SMART HAND FOR DIGITAL TWIN TIMBER WORK

The interactive procedural scanning by industrial arm robot

CHIKA SUKEGAWA[1], ARASTOO KHAJEHEE[2], TAKUYA KAWAKAMI[3], SYUNSUKE SOMEYA[4], YUJI HIRANO[5], MASAKO SHIBUYA[6], KOKI ITO[7], YOSHIAKI WATANABE[8], QIANG WANG[9], TOORU INABA[10], ALRIC LEE[11], KENSUKE HOTTA[12], MIKITA MIYAGUCHI[13] and YASUSHI IKEDA[14]
[1,2,6,7,11,12,14]*Keio University* [3,4,5,8,9,10,13] *Takenaka Corporation*
{[1]c.sukegawa\[2]arastookhajehee\[6]m.shibuya\[7]k.ito\[11]l.alric\[12]hottakensuke\[14]y.ikeda}@keio.jp, {[3]kawakami.takuma\[4]someya.shunsuke\[5]hirano.yuuji\[8]watanabe.yoshiaki \[9]qiang.Wang\[10]inaba.tooru\[13]miyaguchi.mikita}@takenaka.co.jp
[2]0000-0001-5381-7784, [12]0000-0003-3507-570X, [14]0000-0002-2016-4083

Abstract. This paper describes a 3D automated scanning method for building materials, namely *"The Interactive Procedural Scanning"*, in a collaborative environment composed of a human worker and a CNC robot. This procedure aims to translate the observation skill of an experienced carpenter into an intelligent robotic system. The system frames its function on the first stage of a traditional timber examination process, called *'Kidori'*, in which observations and findings are marked on the timber surface to provide hints for the subsequent cutting process. This paper aims to recreate the procedures using an industrial robotic arm, computer vision, and a human worker. A digital twin model of the timber is created with a depth camera serving as a base map to exchange information and receive instruction from the human worker. The margin of a discrepancy between the original processing location and the location of the actual end effector, where the tools are, is minimised in this system.

Keywords. 3D Scanning; Computer Vision; Traditional Technique; Phycology; Machine Learning; SDG 9.

1. Introduction

This research aims to reduce the technical threshold required for a human worker to perform advanced wood processing procedures. In this framework, a human worker with little wood processing knowledge or skill collaborates with a robotic arm to examine and acquire properties information of the material. The robot must be robust enough to function in a dynamic environment with unforeseeable disruptions and

POST-CARBON, Proceedings of the 27th International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2022, Volume 2, 131-140.

encourage interaction with its workers. Disruptions are unpredictable circumstances that include human misjudgement, accidental moves of the material in the middle of the process, or the heterogeneity of materials that will affect the final product's performance. The human worker has two cognition tasks in a traditional CNC machinery workflow. Firstly, they must carefully select an appropriate material suitable for the design purpose, considering the factor of sizing and natural material defects; for example, wood knots should be avoided. Secondarily, the human worker must determine the best orientation and location for the material's placement on the processing table and assign a local-material origin point to confirm the coordinates of CAD/CAM and physical space.

This premise is translated into two predetermined conditions in this demonstration; i) the material is placed at an arbitrarily selected orientation and location within a workable area; ii) the individual materials would have heterogeneous materiality, e.g., the presence of knots wood grain direction. The two conditions reproduce the realistic scenario of a primitive collaboration between a human and robot where cognition tasks are highly dependent on the human counterpart.

The research aims to utilise the unspoiled aspects of human workers' input to robotic behaviour. We research with a skilled carpenter called *'Miya-daiku'* in his observation process on a piece of timber, then the procedures and the embedded techniques are translated into our system. The procedural steps are not copied because machines and humans work require different thinking. However, this deconstruction of a human observation and action process enables us to analyse the behavioural pattern of specialists and provide hints on how to implement the principles in a robotic setting. Our team assumes that machines can perform more efficiently when given implications by conventional human wisdom.

