RESEARCH ARTICLE

Finger Joint Angle Estimation With Visual Attention for Rehabilitation Support: A Case Study of the Chopsticks Manipulation Test

ADNAN RACHMAT ANOM BESARI [1,2], **(Graduate Student Member, IEEE),**
AZHAR AULIA SAPUTRA[1], **(Member, IEEE), WEI HONG CHIN**[1], **(Member, IEEE),**
KURNIANINGSIH[3], **(Senior Member, IEEE),**
AND NAOYUKI KUBOTA[1], **(Senior Member, IEEE)**

[1]Graduate School of Systems Design, Tokyo Metropolitan University, Tokyo 191-0065, Japan
[2]Department of Information and Computer Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya 60111, Indonesia
[3]Department of Electrical Engineering, Politeknik Negeri Semarang, Semarang 50275, Indonesia

Corresponding author: Adnan Rachmat Anom Besari (anom@pens.ac.id)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical, Legal, and Social Issues (ELSI) of Tokyo Metropolitan University, Japan, under Approval No. H21-051.

ABSTRACT Most East Asian rehabilitation centers offer chopsticks manipulation tests (CMT). In addition to impaired hand function, approximately two-thirds of stroke survivors have visual impairment related to eye movement. This article investigates the significance of combining finger joint angle estimation and a visual attention measurement in CMT. We present a multiscopic framework that consists of microscopic, mesoscopic, and macroscopic levels. We develop a feature extraction technique to build the finger kinematic model at the microscopic level. At the mesoscopic level, we propose an active perception ability to detect the position and geometry of the finger on the chopsticks. The proposed framework estimates the proximal interphalangeal (PIP) joint angle on the index finger during CMT using fully connected cascade neural networks (FCC-NN). At the macroscopic level, we implement a cognitive ability by measuring visual attention during CMT. We further evaluate the proposed framework with a conventional test that counts the number of peanuts (NP) which are moved from one bowl to another using chopsticks within a particular time frame. We introduce three evaluation indices, namely joint angle estimation movement (JAEM), chopstick attention movement (CAM), and chopstick tip movement (CTM), by detecting the local minima and maxima of the time series data. According to the experiment results, the velocity of these three evaluation indices could indicate improvement in hand and eye function during CMT. We expect this study to benefit therapists and researchers by providing valuable information that is not accessible in the clinic. Code and datasets are available online at https://github.com/anom-tmu/cmt-attention.

INDEX TERMS Hand-eye interaction, rehabilitation evaluation, first-person vision, eye-gaze, eye-tracking.

I. INTRODUCTION

Chopsticks have become part of the culinary experience in Japan, China, Korea, and worldwide. Historians have mapped their evolution from simple cutlery in ancient times

The associate editor coordinating the review of this manuscript and approving it for publication was Kang Li.

to their status as a symbol of today's culture [1]. Not only in Asian restaurants, but chopsticks are also available in most rehabilitation centers. Occupational therapists use pencils and chopsticks to improve children's fine motor skills [2]. The therapist guides the patient to practice various chopstick activities to improve hand function [3]. In the past, early researchers conducted the chopsticks manipulation test

TABLE 1. The research on hand measurement systems using chopsticks.

No	Research	Sensor Types	Methods	Applications
1.	Okamoto and Yanai (2014) [23]	Smartphone camera	Detect the chopsticks line by probabilistic Hough transform combined with background subtractor.	Real-time eating action recognition system on a smartphone to monitor eating actions during mealtime.
2.	Chia and Saakes (2014) [18]	Potentiometer and accelerometer	Measure the angle between the two sticks by potentiometer and estimate the orientation of the chopsticks by measuring tilt.	Interactive training games that use the chopsticks as a controller to help children develop the skills of eating with chopsticks.
3.	Nakamura et al. (2016, 2017) [19][34]	Accelerometer and gyroscope (IMU)	Detect synthetic acceleration and angular velocity to recognize timing, speed, and eating action (open, close, raise, eat and drop).	Design the rapid prototyping platform for sensoring chopstick activity (SenStick)
4.	Bosch et.al (2018) [38]	Eye-tracker	Measure behavioral performance and visual strategies from eye movement during skill acquisition using motor learning task.	Estimates efficient ways to manipulate objects (moving marbles) with novel tools like chopsticks.
5.	Sawamura et al. (2019) [20]	Near infra-red spectroscopy (NIRS)	Detect brain activity during chopstick operation using statistical analysis.	Examines the relationship between the two chopstick-operation skills (speed and smoothness) with non-dominant hand and brain activity.
6.	Shimomura et. al (2020) [22]	Electromyography (EMG)	Measure task execution times for following muscles: first dorsalis interosseus, flexor pollicis brevis, flexor digiti minimi brevis, flexor digitorum superficialis, and extensor digitorum.	Understand the significance of the traditional manipulation pincers-mode compared to scissors-mode with the non-dominant hand. It has an advantage in terms of high task performance and learning motor control.
7.	Yokubo et al. (2020) [21]	Pressure sensor	Detect pinch force of the chopstick tips by using pressure sensors inside the object.	Examine the effectiveness of chopsticks with different cross-sectional shapes of the upper and bottom chopsticks by focusing on opening operation.
8.	Nakaoka et al. (2021) [24]	Accelerometer-gyroscope (IMU) and camera	Detect the peak in signal to measure eating interval and speed and detect the color and amount of bite object using cameras.	Interactive coloring picture with chopsticks (eat2pic) to encourage people to slow their eating pace and encourage healthier eating habits.
9.	Wang and Zhang (2021) [26]	Camera	Detect correct or wrong behaviors in the use of chopsticks using the HSI color model.	Improved design of chopsticks for dining reminders and construction of civilized table manners to explore a new way of healthy eating.
10.	Namiki and Makino (2021) [25]	Camera and contactless motion sensor	Acquires a trainee's holding posture, then evaluates the motions and represents the correct chopstick activity.	Training support system using augmented reality (AR) technology of holding chopsticks both for beginners and people with the wrong holding style.
11	Our proposed frameworks (2022)	Smart glasses & eye tracker	Estimates finger joint angle by detecting chopsticks features in relation to visual attention.	Perform CMT assessment by measuring finger and eye movement for inexperienced chopsticks user.

(CMT) to assess the skill components of the therapeutic chopsticks task [4] and their relationship with the hand function tests. The CMT has been widely used to evaluate hand dexterity when conducting tabletop activities.

Adults potentially face a decline in hand function after 65 years of age. This decline is primarily due to age-related degenerative changes in the musculoskeletal, vascular, and nervous systems [5]. This issue is often followed by several pathological conditions such as stroke [6] and other neurological diseases. Researchers have examined whether chopsticks manipulation among stroke patients can increase levels of eating independence [7]. On the other hand, about two-thirds of stroke survivors have a visual impairment [8] related to visual field loss, double vision, and perceptual problems. This problem impacts the difficulty some people experience in gripping chopsticks in everyday life [9]. As a result, CMT measurements become less valid when visual attention is not considered [10]. The combination of hand function and eye-tracking measurements is necessary to identify. The upper limbs and eye movement coordination cannot be separated because they support each other in the

study of hand-eye coordination [11]. Recent studies show that gaze features [12] can be helpful in the identification of hand-eye coordination issues. It is necessary to define human visual attention and hand movement measurement during CMT, especially for a stroke patient with hand-eye movement problems. Table 1 shows the research on hand measurement systems using chopsticks.

Research on egocentric videos [13] has the potential to help this study by incorporating the hand image data of the first-person vision. Egocentric, first-person vision records the user's point of view [14] from a wearable camera. An egocentric video provides detailed information on hand posture, movements, and the objects or environment in which the hand interacts. The advantage of egocentric vision is reduced user privacy concerns [15], the ability to monitor people when they are mobile [16], and finding attention during activities [17]. Currently, egocentric vision research has begun to be developed for hand rehabilitation [6]. Unlike egocentric vision, visual attention refers to maintaining the awareness of visual scene's specific locations, objects, or attributes. Visual attention is usually obtained from sensors that detect the
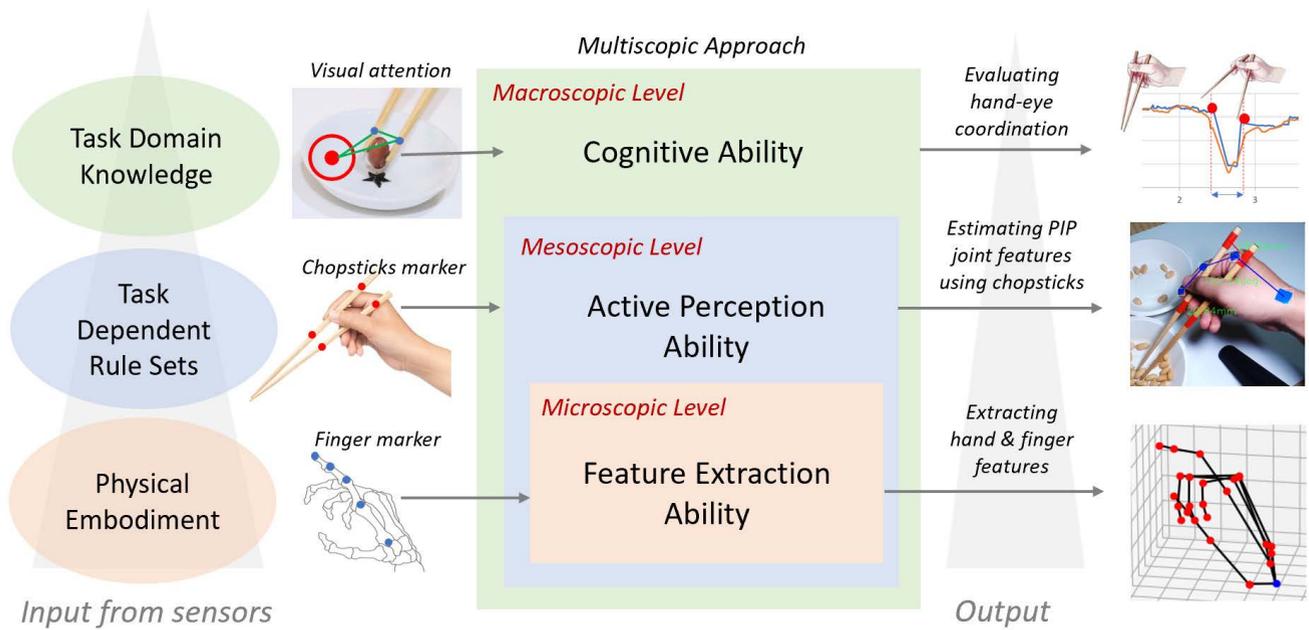
FIGURE 1. Design of multiscopic framework for chopsticks manipulation test with visual attention.

direction of the eyes and map a point on an image plane. Egocentric vision and visual attention shape human perception when performing activities requiring hand-eye coordination. However, its application has not been found in the case of CMT, which considers hand-eye coordination.

Therefore, our research aims to integrate finger measurement and visual attention for CMT assessment using a multiscopic framework. The main contributions of this paper are as follows. First, we apply a CMT measurement using egocentric vision-based feature extraction with marker detection. This framework is adopted to find the model parameters that efficiently describe a physical embodiment of CMT at the microscopic level. Second, we develop the active perception ability of hand-object interaction based on a task-dependent rule set. We use fully connected cascade neural networks (FCC-NN) to estimate the angle features for the proximal interphalangeal (PIP) joint angle on the index finger based on the chopsticks' positions. Third, we propose a system with a cognitive ability at a macroscopic level that can measure visual attention in relation to the chopstick tip. We use eye tracker glasses to detect where the eye is looking, then compare it with chopstick tip movements. We evaluate hand-eye coordination by symbolics hand-object interaction based on task domain knowledge at the macroscopic level. Figure 1 shows the design of a multiscopic framework for the CMT with visual attention.

