A Demonstration-Based Instruction System for Handling Manipulation of Deformable Objects With Overlaying Hand Postures and Contact Forces

Yutaka Takase
Faculty of Engineering

Tomohiro Shintani
Graduate School of Science and Technology

Kimitoshi Yamazaki
Faculty of Engineering

Shinshu University
Wakasato 4-17-1, Nagano, Nagano, Japan
yutaka_takase@shinshu-u.ac.jp

***Abstract*—An effective human resource development system is desirable for mass-production sites, such as sewing factories, which rely on skilled human resources. The manipulation of objects such as clothes, cushions, and stuffed toys during production can deform them, making it difficult to either transfer them directly among people or introduce automation technology to do it. In this study, considering this, we developed a demonstration-based teaching system. The proposed system presents a learner with an instructional video that overlays not only the posture of the skilled operator's hand, but also the contact force. This allows the learner to know the manipulated part of the object, the employed hand posture, and the applied force, even if the occlusion problem occurs. The proposed system does not require learners to wear any auxiliary devices for receiving instructions while learning, making it highly convenient. We implemented the system and confirmed that it could successfully provide instructions for handling the issue of deformable object manipulation.**

***Index Terms*—deformable object manipulation, augmented reality, skill transfer**

I. Introduction

Industrial manufacturing automation has advanced rapidly owing to the use of industrial robots; there are several advantages to mechanized production. For example, when compared with manual production, the worker costs are reduced, and a certain level of product quality is guaranteed. Automation is widely used in the consumer electronics and automobile manufacturing industries, and various automation technologies have been developed accordingly [1], [2], [3]. However, deformable objects such as clothing and stuffed animals are still widely manually manufactured by skilled workers. Performing accurate shape recognition and dealing with the manipulation of deformable objects are both challenging issues. Although automation techniques have been proposed in these areas [4], [5], they have not resulted in the complete elimination of manual work from all the related processes. Digitizing and visualizing the behaviors of skilled operators during the manipulation of deformable objects is a potential way to mitigate this issue.

This study aims to develop a novel demonstration-based training system for visualizing situations during the manipulation of deformable objects. Under the proposed system, it is assumed that the work status of a skilled operator is superimposed on the screen and used as a reference for an unskilled operator to learn the task. Under such an assumption, it is necessary to indicate the posture of the skilled worker's hands during an operation. However, this information is insufficient because the degree of deformation of a flexible object varies depending on the applied force. Considering this, a sensor can be attached to a skilled operator's fingertip for force measurement; preferably, the measurement should not impair the sensation in their fingertip. To achieve this, we introduced a sensor that considered fingernail deformation. It is also important for the system to have a viewing function and interface for training nonskilled operators. In this regard, we are working on the integration of multiple viewpoints and rewind functions to increase the convenience of existing training systems.

This paper is structured as follows. The next section presents the related work. Section III introduces the design goals of the training system. Section IV explains the construction of the proposed training system and the associated operation methods. Section V describes the verification of the system operation and the fundamental evaluation with respect to training for deformable object manipulation. Finally, section VI concludes the paper.

II. Related Work

Most apparel manufacturing facilities rely on human laborers for garment production; the development and assessment of worker performance in these facilities has attracted significant attention [6], [7]. In [6], the authors considered the power consumption of a sewing machine as a feature in assessing the performance and task difficulties of workers. They could adapt the system to a significant number of sewing lines concurrently without needing to recognize complex garment product shapes. In [7], an augmented reality (AR) mobile application was proposed for teaching the way to use sewing machines. This AR application was compared with a traditional learning method using a handout; the results showed that the application could provide a better learning performance, efficiency, and satisfaction in learning experience. The success of this simple method, which employs conventional AR markers and pre-designed teaching videos, indicated the validity of applying immersive technologies to learning systems.

