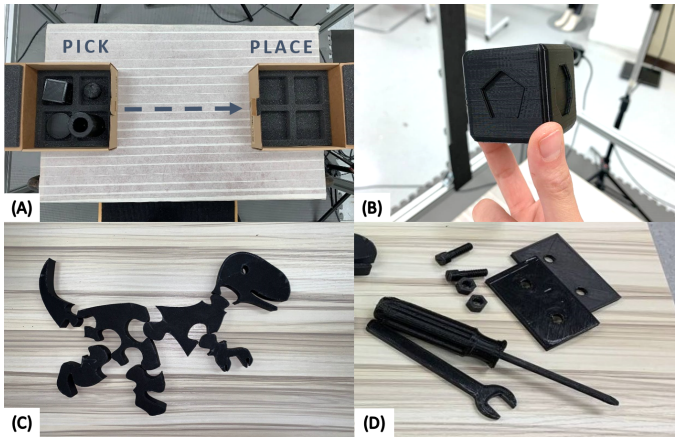

Fig. 3. (A) The two boxes used for pick-and-place.(B) The cubic object to observe during object inspection (C) A seven-piece puzzle to be assembled during the hand assembly task. (D) Assembly components and tools for the last task.

Both for the Cutkosky grasps and for the industry 4.0 tasks, each task is repeated 3 times.

The objects used in the industry 4.0 tasks were all custom designed via computer aided design software and 3D printed in black polylactic acid (PLA) using fuse deposition modelling 3D printing technology. Even the tools (screwdriver and wrench) were 3D printed. This methodological choice was adopted to improve the motion tracking capabilities of both MoCap technologies. In fact, the black rough surface of the 3D printed objects compared to metallic surface minimises light reflection, which could be misinterpreted as passive markers for the camera system. In addition, the use of non-ferromagnetic objects for the manual handling task reduces interferences with the correct functioning of the magnetometers inside the IMU, which are used to evaluate the pose (orientation in the 3D space) of each IMU. An overview of the objects used during the acquisition is given in Fig.3.

E. Data Acquisition Workflow

The participant is informed about the nature of the study and instructed on the different tasks. Any question is answered before informed consent is given. Starting with the camera-based system, the reflective markers are placed on the subject's hand and forearm followed by the actual data acquisition (starting from the taxonomy ending with the industry tasks for a total of 20 tasks). Once all the tasks have been recorded, the markers are removed, and the glove is worn to repeat all the tasks a second time. In both cases, the subject was aided for the task order and hand configuration by the mean of a visual representation of the task (images on a monitor mimicking the task).

F. Feature Extraction

To test the capability of such database for industry task classification, several features have been evaluated from the

industry section of the HANDMI4. Starting from the data provided by the data glove, the quaternions per each hand segment under observation was evaluated using the algorithm described in [12]. As for the data coming from the camera system, the quaternions were already provided. Using the quaternions the joint angle of the proximal interphalangeal, the metacarpophalangeal joints and the wrist were evaluated per each degree of freedom of such joint. From them, some statistical features and one task specific feature were extracted over a time window of 5 seconds of acquisition. The statistical features selected for this work are: mean value, standard deviation, median, interquartile range, covariance, entropy, shape factor. The task specific feature are: peak to peak value and peak count (numbers of peaks that exceed 70% of the maximum value of the joint angle over the selected time window). Such values were then collected in a new dataset of features, where each column represents a feature, with last column left for the label of the industrial task being performed.

Such values were then collected in a new dataset of features, where each column represent a feature and the last column contains the label of the industrial task being performed.

G. AI-based Task Classification

Different supervised machine learning techniques were tested as task classifiers:

- **K-Nearest Neighbors (KNN)**: is a supervised machine learning technique for classification and regression. In Weighted KNN, a variation of the classic KNN algorithm, the voting of the nearest neighbors is weighted based on a certain metric, such as distance. The weights indicate that closer neighbors should have a greater influence on the final prediction than farther neighbors.
- **Bagged Trees**: also known as Bootstrapped Aggregation, is a ML ensemble algorithm that trains several decision trees on random samples of training data. Combining the predictions of these trees yields the final forecast, which reduces variance and instability. As a result, accuracy improves and overfitting is reduced.
- **Support Vector Machine (SVM)**: is a supervised machine learning algorithm, in the case of classification, SVM finds the best decision boundary that separates the data into different classes. The "Medium Gaussian" SVM refers to the use of a Gaussian radial basis function kernel as the similarity measure between two data points, where a medium value of the width parameter (gamma) provides a balance between underfitting and overfitting.