This paper is structured as follows. Section 2 reviews the related works on the use of technology for CMT. Section 3 proposes our method for developing the multiscopic framework. Section 4 examines the results and discusses the effectiveness of the proposed method. Finally, section 5 presents the conclusions and suggestions for future works in relation to this research.

II. RELATED WORKS

A few studies related to CMT measurement have been conducted to examine the complexity of finger measurement analysis. Sensors as contact methods have been developed to support CMT measurement. Chia and Saakes [18] installed movement sensors on chopsticks to interact with children with a tablet. Nakamura *et al.* used an inertia measurement unit to obtain the chopsticks' state on their Senstick [19]. For specific CMT research, Sawamura *et al.* [20] utilized near-infrared spectroscopy (NIRS) to obtain measurements of brain activity in chopsticks-operation skills with the non-dominant hand. Yokubo *et al.* [21] employed force sensors to determine the relationship between the chopstick tips and cross-sectional shape in developmental stages from infancy to early school years. Shimomura *et al.* [22] applied electromyography to measure how the motor controls the hand muscles, clarifying the effect of chopstick manipulation.

The development of cameras as non-contact methods has begun to be used to detect the chopsticks' movement. Okamoto and Yanai [23] use a smartphone camera to monitor chopstick actions during mealtime. Nakaoka *et al.* [24] used a camera attached to chopsticks for coloring paintings to encourage healthier eating habits. Namiki and Makino [25] implemented a camera and a contactless motion sensor with augmented reality to help beginners learn how to hold chopsticks. Wang and Zhang [26] designed smart chopsticks based on HSI color space recognition for dining reminders.

In other fields, psychology researchers believe that good eye movement also supports human hand skills [27]. Inspired by the nature of human vision, they found that humans learn from seeing things, and hand skills are influenced by their perception ability [28]. In other words, human dexterity in using tools [29] depends on hand-eye coordination [30], [31].
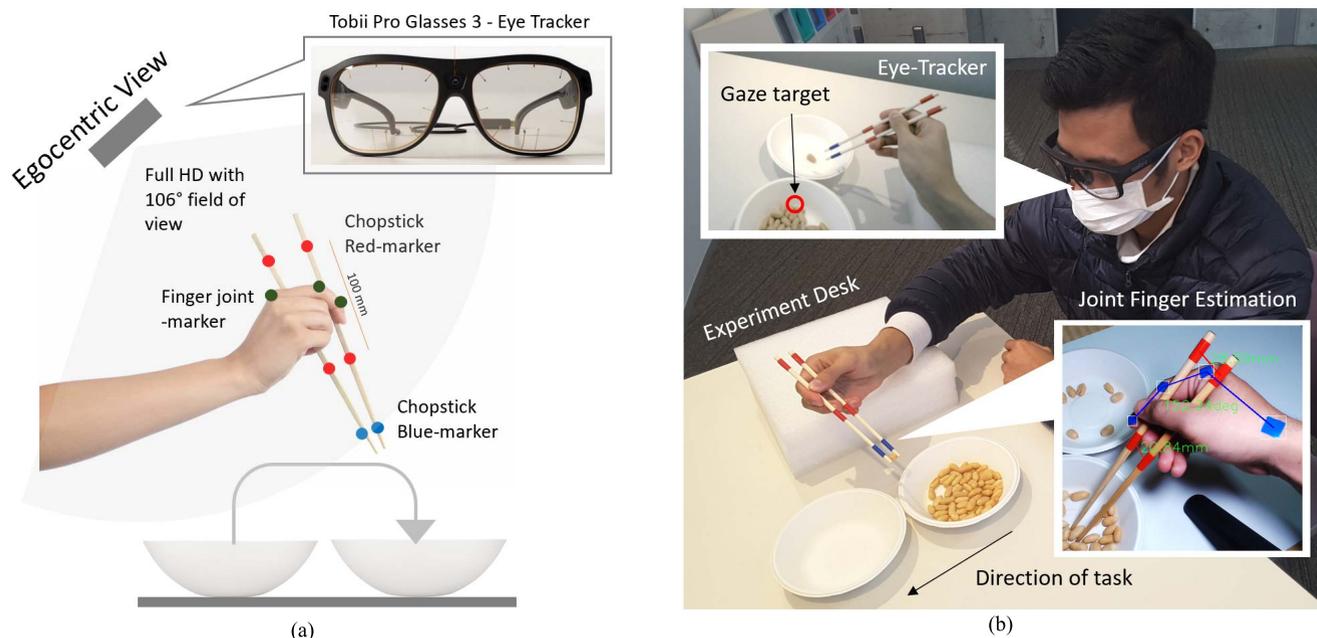
FIGURE 2. Design of chopsticks and visual attention measurement system: (a) Egocentric views system designs; (b) Experimental setup for CMT.

So, the measurement method is more effective when considering visual attention. These two capabilities must mutually support each other to achieve the goal—the eye functions as a sensor coordinate with the hand functions as an actuator. Thus, a suitable measurement method for detecting a hand object manipulation function must consider eye performance. One measurable function is tracking eye movements [32], which implies visual attention.

Currently, it is challenging for therapists to practice CMT following the COVID-19 pandemic. Tele-rehabilitation using CMT is the right choice for monitoring hand function decline and measuring outcomes in patients independently. Several studies have applied cognitive-based measurement systems [33] to monitor daily activity, employing contact methods such as sensors on chopsticks [34]. However, the application of tele-rehabilitation is limited due to privacy issues at the social level, and the complexity of hand measurement analysis causes lower accuracy than the contact method [35] at the physical level.

Many activities involving hand-eye coordination still need to be investigated. Researchers have discussed the relationship between eye and hand movements. Sports science research [36] found that visual attention can indicate the performance of hands in basketball shooting. Another study compares eye movements with finger and hand movements in a mouse-click activity [37]. This study was in line with research on human vision, which shows that eye movement always precedes finger movement, especially the index finger [38]. Changes in eye movement were represented by actions of the point of visual attention, and hand movements were defined by changes in the PIP joint angle of the index finger.

Bosch *et al.* [39] showed that gaze features could be helpful in the identification of behavioral performance and visual strategies during chopstick skill acquisition. However, this study does not estimate the effect of visual attention on finger movement during chopstick manipulation. Thus, it is necessary to define human visual attention and hand movement during CMT, especially for a person with hand-eye movement problems [11]. CMT with visual attention is a new measurement framework that considers eye and finger movement when using chopsticks. Combining finger joint angle measurements and eye movements determines which factors are more dominant in influencing CMT results [38].

Several examples show that the current CMT uses several elements of technology, especially movement sensors, body sensors, and cameras. However, researchers rarely discuss the application of CMT that considers human visual attention as an additional parameter. Designing the CMT framework with visual attention is necessary to support hand function performance. The technology to combine hand measurements and visual attention can be easily captured on smart glasses with eye-tracking features [40]. A physiology study showed that task-oriented tabletop activity training effectively restored upper limb function and visual perception in acute stroke patients [41]. However, machine learning algorithms to process multivariate data incur a computational cost in cyberspace. Combining different methods to deal with the above problems becomes a significant challenge. One appropriate way to address this issue is multi-scope processing [42], which is performed in stages.

Our previous work used a multiscopic approach to address dynamic locomotion in a legged robot [43]. Based on this
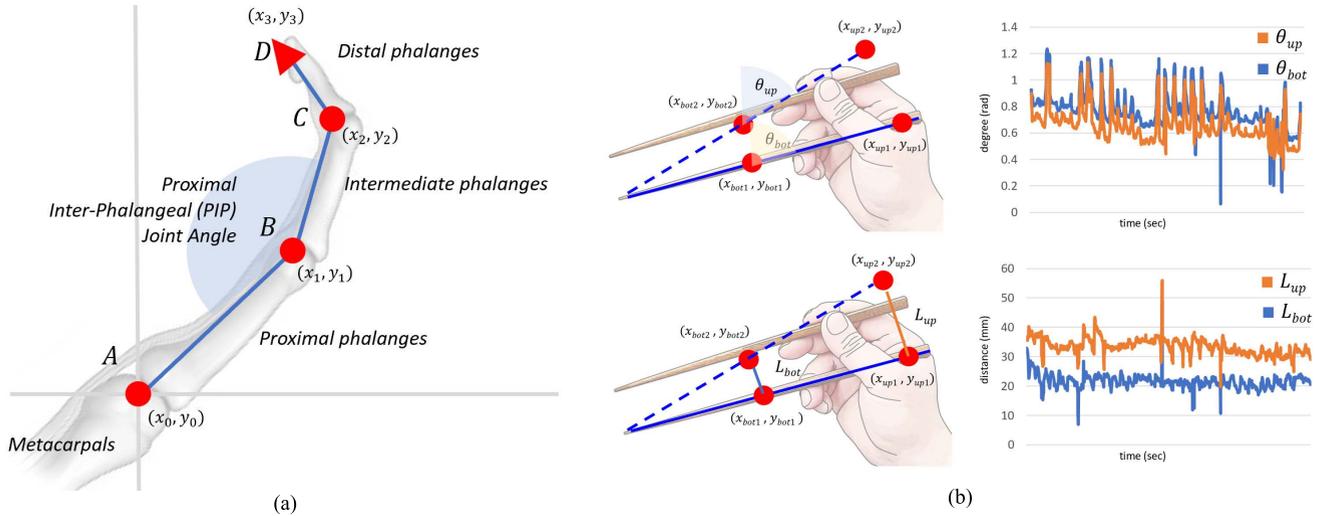
FIGURE 3. Microscopic level: feature extraction process in CMT (a) Finger kinematic model; (b) Chopsticks state model.

experience, we propose a multiscopic framework for developing a CMT finger measurement system with visual attention. Figure 2 shows the design of the chopsticks and the visual attention measurement system. This proposed model integrates intelligent functions from a multiscopic point of view. This approach includes a) modeling finger kinematics and the chopsticks' state for feature extraction at the microscopic level, b) estimating the PIP finger feature [44] using an active perception ability at the mesoscopic level, and c) discovering the human visual attention using cognition ability at the macroscopic level.

III. PROPOSED FRAMEWORK

These sections discuss the multiscopic approach that is used in CMT with attention. We explain the microscopic, mesoscopic, and macroscopic levels in three sub-sections. This section describes our method: setting up the experiments, extracting finger features, detecting the chopsticks' state, estimating the joint angle, and measuring visual attention during the CMT.

A. MICROSCOPIC LEVEL: FEATURE EXTRACTION

Feature extraction is closely related to the data acquisition process and retrieves the attribute. Figure 3 shows the microscopic level: the feature extraction process in CMT, including finger kinematic and the chopsticks' state models. We propose a finger kinematic model and chopsticks state model to represent finger action in CMT to obtain the physical embodiment. In this section, we describe both models in detail.

1) FINGER KINEMATIC MODEL

Finger kinematic models [45] have been widely studied in various research, especially in the industrial sector, such as hand modeling and simulation, hand motion coordination, articulated human hands, prosthetic hands, and robotics. Our previous work [46] used a stereo infrared (SIR) camera from

Leap Motion to capture the position of the hand and fingers. This sensor tracks fingers in a 3D zone extending from 10 cm to 75 cm and from the device in a 170 ° × 170 ° ordinary field of view. The infrared stereo vision sensor takes every joint position and transforms it into a skeleton at 100 fps. This sensor is connected to a single application for hand visualization in the Cartesian space. After setting up the desktop view, the images are processed to remove noise and produce finger models [47]. This sensor is able to reduce occlusion in the finger position with reasonable accuracy. We describe these stages to classify grasp poses taxonomy from an egocentric view.

