Advancing Human-robot Interaction in a Manufacturing Environment by Incorporating Hand Movements for Remote Control

Rim Slama, Ilhem Slama, Pierre Slangen, Oussama Ben-Ammar

Advancing Human-robot Interaction in a Manufacturing Environment by Incorporating Hand Movements for Remote Control

Rim Slama
CESI LINEACT
Villeurbanne, France
rsalmi@cesi.fr

Ilhem Slama
Université de Technologie de Troyes, LIST3N
Troyes, France
ilhem.slama@utt.fr

Pierre Slangen
EuroMov Digital Health in Motion
Univ Montpellier
IMT Mines Ales
Ales, France
pierre.slangen@mines-ales.fr

Oussama Ben-Ammar
EuroMov Digital Health in Motion
Univ Montpellier
IMT Mines Ales
Ales, France
oussama.ben-ammar@mines-ales.fr

***Abstract*—Recently, there has been a notable shift in manual assembly tasks being replaced by robots. These robots are controlled by operators using keyboards. To enhance this practice, we have introduced a remote human-robot collaboration system. This system allows operators to remotely control robots using hand gestures, utilizing an online hand gesture recognition system. A model-driven display system has been implemented to complement the remote robot control system.**

For data acquisition, we have carefully selected Leap Motion technology. The designed approach for remote hand gesture recognition has been implemented and tested for accuracy and response time using the well-known SHREC'21 Gesture Benchmark. The final analysis demonstrates the significant potential of the developed system in improving assembly processes and environments. It optimizes various key performance indicators, leading to enhanced efficiency and effectiveness.

***Index Terms*—Robot remote control, hand gesture recognition, human-robot collaboration, optimization, deep learning.**

I. Introduction

The advent of Industry 5.0 has brought forth a new era of manufacturing, where cutting-edge technologies seamlessly integrate with human expertise. With a focus on harnessing the power of ergonomics and enhancing productivity, one promising avenue lies in developing a manufacturing optimization system through remote robot control [1]. By leveraging hand movements as the interface, this innovative approach enables operators to intuitively guide robots from a distance, streamlining processes and reducing physical strain. The synergy between industry 5.0 principles, ergonomic design, and remote robot control holds immense potential for revolutionizing the manufacturing landscape, empowering workers, and unleashing unprecedented levels of efficiency [2], [3].

Manufacturing systems traditionally rely on human operators to perform various tasks [4]. However, robots can potentially enhance efficiency and optimize productivity by replacing specific tasks, such as tool retrieval or returning, with robots [5]. This substitution not only saves time but also streamlines the overall manufacturing process. Presently, operators control robots using keyboards, which may not provide the most intuitive and efficient means of interaction [6]. To address this limitation, we propose investigating the feasibility of controlling robots through immaterial means, specifically by utilizing hand gesture recognition. Today, practically all industrial robots and electrical equipment use control stations as their interface. Controllers come with many issues, including difficulties replacing them when they are damaged, confusion over the various button layouts, time wasted seeking the correct remote, and more.

One of the earliest and most common forms of human interaction has been through hand gestures. It is regarded as a form of interaction that might offer a more agile, imaginative, and organic way for operators and robots to communicate. A speech recognition system can be used to collaborate and control the robots. However, obtaining an accurate system in our manufacturing context will be difficult since such environments are generally crowded. Therefore, remote control through gesture recognition is particularly relevant in this context [7]. To this end, this paper aims to present a practical transformer-based approach to remote robot control in a manufacturing system.

For this purpose, in this paper (as shown in Fig. 1), a hand gesture recognition system based on the most recent deep learning time series architectures is developed and tested on the gesture benchmark SHREC'21 [8]. Relied on this system, we discussed the benefit of this remote control from different points of view. Our developed recognition system has real-time answers and high precision accuracy, making it reliable for the operator in such a sensitive industrial environment.

The main contributions of this paper are the following:

- Proposing a study for remote robot control, using hand gesture recognition in a manufacturing environment, starting from the selection of technologies, the choice of recognition architecture, the choice of the database, and gestures to learn, to the case study and configuration.
- Introducing an online system for hand gesture recognition based on an accurate neural network architecture;

Fig. 1. The proposed System for remote hand gesture control.

