GMH-D: Combining Google MediaPipe and RGB-Depth Cameras for Hand Motor Skills Remote Assessment

GMH-D: Combining Google MediaPipe and RGB-Depth Cameras for Hand Motor Skills Remote Assessment

Gianluca Amprimo
Politecnico di Torino
Turin, Italy
gianluca.amprimo@polito.it

Claudia Ferraris
National Research Council
Turin, Italy
claudia.ferraris@ieiit.cnr.it

Giulia Masi
University of Turin
Turin, Italy
giulia.masi@unito.it

Giuseppe Pettiti
National Research Council
Turin, Italy
giuseppe.pettiti@ieiit.cnr.it

Lorenzo Priano
University of Turin
Turin, Italy
lorenzo.priano@unito.it

***Abstract*—Impairment in the execution of simple motor tasks involving hands and fingers could hint at a general worsening of health conditions, particularly in the elderly and in people affected by neurological diseases. The deterioration of hand motor function strongly impacts autonomy in daily activities and, consequently, the perceived quality of life. The early detection of alterations in hand motor skills would allow, for example, to promptly activate treatments and mitigate this discomfort. This preliminary study examines an innovative pipeline based on a single RGB-Depth camera and Google MediaPipe Hands, that is suitable for the remote assessment of hand motor skills through simple tasks commonly used in clinical practice. The study includes several phases. First, the quality of hand tracking is evaluated by comparing reconstructed and real hand 3D trajectories. The proposed solution is then tested on a cohort of healthy volunteers to estimate specific kinematic features for each task. Finally, these features are used to train supervised classifiers and distinguish between "normal" and "altered" performance by simulating typical motor behaviour of real impaired subjects. The preliminary results show the ability of the proposed solution to automatically highlight alterations in hand performance, providing an easy-to-use and non-invasive tool suitable for remote monitoring of hand motor skills.**

***Index Terms*—hand tracking, Google MediaPipe, telemedicine, vision-based approach, Azure Kinect**

I. Introduction

The quality of hand function has a significant impact on everyday life. The hands are involved in almost all daily actions (e.g., personal care, work-related and creative activities) that, in most cases, require fine movements and manual skills to grasp and manipulate objects [1]. Therefore, it is evident that an impaired hand function limits independence and autonomy, negatively affecting the overall quality of life. Other than accidental and temporary injuries, several long-term conditions contribute to the deterioration of hand function: structural and functional changes of the joints, muscles, and bones due to aging [2] [3]; symptoms associated with neurodegenerative and chronic pathologies, such as Parkinson's disease (PD) [4] [5]; consequences of acute events, such as stroke [6] [7]. Especially in these contexts, it is essential to detect any hand function alterations and promptly activate treatments capable of mitigating the adverse effects on daily living. Regarding the motor skills of the hand and fingers, the clinical evaluation mainly relies on simple motor tasks (e.g., single and multiple finger tapping, fingers mobility tasks) [8] [9] [10] [11] commonly administered during scheduled outpatient visits. In most cases, the motor examination produces only a qualitative indication of the functional movements but no objective measures. However, the kinematic analysis of the hand motor performance could provide a quantitative characterisation through measures and physical quantities easily comparable over time [12]. The technological advancement of the last few decades has favoured the development of different hand tracking approaches that focus on kinematic analysis and evaluation of hand movements. Many solutions include wearable sensors. For example, [13] employs an IMU-based wireless glove to measure finger movements in rheumatoid arthritis. Concerning PD, wearable inertial units are commonly used to evaluate standard upper limb motor tasks [14] [15] [16]. The same technologies are also used to evaluate and rehabilitate hand function after stroke [17] [18] [19]. However, approaches based on wearable sensors are generally more invasive and less practical in unsupervised settings. Alternative solutions rely on optical devices, particularly 3D cameras and computer vision techniques. For instance, [20] uses a single 3D camera to track specific points of the hand and fingers and evaluate standard hand motor tasks in PD. The same occurs in [21], in this case with the help of a passive glove with colour markers to ensure high tracking robustness. The optical approaches are also widely used in post-stroke subjects for evaluation and rehabilitation purposes [22] [23]. In general, these solutions focus on tracking a reduced number of hand points, thus combining the optimal tracking of specific hand