The application of AR and Virtual reality (VR) technologies in the field of task learning has also drawn attention. In [8] and [9], the authors developed systems to teach the assembly task process using these technologies. They used text instructions to indicate the current or subsequent step in the process. It is necessary to recognize the manipulated target objects or assembly tools to present this information appropriately. To provide users with more detailed information, several studies have aimed to recognize the surrounding environment in real time [10], [11]. In [10], the authors attempted to use deep learning technology to detect working-area tools and provide users with multimodal instructions. They reported that the proposed system helped reduce the time and errors in an assembly task. In [11], the authors proposed a method using wearable AR glasses (Microsoft HoloLens) to reconstruct three-dimensional (3D) replicas of the objects and provide 3D spatial guidance by superimposing these replicas onto the real environment. As mentioned previously, in the manufacturing scenario, support from AR or VR technology can potentially improve workers performance. However, in tasks in which deformable objects are manually handled, as previously discussed, applying such technologies for object recognition is difficult owing to the problems of deformation and hand-hand occlusions. Based on the above discussion, in this study, we adopted an approach to apply AR/MR techniques for training without providing an immersive environment or recognizing users' movements. We constructed an interactive system that provided multi-viewpoint instruction videos in which the instructor's hand movements and the degree of applied force were overlaid according to the task process with the aim of improving the overall learning effect.

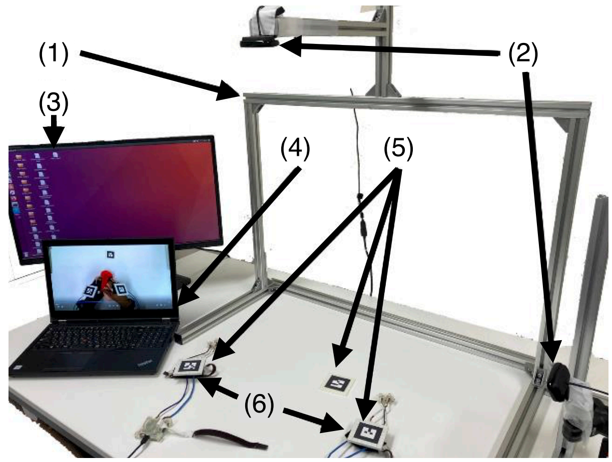


Fig. 1: Overview of the proposed workbench: (1) aluminum frame, (2) web cameras, (3) control server, (4) display, (5) AR markers, and (6) force sensors.

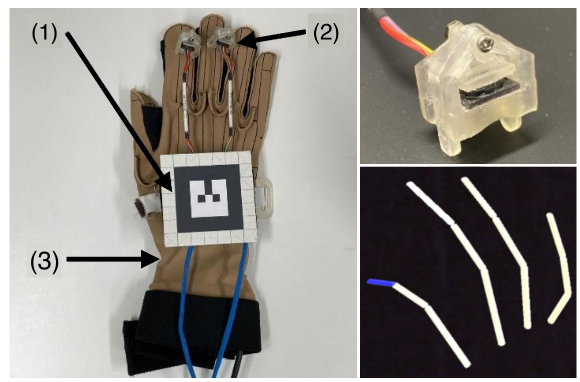


Fig. 2: Left : Overview of the data glove, contact force sensor, and 3D finger model. (1) AR marker, (2) fingertips force sensor, and (3) data glove. Upper right: zoomed-in image of the force sensor. Bottom right: 3D finger model.

III. Design Goals

The following section summarizes the primary features of the proposed demonstration-based instruction system for deformable object manipulation.

1) The system provides instructional videos performed by a skilled instructor in advance. The hand posture that is adopted and the forces exerted on the fingertips by the object during the work are overlaid.
2) The system provides multiperspective views on the instruction videos to reduce the occlusion problem, support depth perception, and observe the details associated with work object deformation. Furthermore, perspectives can be switched whenever the user wishes.
3) The system does not use specific devices and sensors for instructions to reduce the complexity of learning preparation and avoid conflicting with proper practice guidelines.

These features support the use of the system in training an unspecified number of factory workers. Therefore, it is necessary to reduce the burden of learner preparation and the cost of instruction as much as possible. Based on this, we decided to use a general monitor to present instruction videos and avoid measuring or providing feedback about learners' activities. However, skilled instructors should be equipped with sensors to properly measure their operational abilities. Both the measurement method and the parameters to be measured are important considerations. Since the manipulation of deformable objects involves changing the shape of an operator's hand in various ways, the direction of the back of their hand and the joint angles of their fingers are considered to be important information. Additionally, force is often exerted on fingertips to hold them down or induce desirable deformations.