- Both static and dynamic gestures are considered in the proposed recognition system;
- Demonstrating the benefits of using the proposed system in an assembly line configuration.

The rest of the paper is structured as follows: section II gives a brief literature review on robot remote control in a manufacturing environment. Section III introduces the developed vision-based hand gesture recognition system and details some preliminary results. Section IV presents a case study. Finally, a conclusion and perspectives are drawn in section V.

II. Literature review

Remote robot manipulation has recently seen a massive increase in use across several industries. Numerous remote robot manipulation techniques have been used in remote surgery [9]. The TouchMe system, an AR-based framework for remote robot operation, was introduced in [10]. A touch screen serves as the human-side control interface. The user can control the moving robot and its manipulator from a third-person perspective.

Remote robot manipulation has also been incorporated into the robotic assembly. A framework for the remote robotic assembly was presented in [11], where the real-time robot control and monitoring were accomplished using the remote robot manipulation technique. In [7], the authors proposed a remote human-robot collaborative system that can be applied in a hazardous manufacturing environment. The proposed system can provide control tasks from one robot to another. Through two test cases and discussions, the authors indicate an excellent potential for applying the system in different industries. Recently, in [12], a focus was placed on machine learning techniques to make a sensorless robot able to learn and optimize an industrial assembly task. Kung and Chou [13] addressed the cooperation problem between machines and operators using human gesture recognition.

A growing body of literature in the computer vision field investigated the skeleton-based hand gesture recognition research topic and has undergone substantial study, particularly with the advent of deep learning [14]. This inspired the design of numerous advanced skeleton-based methodologies [15]. In particular, there has been a noticeable shift towards employing graph-based representations to capture spatiotemporal information effectively. This approach has gained traction due to its compatibility with the topological structure of the skeleton, which aligns well with the nature of hand gestures. Graph-based convolutions have been widely adopted in this context, as they respect the underlying skeleton's topological characteristics [16], [17], [18]. Furthermore, the rise of transformer models in time series analysis has piqued researchers' interest in exploring their potential combination with graph-based convolutions for improved hand gesture recognition accuracy [19].

This paper studies the potential use of remote hand gesture recognition in human-robot collaboration tasks. Firstly, we introduce a skeleton-based approach that uses Graph Convolutional Networks (GCNs) and a deep learning transformer model. Then, we investigate the impact of integrating remote robot control tasks in a given assembly system scenario.

III. Remote hand gesture recognition method

A. Data representation

Hand gesture data can be acquired by different methods such as marker-based and markerless techniques [4], [20]. In conventional optical-based motion capture, the subject can wear retro-reflective markers to accurately extract the position of the markers, including finger movement (see Fig. 2). Nevertheless, although 3D displacements can be played and transferred in real-time from the MOCAP system, this configuration is quite expensive as it requires at least 6 to 8 cameras to cover the whole scene, coupled with professional labeling and processing software.

Alternatively, markerless systems such as RGB cameras can also provide the displacements of hands and fingers. The process may rely on MediaPipe [21], [22], an open-source framework to detect in real-time the hand key points (x,y) (see Fig. 3).

If this solution is quite simple to implement and ensures real-time results, the relative accuracy of the 3D location of the hand points is not sufficient for our kind of application. Conversely, a new system based on mocap gloves (e.g., from StetchSense Inc) can reach this accuracy, but the system's price ($\sim 10k$€) is too heavy to support. The main goal of the developed set-up is to obtain the accurate coordinates of the hand in real-time, in a dedicated space, i.e., the vicinity of the remote operator.

Hence, this study presumes that implementing a Leap Motion sensor is suitable for capturing hand gestures during robot remote control in manufacturing. Priced at approximately $80, this sensor delivers commendable performance, boasting a frame rate ranging from 50 to 200 frames per second. It also offers exceptional precision, enabling real-time tracking of hands and fingers in a three-dimensional realm with an accuracy of 0.01mm. Furthermore, it occupies minimal space, measuring only 80 x 30 x 12.7mm [23].

