ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 12 (2022) No. 2, pp. 462-476
DOI: 10.33168/JSMS.2022.0225

A Study on Real-time Hand Gesture Recognition Technology by Machine Learning-based MediaPipe

Jeong-Seop Han[1], Choong-Iyeol Lee[2], Young-Hwa Youn[3], Sung-Jun Kim[4]

[1]Department of Big Data Content Convergence, Namseoul University Graduate School, 31020, South Korea

[2]Department of Business Administration, Namseoul University Graduate School, 31020, South Korea

[3]Remian Urban Vista Daycare Center, 14732, South Korea

[4]Department of Big Data Industry Security, Namseoul University Graduate School, 31020, South Korea

mvstar@hanmail.net

Abstract: This study recognizes the movement of hands moving in real time through a camera connected to a computer to be used for children's educational activities. The Ministry of Education of the Republic of Korea has developed a curriculum for children aged three to five through the 2019 revised Nuri curriculum so that they can be applied in the field from 2020. The curriculum was presented in five areas (body exercise, health, communication, social relations, art experience, and nature exploration). Among them, physical exercise, health, and nature exploration are areas where education must be conducted through external activities. However, the reality is that active external activities are difficult. Even now, in January 2022, the world is suffering from the COVID-19 pandemic. The Korean children's education site is also experiencing many difficulties. Therefore, this study attempts to help children learn how to explore nature indoors by using MediaPipe to recognize children's hand movements and using virtual mice that operate through them.

Keywords: MediaPipe, MediaPipe hands, virtual mouse, Nuri curriculum

1. Introduction

The COVID-19 pandemic put the vulnerable children and the elderlies at difficult situations. In particular, the educational environment of children has been severely affected. To solve this problem, various studies related to children's education are being actively conducted. At a time when children's playfulness is threatened by COVID-19, one study argued that education in a true sense should be revived in the educational field and an opportunity for educational transformation during COVID-19 (Oh and Lim 2021). COVID-19 confirmed the possibility of being applied to the field in the future and the importance of mutual supplementation through the connection between field education and remote education, and argued that it is necessary to develop into a connection with the revised Nuri curriculum (Lim 2021). With the development of ICT curriculum in child education, an environment where children can experience various digital play experiences has been provided. Another study argued that safety, rights, and fairness should be addressed (Jo 2021).

In addition, the Ministry of Education of the Republic of Korea declared the vision of "realizing a people-centered future intelligent educational environment" and established a basic plan for educational informatization based on this to implement educational informatization. As detailed action plans, policies such as creating a future ICT-based education and research environment, innovating sustainable educational informatization, realizing customized educational services through ICT, and expanding shared educational informatization resources are implemented. In accordance with the national policy, various studies are being conducted to apply ICT-based technology to the field of education. These include a constructivist-based interaction English learning model to increase the efficiency of early childhood English learning in a home environment using Artificial Intelligence (AI) speech recognition speakers (Hwang 2020) and investigation on the effect of activities using artificial intelligence speakers on infants' interactions and creative problem-solving skills (Yoo and Kim 2021) and language skills (Lee and Oh 2021). Various attempts are being made for early childhood education, but most of them are research related to voice recognition AI speakers. On the other hand, research on gesture recognition from the perspective of ICT is being actively conducted including an interface that does not require space restrictions by analyzing the information of the index and middle fingers using infrared images (Lee et al., 2011), establishing a virtual mouse environment that can move the mouse pointer and execute mouse commands by recognizing the user's hand gestures in the beam projector environment (Seo and Choi 2003; Seo and Choi 2004), and a Two-layer Bayesian Network (TBN) for hand gesture-based virtual mouse interface and real-time hand gesture recognition (Roh 2017).

The Ministry of Education of the Republic of Korea has developed a curriculum for children ages three to five through the 2019 revised Nuri curriculum so that they can be applied in the field from 2020. The curriculum presented through the revised

Nuri curriculum is divided into five areas (physical exercise, health, communication, social relations, art experience, and nature exploration). This study, among the five areas, it was intended to be applied to the nature search area, and the three categories of nature search are explained in Table 1.