A finger has three joints, the first joint *A* connecting the metacarpal and proximal, the second joint *B* connecting the proximal and intermediate phalanges, and the third joint *C* connecting the intermediate phalanges and distal phalanges. The thumb does not have metacarpals. We transform the data into a straightforward representation using finger kinematic models. Figure 3 (a) shows the finger kinematic model. This model is used as a joint rotation configuration [48], which is included in this framework to validate individual-specific finger motion precisely.

We derive a formula to calculate the joint angle of the finger based on the physical relationship of human body. To simplify the data, we acquire a feature to replace the joint's position by obtaining the angle formed by the joints between two bones. We perform three points in the vector-to-angle transformation to obtain these features. If we have Cartesian coordinates for three points, *A*, *B,* and *C,* then we can calculate $\vec{\mathbf{AB}}$ and $\vec{\mathbf{BC}}$. The following are the equations to obtain two vectors from three points in the 3D coordinates:

$$\vec{\mathbf{AB}} = B - A \tag{1}$$
$$\vec{\mathbf{BC}} = C - B \tag{2}$$

Next, we calculate the angle formed by $A \rightarrow B \rightarrow C$ using the *right-hand rule* from *B.* The following is an equation for

the dot product:

$$\overrightarrow{AB} \cdot \overrightarrow{BC} = \left\| \overrightarrow{AB} \right\| \left\| \overrightarrow{BC} \right\| \cos\theta \qquad (3)$$

where $\left\| \overrightarrow{AB} \right\|$ is the length of $\overrightarrow{AB}$, $\left\| \overrightarrow{BC} \right\|$ is the length of $\overrightarrow{BC}$, and θ_{PIP} is the angle between these two vectors. In this way, we can calculate θ_{PIP} by substituting the equation. The following is an equation to find θ_{PIP}:

$$\theta_{PIP} = \arccos\left(\frac{\overrightarrow{AB} \cdot \overrightarrow{BC}}{\left\| \overrightarrow{AB} \right\| \left\| \overrightarrow{BC} \right\|} \right) \qquad (4)$$

We use the standard hand pose in this CMT framework while holding chopsticks. Only the index and middle fingers are used to move the upper chopsticks, and the thumb, ring finger, and little finger are used to hold the lower chopsticks tightly. The position of all fingers in a Cartesian coordinate would lead to many possible joint positions. We record the index finger's trajectory and calculate the PIP joint angle in the chopsticks' activity. The movement of joint positions shows PIP's range of motion (ROM) during CMT. After obtaining the finger position in the Cartesian coordinates, we record each joint's minimum and maximum positions. With this geometric representation of the finger, we are able to represent the kinematics and its motions.

We use two red markers on a single chopstick. The first marker is attached to the bottom handheld chopstick (x_{bot}, y_{bot}), and the second is to the upper (x_{up}, y_{up}). The distance between these two markers is 100 mm. These markers assist in obtaining the angle and length between two upper and two bottom features [49]. These features are essential to obtain the characteristics of the chopsticks' activity. Figure 3 (b) shows the angle and distance between the chopsticks' upper and bottom features. We calculate the estimated upper angle θ_{up} and bottom angle θ_{bot} feature by:

$$\theta_{up} = \arctan\left(\frac{|x_{up1} - x_{bot1}|}{|y_{up1} - y_{bot1}|} \right) \qquad (5)$$

$$\theta_{bot} = \arctan\left(\frac{|x_{up2} - x_{bot2}|}{|y_{up2} - y_{bot2}|} \right) \qquad (6)$$

where x_{bot1}, y_{bot1} and x_{bot2}, y_{bot2} are the respective coordinates of the bottom chopsticks features position then x_{up1}, y_{up1} and x_{up2}, y_{up2} are the respective coordinates of the upper chopsticks features position. This approach is beneficial in reducing the computation in hand recognition processing [50], for example, detecting poses, actions, or gestures.

2) CHOPSTICKS STATE MODEL
In the CMT framework, we use red chopsticks markers and blue finger markers. A red marker is attached to the chopsticks for finger joint angle estimation by obtaining ground truth data to minimize errors. We perform marker detection on the index finger because it is visible in egocentric vision [6], while the middle finger is closed. During the testing stage, the blue marker is removed so that participants feel comfortable

when collecting data. In the mesoscopic setting, this blue marker is used as a marker for the tips of the chopsticks. We obtain the finger features by connecting the center of the blue marker with the center of the nearest blue marker to form a skeletal finger. We use smart glasses as non-contact wearable sensors on an egocentric view. [13] and set them up from the participant's perspective. We install smart glasses with an average of 75 ° facing down to obtain the best results.

After setting up the egocentric view, we prepare the algorithm to detect chopsticks and finger markers. We use HSV color extraction and median-dilate filtering to remove the noise, strengthen the features, and detect the contours. There are two things we can obtain from these feature extraction processes. First, we can identify the feature in a specific image, and second, we can determine the exact location of the feature within the image. We measure the length between these two upper markers and two bottom markers with a pair of chopsticks. We are able to distinguish between the pinch and move mode by measuring two parts of the length of the chopsticks. These features are essential to obtain the characteristics of the chopsticks' activity. We calculate the estimated upper length L_{up} and bottom length L_{bot} feature as follows:

$$L_{up} = \frac{\sqrt{(x_{up1} - x_{up2})^2 + (y_{up1} - y_{up2})^2}}{\sqrt{(x_{up2} - x_{bot2})^2 + (y_{up2} - y_{bot2})^2}} .100 \text{ mm} \qquad (7)$$

$$L_{up} = \frac{\sqrt{(x_{bot1} - x_{bot2})^2 + (y_{bot1} - y_{bot2})^2}}{\sqrt{(x_{up1} - x_{bot1})^2 + (y_{up1} - y_{bot1})^2}} .100 \text{ mm} \qquad (8)$$

The above equation shows how we calculate the length using the Pythagorean formula. In simple terms, the above calculation compares the distance between the center of the marker on one stick of the chopsticks in pixels with the actual marker length. We can take this result of the microscopic level process to the next level, as discussed in the next section.

B. MESOSCOPIC LEVEL: ACTIVE PERCEPTION
Active perception is the ability to capture, process, and actively make sense of the information at the mesoscopic level. To determine an active perception ability, we design a task-dependent rule set using the finger state from feature extraction at the microscopic level. Task dependency is a relationship in which tasks depend on regulations to be performed in a specific order developed in our previous study [51]. This task relates to CMT activities where the fingers interact with the chopsticks. To ensure that the index finger is measured with precision, we attach markers to each joint of the index finger. These markers are connected one by one to form the skeletal finger. However, the marker affixed to the finger is uncomfortable for CMT participants. Therefore, we estimate the PIP joint angle with supervised learning to remove the marker on the finger.

This idea is inspired by the problems associated with therapy assessment, which sometimes makes the patients feel

uncomfortable when sensors or markers are attached to their hands. We chose seamless monitoring in the CMT process by estimating the PIP joint angle with supervised learning using the chopsticks feature to remove the marker on the finger. Our framework prioritizes two main sub-processes: pinching mode (including grip) and moving mode (including hold and release). To separate the two sub-processes, we use the angle inclination parameters on the upper and lower chopsticks (θ_{up}, θ_{bot}) together with the upper and bottom length features (L_{up}, L_{bot}). Figure 4 shows the mesoscopic level: feature extraction to active perception and symbolic cognition.

We use Cheng-type FCC-NN [52] and compare it with conventional architecture such as multi-layer perceptron (MLP). We apply the upper length (L_{up}) and bottom length (L_{bot}) as the input of neural networks to obtain the PIP joint angle (θ_{PIP}) as output. We collect data for 1 minute for all chopstick manipulation poses in different directions. In FCC-NN, a hidden layer is limited to only one node. If we add a new node, it will increase the number of hidden layers. So, the number of hidden layers is always the same as the number of nodes. We employ FCC-NN consisting of one hidden layer as the learning architecture. It has a fully connected layer characteristic, where each neuron is connected to the others. It is different from MLP, in which nodes in the N layer are given input only from the output nodes in the N-1 layer. In FCC-NN, nodes in each layer obtain the input from all output nodes in the previous layer (N-1, N-2, ...). The output of the k-th neuron in FCC-NN is caculated by:

$$x_k = f(z_k) \qquad (9)$$

$$z_k = \sum_{i=0}^{k-1} w_{i,k} x_i + b_k; (i = 1) \qquad (10)$$

From the above equation, the input z_k is processed by an activation function f. The output is x_k, where k is the number of nodes n. We have two layers in the FCC-NN design, so the first layer (layer-1) is the hidden layer that consists of 1 hidden node x_3. The output of this hidden node (x_3) is obtained from x_1 and x_2. Each of the variables is multiplied by $w_{1,3}$ and $w_{2,3}$ with the following equation:

$$x_3 = f(x_1 * w_{1,3} + x_2 * w_{2,3}) \qquad (11)$$

The second layer (layer-2) consists of three input nodes (x_1, x_2, x_3) with the following equation:

$$y = f(x_1 * w_{1,4} + x_2 * w_{2,4} + x_3 * w_{3,4}) \qquad (12)$$

The output node (y) in the output layer is obtained from x_1, x_2, x_3 each of which are multiplied by w_{14}, $w_{2,4}$ and $w_{3,4}$. We determine the backpropagation with a backward process on Cheng-type FCC-NN.

We use the Savitzky–Golay filter, which was successfully implemented on EEG [53], to refine the PIP joint angle data from the error spike. This filter is widely used to increase the time series data's precision without distorting the data's trend. This process is essential to detect the slightest signal

tendency during the chopsticks' movement. The following is the equation for the Savitzky–Golay filter:

$$Y_j = \sum_{i=\frac{1-m}{2}}^{\frac{m-1}{2}} C_i y_{j+i}, \frac{m+1}{2} \le j \le n - \frac{m-1}{2} \qquad (13)$$

The data consists of an array set representing points $\{x_j, y_j,$ where $j = 1, \ldots, n$. The symbol x_j is an independent variable (time) and y_j is an observed value (joint angle). They are treated with a set of m convolution coefficients C_i, according to the expression. We choose the number of coefficients and the polynomial order to fit the samples. These hyperparameters are selected because they are the most suitable base on several experiment results.

In CMT, there are two finger movement postures: pinch and move mode [54]. We found local minima and maxima to detect these modes by comparing neighboring values. We use two parameters in this function, namely height, and distance. The height parameter is used to determine the minimum peak height. We choose a height parameter of Root Mean Square (RMS) as the pinching and moving mode threshold to separate the peaks and the valleys. The minimum horizontal distance required in the sample between adjacent peaks in the distance parameter. This parameter eliminates the smaller peaks first until the condition is met for all remaining peaks.