The sensor possesses a significant advantage in eliminating the need for wearing. Users place their hands in the designated acquisition zone above the sensor to use it. Through the ultrasound response detected from the fingers, the sensor captures detailed three-dimensional information about the hand (See Fig.4).

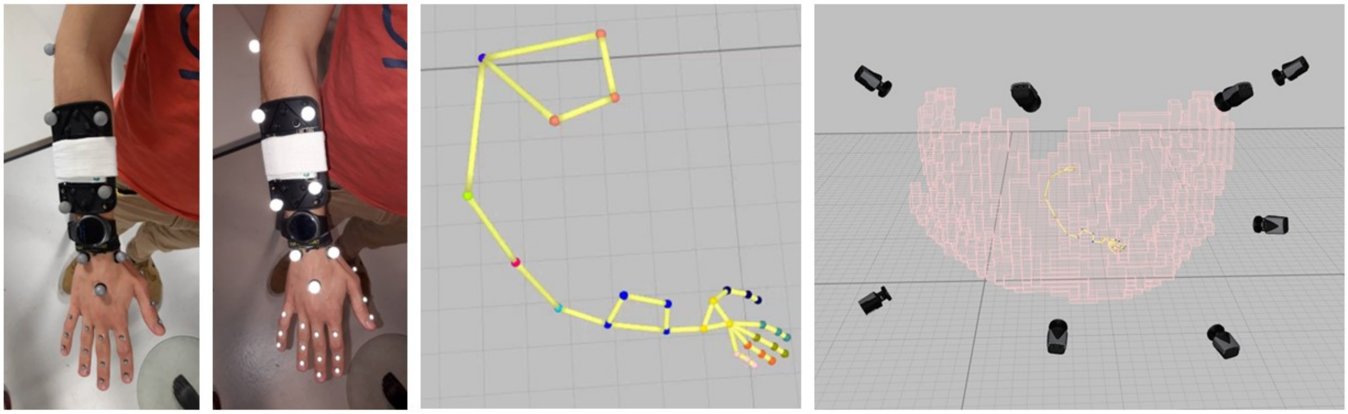

Fig. 2. Example of retroreflective markerset for hand and arm tracking.

Fig. 3. Points of interest for the hands: real-time detection by MediaPipe engine [20].

Fig. 4. Example of hands representation by Leap Motion sensor (source: Ultraleap).

Data obtained from this sensor comprises a hand skeleton composed of 23 joints for each hand. Throughout the skeletal tracking process, the three-dimensional position of each of these 23 joints is transmitted at a frequency of 15 frames per second.

B. Skeleton-based Recognition system method

This preliminary work introduces a graph-based approach that combines (I) a Spatial Graph Convolutional Network (S-GCN) [24] to represent the spatial interdependence among the joints and (ii) a Transformer architecture [25] to represent the temporal connection.

Thus, we can exploit a fully graph-based approach within the hand skeleton sequences acquired by the Leap Motion device (see Fig. 4).

Fig. 5 shows the architecture of our recognition system. We create a graph sequence using an adjacency matrix and a 3D hand skeletons sequence. Firstly, the S-GCN technique is utilized to extract hand features from each frame by leveraging the inherent graph structure of the hand skeleton in the spatial domain.

The spatial graph convolution operation performs a weighted average aggregation of node features, incorporating the current node's features and neighboring nodes within the same frame. This process generates a new feature vector for each node, encompassing information about the node itself, its neighbors, and the strength of their connections within the hand structure.

Subsequently, we capture relevant inter-frame information in the temporal domain using a transformer architecture.

Finally, the transformer's output undergoes a global pooling operation before inputting into our classifier. The encoder block described in [25] is the foundation for this module.

In the context of an online hand gesture dataset, which differs from offline scenarios, the gesture samples are not pre-segmented. The provided sequences are continuous and consist of sub-sequences containing both gestures and instances of "No gesture."