Table 1: Categories of nature search

Category	Overview
Enjoy the exploration process	Be constantly curious about the world and nature around
	Have fun participating in the process of exploring questions
	Take interest in different ideas in the process of inquiry
Exploring things in your daily life	Explore the properties and changes of objects in various ways
	Count objects to find out the quantity
	Recognize and distinguish the position, direction, and shape of objects
	Compare properties such as length and weight in daily life
	Look for repeating rules around
	Classify data collected from daily life according to criteria
	Be interested in tools and machines
Living with nature	Pay attention to the flora and fauna around
	Value life and the natural environment
	Relate changes in weather and seasons to life

The purpose of this study is to help children learn how to explore nature indoors by using MediaPipe to recognize children's hand movements and use virtual mice that operate through them. In addition, in connection with the education presented in the revised Nuri curriculum, it is to find out whether there is a possibility of using MediaPipe in the remaining areas (physical exercise, health, communication, social relations) other than the field of nature exploration.

2. Related Works

Image processing technologies tended to have their accuracy dependent on the input image generation conditions, but as deep learning-related research is actively progressing, deep learning structures such as Faster R-CNN and YOLO have begun to be proposed as technologies for this purpose (Yu et al., 2020). Until now, many studies have been conducted to recognize human movements and apply them to various fields. A home training system was proposed to classify users' poses in real time through a learning-based pose classification model, extract joint coordinates of poses using the MediaPipe Pose API, and analyze poses based on human joint points to correct incorrect poses (Lee and Kim 2021). After extracting the key points of the joint using Mediapipe’s real-time Hand Skeleton Tracking, sign language was

predicted by learning CNN using residual structure and models using BI-DIRECTIONALLSTM, and Bahdanau Attention (Gil et al., 2011). Using OpenCV library and MediaPipe, we studied a system that recognizes sign language movements by tracing hands and fingers, and converts the meaning of sign language into text-type data using CNN technology and provides it to learners (Kim 2021). By utilizing the Python-based MediaPipe Framework, hand tracking is used to detect hands and recognize gestures so that users can control the kiosk without touching it (Noh et al., 2021). Based on the open-source MediaPipe, a model representing the finger state, a model representing a hand posture, and a model representing the hand posture through this model are also applied to number recognition in sign language using a hierarchical model using the bending of one finger and the touch of two fingers to verify its usefulness (Heo et al., 2021).

Recently, as various functions have been added to smartphones, the prevalence of smartphones with improved performance is increasing. Accordingly, research to recognize motion using a smartphone is being actively conducted. A study was conducted to detect the user's movement using the tilt sensor of the smartphone and to detect the user's motion and motion recognition by applying the tilt reduction algorithm (Lee and Lee 2014). It detects the user's movement using the sensors built into the smartphone and uses the Kalman Filter algorithm to correct the user's movement data to compare whether the user is exercising with the correct posture so that the user can maintain the correct exercise posture. Research was conducted to help (Kwon et al., 2016).

Education using ICT is already being conducted in the educational field for infants and toddlers, and many studies are being conducted in this regard. In a study on the perception and operation status of early childhood teachers on ICT utilization education according to early childhood age, there was no significant difference in how to operate ICT utilization education according to age, but it was found that teachers were having difficulty in operating ICT utilization education (Kim 2017). Creative problem-solving activities using ICT (appreciating fairy tales and nursery rhymes related to life topics) were conducted for 8 weeks to find out what effect the creative problem-solving activities using ICT have on children's creativity by applying them to the educational field. It was confirmed that creativity was significantly improved through solution activities.

Therefore, in this study, we plan to use OpenCV and MediaPipe Hand models in the Python operating environment. We will develop an interface that can operate a virtual mouse to replace a computer mouse by detecting the hand shape through the implemented model. In addition, it is intended to determine whether or not it can be applied to the educational field by allowing children to directly experience it using the balloon-popping game that children will like and an application program that allows them to observe natural objects.

3. Research Methodology

3.1. MediaPipe

MediaPipe is an artificial intelligence framework provided by Google. This is a service that provides a solution-type library so that the human body recognition function model included in the image data can be developed and learned and used easily. MediaPipe's representative solutions are configured as shown in Table 2, and continue to provide new solutions. MediaPipe supports various development environments such as web pages, Android, and iOS. And supported languages include Android, iOS, C++, Python, JS, and Coral. In addition, the media pipe is an open-source project, and the program source is disclosed, so you can modify what you want and use it for development. In this study, the Hands solution, which recognizes the shape and movement of the hand among the MediaPipe solutions, was used.