2. Background

In 2021, there are many examples of computer vision or machine vision applications in the market, such as object detection in autonomous driving or image recognition in industrial robots (Klette, 2014). Most carmakers attempt to improve the scanning capacity by utilising multiple sensors, distance-related sensors include radar, laser rangefinder 3D-lider, RGB camera, other sensors are geography-based such as GPS and gyrocompass. In recent years, there have been the following three trends. The first is the development of cognitive image processing by machine learning. In the 2010s, most automakers used multiple sensors to make up for their weaknesses, but tech companies have begun to develop camera-only technology again (Mobileye, 2020). The second is the utilisation of digital twins. Some companies are training in virtual space to quickly acquire the results of the above-mentioned geometric learning, which is dozens of times more efficient than in real space. (Ohnsman, 2018). Virtual space is useful for training robot too. (Matsas and Vosniakos, 2017.)The third is cooperative: automobiles have become connected to the network with infrastructure development. There is a flow of sharing information and synchronising with a conventionally operating system for everyone, with each other, or with a central concept such as a global map, in some companies such as Honda, Komatsu, Google. Also, there is an example of training robotic arm by using virtual and physical environment. (Luis et al., 2019)

3. Limit of Current Research and Problem Statement

There are two problem statements: 1) Similar to applications in other industries, the role of robots will shift from executing monotonous, repetitive commands to more intelligent ones that involve cognitive and decision making in the construction field. This paper suggests an adaptive setup for a robotic arm on-site that accommodates accidental occurrence and heterogeneous materials. The use of computer vision is the first step in understanding the situation in front of the robot on-site. 2) A few examples of transferring human specialist skills to a robotic system (Madhura et al., 2021). The mechanical resemblance of a robotic arm and that of a human arm provides a good starting point on adapting the experience and wisdom of a skilled specialist to a computer-assisted robotic system.

4. Aim of This Paper

The first aim is to dissect and understand the workflow of skilled Japanese carpenters; The subject of this research, specialist carpenters, are also known as palace carpenters in Japan. It is not easy to replace their implicit knowledge with a robotic system because a significant part of the knowledge is embedded into the specialists' intuition developed from years of experience (Nishioka, 1984). As the author of this paper possesses no carpentry knowledge, a documentary of a palace carpenter's workflow was filmed for further investigation. We extracted points of interest from an amateur's point of view and studied the unique abilities of a palace carpenter. Initially, we tried to replicate the entire wood process but later restricted the scope to only the "wood inking" stage, where the most exciting features have been observed in experiment 1.

The second aim is to reconstruct a human's workflow into a robot. There are a few examples of robotic applications in wood processing for construction. Some robotic arms are fitted with sensors on the end-effector to provide a feedback loop. The accuracy of the industrial robotic arm ensures the precision of coordinates interpolated by the computer vision program. Secondly, the idea of digital twins is applied to wood processing robots. The virtual twin space is used for obtaining training data for machine learning engines through simulation. The physical space counterpart is used to conduct interactive experiments with human agents. As this study aims to synchronise knowledge between all agents, global information obtained from the sensors of other robots and even human operators is to be shared. We envision a system that can share experiences through connections. The authors have chosen a case study in which a robotic arm processes a wood part. Machine vision technology may be used to its full potential as most of the wood process has been replaced by electrically driven tools. Thus, two experiments will prove the concept for the above two points.

5. Hypothesis and Research Questions

The overall hypothesis is that by closely observing the working process of palace carpenters, hints can be obtained to design an autonomous and adaptive control system to handle non-homogeneous materials with a robotic arm and computer vision.

- **Hypothesis for experiment 1:** From the palace carpenters' work analysis and linguistic analysis, the skilful points for observing their trees (raw parts) can be extracted. Also, what kind of information is being read, what resolution is being read, etc., can be extracted and used when implementing them in computers and robots?