movements and the real-time needs. A first effective attempt to track the complete movement of the bare hand is Leap Motion Controller [24]: the potentiality of this depth camera has been investigated in several studies, thanks to its ability in estimating a complete and articulated hand skeletal model [25] [26] [27] [28]. The main drawbacks of this device are the restricted hand tracking area and the loss of tracking accuracy in high dynamic conditions [29] [21]. Therefore complete, non-invasive and real-time hand tracking on a less restricted workspace is still an open challenge. A promising and interesting methodology to address these issues is Google Mediapipe Hands (GMH), a recent open-source framework with a simple pipeline that combines computer vision and machine learning techniques to recognise bare hands and provide a complete skeletal model of hand and fingers in real-time [30] [31]. Currently, the potentiality of GMH has been mainly explored for gesture-based interaction purposes [32] [33] [34] [35] and in sign language recognition applications [36] [37] [38], which are static or low-dynamic scenarios. On the contrary, the state-of-the-art highlights that GMH has not yet been applied in more dynamic scenarios, as required for the analysis and automatic evaluation of hand motor skills.

In this context, the paper presents an innovative mixed approach and non-invasive pipeline for the remote assessment of the hand motor function. The proposed mixed-approach combines GMH with a single RGB-Depth camera (i.e., Microsoft Azure Kinect DK), using depth information to implement a three-dimensional solution (GMH-D) to analyse the 3D trajectories of hand movements and characterise them through objective features. To this end, the study focuses on two hand motor tasks (i.e., finger tapping and hand opening-closing) commonly used in clinical practice to evaluate the hand motor dysfunctions in the elderly and pathological/neurological subjects.

The main goals of this study are the following: to verify the accuracy and robustness of GMH-D hand tracking by comparing reconstructed and real hand trajectories; to estimate specific kinematic features to characterise hand motor skills on a cohort of healthy volunteers; to train supervised classifiers to distinguish between "normal" and "altered" performance by simulating typical motor behaviour of real impaired subjects. The preliminary results suggest that the proposed solution could be a useful tool to detect hand function alterations, particularly suitable for remote monitoring applications thanks to its non-invasiveness, portability, and usability. The rest of the paper is organised as follows: Section II describes GMH and the problem of uplifting its 2D tracking coordinates to 3D; Section III describes the pipeline implemented, GMH-D and its validation; Section IV details the experimental protocol and the machine learning approach used to evaluate hand motor skills; Section V presents the preliminary results of the study; Section VI discusses key findings and future developments.

II. 3D Tracking of Hand Movements

The 3D tracking of hand movements is still an open challenge. However, the recent development of better-performing

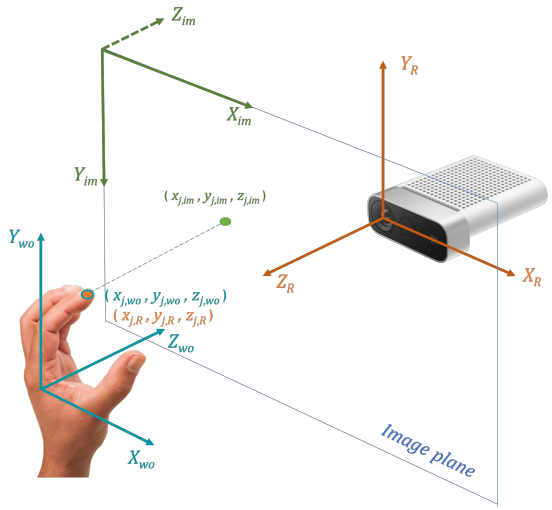

Fig. 1. Reference systems of coordinates extracted by GMH

sensors and algorithms could lead to innovative solutions to address the problem. GMH, for example, implements a computationally cheap, easy-to-use, and high-fidelity tool for hand tracking from a monocular RGB video. For each captured frame, GMH applies two models: the first infers the position and orientation of the hand palm in the image; the second regresses from the palm bounding box the whole hand skeleton, composed of 21 virtual joints related to fundamental hand articulations.