After distinguishing the pinching and moving peaks, we propose joint finger measurement values in CMT [20], namely Joint Angle Estimation Movement (JAEM). This movement parameter includes a variable, namely the T_{JAEM} which is the average time required for chopsticks to open and close in milliseconds. We can obtain the T_{JAEM} by substracting the nearest peaks and valleys time peaks $T_{V.JAEM}$ with the time peaks $T_{P.JAEM}$, base on the following equation:

$$\Delta T_{JAEM} = \sum_{n=0}^{N} T_{V.JAEM}(n) - T_{P.JAEM}(n) \qquad (14)$$

Then, the ROM_{JAEM} PIP joint angle is estimated for chopsticks when open and close in degrees (°). This second variable is obtained by finding each signal's average distance from peak to valley. Finally, we determine ω_{JAEM}, the angular velocity of JAEM, by dividing the change of its Range of Motion (ROM_{JAEM}) by the change of T_{JAEM}. The following is the equation for finding ω_{JAEM}:

$$\omega_{JAEM} = \frac{\Delta ROM_{JAEM}}{\Delta T_{JAEM}} \qquad (15)$$

The ω_{JAEM} show how fast the PIP joint angle changes during CMT in angle degree per millisecond (°/ms). This variable measures the development of a person's finger motor function during rehabilitation with CMT.

We design symbolic cognition from the task domain knowledge at this level to support our previous study [55]. This task is related to CMT activities where the fingers interact with chopsticks. After obtaining the angle value in the upper and bottom of the chopsticks, we detect the two
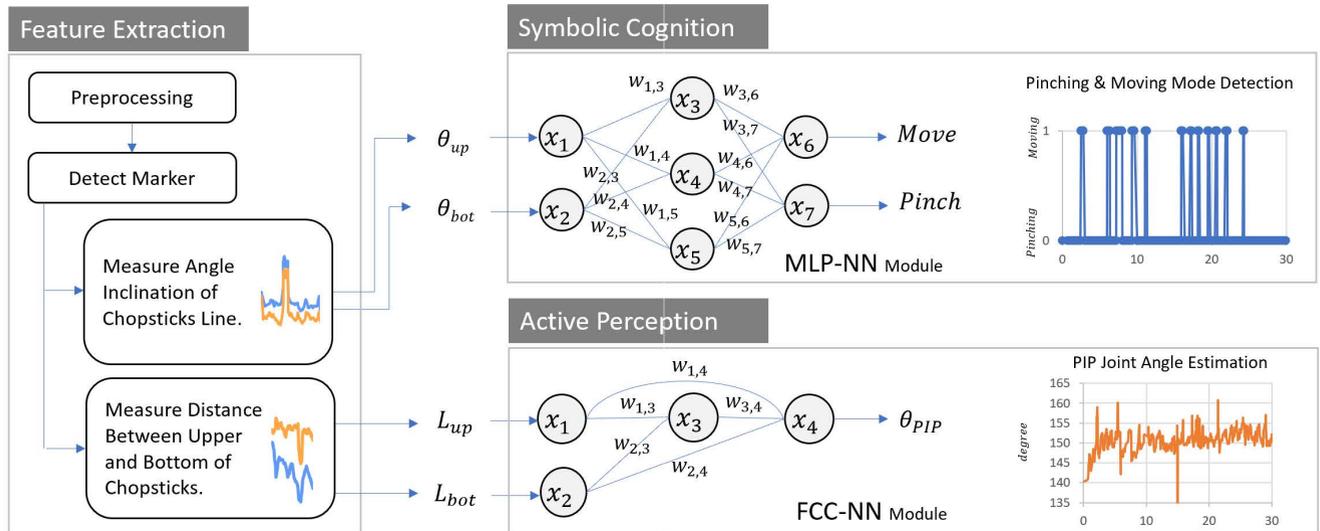
FIGURE 4. Mesoscopic level: Feature extraction to active perception and symbolic cognition.

chopsticks' modes with supervised learning. We use the MLP module as the learning architecture. The input layer contains two nodes ($\theta_{up}, \theta_{bot}$), while the output layer has two nodes (y_0, y_1), a class of chopsticks grip poses (pinching and moving mode). We use fully connected networks, so every unit receives a connection from all the units in the previous layer. Our model uses one hidden layer consisting of three hidden nodes.

We employ a training loop similar to previous models: we establish an optimizer, feed the inputs to the model, calculate the loss, and then optimize it using the autograd function. The following are the training data steps: (a) standardize the data structure so that it is accepted as an MLP input with node and weight properties, (b) perform a linear transformation on the incoming data, and (c) employ the sigmoid function element-wire during the learning process. (d) The SoftMax layer generates the final classification and decides. The aim is to determine which input data class belongs to the two CMT pinch or move mode categories. Each node in the output layer receives output from the model. We obtain information via the SoftMax function.

C. MACROSCOPIC LEVEL: COGNITIVE ABILITY

Cognitive ability involves human attention to a particular object and expanding on it over an extended time. We design this ability from the task domain knowledge at the macroscopic level. This task is related to CMT activities where the chopsticks' movement is based on visual attention. Figure 5 shows the macroscopic level: cognitive ability design using eye movement and chopstick tip tracking. The direction of attention is usually focused on the object and the tip of the chopsticks. The movement of the chopstick tip also follows the person's attention. We utilized Tobii Pro Glasses 3 to get attention on the image plane [56]. This intelligent eyeglass gives us information about the attention location represented

by a red circle in the image area. The location of the center of attention can rapidly change according to the movement of the eyes. We also put blue markers on both tips of the chopsticks to determine the location of the chopstick tips in the image coordinate.

We detect the red circles and the blue marker using preprocessing techniques as the feature extraction at the microscopic level. We apply HSV color extraction, median-dilate filtering, and contour detection [49]. Then, we obtain the centroid of the attention in the x and y coordinates of the image. The placement of the peanut bowl is made vertically. The source bowl is located at the bottom, and the target bowl is above the image plane. We simplify the movement tracking on the y coordinate.

This system gives two output signals. First, the red signal indicates the Chopstick's Attention Movement (CAM), while the blue signal specifies the Chopstick's Tip Movement (CTM). At first glance, these two signals look the same because the attention movement moves almost simultaneously with the tip of the chopsticks. We measure the time required to perform CAM and CTM and compare these data. These two evaluation indices simultaneously determine the speed of the hand and eye movements. Figure 5 shows the red and blue signals with several peaks due to eye and chopstick tip tracking. We measure the time from valley to signal peak and return to the second valley indicated by the arrow symbol. The time series data represents the activity that succeeded in picking up peanuts and putting them down until starting to pick them up again.

We measure time by finding the peaks and valleys for each CAM and CTM signal using the following discriminant equation:

$$\Delta \text{CAM} = \sum_{t=0}^{N-1} \text{CAM}(t+1) - \text{CAM}(t) \tag{16}$$
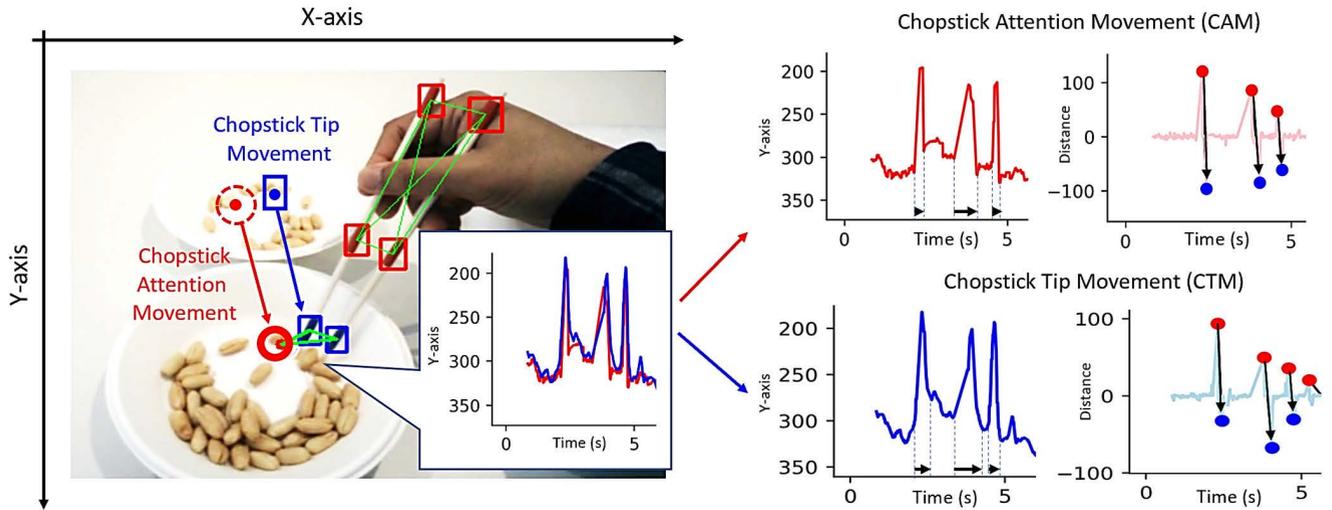
FIGURE 5. Macroscopic level: Cognitive ability design using eye movement and chopstick tip tracking.

$$\Delta \text{CTM} = \sum_{t=0}^{N-1} \text{CTM}(t+1) - \text{CTM}(t) \qquad (17)$$

The above discriminant equation is obtained by comparing the data at t and $t+1$. After we obtain the discriminant results ΔCAM from the CAM and ΔCTM from the CTM, we determine the peaks and valleys in the new time series data with several parameters such as distance and prominence. We obtain the best results at a distance value and prominence by conducting several experiments. After obtaining the peaks and valleys of ΔCAM, we look for the nearest peaks and valleys by subtracting each valley time $T_{\text{V.CAM}}$ from the peak time $T_{\text{P.CAM}}$. For the ΔCTM, we obtain the nearest peaks and valleys by subtracting each valley time $T_{\text{V.CTM}}$ from the peak time $T_{\text{P.CTM}}$, based on the following equation:

$$\Delta T_{\text{CAM}} = \sum_{n=0}^{N} T_{\text{V.CAM}}(n) - T_{\text{P.CAM}}(n) \qquad (18)$$

$$\Delta T_{\text{CTM}} = \sum_{n=0}^{N} T_{\text{V.CTM}}(n) - T_{\text{P.CTM}}(n) \qquad (19)$$

After obtaining the time ΔT_{CAM} and ΔT_{CTM}, we measure the distance D_{CAM} and D_{CTM} to calculate the linear velocity of each. Both variables use the pixel as units in the 2D image plane. Using a similar equation to find the time parameter, we obtain ΔD_{CAM} and ΔD_{CTM} with the following equation:

$$\Delta D_{\text{CAM}} = \sum_{n=0}^{N} D_{\text{P.CAM}}(n) - D_{\text{V.CAM}}(n) \qquad (20)$$

$$\Delta D_{\text{CTM}} = \sum_{n=0}^{N} D_{\text{P.CTM}}(n) - D_{\text{V.CTM}}(n) \qquad (21)$$

After obtaining the distance and time parameters, we then obtain the linear velocity v_{CAM} and v_{CTM} using the following

TABLE 2. The comparison of FCC-NN with MLP in the 10.000$^{\text{th}}$ epoch.

No.	The architecture of neural networks	Error Rate (degree)	Learning Time (sec.)
1.	MLP 1 hidden node	2.044	4.200
2.	MLP 2 hidden nodes	2.040	4.875
3.	MLP 3 hidden nodes	2.016	5.185
4.	FCC-NN 1 hidden node	1.947	4.915

equation:

$$v_{\text{CAM}} = \frac{\Delta D_{\text{CAM}}}{\Delta T_{\text{CAM}}} \qquad (22)$$

$$v_{\text{CTM}} = \frac{\Delta D_{\text{CTM}}}{\Delta T_{\text{CTM}}} \qquad (23)$$

The ΔT_{CAM} and ΔD_{CAM} are the average time and distance achieved by CAM from one bowl to another and the ΔT_{CTM} and ΔD_{CTM} is the average time and distance CTM achieves from one bowl to another. The v_{CAM} and v_{CTM} both have units of pixels/ms. These two variables are evaluated in the CMT experiment with scenarios using regular and trainer chopsticks. In the following sections, we discuss the results of our proposed method.