Table 2: Solutions of MediaPipe

Solutions	Overview
Face Detection	A high-speed face detection solution equipped with multiple face detection functions.
Face Mesh	Face geometry solution that estimates real-time 3D facial landmarks.
Iris	A solution that estimates human iris and tracks landmarks in real time.
Hands	A solution that recognizes the shape and movement of your hands.
Pose	A solution that detects body movements and tracks body gestures.
Holistic	A solution that integrates and provides individual models for the components of the whole body pose, face, and hand.
Hair Segmentation	A solution that distinguishes human hair.
Object Detection	A solution that detects 3D objects in real time.
Box Tracking	A solution for calculating the position of a detected object.
Instant Motion Tracking	A solution that tracks AR through a platform.

3.2. MediaPipe Hand Tracking

MediaPipe hand tracking is a model that uses a machine learning pipeline to detect and track recognized hands and fingers in images. After detecting the palm of the entire image using machine learning, as shown in Figure 1, the method uses a method of mapping 21 coordinates to define 3D coordinates to display the joints of the palm and obtain real data. MediaPipe hand tracking provides real-time performance that

can be used not only on desktop computers but also on smartphones, and can be expanded to recognize multiple hands at once.

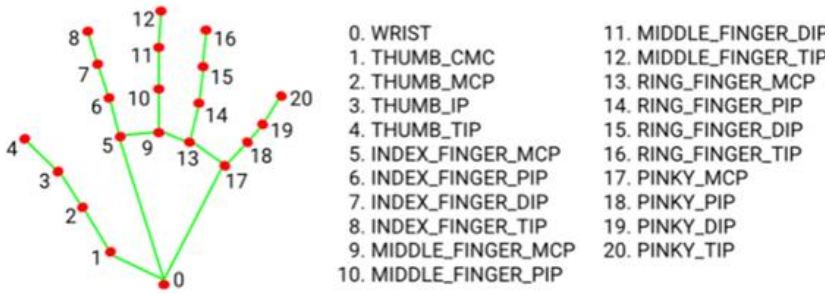

Fig. 1: Hand landmarks

3.3. ML Pipeline

MediaPipe Hands utilizes a multi-model ML pipeline. Representative examples are the palm detection model, which operates on the entire image and returns an oriented hand bounding box, and the hand landmark model, which operates on the cropped image region defined by the palm detector and returns high-fidelity 3D hand keypoints.

3.4. K-NN (K-Nearest Neighber) Algorithm

To increase the accuracy of motion, K-NN algorithm was used. First, the K-NN algorithm was trained to distinguish the palm from the fist. The files used for learning consist of 15 coordinates and 1 label item, and a total of 110 data were used. The K-NN algorithm is the simplest classification algorithm among machine learning algorithms. Among the images recognized through the image, the coordinates of the palm and fist shapes and the coordinate values of the learning data were compared to determine the label value of the closest data. Although the K-NN algorithm is a simple algorithm, it is widely used in computer vision sales such as image processing, character recognition, and face recognition. The advantage of the K-NN algorithm is that it is fast in implementation and speed because it does not create a model through special training using training data.

3.5. Research Design

This study was conducted according to the procedure defined in Fig. 2. Python was used as the environment for driving MediaPipe. Python is a high-level programming language, platform-independent, interpreted, object-oriented, and dynamic typing interactive language that is actively used in various fields such as data analysis, artificial intelligence, and web development. I installed the library required for image processing in Python and the library required to use the media pipe, respectively (pip install opencv-python / pip install mediapipe) and configured the environment.

First, the MediaPipe Hands model was applied. After importing the MediaPipe module to Python, it was confirmed that the shape and movement of the hands were normally recognized, and landmarks were applied to the finger joints. At this time, we started working after setting the configuration of MediaPipe Hand Tracking. The STATIC_IMAGE_MODE setting is set to False to process the video coming through the laptop camera as a stream, and the MAX_NUM_HANDS value is set to 1 to specify the number of hands that can be detected in the video. The value of MIN_DETECTION_CONFIDENCE was set to 0.5 to specify the minimum confidence value of the judging model, and the value of MIN_TRACKING_CONFIDENCE was set to 0.5 to specify the minimum confidence value of the model judging that the hand landmark was successfully tracked.