For each tracked joint *j*, GMH provides two types of coordinates, here denoted as *image coordinates* and *world coordinates* (Fig. 1). The former is composed of a pair $(x_{j,im}\,,\ y_{j,im})$ that locates *j* inside the image reference frame. Axes X_{im} and Y_{im} are normalised between 0 and 1 by image height and width. A third parameter $z_{j,im}$, with roughly the same scale as the other two axes, provides a depth estimation of *j* with respect to the wrist reference ($z_{wrist,im} \simeq 0$). The function to estimate such parameter was learned by the machine learning model from training on synthetic hand images [31]. On the contrary, world coordinates $(x_{j,wo}\ ,\ y_{j,wo}\ ,\ z_{j,wo})$ are measured in metres and refer to a reference system with its origin at the centre of the palm bounding box. This additional set of 3D coordinates allows estimating real relative distances between tracked joints. However, no validation of its reliability for computing such distances has been carried out in the literature so far. Moreover, no information of joints positioning with respect to the real-world reference frame centred in the RGB-camera $(X_R\,,\ Y_R\,,\ Z_R)$ can be inferred from GMH.

Exploiting only 2D *image coordinates*, so neglecting the depth estimation provided by GMH, mapping to 3D could be realised through calibration techniques like Perspective-n-Point (PnP) as done in [32], but these procedures require complex setup and pre-processing. Another alternative for 2D coordinates *uplifting* is employing an additional camera and performing triangulation [39]. A third approach relies upon combining colour and depth video streams (i.e, using

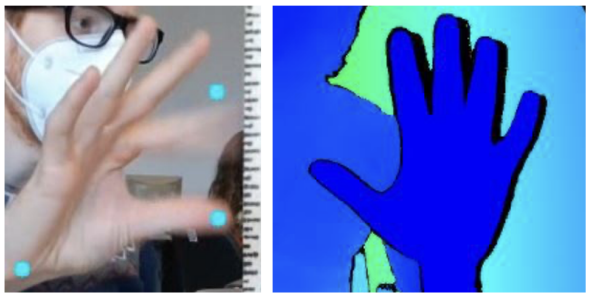

Fig. 2. a) Example of tracking error by GMH which could alter depth estimation of index finger; b) depth interference (black contour) at hand border

RGB-Depth sensors): this is the procedure implemented, for example, in the Body Tracking algorithm of Kinect cameras, whose quality in estimating human motion features has often been successfully compared to gold standard motion capture (MOCAP) systems [40]. In this case, given a virtual joint 2D position in the image and the corresponding depth value measured by the sensor, 3D coordinates are estimated by a 2D-to-3D projection procedure based on the intrinsic matrices of the two embedded cameras (colour and depth) and the standard pinhole camera model [41]. Even though a more complex hardware is required, this solution is almost straightforward and does not entail any complex preliminary setup, as in most cases 2D-to-3D projection routines are already implemented and ready-to-use for such devices in their Software Development Kit (SDK). Moreover, recent years have shown a positive trend towards the embedding of depth cameras also in consumer-grade smartphones and tablets, to favour the spread of Augmented Reality applications. A pure use of this approach, however, would not leverage the depth estimation model embedded in GMH and it is sensitive to phenomena such as those shown in Fig. 2: due to reduced finger surface and motion artifacts, caused by finger blurring during high-speed movements, the virtual joints could be placed on the borders or even completely outside them by the regression model, thus reading a wrong corresponding value from the associated depth-map. Furthermore, depth estimations for a single video frame are usually performed averaging several measurements obtained by Time-Of-Flight (TOF) technology: hence, in case of objects moving at high dynamics as fingers, depth estimations could have reduced accuracy, especially at the borders [42] (Fig. 2b). This is reflected in unstable 3D tracking and the subsequent corruption of all the distances measured between wrongly reconstructed joints.

III. Acquisition Pipeline Description

A. A mixed approach to 3D point estimation

This study proposes GMH-D, an innovative mixed approach that tries to overcome the previously mentioned limitations by combining the facilities of an RGB-Depth camera with GMH coordinates. The implemented technique is based on three fundamental assumptions, which have been empirically observed:

- the relation $d_j = \mathrm{f}(z_{j,im})$, with d_j the depth of virtual joint *j* with respect to the camera centre, is approximately linear for $0.3m \leq d_j \leq 1m$;
- the GMH wrist virtual joint will never fall outside the body surface due to the wide wrist region, thus avoiding inconsistent depth associations;
- wrist movements, even when studying the mobility of the hand and its fingers, are limited: the wrist joint will be in general visible, hence its depth estimation, performed using TOF technology, will be consistent throughout the whole hand movement.