Thus, it is preferable to measure the force applied to the fingertips. In this study, a combination of a data glove and a fingertip force sensor was used. Further details are presented in the next section.

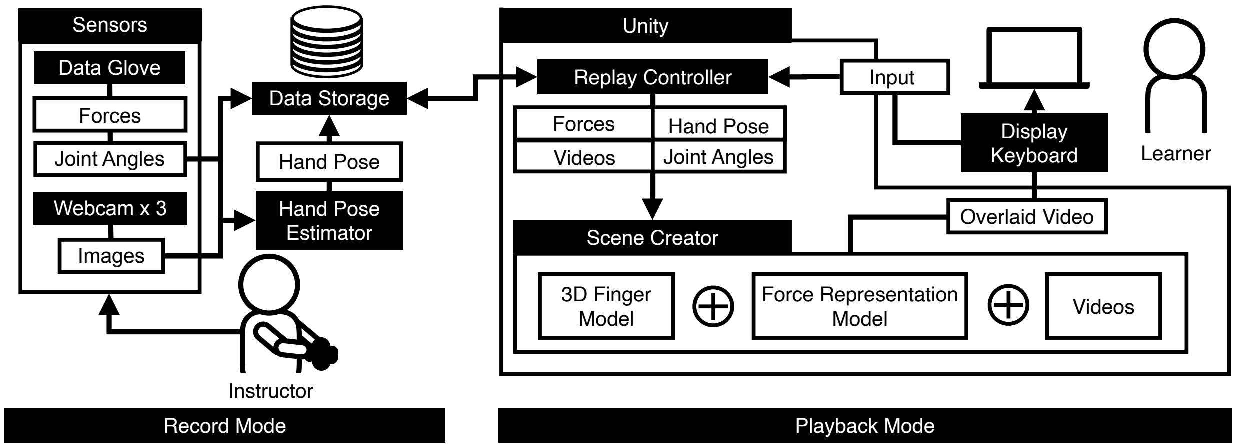


Fig. 3: The Schematic of the proposed system.

IV. Operation Instruction System

A. System Appearance

Fig. 1 shows the proposed instruction system. The cameras are attached to an aluminum frame that can be placed on a desk. An operator's hands can be observed from multiple viewpoints with a maximum of three HD web cameras (Logitech C615n); one is set at the top, and the others are set on either side of the frame. Each camera can record videos with a 1920×1080 resolution at 30 fps. Additionally, AR markers were used; one was placed on the desk and to calibrate the external camera parameters. These markers could precisely calculate the new positions of cameras even after they only moved slightly. Furthermore, two other AR markers are used to collect data from skilled instructors. These markers are placed on the back of a user's hands and calculate the root pose of the hand that is viewed from the camera. In addition to the markers, data gloves and fingertip force sensors are used to measure the work status of skilled instructors. The display, which is placed next to the work area, shows a video of a skilled instructor's work. With the superimposition of the work status on the image, the learner watches the video and performs their work through trial and error. The control server is connected to another computer that monitors the system program, and acquires sensor information, and performs other related operations.

B. Work Status Measurement of Hands

The left side of Fig. 2 shows the data glove (CyberGlove Systems CyberGlove) that is employed to capture the finger joint angles during the task performance. It has 18 sensors with a resolution of less than 1 degree and captures finger joint angles with respect to the wrist. The glove does not cover the fingertips, enabling the instructor to demonstrate the task without losing sensation in them. This is important in the context of the manipulation of deformable objects. The forces applied to the fingertips need to also be measured; we used the sensor proposed in [6]. This sensor can measure the deformation of the cross section of a fingertip when the finger pad is pressed against something; the small arms are set along the side of the fingertip, and when deformation occurs, the displacement is transmitted to a tactile sensor on the nail side. A Shokac Chip, manufactured by Touchence Inc.[1] has been used as a tactile sensor. Small arms and other parts were created using a digital light processing 3D printer. The instructor wore these sensors to capture the hand status. In summary, an instructor can simultaneously record three perspective images, hand poses, the joint angles of each finger, and the forces exerted on the fingertips using this system.