- **Hypothesis for experiment 2** (implementation on a robot): Is it possible to increase the accuracy of the computer vision of the robotic arm and the accuracy of machining by dividing the scan into different stages?

6. Experiment-1: Observation of Well-skilled Carpenter

6.1. INTRODUCTION

One of the most critical parts of a palace carpenter's working process is 'inking', in which they inspect the timber and makes markings on the surface for further processing. To design a parallel mechanism for the robot to acquire the characteristics of the subject material in collaboration with an amateur human worker. To determine the necessary knowledge, it is essential to understand the thinking process of palace carpenters. Table 1 shows how an experienced carpenter looks at their subject.

	What to look for	**What can be understood**
Pattern /Grain	Distance of lines, Comparing the pattern of each surface	Tree species / weak direction / dry deformation direction / creep direction
Annual Ring	Distance of ring, Pattern, Centre direction	Tree species / address (weak direction) / original up / down / expected dry shrinkage / natural tree or tree planting
Colour	Sapwood, Heartwood	Heartwood / Sapwood, Compression / Tension material, Internal stress, Bottom/ End of log
Knot	The density of fibre, Texture	Top and bottom, Brittleness, appearance, Knots appear.

Table 1. Observed Items Respectively

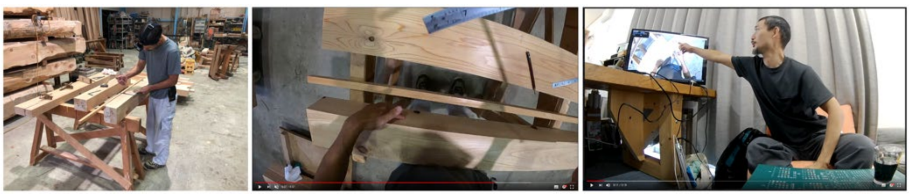


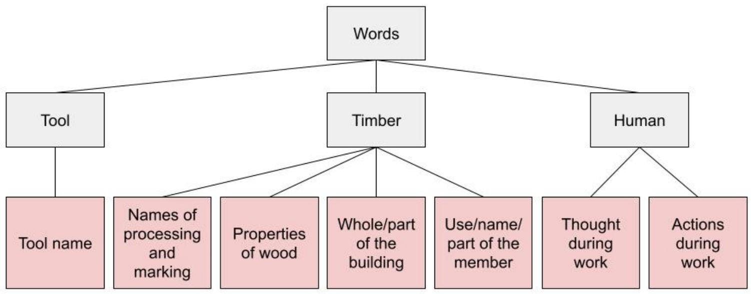
Figure 1. (left) Work environment (centre) Screenshot of the head-mounted camera for Experiment-1. (right) the place carpenter explained what he was thinking while watching the video

6.2. METHOD

The proficiency of the carpenters is based on implicit thinking. Protocol Analysis is a way to analyse their thoughts and existing physiological methods to reveal the process of intellectual action (Suwa and Barbara, 1997), (Ericsson, K and Simon, H. A, 1984). Here, we recorded the process of 'inking' as part of the construction of a small temple.

To document the eye movements of an experienced palace carpenter, a camera is mounted to the palace carpenter's forehead, and a first-person view video is recorded. After that, while watching the video, the carpenter was asked to explain what they were thinking of. The video footage of the explanation is transcribed, keywords were extracted and categorised. Finally, the number of appearances is graphed in chronological order for each category— to understand the logistics of the thinking flow. The video recorded is about the inking work for four pillars in the small temple (Figure 1).

Since no existing quantitative method can systematically classify the terms, the carpenter explained his working process. Yet hierarchy and the level of abstraction is essential (Suwa, 2016). Then here, we classified key terms into three categories: 'tools', 'lumbers', and 'humans'. Languages related to 'timber' have been categorised into “processing / marking names”, “wood properties”, “whole/part of members”, and “uses/names of members/parts of members”. Determine the "classification of inking" after having "the function (pillar) of the result". 'Human' was also fixed in the hierarchy of "thinking at work" and "behaviour at work".