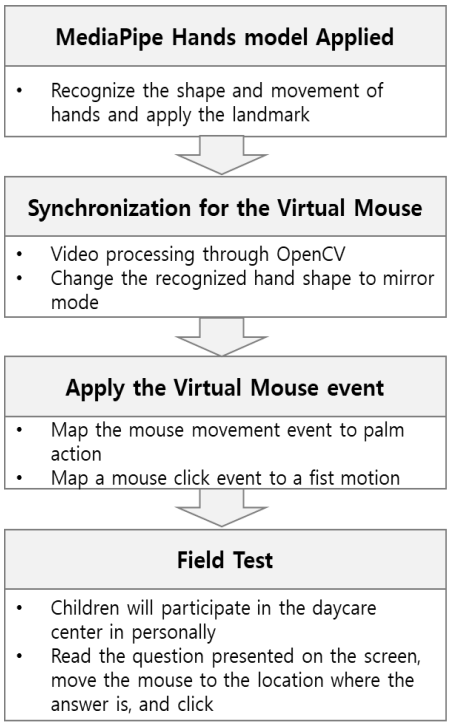

Fig. 2: Research design

Second, we performed virtual mouse synchronization. OpenCV was used to capture the live stream of the video transmitted from the camera installed on the laptop. OpenCV provides a very simple interface, so it is easy to use. To create a 'VideoCapture' object, the first camera was selected using a simple one-line command (cv2.VideoCapture(0)). By running the program in the Python environment, the image transmitted through the camera installed in the laptop was controlled through OpenCV and switched to mirror mode so that the recognized hand shape does

not work asymmetrically. Mirror mode conversion was performed using the cv2.Flip() function provided by OpenCV. This is because the users are children, so if the hand movements are reversed, the children may be confused.

Third, an event was applied to the virtual mouse. Among the 21 landmarks provided by the MediaPipe Hands model, the program was changed to recognize only the palm and fist shapes required for research. In order to control the virtual mouse, a mouse drag event was mapped when the recognized hand shape was a palm, and a mouse click event was mapped when the recognized hand shape was a fist. For the division of palm and fist, when coordinates 4, 8, 12, 16, and 20 among the numbers displayed in Hand landmarks were all recognized, the palm was recognized as a fist when coordinates 4, 8, 12, 16, and 20 were not recognized. The module used at this time is autopy, and the environment was set with the pip install autopy command. AutoPy is a GUI automation library that can be used in a Python environment and provides functions to control the keyboard and mouse.

Fig. 3: Demonstration scene

Fourth, a field test was conducted. As shown in Fig.3, we visited the daycare center and received support for an environment in which children can participate in person. The tests were conducted in two ways. The first test is a quiz. We have created a problem in advance that the children will be interested in. When children asked a question, they were asked to choose the correct answer from among the views displayed on the screen. The test environment was configured to display the image coming through the computer's camera on the right side of the beam projector screen. The children who participated in the test were configured to increase their interest by allowing them to check their appearance and landmarks displayed on their hands on the left side of the screen. The children moved to the coordinates with the correct answer with their palms open, and when they clenched their fists, a mouse click event occurred, allowing them to select the correct answer and examining the children's reactions. The second test is a balloon popping game. The test environment was conducted in the same environment as the first test, and by dynamically creating

balloons on the right side of the screen, children manipulated the virtual mouse to pop the balloons to score points. Fig. 4 shows the mapping procedure for recognizing the shape of the hand and executing the virtual mouse using the MediaPipe Hand Tracking model in the video entered through the camera.