IV. RESULT AND DISCUSSION

First, we discuss the feature extraction ability at the microscopic level. The finger model kinematic is used to estimate the finger's features [59]. However, not all feature extractions have precise results because the position of the fingers is not always perpendicular to the camera. This issue was addressed by comparing the length of the joint position with the length of the chopsticks. Therefore, we can only use RGB cameras rather than depth cameras to obtain acceptable results. The marker detection based on contour detection was successfully applied to obtain the position of the chopsticks and index
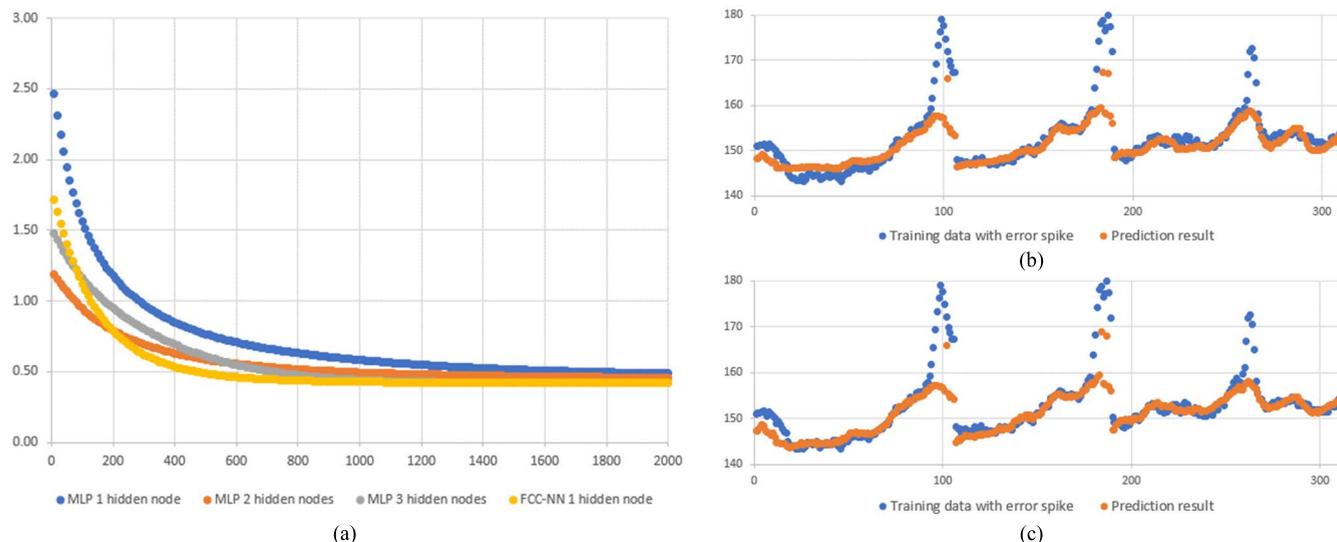
FIGURE 6. PIP joint angle estimation training and testing results using MLP and FCC-NN: a) Comparison of training loss until 2.000 epochs; b) Testing result by MLP with 3 hidden nodes; c) Testing result by FCC-NN with 1 hidden node.

fingers compared to surface EMG [60]. However, marker detection did not always provide a high level of accuracy due to various factors such as occlusion and lighting. Minor errors occur when the hand moves quickly and produces a blurry image. The background color of the environment was limited, and the area had uniform lighting to obtain the best results. The CMT video result is available at our github.
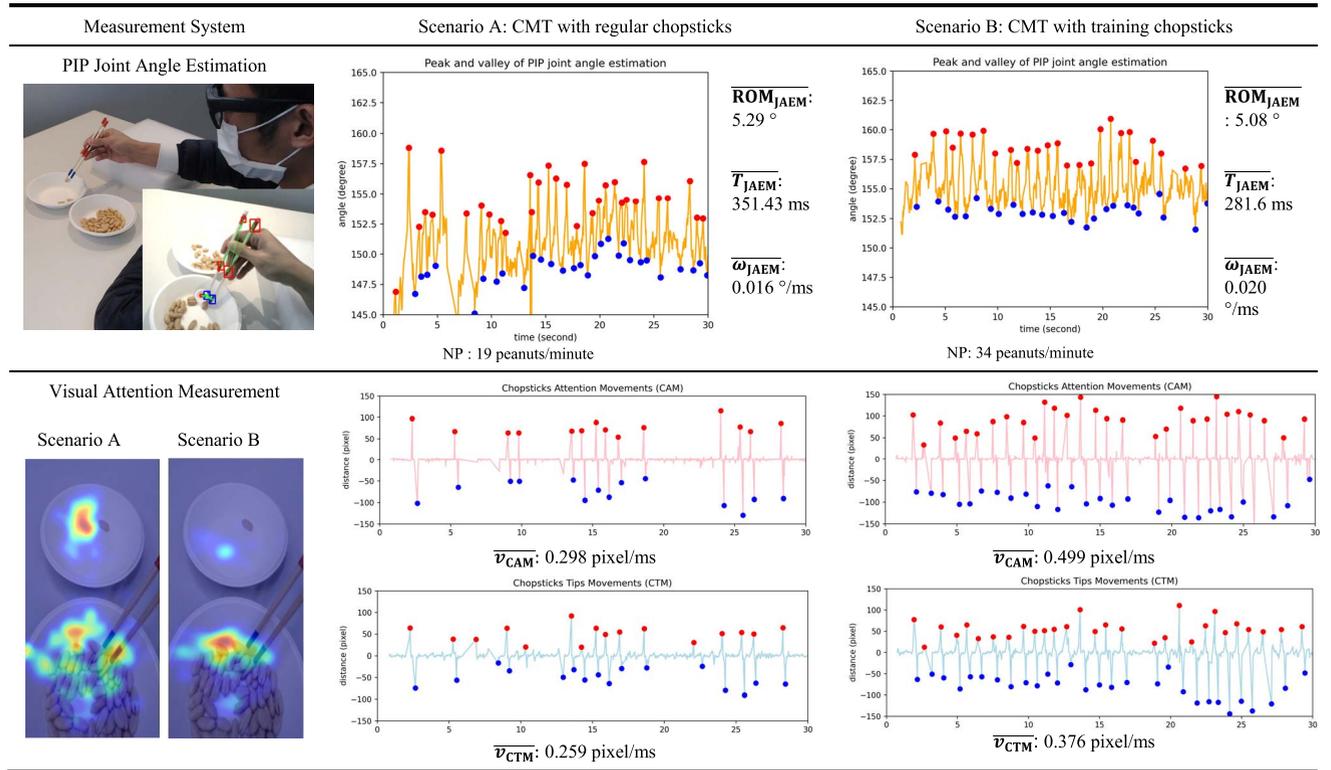
Second, we examine the participants' active perception ability at the mesoscopic level. We collected data from the *One-Minute Chopsticks Challenge* for ten participants who are not habitual chopstick users. Using chopsticks as a novel tool, we are inspired to test the behavioral performance improvement and the visual strategies used during skill acquisition [39]. Before trying it on rehabilitation patients, we obtained permission to test CMT on international students. We collected data from ten users as participants. All the participants gave their informed consent before this study. The following is the Standard Operating Procedure for conducting experiments:

a) Experiment ethics: The researcher informs the participants that this study was approved by the Ethical, Legal, and Social Issues (ELSI) of Tokyo Metropolitan University, Japan. The eye-tracking data collected in the experiment will be stored and used in accordance with the eye-tracking data transparency policy [57], which relates to: (i) what data to collect, (ii) how the data will be used, (iii) what the eye-tracking data will not be used for, and (iv) how the data will be helpful to users.

b) Chopsticks trial: After the participants have agreed to participate, the researcher gives them one minute to read the guidelines about using regular and adult-sized Edison training chopstick, which is commonly used to

help people who have difficulty gripping in tabletop activities [40]. The training chopsticks have two holes on the upper chopstick into which the participants place their index and middle fingers so their fingers can grip correctly. The chopstick trainer is selected so that the participants, who are not habitual chopsticks users, could engage in error-less learning. We can assume linearity in the time-related variables by ensuring that the participants practice using chopsticks under the paradigm of error-less learning. It will ensure that all participants succeed in their initial task because many failures will cause participants to give up.

c) Preparation and setup: The first step in the preparation and setup is to condition the participants to be ready for the experiment [58]. The preparation and setup that was undertaken are as follows: (i) peanuts are placed in the bowl closest to the participant, then the empty bowl is placed on the far side; (ii) participants wear eye trackers, and they can request additional lenses if needed; (iii) the dominant hand holds the standard chopsticks, then places them on the square pad; (iv) the non-dominant hand is folded inward close to the body (or it may be placed on the table); (v) the computer system is connected to the eye-tracker wirelessly; (vi) the trainer chopsticks are prepared; (vii) participants calibrate the eye-tracker with a calibration card; (viii) the battery is checked every time the experiment is finished, and the battery is replaced if the capacity is <30%; (ix) pictures are taken from behind the participants.

d) Experiment: We ask participants to take as many peanuts as possible in one bowl and then transfer them to another. Each participant took part in two experiments: (i) Scenario A using regular chopsticks

TABLE 3. The CMT results for participant 1 (P1) in two scenarios.

At the mesoscopic level, we also propose a Joint Angle Estimation Movement (JAEM) evaluation indices to predict the movement of the PIP joint angle when using chopsticks. We measure the differences in a pinch and move from these two scenarios. The experiment results indicate that all participants achieve better JAEM results in less operating time in scenario B than in scenario A. We found the angular velocity of JAEM (ω_{JAEM}) can be used for further finger analysis [61]. The results show that all the participants had a higher ω_{JAEM} value in scenario B. This result shows that participants are able to perform the chopstick activity better when using the training chopsticks. This result is based on the fact that more peanuts were moved in scenario B than in scenario A. We assume that the ω_{JAEM} evaluation index can be used to measure the progress of the participants in their use of chopsticks by measuring the angular velocity of PIP joint angle estimation movement.

Third, we evaluate cognitive ability during CMT at the macroscopic level. Table 3 shows the CMT results of Participant 1 (P1) in the two scenarios. Two evaluation indices, v_{CAM} and v_{CTM}, were measured by calculating the average time ΔT_{CAM} and ΔT_{CTM}, and the distance ΔD_{CAM} and ΔD_{CTM}. The experiment results at the macroscopic level show some change in the attention heatmap. In scenario A, most participants focused more on the source bowl than the destination bowl. In scenario B, the participants were more confident and focused on picking the peanuts instead of placing peanuts in the other bowl [62]. The behavioral

(1 minute); (ii) Scenario B using trainer chopsticks (1 minute). In scenario A, the participants are required to use chopsticks. In scenario B, they are required to use training chopstick. Both scenarios are used to test the system at the mesoscopic and macroscopic levels.

e) Closing: Participants remove the eye-tracker and lens and return these to their original place; researchers return the nuts to their original place; all equipment is ready for the next new participant.