Classification of Words	Explanation of Words, and examples	The number of appearances
①Name of tools	Toolbox・Carpentor’s square・Long mesure・Inking parallel ruler	58
②Feature of Timbers	Knot・Straight-grain・Sapwood・Swirl・Straigt	66
③The name of the process, and Inking	Grid・Mortise・Inking reference line etc.	57
④usage, the name/part of the building,	Column・Horizontal interior timber・Front side of building・bottom part of wood・Top part of the wood	43
⑤the whole building, part	Building・Foundation・a side of facade people pass by most	27
⑥action in operation	Roll (wood)・draw (inking) lines・look (drawing)	201
⑦thought, related words	To think (about the arrangement of buildings and parts) ・ To decide (to replace the arrangement)	94

Figure 2 & Table 2. Classification of words and their number of appearances

6.3. RESULT

The verbal explanation was then transcribed. Its length was about 15,200 words, including the experimenter's remarks. Based on the transcripts, we classified the words into seven categories according to their contents and analysed the transition of thoughts. Thus, from the first-person video of the palace carpenter and the analysis of his remarks, it turned out that there are 3 phases following: 1) the period of observing multiple surfaces while rolling multiple pieces of wood, 2) the period of observing while rolling one piece of wood, 3) the period of inking. In this paper, each period is tentatively named *'overall observation period'*, *'individual observation period'*, and *'inking period'*, respectively.

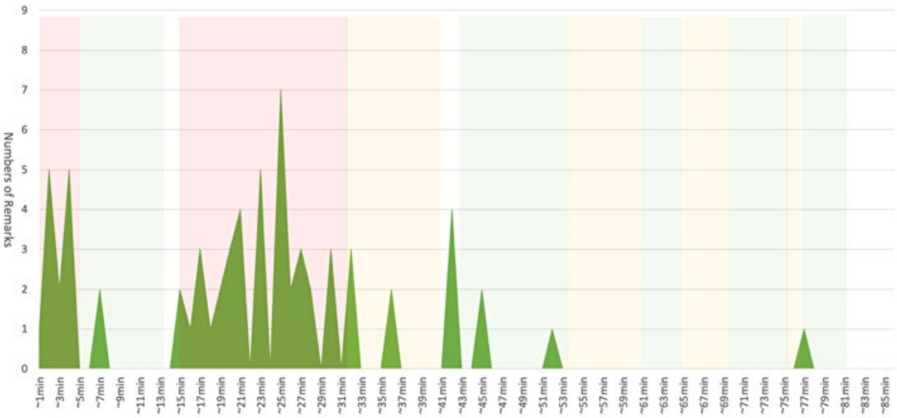


Figure 3. The number of remarks regarding 'Feature of Timbers' on the table in the working period; the number of remarks about 'Feature of Timbers' peaks immediately after the beginning of work and between 15 and 30 min from the start. On the contrary, the number of times decreases significantly after 35 min, and there are almost no remarks after 55 min.

To sum up from experiment 1: protocol analysis reveals two phases within the observation stage, called 'Sumitsuke' in which the timber is prepared for further processing. The first phase is "comprehensive inspection". All material sides are quickly inspected visually, sometimes by rotating the wood. Then the carpenter decides on the final orientation of the material based on the inspection result. The second phase is the "local observation" period, where material parts are scrutinised, and precise cutting location is decided. From the above, human observe of wood, there is a "period to compare multiple timbers and roughly grasp their characteristics" and a "period to determine detailed processing based on the rough observations".