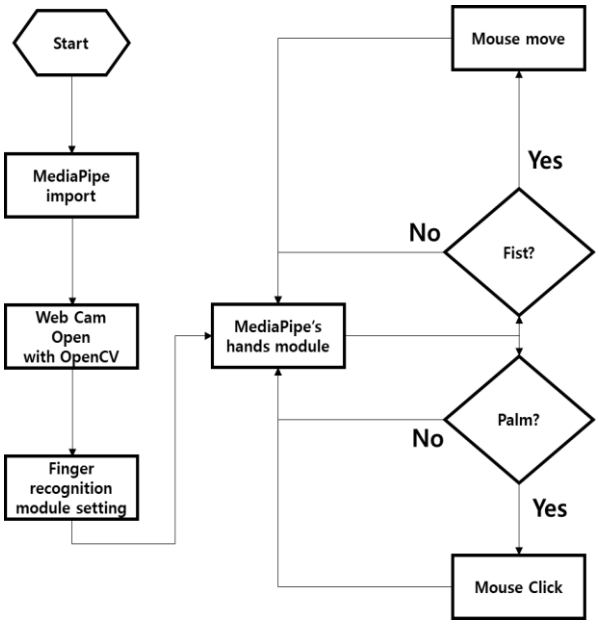

Fig. 4: Hand Tracking Process

Fig. 5 summarizes the conceptual diagram of the system constructed for this study. Among the modules of the media pipe, OpenCV and AutoPy modules were added based on the Hand module, and by using the k-NN algorithm as a machine learning algorithm, the hand shape from the image transmitted through the camera is recognized and mapped with a virtual mouse event so that infants and young children can Using the hand gestures of , the educational program was directly used. Toddlers were allowed to use educational programs while checking their movements through the beam project.

4. Conclusion

In this paper, we tried to check whether a program that operates a virtual mouse by recognizing the shape of a hand as an image can be applied to the educational field of children three to five years old through indoor nature exploration activities. In the test conducted on adults who are the developers of the program, it was confirmed that the movements and clicks of the virtual mouse were processed naturally according to the movement of the hands, but when applied to real children, an unexpected situation appeared. As a result of failing to consider the adult hand size and children's hand

size, the accuracy decreased as the distance between the child participating in the test and the camera increased. This is because, unlike adults who conduct tests in place, the characteristics of children moving freely in a wide space were not considered.

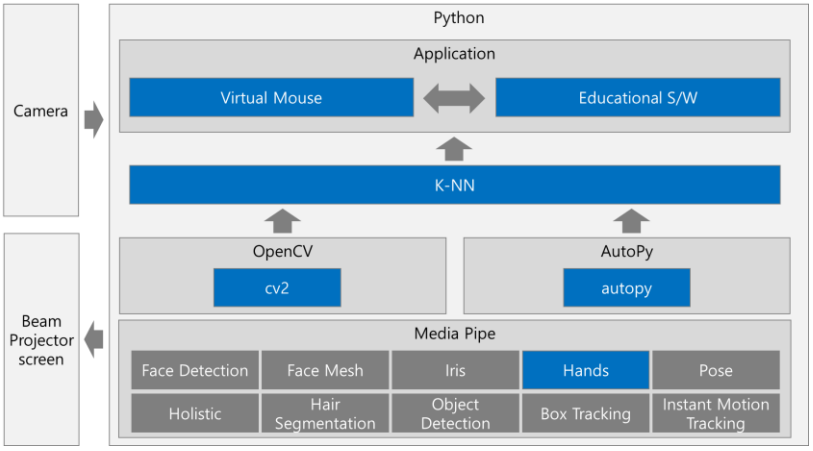

Fig. 5: System Schematic

However, it was confirmed that the closer the distance between the child and the camera, the higher the accuracy. Considering these results, children's curiosity about the virtual mouse was confirmed, but it can be concluded that the current model is somewhat difficult to apply directly to the educational field, so it needs to be supplemented. Therefore, in future research, if a hand recognition algorithm and model are developed that consider the distance and angle to the camera based on a child user, it is expected that the child's hand shape recognition performance will be improved and can be applied educationally. It is judged that it is necessary to study an algorithm that can recognize children's hand shapes even when several children participate. In addition, if children and young children are given the opportunity to experience ICT in all areas of the revised Nuri course, it is expected that positive effects can be obtained in both fun and educational aspects. In Table 2, 18 items that can be educated using a virtual mouse in the revised Nuri course are identified and displayed. In the future, additional research is needed to see if education using ICT is possible for 18 detailed educational courses classified based on the results of this study.