Starting from these premises, GMH-D computes the 3D coordinates of each hand joint *j* in the real-world reference system (origin in the camera centre, Fig. 1) $(x_{j,R},\, y_{j,R},\, z_{j,R})$ using the following steps:

1) Unnormalise image coordinates $(x_{j,im}\,,\, y_{j,im})$ to get virtual marker position in pixels $(p_{x,j}\,,\, p_{y,j})$
2) From the depth map aligned over the current RGB frame, extract only the depth of the wrist virtual joint d_{wrist}
3) Compute approximate depth $\hat{d}_j$ using (1)

$$\hat{d}_j = d_{wrist} + z_{j,im} d_{wrist} \tag{1}$$

4) Convert $(p_{x,j}\,,\, p_{y,j}, \hat{d}_j)$ to $(x_{j,R},\, y_{j,R},\, z_{j,R})$ using pre-defined internal 2D-to-3D projection procedures

This procedure removes the dependency of the reconstructed 3D hand from the depth value associated to the finger joints, which could be unstable as discussed in Section II. At the same time, it embeds also information coming from the complete 3D hand model provided by GMH.

B. System hardware and software components

The pipeline designed to test GMH-D is suitable for a remote scenario in which hardware complexity should be minimised while producing accurate and reliable results for clinical assessment of the hand motor skills. The required hardware consists of an RGB-Depth (RGB-D) camera, working at or above 30 frames-per-seconds (fps), and a processing unit to elaborate the data stream. In this preliminary experiments, an Azure Kinect camera connected to a Zotac minipc (Intel i5 9th generation processor, 16 GB RAM) was employed, even though the same pipeline could be implemented on devices having similar hardware specifics. Azure Kinect was selected also because its SDK already implements all the necessary routines to perform 2D-to-3D projection using depth information estimated through TOF technology.
Tracking data captured by GMH-D are then processed by custom-written scripts for the automatic extraction of task-specific kinematic features. Finally, these features are fed to supervised machine learning models to identify normal or

altered task executions. Feature extraction and classification were implemented in MATLAB and Python respectively. Fig. 3 summarises the implemented pipeline for acquisition and processing.

C. Evaluation of GMH-D Hand tracking solution

The first step is to verify if GMH-D infers coordinates that allow the correct reconstruction of relative distances between tracked joints. In this assessment, a comparison with the trajectories estimated by GMH *world coordinates*, as well as those obtained by recovering the depth information for all 21 virtual hand joints directly from Azure Kinect was performed. These two alternative methods are here denoted as "World Coordinates Approach" (WCA) and "Kinect Approach" (KA). For a fair comparison, KA needs some correction mechanism to cope with the limitations described in Section II. A brute but effective approach consists in applying the following procedure for each virtual joint *j*:

1) Unnormalize image coordinates $(x_{j,im}\ ,\ y_{j,im})$ to get virtual marker position in pixels $(p_{x,j}\ ,\ p_{y,j})$
2) Take a circle $\mathcal{C}$ in the depth map of radius r and centre $(p_{x,j}\ ,\ p_{y,j})$
3) Consider only pixels $c \in \mathcal{C}$ having a value of depth d_c such that $d_{min} < d_c < d_{max}$, where d_{min} and d_{max} are two threshold values defined with respect to hand positioning in front of the camera
4) Select depth $\hat{d}_j$ as in (2)

$$\hat{d}_j = \min\{d_{c,Px} \in \mathcal{C}\} \tag{2}$$

5) Convert $(p_{x,j}\ ,\ p_{y,j}, \hat{d}_j)$ to $(x_{j,R},\ y_{j,R},\ z_{j,R})$ using pre-defined internal 2D-to-3D projection procedures

This correction procedure, repeated for each joint in each frame, is time-consuming and dependant on the value chosen for *r*, d_{min} and d_{max}, making KA best suited as an offline processing methodology. On the contrary, both GMH-D and WCA can run in our proposed system at 30 fps during live video recordings. Furthermore, this correction procedure could still fail in extreme cases.

The results produced by the three approaches should be compared with a real measurement or a trustworthy approximation; this normally occurs through a MOCAP system, the gold standard for human motion analysis. However, this kind of evaluation is not feasible in this study, for two main reasons: such system requires the use of passive or active markers, which, in this case, could alter the appearance of the hand and consequently the capability of GMH-D to estimate the hand virtual model. In addition, MOCAP systems could interfere with TOF technologies [43], deteriorating the quality of the uplifting procedure of our method.