C. System Configuration

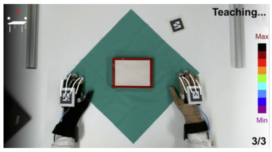
Fig. 3 shows an overview of the instruction system. The system consists of two modes with different functions: the left side of the figure shows the record mode that captures the demonstration by a skilled instructor, and the right side shows the playback mode that produces an interactive instruction video constructed from the recorded data. The black rectangles in Fig. 3, indicate the functional modules. Details of the implementation are described in the following subsections.

1) Record Mode: In the record mode, a skilled instructor's task demonstration is captured using the aforementioned sensors. In the hand pose estimator, the poses of the backs of both hands are estimated from each image that is captured with cameras using the AR markers fixed on the data gloves. Here, the top-view camera frame is defined as the reference coordinate system, and the AR marker frame fixed on the working desk is used to derive the pose of other side-view cameras with respect to the reference coordinate system by using the following transform equations:

$$\mathbf{P}^m = \mathbf{H}_s^m \mathbf{P}^s, \tag{1}$$
$$\mathbf{H}_s^m = \mathbf{H}_r^m \mathbf{H}_r^{s-1}, \tag{2}$$
$$\mathbf{R}_s^m = \mathbf{R}_r^m \mathbf{R}_r^{s-1}. \tag{3}$$

[1]Touchence Inc. : http://touchence.jp/en/

TABLE I: Visualization of hand postures and applied forces exerted on the fingertips on the wrapping the Furoshiki cloth. In the playback mode, the user interfaces shown in (1) can turn off according to user preference, as shown in (2)-(8).



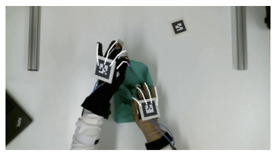
(1)

(2)

(3)

(4)

(5)



(6)

(7)

(8)

where $\mathbf{P}^i$ is a homogeneous representation of a vector $\mathbf{p}^i$, which denotes a position with respect to coordinate system i, implying that $\mathbf{P}^i = (\mathbf{p}^i, 1)^T$; $\mathbf{H}_1^0$ represents a homogenous transform matrix that transforms a vector from frame 1 to frame 0; $\mathbf{R}_1^0$ represents the orientation of frame 1 with respect to 0; and m, s, and r denote the reference, side camera, and reference AR marker coordinates, respectively. If a side camera recognizes the hand position $\mathbf{p}^s_{\text{hand}}$ and the orientation $\mathbf{R}^s_{\text{hand}}$, the position and the orientation of the hand with respect to the main camera can be derived as follows:

$$\mathbf{P}^m_{\text{hand}} = \mathbf{H}^m_s \mathbf{P}^s_{\text{hand}}, \quad (4)$$
$$\mathbf{R}^m_{\text{hand}} = \mathbf{R}^m_s \mathbf{R}^s_{\text{hand}}. \quad (5)$$

The postures derived from the hand pose estimator are stored in data-storage module described below. Even if an occlusion problem occurs in a top-view camera image, the values derived from other perspective images are used in the same manner.

The data storage module stores the dataset obtained in the record mode. This dataset includes data on the finger joint angles, forces applied to the fingertips, poses of both hands, and three perspective videos created from the images captured by each camera. This module also sends data according to the learner's demands in the playback mode.