7. Three Directions of Utilising the Human Analogy into a Robot

The above conclusion implies the potential to 1) utilise the reductionistic analysis approach to reconfigure the workflow. 2) Prepare various sensors, depending on an interview that reveals implicit knowledge. Moreover, the mechanical resemblance between robot and the human eye, and the human hand can now combine into one device as an end-effector. Concretely, 3) the relative position of the eye (camera) and material helps effectively detect and recognise features of the material and simultaneously can move to an operation that the hand does.

8. Experiment-2; The Implementation for Robot and its Precision

8.1. INTRODUCTION

From the behavioural analysis of the palace carpenter in Experiment 1, the purpose of each observational action had been linked to a specific piece of information. This experiment constructs the input and output logic of an automated program like the set of activities performed by the carpenter. Depending on the subject, the robotic arm's end-effector, the scanning method, and the workflow must be designed accordingly. The following is an account of the necessary functional requirements of the system; 1) The system recognises the location and shape of the object that is placed arbitrarily on the table. 2) Determine if the object has been registered in the database. 3)Recognize and identify material parts that may cause performance problems in the final product (limited to timber knots in this study). 4)With computer vision, the end-effector's position can be controlled with an accuracy of a millimetre for machining. 5)Perform a scan of the object with a resolution optimised for quick inspections inside the computer environment.

8.2. CONCEPT OF PROCEDURAL SCANNING

Attempts to scan the timber without a plan had encountered several difficulties. The depth sensor had involuntarily acquired data irrelevant to the wood in the subject, unnecessarily increasing the computational time since the algorithm cannot distinguish the object from the immediate environment; hence, a low-resolution map that provides an overview of the material is to be obtained first, then the scanning resolution can be increased for areas of interest for closer inspection. Different scanning scales can yield a comprehensive understanding of the working environment. Getting useful information from the environment selectively and quickly will be an essential subject for future development in computer vision.

STEP	Title	Entry	Sensor/Device	Purpose
0	**Pre-training for ML**	Image (RGB)	RGB-camera	ML training
1	**Surrounding Scan**	Geometry (very rough)	HoloLens, (+Lider)	Human safety
2	**Rough Scan**	Geometry, location	D-camera	Workable area, Material position, -axis
3	**Individual-Timber Recognition**	Image (RGB, Outline enhancement)	RGB-camera + ML(cloud)	To pull information from DB
4	**Individual's Surface-recognition**	RGB Image (Outline enhancement)	RGB-camera + ML(cloud)	To recognise which side is up and down
5	**Detailed Scan**	Geometry (high resolution)	D-camera + robot arm (arbitrary position)	To get the latest geometry
6	**Re-calibration for process**	Points and their line	D-camera + robot arm (arbitrary position)	Minimise physical deviation

Table 3. Seven procedures of Scanning

8.3. TOOLS AND SETTINGS

Microsoft Hololens2® as a data communication tool for human workers, Intel Relasense® as a composite computer vision for the robot, and an RGB-D camera was used in this experiment. Since the digital twin model is projected and placed in the MR, the human workers can superimpose virtual objects on the physical space and perform a real-time evaluation. In addition, a mapping is projected from the robot's end-effector onto the material surface to display its properties in a visually intuitive way. In other words, both the human and the robot are equipped with a channel to generate input and obtain output in the system. All the scan stages are used to gradually achieve an increasingly accurate representation of the material in the computer. The level of detail achievable using each scan stage corresponded with what was required by the next step. The resulting system can operate on a given timber block with a repeatable coordination accuracy of the timber’s local origin point, the deviation between the end of milling and scanned timber 5mm to 10mm. The projection mapping function can also instruct and interact with the robot to increase the accuracy to the desired degree. In addition, human has a role in authenticating whether the robot's understanding of the material is correct.

8.4. PERFORMANCE AND EVALUATION METHOD

The purpose of experiment 2 is to confirm if the proposed system works effectively to satisfy the required performance. Various factors work effectively, but here we set the accuracy of the entire robot arm system using computer vision tentatively (potentially it can be the amount of data, time, labour of humans). Precisely, the difference between the ideal target point and the tip of the mill of the end effector was measured (Figure 3). We also indirectly justify the above by saying that this difference is exacerbated when the "step scan" known below is not used.