To address this, we employ the widely used sliding window approach. At each time step, denoted as t, we extract a small window from the sequence, spanning the interval [t, t+window

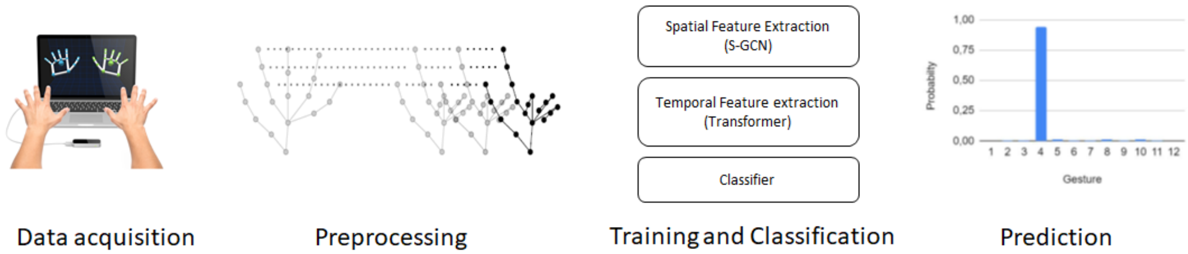

Fig. 5. The architecture of the recognition system.

size]. These windows are sampled with a stride of 5, meaning that windows are selected every five frames. Consequently, we obtain per-frame labeling for each window, allowing for gesture recognition within the window. However, assigning the label with the maximum probability as the prediction for the entire window is not straightforward. This is because the window can contain frames belonging to both gestures and the ”No gesture” class.

A significant class imbalance exacerbates this issue, as most extracted windows belong to the ”No gesture” class and could be classified as false positives. We introduce a prediction filtering system based on probability thresholds to mitigate this.

Windows predicted to belong to a gesture class C, but having a probability score $P(C)$ below a threshold $\alpha(C)$, are deemed false positives and assigned to the ”No gesture” class.

During the validation phase, we learn these probability thresholds. The threshold $\alpha(C)$ for gesture class C is computed as the average probability the classifier estimates for gestures correctly predicted as class C. This calculation only considers cases where the classifier accurately predicts the gesture class, thus excluding false positives from contributing to the average threshold score.

By incorporating this threshold filtering system, the number of false positives is significantly reduced exponentially.

C. Dataset, experimental settings and results for hand gesture recognition

The Skeleton-based Hand Gesture Recognition in the Wild track SHREC’21 [8] dataset is specifically designed as a benchmark for evaluating methods and algorithms that aim to detect and recognize hand gestures in a continuous stream of captured hand joints.

This database encompasses gestures that can be translated into actions that a robot can perform. It includes both static and dynamic gestures with distinct characteristics.

Skeleton sequences are acquired using a Leap Motion devise. The total dataset offers a diverse and extensive gesture dictionary consisting of 18 gesture classes categorized into three categories (see Fig. 6).

To assess the performance of our approach, we evaluate it using the False Positive Rate (FPR) metric. This metric quantifies the ratio of false predictions (FP) that do not correspond to ground truth samples of a specific class, divided by the total number of ground truth samples belonging to that class. Mathematically, the false positive rate can be expressed as follows:$FPR = \frac{FP}{(TP+FN)}$, where TP refers to true positives and FN refers to false negatives.

Our experiments used PyTorch as the framework, with Adam optimizer and Cross-entropy loss function. The chosen hyper-parameters include a batch size of 32, 6 encoders, and eight heads for the multi-head attention mechanism.

In terms of performance, we achieved an impressive false positive rate (FPR) of 0.083%. Additionally, our real-time gesture recognition system provides prompt results, with an average processing time of 0.089 seconds for each window size.

IV. Hand Gesture Remote Control in an Assembly Line

In the following, we study the possibility of applying gesture recognition in the human-robot collaborative manufacturing environment.

We propose as an example of the assembly line shown in Fig. 7 where operators can collaborate with cobots to accomplish specific tasks efficiently. For instance, let’s consider the operator at Station 1, who needs to utilize a particular tool. He has two options to interact with the cobot:

(1) Using a keyboard: The operator can press suitable keys and provide instructions to the cobot regarding the desired actions. These instructions may include commands to locate the tool, grasp it, and bring it to the operator.

(2) Gestural commands: Alternatively, the operator can issue commands to the cobot through a succession of gestures. For example, the operator can use the gesture “point” to indicate the tool’s location, followed by the “grasp” gesture to communicate the action of grasping the tool. The cobot will then execute these gestures accordingly, bringing the tool to the operator.