We successfully implemented FCC-NN as a learning algorithm to estimate the PIP joint angle in the preliminary test. The upper length (L_{up}) and bottom length (L_{bot}) were the inputs for the networks, and the PIP joint angle (θ_{PIP}) was the output. We compare this result with MLP-NN using several hidden nodes and evaluate the performance. Figure 6 shows the PIP joint angle estimation training using MLP and FCC-NN testing results. FCC-NN training loss decreases more rapidly than the other architecture. We found that learning using FCC-NN has slightly better accuracy with less learning time than MLP with three hidden nodes. The accuracy obtained is 98.9 %, with an average error of 0.0339 radians (1.947 °). Therefore, FCC-NN provides slightly better results and speeds up the process [52]. However, there is no guarantee that the accuracy and learning time of the above neural networks will be the same on different data sets and under different conditions. This can be due to differences in hand size, a non-standard way of holding chopsticks, sensor readings, or any environmental changes. Table 2 compares FCC-NN with MLP in the 10.000th epoch.

TABLE 4. The comparison of evaluation indices for ten participants in two scenarios.

Scenario	Subject	Evaluation indices			
		NP (peanuts/minute)	v_{CAM} (pixel/ms)	v_{CTM} (pixel/ms)	ω_{JAEM} (°/ms)
A: Regular chopsticks	P1	19	0.298	0.259	0.016
	P2	12	0.201	0.159	0.026
	P3	9	0.232	0.167	0.018
	P4	12	0.209	0.185	0.024
	P5	8	0.206	0.135	0.026
	P6	34	0.281	0.232	0.020
	P7	19	0.152	0.221	0.017
	P8	16	0.154	0.128	0.019
	P9	17	0.144	0.175	0.024
	P10	17	0.161	0.234	0.021
	Average	**16.3**	**0.204**	**0.190**	**0.021**
B: Trainer chopsticks	P1	34	0.499	0.376	0.020
	P2	33	0.380	0.305	0.029
	P3	15	0.328	0.192	0.023
	P4	26	0.355	0.208	0.034
	P5	17	0.254	0.220	0.032
	P6	41	0.440	0.343	0.024
	P7	22	0.260	0.268	0.025
	P8	16	0.136	0.138	0.021
	P9	17	0.104	0.151	0.033
	P10	21	0.216	0.224	0.026
	Average	**24.2**	**0.297**	**0.243**	**0.027**
Progress of CMT	P1	+15	+0.201	+0.117	+0.004
	P2	+21	+0.179	+0.146	+0.003
	P3	+6	+0.096	+0.025	+0.005
	P4	+14	+0.146	+0.023	+0.010
	P5	+9	+0.048	+0.085	+0.006
	P6	+7	+0.159	+0.111	+0.004
	P7	+3	+0.108	+0.047	+0.008
	P8	0	-0.018	+0.010	+0.002
	P9	0	-0.040	-0.024	+0.009
	P10	+4	+0.055	-0.010	+0.005
	Average	**+7.9**	**+0.093**	**+0.053**	**+0.006**
	Percentages (%)	**+48.47**	**+45.83**	**+27.97**	**+26.54**

TABLE 5. The pearson correlation coefficient (R) for ten participants in two scenarios.

Scenario	Pearson Correlation Coefficient (R)					
	$R(v_{CAM}, NP)$	$R(v_{CTM}, NP)$	$R(v_{CAM}, v_{CTM})$	$R(\omega_{JAEM}, NP)$	$R(v_{CAM}, \omega_{JAEM})$	$R(v_{CTM}, \omega_{JAEM})$
A	0.617	0.332	0.438	-0.381	-0.542	-0.274
B	0.878	0.805	0.872	-0.182	-0.340	-0.286

changes in eye focus coincided with the CAM and CTM evaluation indices. We can see that all respondents experienced an increase in v_{CAM} and v_{CTM} in scenario B compared to scenario A.

This evaluation discusses the final results of the three proposed evaluation indices compared to the conventional method of counting the number of peanuts moved (NP). Table 4 compares the evaluation indices for ten participants in the two scenarios. The average increase in the number of peanuts moved by all participants is 48.47 %. Thus, the PIP joint index finger performance increases for all participants before and after training. At the mesoscopic level, the average increase for ω_{JAEM} is 26.79 % or, in other words, the average angular velocity at the PIP joint angle increases by 0.0056

°/ms. At the macroscopic level, the average increase for v_{CAM} is 42.13 %, and the average increase for v_{CTM} is 28.63 %. The experiment results show that the eye tracker moves earlier than the chopstick tip. Hence, we conjecture that visual attention influences finger movement significantly. The behavioral changes in eye focus coincided with the chopsticks' attention movement CAM and CTM evaluation indices. Thus, we estimate that visual attention provides finger joint angle estimation to interpret the behavioral performance and visual strategy better than in the previous study [39].

This result is evidenced by regression and correlation analysis between finger movement speed and eye movement angular velocity. We performed regression analysis [63] to verify the linearity between the v_{CAM}, v_{CTM} and ω_{JAEM} as the
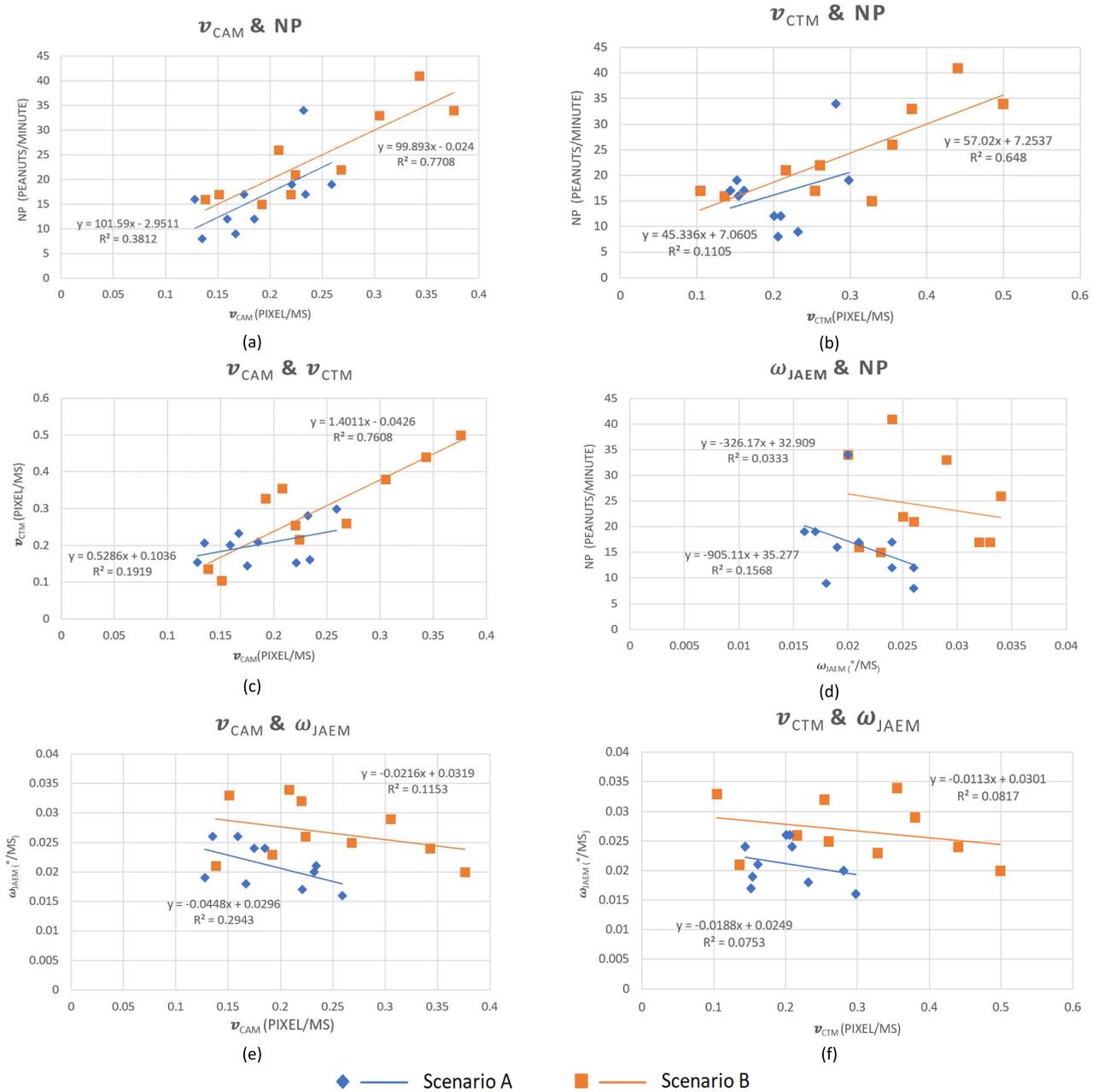
FIGURE 7. Regression and correlation analysis between scatter plots: (a) relationship between vCAM and NP; (b) relationship between vCTM and NP; (c) relationship between vCAM and vCTM; (d) relationship between ωJAEM and the NP; (e) relationship between vCAM and ωJAEM; (f) relationship between vCTM and ωJAEM.

three proposed variables with number of peanuts (NP) moved as a conventional method in scenarios A and B. Table 5 shows the Pearson correlation coefficient for ten participants in two scenarios. The regression results show that these two scenarios do not have many different gradient values. The constant value of the regression line equation in Scenario B moves to the top right in Scenario A with a positive value. This result shows an increase in the value for the three proposed variables. The three proposed evaluation indices have linear effects with the NP moved to the traditional method. We con-

clude that the system has successfully estimated the finger measurement together with visual attention. Thus, these three evaluation indices can be proposed as indicators of the finger and eye movement progress during CMT.

Next, we undertake a correlation analysis to verify each indices's relationship strength. We analyzed each evaluation indices using the Pearson correlation coefficient (R) [64]. The R-value of the three relationships, R(v_{CAM}, NP), R(v_{CTM}, NP), and R(v_{CAM}, v_{CTM}), have a positive correlation with each other. The R-value shows that v_{CAM} and v_{CTM} affect

the increase in chopsticks skill, indicated by the NP taken in one minute in both scenarios. Likewise, if the visual attention is high, it will reduce the angular finger joint angle because it will be more careful in pinching and moving the peanut. Thus, visual attention generally dominates skill improvement compared to finger movements. These results indicate that visual attention affects the progress of hand-eye coordination performance, especially in CMT activities.

However, another R-value of the three relationships, $R(\omega_{JAEM}, NP)$, $R(v_{CAM}, \omega_{JAEM})$, and $R(v_{CTM}, \omega_{JAEM})$, has a negative correlation with each other. The R-value shows that ω_{JAEM} has a negative correlation with the NP. This means the increase of ω_{JAEM} will reduce the NP. Likewise, increasing v_{CAM} and v_{CTM} will also relieve the ω_{JAEM}. If the finger moves faster, it will reduce the precision of pinching objects like robotic imitation learning cases [65]. If the angular velocity of the finger is high, it will be more challenging to pick up and move the peanuts. In the case of a person skilled with chopsticks, we can assume linearity. However, in the case of an inexperienced user, it is challenging to assume linearity because the speed increases will exceed the upper limit of saturation, resulting in a reduced success rate. This assumption follows the experiment results, which show that ω_{JAEM} has a negative correlation with NP. Figure 7 shows the regression and correlation analysis in a scatter plot.

Compared to the contact method, the multiscopic approach has several advantages. First, this method estimates finger joint angle by simply using a camera to detect markers on chopsticks. The contact methods might use a sensor representing the CMT on the object, chopsticks, or body. Second, this method may replace the electric glove, which is considered less comfortable to describe the kinematics finger model. Third, the egocentric vision is regarded as the most appropriate for maintaining privacy compared to the other method, and it is more suitable for independent hand rehabilitation using smart glasses. Classifying pinch and move mode is necessary based on time-series gaze analysis [66]. The real-time CMT results can be suggested as an input or sensor for robotic hand prostheses, partner robots, or other health care systems.