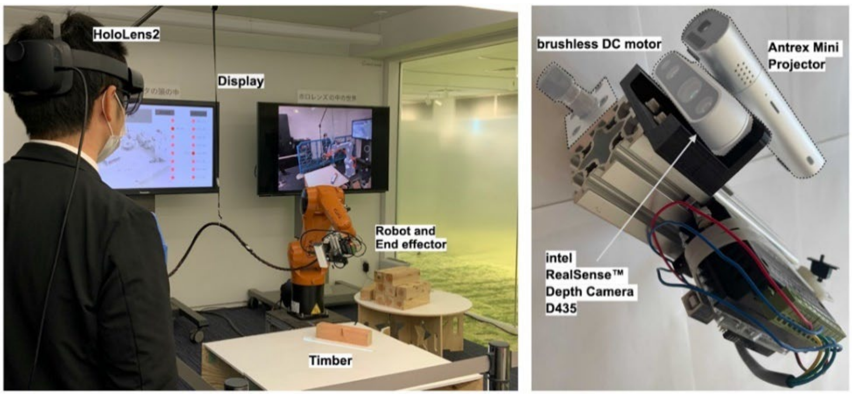


Figure 4. Developed end effector consists of milling, computer vision, projector. The design of this end-effector is made an aiming with lightweight and small size. Our team used kinectV2 and KinectAzure, but Intel Realsense® is used mainly because of size and weight.

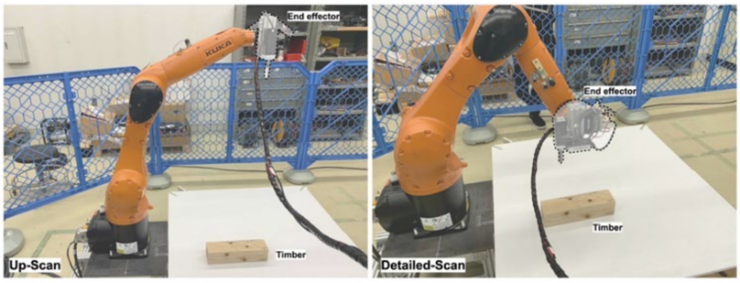


Figure 5. The screenshot of each scanning moment and robot attitude ;(case-1) Up-scan is the rough scan of the procedure following 0, 1, 2,(see table 3) ;(case-2) Detailed-Scan is the procedure following 0,1,2,(3,4), 5 (see table 3)

8.5. STATISTICAL METHOD

Experiments were performed ten times for each timber. In addition, five pieces of wood called A, B, C and D were used to smoothen the differences between individuals. As for scanning, the above cases 1 and 2 were tried, so 40 samples were obtained. Here, we wanted to see the effectiveness of the procedural scanning method, so the difference distribution was taken separately for case-1 and case-2 (figure5).

The below boxplot (Figure6), The left side (blue) of each table is the scanning method of case 1 ('Up-scan'), and the plot (orange) on the right is an example of performing a stepwise scan in case 2 ('Detailed-scan') to some extent. Case- 2 has a more negligible average difference at any point. For points 1 and 2, the average difference is about 12 mm or less, but for points 3 and 4, the absolute value of the difference is 18 mm or less. It is inferred that this is due to the poor accuracy of the peripheral edge of the angle of view of the depth camera. Again, the benefits of stepped scanning come out.