To avoid overfitting and improve the robustness of the approaches used, it was chosen to train the networks using the entire dataset with the 'k-fold cross validation' technique. The goal of k-fold cross-validation is to split the available data into "k" folds, or parts, and train the model on k-1 folds while evaluating it on the remaining fold. This process is repeated k times, with each fold serving as the validation set once. The results of the k evaluations are then averaged to produce a single performance metric, such as accuracy, precision, recall, or F1-score.

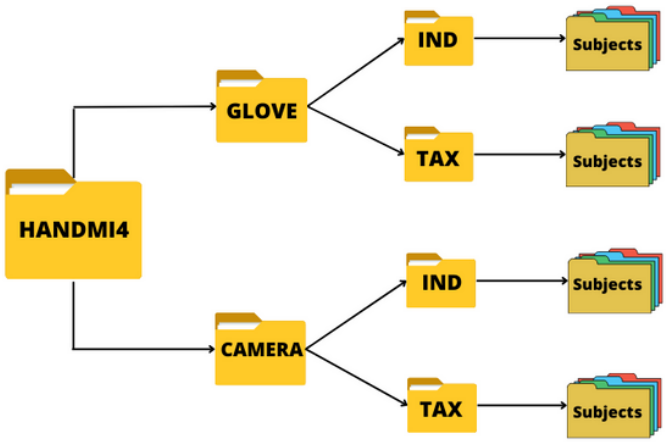

Fig. 4. Folders' organization in the HANDMI4.

III. Results and Discussion

A. HANDMI4 Dataset

The resulting database, called HANDMI4, is a collection of comma separated value (CSV) files, containing either hand's segments positions in the 3D space, for the camera-based system, or hand's segments poses in the 3D space, for the data glove. The folder tree structure, shown in Fig. 4, divides the database according to the MoCap technique used and by the type of task. A coded labelling system has been implemented to improve data navigation and consistency. Each CSV name is composed by an alphanumeric label, which represents the subject number, the MoCap technique, the task number and the task repetition number (Table II). As an example, in the label 000GLVTAX0201: 000 identifies the first participant, GLV stands for data glove, TAX refers to the Cutkosky's taxonomy, 02 refers to the second grasp of the Cutkosky's taxonomy, 01 means that this is the second repetition.

TABLE II
File's Labelling Code

Label	Code	Description
ID	000	3-digit numeric participant identifier
Technology	GLV or OPT	3 letters which refer to the motion capture technology used
Task	TAX or IND	3 letters which refer to task performed
i-Task	00	2-digit numeric for the type of task
j-Repetition	00	2-digit numeric for the number of repetition (starting form 00)

The database includes text file with instruction on how to handle and navigate the data. Similar metadata are included in the main folders of the database in case of partial download. This database represents a valuable tool for researchers involved in AI-based task classification and biomechanics. As shown by making the HANDMI4 openly accessible and providing detailed instructions on how the data have been collected and organised we hope to stimulate other researchers to implement and enlarge the dataset.

B. Industry Tasks Classification

The task classification performance of the best 3 classifiers are reported in Table III and Table IV. Overall, the SVM technique performed better for both MoCap technology with accuracy over the test dataset reaching 86.5% and 90.9% for the data glove and the camera system, respectively. Being preliminary results, it was decided not to change the hyper-parameters of the networks in order to show as neutral results as possible.

TABLE III
PERFORMANCE OF CAMERA DATASET

Method	Accuracy_Test	Training time	Prediction Speed
SVM	90.9 %	7.32 s	∼1500 obs/s
WNN	89.2 %	30.1 s	∼3600 obs/s
Bagged Trees	87.1 %	35.3 s	∼1300 obs/s

TABLE IV
PERFORMANCE OF GLOVE DATASET

Method	Accuracy_Test	Training time	Prediction Speed
SVM	86.5 %	28.2 s	∼1500 obs/s
WNN	77.0 %	70.6 s	∼85 obs/s
Bagged Trees	76.9 %	4.42 s	∼13000 obs/s

It was also highlighted that the dataset's lack of left-handers affected overall performance. In fact, employing cross-validation helped us to acquire consistent performance with the dataset and hence spread the borderline cases by obtaining averaged performance.