The feasibility of remotely controlling robots in a collaborative environment is demonstrated in the proposed setting and with the developed approach incorporating various gesture classes. This collaborative interaction yields several benefits:

Ergonomics: The remote control aspect of human-robot collaboration reduces the mental load on operators. They no longer need to comprehend complex instructions for robot control, as they can utilize natural gestures instead. This allows operators to focus on the environment they are interacting

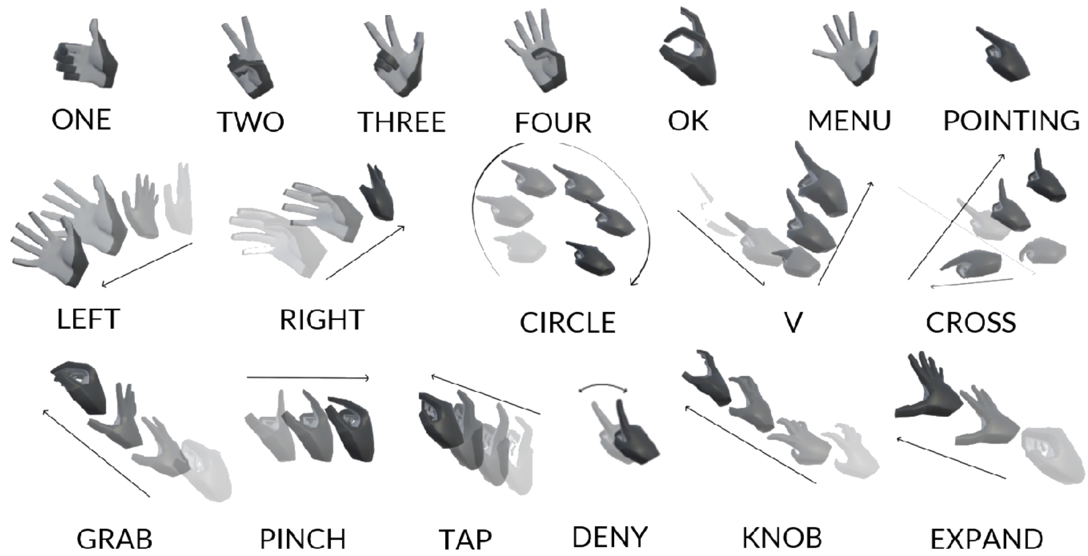

Fig. 6. The illustration of gestures from the SHREC'21 dataset [8].

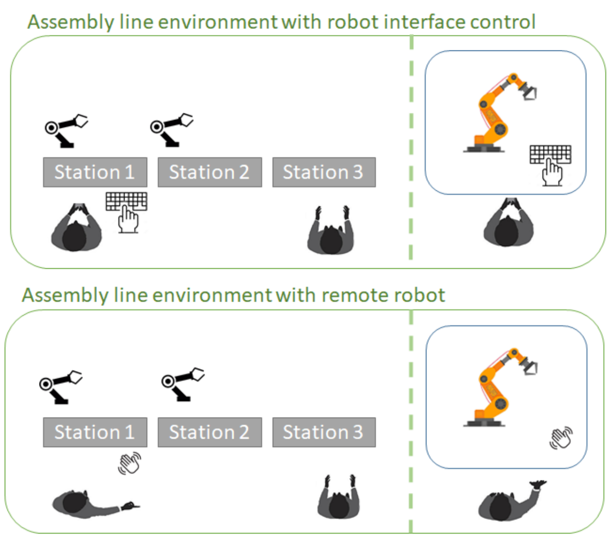

Fig. 7. Illustration of a simple assembly line configuration with two scenarios and a human-robot interaction.

with without needing constant visual attention shifts between the robot and the control unit. This ergonomic improvement enhances operator comfort and reduces the potential for errors.

Enhanced safety: Hand gesture control reduces the need for physical interaction with the robot, minimizing the risk of accidents or injuries. Operators can maintain a safe distance while still effectively controlling the robot's actions, reducing the potential for workplace incidents.