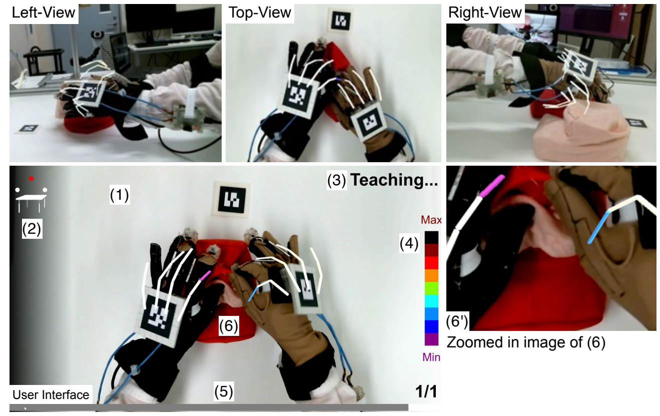


Fig. 4: Overview of the user interface. (1) Top view of the workbench, (2) perspective indicator, (3) current mode, (4) contact force indicator, (5) seek bar, (6) 3D finger model, (6') zoomed-in image of (6). The color of the fingertips is changed according to the contact forces.

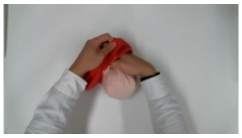
2) Playback Mode: In the playback mode, the system displays an instruction video to the learner, who conducts their study according to the instructor's actions in the video. The system has functions to switch the video perspective, turn the user interface on/off as described below, and overlay the information. The Unity 3D game engine [12] was used to construct an interactive system with these features.

The replay controller module accepts inputs from the learner, retrieves instruction videos and other information to overlay the video from the data storage module, and sends them to the scene creator module. Learners can control instruction videos during the learning process in different ways, with the easiest being keyboard inputs. Other interfaces, such as using leg movements or voice, can also be options. We implemented a movie player with the following basic functions: the play, pause, and change playback positions. The user can also change the hidden/visible view of the user interface and the overlaid information.

Fig. 4 shows the user interface. A 3D finger model is used to illustrate the instructor's hand posture and the approximate amount of force acting on the fingertips. The finger models represent a finger with four joints each, and four fingers, except the thumb, are used to increase visibility. Using Unity, the models are superimposed on the hand positions in the instructional video. The colors bar in Fig. 4 (4) shows the relationship between the amount of force and the color of the fingertip. The system internals, except for the user interfaces, were implemented with a robot operation system (ROS) [13], allowing the quick implementation of the time synchronization of multiple sensor data or coordinate transformation using AR markers.

In the next section, the experiments conducted to explore the potential usability of the proposed system are discussed.

TABLE II: Steps in the cover-on-a-cushion task

				
Initial State	Step 1	Step 2	Step 3	Step 4
				
Step 5	Step 6	Step 7	Step 8	Step 9

V. Experiments

A. A Visualization Example

First, using a Furoshiki cloth, we attempted to intuitively understand the steps involved in measurement and visualization using the proposed system. Both hands are used to simultaneously manipulate the object to be wrapped and the Furoshiki cloth. In other words, there are multiple instances in which the fabric and the object are touched. Additionally, there are many situations where the fingertips are hidden by the cloth or the back of the hand are hidden from the camera. The system should be capable of providing clear presentations to learners, even under these circumstances.

Table I shows snapshots of the instructor manipulating the Furoshiki cloth. To obtain the measurement results, a superimposed hand model was also displayed. Fingertip force sensors were attached to the index and middle fingers of each hand, and the force applied to each finger was visualized by changing the color of the tip of each finger model.

The panels in the table show that the shape of the fingers and force applied to the fingertips in areas that were not directly visible have been appropriately presented. For example, in (2), both hands pick a cloth, and force is applied to the index finger of each hand. In (7), the participant picks and moves the Furoshiki cloth with his left hand while holding the object with his right hand; this can be inferred from the exertion of force on his index fingers. Thus, we can intuitively obtain information that cannot be understood by simply by watching videos. However, to deal with problems such as difficulties in grasping a sense of depth, methods such as viewing images from different viewpoints need to be actively devised for users.

B. Experiment for Usefulness Assessment

A basic comparison experiment was conducted to confirm the applicability of the system using the more complex deformable object manipulation task described below; Eleven university students (10 males and one female) were recruited from our laboratory. They were divided into experimental (Exp, $n = 6$) and control (Ctrl, $n = 5$) groups. The subjects in the Exp group used the proposed system to learn the tasks, enabling them to select and switch perspectives and change the overlaid information during learning. In contrast, the Ctrl group learned a task with a top-view instruction video generated using the record mode of our system without overlaid information. Both groups were permitted to freely pause, resume, or change the playback points on the videos.