IV. LIMITATIONS AND FUTURE WORKS

As the data collection protocol is now tested and made available, the next step would involve the inclusion of more subjects, with a focus on increasing the number of left-handers. Although the number of the initial cohort (20 subjects) is in line with similar datasets [14]–[16], by increasing the subjects and their variability (e.g. age range, dominant hand) any task classification algorithm can be improved. The number of cameras of the camera-based system was found to generate some limitations in the tracking capabilities, especially during movements where some hand segments overlap or go behind some of the objects. To mitigate such problem, increasing the number of cameras distributed around the tracking volume and reducing any obstruction (e.g. reducing the table size) could offer some advantages. Furthermore, the use of smaller markers can reduce overlapping and increase wearability and ergonomics for the subject. Smaller markers might also enable the creation of single rigid bodies per each phalangeal segment, eliminating the needs for an hand model, like the one used in this work. In fact such model operates a data reconstruction based on an anatomical model which has an impact on the accuracy of the actual 3D position of each hand segment. As for the data glove, the hardware design was found to have a critical role in ensuring an adequate adhesion of the sensors to the hand's segment and wearability. In the current design, these two elements are in opposition (high adherence-low wearability and vice versa). Different data glove design will be tested in future steps. In conclusion, with this new open access database of hand motion capture data we aim to provide the research field with a development tool for task classification and biomechanical modelling. The potential of HANDMI4 as data training and testing for AI-based task classification was also tested and presented in this work, encouraging the researchers for further usage and optimization.

REFERENCES

[1] ISO/TS 15066:2016: Robots and robotic devices – collaborative robots. Standard ISO/TS 15066:2016, International Organization for Standardization, Geneva, CH (2016). https://www.iso.org/ standard/62996.html

[2] O. Kozar, "Towards Better Group Work: Seeing the Difference between Cooperation and Collaboration," in English Teaching Forum, 2010, vol. 48, no. 2, pp. 16-23: ERIC.

[3] B. Chandrasekaran and J. M. Conrad, "Human-robot collaboration: A survey," presented at the Southeastcon 2015, Fort Lauderdale, FL, USA, 2015.

[4] R. Gervasi, L. Mastrogiacomo, and F. Franceschini, "A conceptual framework to evaluate human-robot collaboration," The International Journal of Advanced Manufacturing Technology, vol. 108, no. 3, pp. 841-865, 2020.

[5] K. Ikuta, H. Ishii, and M. Nokata, "Safety Evaluation Method of Design and Control for Human-Care Robots," The International Journal of Robotics Research, vol. 22, no. 5, pp. 281-297, 2016.

[6] Mukherjee, D., Gupta, K., Chang, L.H. and Najjaran, H., 2022. A survey of robot learning strategies for human-robot collaboration in industrial settings. Robotics and Computer-Integrated Manufacturing, 73, p.102231.

[7] A. De Luca and F. Flacco, "Integrated control for pHRI: collision avoidance, detection, reaction and collaboration," in Proceedings of the 2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob), Jun. 2012, pp. 288–295, doi: 10.1109/BioRob.2012.6290917

[8] Ziaeefard, M. and Bergevin, R., 2015. Semantic human activity recognition: A literature review. Pattern Recognition, 48(8), pp.2329-2345.

[9] Menolotto, M., Komaris, D.S., Tedesco, S., O'Flynn, B. and Walsh, M., 2020. Motion capture technology in industrial applications: A systematic review. Sensors, 20(19), p.5687.

[10] Cutkosky, M. R., and Howe, R. D. (1990). Human grasp choice and robotic grasp analysis. Dextrous robot hands, 5-31.

[11] Wilk, M.P., Torres-Sanchez, J., Tedesco, S. and O'Flynn, B., 2018, August. Wearable human computer interface for control within immersive VAMR gaming environments using data glove and hand gestures. In 2018 IEEE Games, Entertainment, Media Conference (GEM) (pp. 1-9). IEEE.

[12] O'Flynn, B., Sachez-Torres, J., Tedesco, S., Downes, B., Connolly, J., Condell, J. and Curran, K., 2015. Novel smart glove technology as a biomechanical monitoring tool. Sensors and Transducers, 193(10), pp.23-32.

[13] F. Semeraro, A. Griffiths, and A. Cangelosi, "Human–robot collaboration and machine learning: A systematic review of recent research," Robotics and Computer-Integrated Manufacturing, vol. 79, p. 102432, 2023.

[14] Sun, X., Wei, Y., Liang, S., Tang, X. and Sun, J., 2015. Cascaded hand pose regression. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 824-832).

[15] Tang, D., Chang, H.J., Tejani, A. and Kim, T.K., 2016. Latent regression forest: structured estimation of 3d hand poses. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(7), pp.1374-1387.

[16] Tompson, J., Stein, M., Lecun, Y. and Perlin, K., 2014. Real-time continuous pose recovery of human hands using convolutional networks. ACM Transactions on Graphics (ToG), 33(5), pp.1-10.