Intuitive and user-friendly interface: Hand gestures provide a natural and intuitive way of interacting with robots. Operators can quickly learn and adapt to the gesture-based control system, eliminating the need for extensive training or specialized technical knowledge. This user-friendly interface facilitates the adoption of robotic technology in manufacturing systems.

Rapid task switching: Hand gesture control enables operators to switch between tasks or robot actions seamlessly. With a simple change in gestures, operators can instruct the robot to perform a new action or move to a different location, allowing for efficient task switching and increased productivity.

Improved collaboration: Hand gesture control promotes a collaborative environment between operators and robots. Using gestures to communicate instructions, operators can easily convey their intentions, leading to better understanding and coordination between human and robotic workers. This collaboration enables efficient task allocation, where robots can assist operators in performing complex or physically demanding tasks.

Reduced equipment costs: Hand gesture control eliminates the need for additional hardware or complex control panels. Operators can control robots using their hand movements, minimizing the cost and maintenance associated with specialized control devices. This cost reduction makes robotic systems more accessible and financially viable for manufacturing businesses.

Adaptability to diverse tasks: Hand gesture control offers flexibility in controlling a wide range of robotic applications. Operators can employ different gestures to instruct robots in tasks such as picking and placing objects, assembly operations, or machine maintenance. This adaptability allows for the integration of robotic systems into diverse manufacturing processes.

Flexibility: Remote human-robot collaboration offers increased flexibility in task execution. Establishing effective communication channels between humans and robots makes it possible to allocate heavy or strenuous tasks to robots, thereby relieving human workers. This delegation of tasks results in better overall performance, higher speed production, enhanced repeatability, and improved productivity.

V. Conclusion

This paper proposes and studies an online hand gesture recognition system as part of a manufacturing system. A dual convolution graph architecture and attention mechanism are adopted for hand gesture classification. The sliding windows technique proves its efficiency for online recognition. Tasks commanded remotely by the operator provide a significant benefit in terms of productivity and facilitate human-robot collaboration. The impact of using hand gestures as an alternative in an assembly line is studied.

In future works, we propose to develop a user-friendly graphical interface for the proposed framework. We will also investigate the use of the digital twin concept coupled with our system. Finally, we intend to deploy and apply our system to real industrial scenarios such as disassembly.

References

[1] R. Jahanmahin, S. Masoud, J. Rickli, and A. Djuric, "Human-robot interactions in manufacturing: A survey of human behavior modeling," *Robotics and Computer-Integrated Manufacturing*, vol. 78, p. 102404, 2022.

[2] E. Coronado, T. Kiyokawa, G. A. G. Ricardez, I. G. Ramirez-Alpizar, G. Venture, and N. Yamanobe, "Evaluating quality in human-robot interaction: A systematic search and classification of performance and human-centered factors, measures and metrics towards an industry 5.0," *Journal of Manufacturing Systems*, vol. 63, pp. 392–410, 2022.

[3] S. Panagou, W. P. Neumann, and F. Fruggiero, "A scoping review of human robot interaction research towards industry 5.0 human-centric workplaces," *International Journal of Production Research*, pp. 1–17, 2023.

[4] R. Slama, O. Ben-Ammar, H. Tlahig, I. Slama, and P. Slangen, "Human-centred assembly and disassembly systems: a survey on technologies, ergonomic, productivity and optimisation," *IFAC-PapersOnLine*, vol. 55, no. 10, pp. 1722–1727, 2022.

[5] J. Hughes, K. Gilday, L. Scimeca, S. Garg, and F. Iida, "Flexible, adaptive industrial assembly: driving innovation through competition: Flexible manufacturing," *Intelligent Service Robotics*, vol. 13, pp. 169–178, 2020.

[6] H. Fang, J. Wen, X. Yang, P. Wang, and Y. Li, "Assisted human-robot interaction for industry application based augmented reality," in *Virtual, Augmented and Mixed Reality: Applications in Education, Aviation and Industry: 14th International Conference, VAMR 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part II*, pp. 291–301, Springer, 2022.

[7] H. Liu and L. Wang, "Remote human–robot collaboration: A cyber–physical system application for hazard manufacturing environment," *Journal of manufacturing systems*, vol. 54, pp. 24–34, 2020.