In this experiment, all participants attached contact force sensors to both their left and right index and middle fingertips, as done in the recording phase by the instructor. Note, that the data obtained from the sensors were only used to compare the differences in conditions. No other sensors or AR markers were attached to the participants' bodies.

Fig. 5: Left: cushion, middle: cover, and right: target state

C. Task

We designed a cover-on-a-cushion task as a deformable object manipulation task for the experiment. As shown in Fig. 5, the cushion resembles a stack of two cuboids of different sizes. Owing to the specific shape, it needs to be deformed according to a specific manual procedure to place the cover on it. The participants tried to do this using instructions for the experimental condition or control condition stated above. Table II lists the procedure of the designed task, and the details are as follows.

- Step 1 The cover's upper part is taken and folded to the inside.
- Step 2 The lower portion of the cushion is grasped and inserted into the cover.
- Step 3 The cushion is placed into the space prepared in step 1.
- Step 4 The lower part of the cover is fitted to the lower part of the cushion.
- Step 5 The lower part of the cushion is tucked inside the lower part of the cover.
- Step 6 The cushion's upper-side fillers are pushed to the lower part.
- Step 7 The cover's upper part that has been folded in step 1 should be withdrawn.

Step 8 The upper portion of the cushion should be fitted with the upper part of the cover that has been withdrawn in step 7.
Step 9 The shape of the cushion should be adjusted.

The demonstrator completed this task in approximately 150 seconds.

D. Procedure

The following procedure was followed by the participants in the experiment. First, all four contact force sensors were calibrated to be comparable to the values among subjects or groups. Second, the subjects performed 5 min of simple papercraft work to tune themselves to use their hands with the force sensors. Third, the participants were instructed on the task objectives and the instructions for using each system, following which they performed the task. During the task, the participants freely controlled the learning system. After completion, the participants self-reported their results. The task was repeated thrice.

E. Measures

We designed two quantitative scales to evaluate the performance of the participants in the task.

1) Completion Time: The completion times were measured before each subsequent trial.

2) Understanding Score: The score aims to measure the degree of understanding at each step of the task mentioned in subsection C. We determined whether each step was completed correctly in the following three categories per trial. Subject A successfully completed the step by following the correct process; subject B failed to complete the step despite following the correct process; and subject C either skipped the step or finished it incorrectly. We defined the understanding score as the number of steps ranked in category A; the range of this score was 0-9 (equal to the number of steps) per trial. The contact force sensor data were excluded from the analysis because of the frequent detachment of the sensors from the fingertips during the experiment.

F. Result

The completion time and the understanding score were analyzed using two-way ANOVA (one factor was groups and the other was the number of trials). The results are summarized in Fig. 6 and Table III. Fig. 6 shows that the task completion time (left) decreased with each trial in both groups. By contrast, the understanding score (right) varied less over the trials. The results of the statistical analysis (Table III) showed no interaction effect between the groups and number of trials for both features. Regarding the completion time, we observed that the groups ($F(1, 27) = 19.95$, $p < 0.001$) and number of trials ($F(2, 27) = 5.25$, $p = 0.012$). produced significant effects overall. Additionally, we observed that the groups had a significant effect on the understanding score ($F(1, 27) = 9.52$, $p = 0.005$).

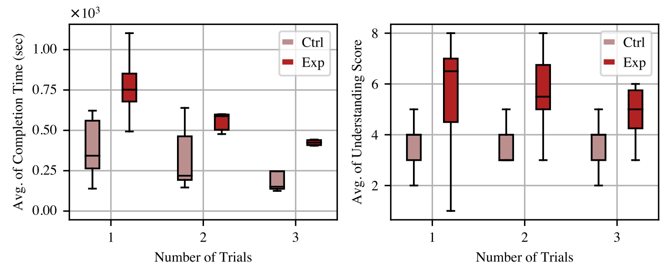


Fig. 6: Difference in average values between groups and number of trials. Left: completion time, and right:understanding score