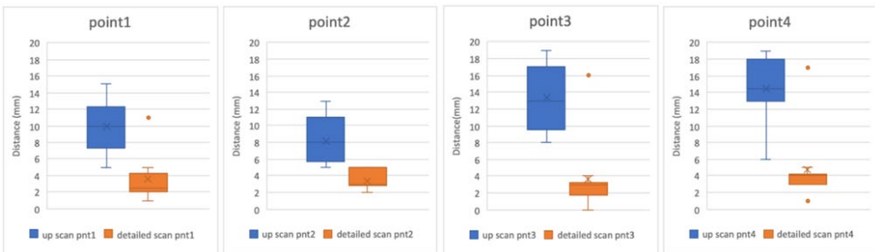


Figure6. The result of the experiment. Comparison of the error between the ideal position of timber and physical place.pt1~4 indicates four corners of timber's top surface.

8.6. SUMMARY

To realise the concept of *'Procedural Scanning'*, it is necessary to properly design the scanning method and its order in consideration of LOD. From experiment-2, the accuracy tends to improve with this scanning method, but other evaluation criteria must be considered to measure the multifaceted effectiveness. The concept has been proven valid, but the methodology still must be refined. To evaluate the system's practicality, it is necessary to meet the requirement inaccuracy and realistic criteria such as time for system execution, amount of human labour, power consumption, and data size. As the system has gone through rounds of iterations, issues that would likely arise in practical applications have been noticed and rectified. For example, in the initial game, the amount of data obtained from scanning had exceeded the processing computer's capacity and crashed the system. It is essential to ensure the data is lightweight enough for rapid computation for real-time processing.

9. Conclusion

The contribution of this research is the development of an effective scanning system for half automated and interactive CNC robotic arms as a part of wood processing. The system reduces the time and resources usually required to create a meaningful digital information profile of the material, with an adaptive scanning resolution in parts of interest determined by the human user during the interactive process. In conclusion, the computer vision system developed in this paper creates an interactive workflow in which the robotic movement is visualised beforehand and presents a current understanding of the material to the user in real-time with gradually increasing accuracy, demonstrating the potential of machine vision in the AEC industry, not as a replacement for human workers but as a complementary counterpart.

Acknowledgements

Takenaka Corporation and Keio University hold this co-research, also thank the master carpenter firm, Hakuo-syaji.

Selected References

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. US; The MIT Press.

Klette, R. (2014). *Concise Computer Vision*. Springer. ISBN 978-1-4471-6320-6.

Luis Perez, Eduardo Diez, Ruben Usamentiaga, & Daniel F. Garcia. (2019). Industrial robot control and operator training using virtual reality interfaces, *Computers in Industry*, 19, 114-120.

Madhura Jayaratne, Damminda Alahakoon, & Daswin de Silva.(2021). Unsupervised skill transfer learning for autonomous robots using distributed Growing Self Organizing Maps, *Robotics and Autonomous Systems*, 144, 103835.

Matsas, E., Vosniakos, GC. (2017). Design of a virtual reality training system for human–robot collaboration in manufacturing tasks, *Int J Interact Des Manuf* , 11, 139–153. https://doi.org/10.1007/s12008-015-0259-2

Mobileye. (2020). *CES2020* and Retrieved December 1, 2021, from https://www.mobileye.com/

Nishioka,Y (1984). 木に学べ[*Lean from Tree*]. 1st ed. Tokyo, Japan; Shogakukan publish.

Ohnsman, A., (2018). *Waymo Is Millions Of Miles Ahead In Robot Car Tests; Does It Need A Billion More?* Retrieved December 1, 2021, from https://www.forbes.com/sites/alanohnsman/2018/03/02/waymo-is-millions-of-miles-ahead-in-robot-car-tests-does-it-need-a-billion-more/?sh=7a7f080b1ef4

Suwa M. (2016). 「こつ」と「スランプ」の研究 身体知の認知科学*[Research on Tip and Slump Cognitive Science of Embodied knowledge],* Tokyo, Japan; Kodan-Sya.

Suwa, M., Barbara, T.(1997). What do architects and students perceive in their design sketches? A protocol analysis, *Design Studies*, 18(4), pp385-403.