[8] A. Caputo, A. Giachetti, S. Soso, D. Pintani, A. D'Eusanio, S. Pini, G. Borghi, A. Simoni, R. Vezzani, R. Cucchiara, *et al.*, "Shrec 2021: Skeleton-based hand gesture recognition in the wild," *Computers & Graphics*, vol. 99, pp. 201–211, 2021.

[9] K. E. Kaplan, K. A. Nichols, and A. M. Okamura, "Toward human-robot collaboration in surgery: performance assessment of human and robotic agents in an inclusion segmentation task," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 723–729, IEEE, 2016.

[10] S. Hashimoto, A. Ishida, M. Inami, and T. Igarashi, "Touchme: An augmented reality based remote robot manipulation," in *The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011*, vol. 2, 2011.

[11] L. Wang, A. Mohammed, and M. Onori, "Remote robotic assembly guided by 3d models linking to a real robot," *CIRP annals*, vol. 63, no. 1, pp. 1–4, 2014.

[12] L. Roveda, M. Magni, M. Cantoni, D. Piga, and G. Bucca, "Human–robot collaboration in sensorless assembly task learning enhanced by uncertainties adaptation via bayesian optimization," *Robotics and Autonomous Systems*, vol. 136, p. 103711, 2021.

[13] C. Y. Kung and T.-Y. Chou, "Skeleton based action analysis on manufacturing assembly site," in *Pacific Asian Conference on Information Systems-(PACIS 2022)*, Association for Information Systems (AIS), 2022.

[14] F. Al Farid, N. Hashim, J. Abdullah, M. R. Bhuiyan, W. N. Shahida Mohd Isa, J. Uddin, M. A. Haque, and M. N. Husen, "A structured and methodological review on vision-based hand gesture recognition system," *Journal of Imaging*, vol. 8, no. 6, p. 153, 2022.

[15] S. Sarker, S. Rahman, T. Hossain, S. Faiza Ahmed, L. Jamal, and M. A. R. Ahad, "Skeleton-based activity recognition: preprocessing and approaches," in *Contactless Human Activity Analysis*, pp. 43–81, Springer, 2021.

[16] T. Ahmad, L. Jin, X. Zhang, S. Lai, G. Tang, and L. Lin, "Graph convolutional neural network for human action recognition: A comprehensive survey," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 2, pp. 128–145, 2021.

[17] L. Feng, Y. Zhao, W. Zhao, and J. Tang, "A comparative review of graph convolutional networks for human skeleton-based action recognition," *Artificial Intelligence Review*, pp. 1–31, 2022.

[18] R. Slama, W. Rabah, and H. Wannous, "Str-gcn: Dual spatial graph convolutional network and transformer graph encoder for 3d hand gesture recognition," in *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*, pp. 1–6, IEEE, 2023.

[19] W. Xin, R. Liu, Y. Liu, Y. Chen, W. Yu, and Q. Miao, "Transformer for skeleton-based action recognition: A review of recent advances," *Neurocomputing*, 2023.

[20] H. Bayd, P. Guyot, B. Bardy, and P. R. Slangen, "Automatic scoring of synchronization from fingers motion capture and music beats," in *Image Analysis and Processing. ICIAP 2022 Workshops: ICIAP International Workshops, Lecce, Italy, May 23–27, 2022, Revised Selected Papers, Part I*, pp. 235–245, Springer, 2022.

[21] Z. Kaoudi, J.-A. Quiané-Ruiz, B. Contreras-Rojas, R. Pardo-Meza, A. Troudi, and S. Chawla, "Ml-based cross-platform query optimization," in *2020 IEEE 36th International Conference on Data Engineering (ICDE)*, pp. 1489–1500, IEEE, 2020.

[22] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," *arXiv preprint arXiv:2006.10214*, 2020.

[23] S. Ameur, A. B. Khalifa, and M. S. Bouhlel, "A comprehensive leap motion database for hand gesture recognition," in *2016 7th international conference on sciences of electronics, technologies of information and telecommunications (SETIT)*, pp. 514–519, IEEE, 2016.

[24] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.