Moreover, integrating finger measurement and visual attention in the CMT assessment can result in several benefits that may not be achieved from the conventional method. We obtain more detailed information regarding the causes of success or failure when the patient undertakes a CMT assessment. The causative factor may be seen in the dominant movement of the fingers and eyes [67]. We can acquire information on the increase or decrease in the function of the hand and eye movements simultaneously. This information may help the therapist develop a rehabilitation plan to improve hand-eye coordination. However, there are physical limitations where practitioners must deal with the 2D position of the chopsticks in data collection. So, we need to implement a data acquisition strategy by reducing the degrees of freedom. As a result, this proposed method must still be validated with existing CMT practices in rehabilitation facilities. These

three proposed evaluation indices are expected to provide added value to the CMT measurement.

V. CONCLUSION AND FUTURE WORKS

This paper proposes a finger joint angle estimation with visual attention in the CMT using a multiscopic framework. Feature extraction based on the finger model kinematic is developed at the microscopic level. The marker detection algorithm is applied to determine the status of chopsticks and joint fingers at the mesoscopic level. The proposed FCC-NN module is able to learn the bottom and upper chopsticks' length patterns for estimating the PIP joint angle and pinching-moving detection. The learning algorithm increased the accuracy compared to conventional methods and reduced hidden units at the mesoscopic level. Based on the experiment results, we can use a small amount of the angular velocity of JAEM with acceptable accuracy and reduce the processing time. In addition, we investigate the local maxima and minima at the macroscopic level to obtain the velocity of CAM and CTM indices for visual attention and chopstick movement evaluation. We expect this study to benefit therapists and researchers by providing valuable information that is not accessible in the clinic. For future work, we will collect more human samples for further validation and deploy the system for rehabilitation purposes.

REFERENCES

[1] Q. E. Wang, *Chopsticks: A Cultural and Culinary History*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[2] S.-N. Ahn, "Effects of occupational-based intervention on chopsticks skill in children with autism spectrum disorder," *Int. J. Adv. Culture Technol.*, vol. 6, no. 4, pp. 80–86, 2018.

[3] Y.-H. Chen and J.-H. Chang, "Establishment and discussion of the design criteria for training chopsticks for children," in *Proc. 21st Congr. Int. Ergonom. Assoc. (IEA)*, Cham, Switzerland, 2021, pp. 63–70.

[4] H. M. Chen and J. J. Chang, "The skill components of a therapeutic chopsticks task and their relationship with hand function tests," *Kaohsiung J. Med. Sci.*, vol. 15, no. 12, pp. 704–709, Dec. 1999.

[5] E. Carmeli, H. Patish, and R. Coleman, "The aging hand," *J. Gerontol., A*, vol. 58, no. 2, pp. M146–M152, Feb. 2003.

[6] M.-F. Tsai, R. H. Wang, and J. Zariffa, "Identifying hand use and hand roles after stroke using egocentric video," *IEEE J. Translational Eng. Health Med.*, vol. 9, pp. 1–10, 2021.

[7] K. Iokawa, T. Sone, T. Fujita, T. Tsukada, M. Kaneda, and K. Hasegawa, "Functional and cognitive variables predicting successful use of chopsticks or a spoon by the paretic upper extremity in patients following stroke: A cross-sectional study," *Topics Stroke Rehabil.*, vol. 26, no. 1, pp. 1–5, Jan. 2019.

[8] T. M. Smith, M. R. Pappadis, S. Krishnan, and T. A. Reistetter, "Stroke survivor and caregiver perspectives on post-stroke visual concerns and long-term consequences," *Behavioural Neurol.*, vol. 2018, pp. 1–8, Oct. 2018.

[9] T. Singh, C. M. Perry, S. L. Fritz, J. Fridriksson, and T. M. Herter, "Eye movements interfere with limb motor control in stroke survivors," *Neurorehabilitation Neural Repair*, vol. 32, no. 8, pp. 724–734, Aug. 2018.

[10] B.-S. Lin, J.-L. Chen, and H.-C. Hsu, "Novel upper-limb rehabilitation system based on attention technology for post-stroke patients: A preliminary study," *IEEE Access*, vol. 6, pp. 2720–2731, 2018.

[11] K. L. Meadmore, T. A. Exell, J. H. Burridge, A.-M. Hughes, C. T. Freeman, and V. Benson, "Upper limb and eye movement coordination during reaching tasks in people with stroke," *Disability Rehabil.*, vol. 40, no. 20, pp. 2424–2432, Sep. 2018.

[12] Z. Zuo, L. Yang, Y. Peng, F. Chao, and Y. Qu, "Gaze-informed egocentric action recognition for memory aid systems," *IEEE Access*, vol. 6, pp. 12894–12904, 2018.

[13] A. Bandini and J. Zariffa, "Analysis of the hands in egocentric vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Apr. 13, 2020.

[14] A. Cartas, P. Radeva, and M. Dimiccoli, "Activities of daily living monitoring via a wearable camera: Toward real-world applications," *IEEE Access*, vol. 8, pp. 77344–77363, 2020.

[15] S. Bano, T. Suveges, J. Zhang, and S. J. Mckenna, "Multimodal egocentric analysis of focused interactions," *IEEE Access*, vol. 6, pp. 37493–37505, 2018.

[16] Y. Saito, R. Hachiuma, H. Saito, H. Kajita, Y. Takatsume, and T. Hayashida, "Camera selection for occlusion-less surgery recording via training with an egocentric camera," *IEEE Access*, vol. 9, pp. 138307–138322, 2021.

[17] K. Seo, H. Cho, D. Choi, and T. Heo, "Stereo feature learning based on attention and geometry for absolute hand pose estimation in egocentric stereo views," *IEEE Access*, vol. 9, pp. 116083–116093, 2021.

[18] F.-Y. Chia and D. Saakes, "Interactive training chopsticks to improve fine motor skills," in *Proc. 11th Conf. Adv. Comput. Entertainment Technol.*, Nov. 2014, pp. 1–4.

[19] Y. Nakamura, Y. Arakawa, T. Kanehira, M. Fujiwara, and K. Yasumoto, "SenStick: Comprehensive sensing platform with an ultra tiny all-in-one sensor board for IoT research," *J. Sensors*, vol. 2017, pp. 1–16, Oct. 2017.

[20] D. Sawamura, S. Sakuraba, Y. Suzuki, M. Asano, S. Yoshida, T. Honke, M. Kimura, Y. Iwase, K. Yoshida, and S. Sakai, "Acquisition of chopstick-operation skills with the non-dominant hand and concomitant changes in brain activity," *Sci. Rep.*, vol. 9, no. 1, p. 20397, Dec. 2019.

[21] Y. K. Yokubo, T. Ota, and K. Shibata, "Relationship between chopstick manipulation and cross-sectional shape in the developmental stages from infancy to early school age," *Appl. Ergonom.*, vol. 97, Nov. 2021, Art. no. 103507.

[22] Y. Shimomura, T. Ohsawa, M. Shimura, Y. Xia, K. Iwanaga, and T. Katsuura, "What is the significance of the traditional pinching mode of holding chopsticks?" *J. Physiol. Anthropol.*, vol. 39, no. 1, p. 13, Dec. 2020.

[23] K. Okamoto and K. Yanai, "Real-time eating action recognition system on a smartphone," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW)*, Chengdu, China, Jul. 2014, pp. 1–6.

[24] R. Nakaoka, Y. Nakamura, Y. Matsuda, S. Misaki, and K. Yasumoto, "Eat2pic: Food-tech design as a healthy nudge with smart chopsticks and canvas," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops Affiliated Events (PerCom Workshops)*, Kassel, Germany, Mar. 2021, pp. 389–391.

[25] M. Namiki and M. Makino, "A contactless training support system of holding chopsticks by posture estimation of chopsticks, hand, and fingers," *Proc. SPIE*, vol. 11766, Mar. 2021, vol. 11766.

[26] H. Wang and Y. Zhang, "Research on the design of chopsticks for dining reminders based on HSI color space recognition," in *Proc. 4th Int. Conf. Data Sci. Inf. Technol.*, Shanghai China, Jul. 2021, pp. 317–322.

[27] A. Kovari, J. Katona, and C. Costescu, "Quantitative analysis of relationship between visual attention and eye-hand coordination," *Acta Polytechnica Hungarica*, vol. 17, no. 2, pp. 77–95, 2020.

[28] Z. Li, S. Zhao, L. Su, and C. Liao, "Study on the operators' attention of different areas in university laboratories based on eye movement tracking technology," *IEEE Access*, vol. 8, pp. 8262–8267, 2020.

[29] S. John and V. G. Renumol, "Design and development of an Android app (HanDex) to enhance hand dexterity in children with poor handwriting," *IEEE Access*, vol. 10, pp. 48973–48993, 2022.

[30] A. D. Sonar, N. Sawant, J. Salunkhe, and S. S. Baraskar, "Design, development, and validation of hand-eye coordination equipment," *IETE J. Res.*, pp. 1–9, Apr. 2022.

[31] M. Haslgrübler, B. Gollan, C. Thomay, A. Ferscha, and J. Heftberger, "Towards skill recognition using eye-hand coordination in industrial production," in *Proc. 12th ACM Int. Conf. Pervasive Technol. Rel. Assistive Environ.*, Rhodes, Greece, Jun. 2019, pp. 11–20.

[32] L. Leveque, H. Bosmans, L. Cockmartin, and H. Liu, "State of the art: Eye-tracking studies in medical imaging," *IEEE Access*, vol. 6, pp. 37023–37034, 2018.

[33] A. John, S. Mohan, and D. M. M. Vianny, "Cognitive cyber-physical system applications," in *Cognitive Engineering for Next Generation Computing*, 1st ed., K. B. Prakash, G. R. Kanagachidambaresan, V. Srikanth, E. Vamsidhar, Eds. Hoboken, NJ, USA: Wiley, 2021, pp. 167–187.

[34] Y. Nakamura, Y. Arakawa, and K. Yasumoto, "Senstick: A rapid prototyping platform for sensorizing things," in *Proc. 9th Int. Conf. Mobile Comput. Ubiquitous Netw. (ICMU)*, Kaiserslautern, Germany, Oct. 2016, pp. 1–6.

[35] E. Rahimian, S. Zabihi, A. Asif, D. Farina, S. F. Atashzar, and A. Mohammadi, "FS-HGR: Few-shot learning for hand gesture recognition via electromyography," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1004–1015, 2021.

[36] M. Sirnik, F. Erčulj, and J. Rošker, "Research of visual attention in basketball shooting: A systematic review with meta-analysis," *Int. J. Sports Sci. Coaching*, vol. 2022, Feb. 2022, Art. no. 174795412210757.

[37] S. Egner, S. Reimann, R. Höger, and W. H. Zangemeister, "Attention and information acquisition: Comparison of mouse-click with eye-movement attention tracking," *J. Eye Movement Res.*, vol. 11, no. 6, Nov. 2018.

[38] C. Cavina-Pratesi and C. Hesse, "Why do the eyes prefer the index finger? Simultaneous recording of eye and hand movements during precision grasping," *J. Vis.*, vol. 13, no. 5, p. 15, Apr. 2013.

[39] T. J. Bosch, T. Hanna, K. A. Fercho, and L. A. Baugh, "Behavioral performance and visual strategies during skill acquisition using a novel tool use motor learning task," *Sci. Rep.*, vol. 8, no. 1, p. 13755, Dec. 2018.