Table 2: Expandable targets in the Nuri Curriculum

Area	Category	Overview	Expansion target
Physical exercise and health	enjoy physical activity	Recognizes and moves the body	-
		control body movements	-
		Perform basic movement	-

		movements, stationary movements, and exercises using tools.	
		Voluntary participation in indoor and outdoor physical activities	-
	to live healthy	Clean yourself and your surroundings.	-
		Take an interest in food that is good for your body and eat it happily with the right attitude.	○
		Get adequate rest from your daily routine.	-
		Know and practice how to prevent disease.	○
	live safely	Play and live safely in everyday life.	○
		Use the TV, computer, smartphone, etc. properly.	○
		Observe traffic safety rules.	○
		Experience how to deal with accidents, fires, disasters, abuse, and kidnappings.	○
Communication	listening and speaking	Listens to words or stories with interest.	-
		Talk about your experiences, feelings, and thoughts.	-
		Speak using words appropriate to the situation.	-
		Listen to what the other person is saying and talk about it.	-
		Listen and speak with the right attitude.	-
		Use soft words	○
	Take an interest in reading and writing	Pay attention to the relationship between speech and writing.	-
		They are interested in reading the symbols and letters around them.	○
		Express your thoughts in a form similar to letters.	-
	enjoy books and stories	Interested in books and enjoys imagining.	-

		A fairy tale and at the same time feel the fun of words.	○
		Enjoy playing with horses and telling stories.	○
Social relations	know me and respect me	know me and cherish me	-
		I know my feelings and express them appropriately.	-
		I do what I can myself.	-
	live together	Know the meaning of family and live in harmony.	-
		Helping each other and getting along with friends.	-
		Resolve conflicts with friends in a positive way.	-
		Respect different feelings, thoughts and actions.	-
		Be courteous to friends and adults.	○
		Knowing and keeping promises and rules.	-
	take an interest in society	Ask questions about where you live.	-
		Take pride in our country.	-
		Interested in various cultures	○
Art experience	look for beauty	Feel and enjoy the beauty of nature and life.	-
		Interested in and looking for artistic elements	-
	be creative	enjoy singing	-
		Create simple sounds and rhythms with bodies, objects, and instruments.	-
		Express yourself freely through movement and dance using your body or tools.	-
		Express your thoughts and feelings with a variety of art materials and tools.	-
		Expressing experiences or stories through play	-
	appreciate art	Enjoys imagining and appreciating	○

		a variety of art.	
		Respect different artistic expressions.	-
		Become familiar with Korean traditional art	○
Nature search	Enjoy the exploration process	Be constantly curious about the world and nature around	-
		Have fun participating in the process of exploring questions	-
		Take interest in different ideas in the process of inquiry	-
	Exploring things in your daily life	Explore the properties and changes of objects in various ways	-
		Count objects to find out the quantity	○
		Recognize and distinguish the position, direction, and shape of objects	○
		Compare properties such as length and weight in daily life	○
		Look for repeating rules around	-
		Classify data collected from daily life according to criteria	-
		Be interested in tools and machines	-
	Living with nature	Pay attention to the flora and fauna around	○
		Value life and the natural environment	-
		Relate changes in weather and seasons to life	-

References

Al-akashi Falah. (2021). Improving Learning Performance in Neural Networks. *International Journal of Hybrid Innovation Technologies*, 1(2), 27-42, doi:10.21742/IJHIT.2021.1.2.02.

Barak, F. & Kaplan, K. (2021). The Study of Handwriting Recognition Algorithms based on Neural Networks. *International Journal of Hybrid Innovation Technologies*. 1(2), 63-74. DOI:10.21742/IJHIT.2021.1.2.04.

Gil S. H., Lee S. H., Oh C. Y., Yoo S. B., & Han Y. H. (2021). Design and implementation of a hospital sign language translation program using Deep Learning-based posture and hand motion recognition technology. *The Journal of Korean Institute of Communications and Information Sciences*. 2021(11), 1015-1016.

Heo G. Y., Song B. D., & Kim J. H. (2021). Hierarchical Hand Pose Model for Hand Expression Recognition, *Journal of the Korea Institute of Information and Communication Engineering*. 25(10), 1323-1329. DOI:10.6109/jkiice.2021.25.10.1323.

Hwang J. H. (2020). A study on the development of interaction English learning model for children at home using Artificial Intelligence (AI) speech recognition speakers. *Department of Education Graduate School Korea University*.