TABLE III: Results of two-way ANOVA

Features	Factor	df	F	p
Completion time[sec]	Groups	1	19.95	< 0.001*
	Number of trials	2	5.25	0.012*
	Groups * No. of trials	2	0.875	0.428
Understanding score	Groups	1	9.52	0.005*
	Number of trials	2	0.24	0.976
	Groups * No. of trials	2	0.024	0.976

* significant at $p < 0.05$

VI. Conclusions

We present a novel training system for the manipulation of deformable objects. The system provides instructions by visualizing a skilled instructor's hand posture and the degree of applied force, without directly recognizing the deformation of soft object shapes. The experiment results revealed that the overlaying of skilled instructors' hand postures and contact forces on the instruction videos was effective in encouraging the understanding of learning in deformable object manipulation. However, it was also found that the task completion time increased with problems in the system. This suggested that issues existed in the intuitive presentation of information. Therefore, improving the intuitiveness and useability of the system is an important issue to be addressed in future studies.

Acknowledgement

This work was partially supported by NEDO and JST [Moonshot R&D][Grant Number JPMJMS2034].

References

[1] S. Proia, R. Carli, G. Cavone, and M. Dotoli, "Control techniques for safe, ergonomic, and efficient Human-Robot collaboration in the digital industry: A survey," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 3, pp. 1798–1819, July 2022.

[2] Y. Domae, "Recent trends in the research of industrial robots and future outlook," *Journal of Robotics and Mechatronics*, vol. 31, no. 1, pp. 57–62, 2019.

[3] V. Vyatkin, "Software engineering in industrial automation: State-of-the-Art review," *IEEE Trans. Ind. Inf.*, vol. 9, no. 3, pp. 1234–1249, Aug. 2013.

[4] H. He, G. Saunders, and J. T. Wen, "Robotic fabric fusing using a novel electroadhesion gripper," in *2022 IEEE 18th International Conference on Automation Science and Engineering (CASE)*, Aug. 2022, pp. 2407–2414.

[5] K. Yamazaki and T. Abe, “A versatile End-Effector for Pick-and-Release of fabric parts,” *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1431–1438, Apr. 2021.

[6] Y. Mo, Y. Nakagawa, K. Nagahama, and K. Yamazaki, “Motion measurement and segmentation toward automated sewing operations,” in *2021 IEEE International Conference on Mechatronics and Automation (ICMA)*, Aug. 2021, pp. 790–796.

[7] J. Yip, S. H. Wong, K. L. Yick, K. Chan, and K. H. Wong, “Improving quality of teaching and learning in classes by using augmented reality video,” *Comput. Educ.*, vol. 128, pp. 88–101, Jan. 2019.

[8] S. Büttner, M. Prilla, and C. Röcker, “Augmented reality training for industrial assembly work - are projection-based AR assistive systems an appropriate tool for assembly training?” *Conference on Human Factors in Computing Systems - Proceedings*, Apr. 2020.

[9] L. C. Wu, I. C. Lin, and M. H. Tsai, “Augmented reality instruction for object assembly based on markerless tracking,” *Proceedings - 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, I3D 2016*, pp. 95–102, Feb. 2016.

[10] Z. H. Lai, W. Tao, M. C. Leu, and Z. Yin, “Smart augmented reality instructional system for mechanical assembly towards worker-centered intelligent manufacturing,” *Journal of Manufacturing Systems*, vol. 55, pp. 69–81, 2020.

[11] K. B. Park, M. Kim, S. H. Choi, and J. Y. Lee, “Deep learning-based smart task assistance in wearable augmented reality,” *Robot. Comput. Integr. Manuf.*, vol. 63, p. 101887, 2020.

[12] A. Juliani, V.-P. Berges, E. Teng, A. Cohen, J. Harper, C. Elion, C. Goy, Y. Gao, H. Henry, M. Mattar, and D. Lange, “Unity: A general platform for intelligent agents,” Sept. 2018.

[13] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, “ROS: an open-source robot operating system,” in *ICRA workshop on open source software*, vol. 3, 2009.