[40] W. Choi, "The effect of task-oriented training on upper-limb function, visual perception, and activities of daily living in acute stroke patients: A pilot study," *Int. J. Environ. Res. Public Health*, vol. 19, no. 6, p. 3186, Mar. 2022.

[41] I. Pysmennyi, R. Kyslyi, and A. Petrenko, "Edge computing in multi-scope service-oriented mobile healthcare systems," *Syst. Res. Inf. Technol.*, no. 1, pp. 118–127, Mar. 2019.

[42] J. Lv, Q. Sun, Q. Li, and L. Moreira-Matias, "Multi-scale and multi-scope convolutional neural networks for destination prediction of trajectories," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 8, pp. 3184–3195, Aug. 2020.

[43] A. A. Saputra, K. Wada, S. Masuda, and N. Kubota, "Multi-scopic neuro-cognitive adaptation for legged locomotion robots," Tech. Rep., Aug. 2021.

[44] A. F. Zhu, P. Rahgozar, and K. C. Chung, "Advances in proximal interphalangeal joint arthroplasty," *Hand Clinics*, vol. 34, no. 2, pp. 185–194, May 2018.

[45] J.-S. Won and S. Lee, "Geometry-based finger kinematic models for joint rotation configuration and parameter estimation," *Int. J. Adv. Robotic Syst.*, vol. 17, no. 4, Jul. 2020, Art. no. 172988142093057.

[46] A. R. A. Besari, A. A. Saputra, W. H. Chin, N. Kubota, and Kurnianingsih, "Feature-based egocentric grasp pose classification for expanding human-object interactions," in *Proc. IEEE 30th Int. Symp. Ind. Electron. (ISIE)*, Kyoto, Japan, Jun. 2021, pp. 1–6.

[47] P. B. Shull, S. Jiang, Y. Zhu, and X. Zhu, "Hand gesture recognition and finger angle estimation via wrist-worn modified barometric pressure sensing," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 724–732, Apr. 2019.

[48] I. Imbinto, F. Montagnani, M. Bacchereti, C. Cipriani, A. Davalli, R. Sacchetti, E. Gruppioni, S. Castellano, and M. Controzzi, "The S-finger: A synergetic externally powered digit with tactile sensing and feedback," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 6, pp. 1264–1271, Jun. 2018.

[49] J. Mazumder, L. N. Nahar, and M. Md Uddin Atique, "Finger gesture detection and application using hue saturation value," *Int. J. Image, Graph. Signal Process.*, vol. 10, no. 8, pp. 31–38, Aug. 2018.

[50] W. B. Dou, W. H. Chin, and N. Kubota, "Hand gesture communication using deep learning based on relevance theory," in *Proc. Joint 11th Int. Conf. Soft Comput. Intell. Syst. 21st Int. Symp. Adv. Intell. Syst. (SCIS-ISIS)*, Dec. 2020, pp. 1–5.

[51] A. R. A. Besari, W. H. Chin, N. Kubota, and Kurnianingsih, "Ecological approach for object relationship extraction in elderly care robot," in *Proc. 21st Int. Conf. Res. Educ. Mechatronics (REM)*, Cracow, Poland, Dec. 2020, pp. 1–6.

[52] Y. Cheng, "Backpropagation for fully connected cascade networks," *Neural Process. Lett.*, vol. 46, no. 1, pp. 293–311, Aug. 2017.

[53] P. Gajbhiye, N. Mingchinda, W. Chen, S. C. Mukhopadhyay, T. Wilaiprasitporn, and R. K. Tripathy, "Wavelet domain optimized Savitzky–Golay filter for the removal of motion artifacts from EEG recordings," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[54] R. Chen, M. Wang, and J.-H. Chang, "A study of the pinch performance and subjective satisfaction with chopsticks of different lengths," in *Proc. MATEC Web Conf.*, vol. 119, 2017, p. 1017.

[55] A. R. A. Besari, W. H. Chin, N. Kubota, and Kurnianingsih, "Ecological approach for object relationship extraction in elderly care robot," in *Proc. 21st Int. Conf. Res. Educ. Mechatronics (REM)*, Dec. 2020, pp. 1–6.

[56] Y. J. R. De Kloe, I. T. C. Hooge, C. Kemner, D. C. Niehorster, M. Nyström, and R. S. Hessels, "Replacing eye trackers in ongoing studies: A comparison of eye-tracking data quality between the Tobii pro TX300 and the tobii pro spectrum," *Infancy*, vol. 27, no. 1, pp. 25–45, Jan. 2022.

[57] E. P. Larsen, J. M. Kolman, F. N. Masud, and F. Sasangohar, "Ethical considerations when using a mobile eye tracker in a patient-facing area: Lessons from an intensive care unit observational protocol," *Ethics Hum. Res.*, vol. 42, no. 6, pp. 2–13, Nov. 2020.

[58] D. C. Niehorster, T. Santini, R. S. Hessels, I. T. C. Hooge, E. Kasneci, and M. Nyström, "The impact of slippage on the data quality of head-worn eye trackers," *Behav. Res. Methods*, vol. 52, no. 3, pp. 1140–1160, Jun. 2020.

[59] C. Chen, W. Guo, C. Ma, Y. Yang, Z. Wang, and C. Lin, "SEMG-based continuous estimation of finger kinematics via large-scale temporal convolutional network," *Appl. Sci.*, vol. 11, no. 10, p. 4678, May 2021.

[60] C. Dai and X. Hu, "Finger joint angle estimation based on motoneuron discharge activities," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 3, pp. 760–767, Mar. 2020.

[61] H. Eschmann, M. E. Héroux, J. H. Cheetham, S. Potts, and J. Diong, "Thumb and finger movement is reduced after stroke: An observational study," *PLoS ONE*, vol. 14, no. 6, Jun. 2019, Art. no. e0217969.

[62] S. N. Emhardt, M. Wermeskerken, K. Scheiter, and T. Gog, "Inferring task performance and confidence from displays of eye movements," *Appl. Cognit. Psychol.*, vol. 34, no. 6, pp. 1430–1443, Nov. 2020.

[63] X. Fang, F. Yu, G. Yang, and Y. Qu, "Regression analysis with differential privacy preserving," *IEEE Access*, vol. 7, pp. 129353–129361, 2019.

[64] H. Zhu, X. You, and S. Liu, "Multiple ant colony optimization based on Pearson correlation coefficient," *IEEE Access*, vol. 7, pp. 61628–61638, 2019.

[65] L. Ke, A. Kamat, J. Wang, T. Bhattacharjee, C. Mavrogiannis, and S. S. Srinivasa, "Telemanipulation with Chopsticks: Analyzing human factors in user demonstrations," Tech. Rep., Jun. 2022.

[66] A. Murata, T. Doi, K. Kageyama, and W. Karwowski, "Development of an eye-gaze input system with high speed and accuracy through target prediction based on homing eye movements," *IEEE Access*, vol. 9, pp. 22688–22697, 2021.

[67] B. Banire, D. Al-Thani, M. Qaraqe, K. Khowaja, and B. Mansoor, "The effects of visual stimuli on attention in children with autism spectrum disorder: An eye-tracking study," *IEEE Access*, vol. 8, pp. 225663–225674, 2020.

ADNAN RACHMAT ANOM BESARI (Graduate Student Member, IEEE) received the Bachelor of Applied Science degree in electronics engineering from the Politeknik Elektronika Negeri Surabaya (PENS), in 2008, and the Master of Science degree in robotics and automations from the Faculty of Manufacturing Engineering, Universiti Teknikal Malaysia Melaka (UTeM), in 2011. He is currently pursuing the Ph.D. degree with Tokyo Metropolitan University (TMU), Japan. He is also a Lecturer at the Department of Informatics and Computer Engineering, PENS. He previously served as the Head of the Real-Time Programming Laboratory, from 2017 to 2019. His current research interests include computer vision, educational robotics, the Internet of Things, and elderly care systems.

AZHAR AULIA SAPUTRA (Member, IEEE) received the Bachelor of Applied Science degree in electronic engineering from the Politeknik Elektronika Negeri Surabaya, Indonesia, in March 2014, and the Master of Engineering and Doctoral of Philosophy degrees from Tokyo Metropolitan University, Japan, in March 2018 and 2021, respectively. He is currently an Assistant Professor at Tokyo Metropolitan University. He received a Bronze Award at the Capstone Design Fair International Session, South Korea, and several achievements in national and international robotic competitions. He was a Ph.D. Student at Tokyo Metropolitan University under JSPS Research Fellow—DC1 (2018–2021). He received an Excellent Graduate School Research Award from the Japan Society of Automotive Engineers (JSAE), in 2020. His research interests include intelligent control systems, neural-based locomotion systems, and robotics.

WEI HONG CHIN (Member, IEEE) received the bachelor's degree (Hons.) in electronics engineering, majoring in robotics and automation from Multimedia University, Malacca, Malaysia, in 2011, the master's degree in computer science from the University of Malaya, Malaysia, in 2015, and the Ph.D. degree from Tokyo Metropolitan University, Japan, in 2019. He is currently an Assistant Professor at Tokyo Metropolitan University. His research interests include biologically inspired incremental learning, robot navigation, and human–robot interaction.

KURNIANINGSIH (Senior Member, IEEE) received the B.Eng. degree in informatics engineering from Telkom University, Indonesia, the M.Eng. degree in electrical engineering from North Sumatera University, Indonesia, and the Ph.D. degree in electrical engineering from Universitas Gadjah Mada, Indonesia. She is currently an Assistant Professor with the Department of Electrical Engineering, Politeknik Negeri Semarang, Indonesia. Her current research interests include sensor networks, machine learning, and computational intelligence. She has been an Executive Committee Member in the IEEE Region 10 (Asia-Pacific Region), since 2018, and she has appointed as the Information Management Committee Chair. She was the Past Vice-Chair of the IEEE Indonesia Section. She received the IEEE MGA Achievement Award, in 2020, and the IEEE Region 10 Young Professionals Award in Academician, in 2018. She is a member of the IEEE Computational Intelligence Society (IEEE CIS) and the IEEE Systems, Man, and Cybernetics Society (IEEE SMCS).

NAOYUKI KUBOTA (Senior Member, IEEE) received the Graduate degree from Osaka Kyoiku University, in 1992, the M.E. degree from Hokkaido University, in 1994, and the D.E. degree from Nagoya University, Japan, in 1997. He is currently a Professor with the Department of Mechanical Systems Engineering, Graduate School of Systems Design, Tokyo Metropolitan University, Japan. He currently serves as the Vice Director of the Tokyo Biomarker Innovation Research Association, Japan, and the Chair for IEEE Society on Systems, Man, and Cybernetics, Japan Chapter. He was a Visiting Professor at the University of Portsmouth, U.K., in 2007 and 2009, and an Invited Visiting Professor at Seoul National University, from 2009 to 2012. His current research interests inlcude the fields of coevolutionary computing, fuzzy computing, topological mapping, cognitive robotics, social robotics, and informationally structured space. He has published more than 500 refereed journals and conference papers in the above research fields. He received the Best Paper Award of IEEE IECON'96 and the Best Paper Award of IEEE CIRA'97, amongst others. He was an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS, from 1999 to 2010, the Robotics Task Force Chair of the IEEE CIS Intelligent Systems Applications Technical Committee, from 2007 to 2014, and the Japan Chapter Chair of the IEEE Systems, Man, and Cybernetics Society, since 2018.

• • •