Jo W. J. (2021). Remote classes in early childhood education and the use of digital technology. *Journal of Korea Institute of Child Care and Education*. 69, 7-14.

Kim B. H. (2017). Early childhood teachers' perceptions and practices of ICT education by age of children. *Korean Association for Learner-centered Curriculum and Instruction*. 17(5), 287-314. DOI:10.22251/jlcci.2017.17.5.287.

Kim J. Y. & Sim H. (2021). Development of a Sign Language Learning Assistance System using Mediapipe for Sign Language Education of Deaf-Mutility. *The Journal of Korea Institute of Electronic Communication Sciences*. 16(6), 1355-1361. DOI:10.13067/KIECS.2021.16.6.1355.

Kim Y. J. (2010). The Effect of Creative Problem Solving Activity using ICT on Young Children's Creativity. *Korean Journal of Early Childhood Education Research*. 12(), 57-74.

Kwon S. H., Choi Y. S., Lim S. J., & Joung S. T. (2016). A Implementation of User Exercise Motion Recognition System Using Smart-Phone. *Journal of the Korea Academia-Industrial cooperation Society*. 17(10), 396-402. DOI:10.5762/KAIS.2016.17.10.396.

Lee J. E. & Oh S. K. (2021), The effects of activities using artificial intelligence speakers on the language skill of young children. *The Journal of Korea Open Association for Early Childhood Education*. 26(5), 185-208. DOI:10.20437/KOAECE26-5-08.

Lee K. Y., Lim M. J., Kim K. H., Lee M. K., & Kim J. L. (2011). A Study on the Virtual Mouse Interface System, *The journal of the Institute of Internet Broadcasting and Communication*, 11(2), 57-62. DOI:10.7236/JIWIT.2011.11.2.057.

Lee, Y. C. & Lee C. W. (2014). Motion Recognition of Smartphone using Sensor Data. *Journal of Korea Multimedia Society*. 17(12), 1437-1445. DOI:10.9717/kmms.2014.17.12.1437.

Lee, Y. J. & Kim T. Y. (2021). Development of an efficient home training system through deep learning-based pose recognition and correction. *The Journal of Korean Institute of Next Generation Computing*. 17(6), 89-99. DOI:10.23019/kingpc.17.6.202112.008.

Lim, E. M. (2021). Early childhood teachers' perception and needs on distance education after COVID-19. *The Journal of Learner-centered curriculum and Instruction*. 21(24), 631-645.

Noh H. S., Kim J. J., Won J. U., Lim H. H., Li J. Y., & Jung S. H. (2021). Hand Tracking Kiosk System Using Mediapipe. *Journal of the Korea Institute of Information and Communication Engineering*. 28(2), 1080-1183. DOI:10.3745/PKIPS.Y2021M11A.1180

Oh H. J. & Lim B. Y. (2021). An Exploration on Educational Values of Outdoor play in Early Childhood Education With the Age of Covid-19. *The Journal of Learner-centered curriculum and Instruction*. 21(1), 1237-1267. DOI:10.22251/jlcci.2021.21.1.1237 .

Roh M. C. (2017), A virtual mouse interface with a two-layered Bayesian network, *Multimedia tools and applications*, 76(2), 1615-1638. DOI:10.1007/s11042-015-3144-x

Seo M. H. & Choi W. Y. (2003). A Study on the Virtual Mouse Interface System. *Journal of the Research Institute of Industrial Technology*. 22, 84-89.

Seo M. H. & Choi W. Y. (2003). Development of virtual mouse in LCD beam projector environment. *Journal of the Research Institute of Industrial Technology*. 23, 199-202.

Yoo G. J. and Kim S. R. (2021). Interaction analysis and pattern between Artificial Intelligence (AI) speakers and infants. *The Journal of Learner-centered curriculum and Instruction*. 26(5), 209-244. DOI:10.20437/KOAECE26-5-09.

Yu Y. J., Moon S. H., Sim S. J., & Park S. H. (2020). Recognition of License Plate Number for Web Camera Input using Deep Learning Technique. *Journal of Next-generation Convergence Technology Association*. 4(6), 565-572. DOI:10.33097/JNCTA.2020.04.06.565.