As an alternative solution, the hand motor tasks were executed with the hand next to a ruler (ticks every 5 mm), parallel to the camera plane. The real distance (RD) between joints of interest, as read from the ruler, was manually annotated from collected videos using Kinovea™ software. RD values were then compared with the values estimated by GMH-D, WCA, and KA. The experimental tests were carried out such that the hand was, as much as possible, parallel to the ruler plane and the camera plane at evaluation instants (maximum and minimum motion peaks), thus reducing distortion effects that could alter the visual assessment of RD. Even though the ruler precision is reduced, its characteristics allow RD estimation from a 720p video even at the maximum distance of 1 m. Furthermore, this level of precision could be sufficient to provide a clinically relevant remote estimation of hand motor abilities.

The validation procedure was completed considering also the effects of two external factors: distance from camera (DF) and velocity of hand movements (VF). To test the effects of DF, the hand motor tasks were repeated at different distances, respectively 30 cm (near), 60 cm (middle), and 1 m (far). The near distance is the one that guarantees correct depth estimation by Azure Kinect camera, whereas the far distance was selected considering that, for an evaluation task focusing on the hand, this is likely the farthest position at which the hand will be placed and recorded. To test VF, only the optimal distance from DF validation was considered. The motor tasks were repeated at low, medium, and high speed. In particular, these executions were guided by a metronome running respectively at 45 bpm, 80 bpm, and 120 bpm.

IV. Characterisation of Hand Motor Skills

A. Evaluating hand motor skills

For evaluating the hand motor skills, two motor tasks commonly used in clinical practice were considered: finger tapping (FT) and opening-closing of the hand (OC). Both exercises are part of the Unified Parkinson's Disease Rating Scale (UPDRS) [11] [44] and are used in clinical practice to evaluate the effects of symptomatology on fine movements of upper limbs [25]. Moreover, the same tasks are used to evaluate motor decline in older people and in other movement disorders [9] [45].

FT and OC are challenging tasks for GMH-D due to the high dynamic profile involving hand and fingers movements. FT consists of repetitive tapping of thumb and index fingers, generally to be performed at the maximum amplitude and speed. OC consists of repetitive opening and closing movements of the hand, generally to be performed at the maximum amplitude and speed.

B. Participants and experimental procedure

To validate the measures provided by the system and to preliminarly check whether the parameters extracted by GMH-D joints can discriminate altered FT and OC executions, only healthy volunteers were recruited, without specific constraints. Nevertheless, we tried to involve subjects having different hand features (small vs big hand, male vs female) to test the robustness of the tracking methodology under this variability condition. Overall, 120 trials for each motor task were performed by 10 subjects (5 males) recruited for the study. Mean

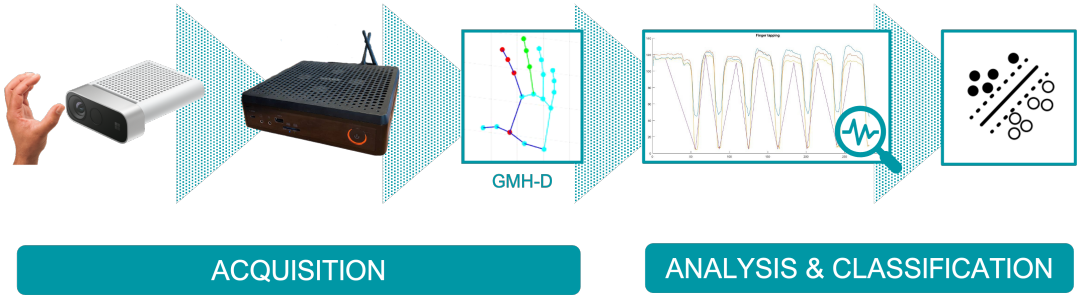

Fig. 3. Data acquisition and processing pipeline

hand length (std) for the male and female groups was 18.3 (0.6) cm and 16.5 (0.4) cm, respectively. Participants were free to choose either right or left hand to perform the task, considering this information not relevant for the test.

For DF validation, 5 subjects were involved. They were asked to perform two trials of FT and two trials of OC for each of the three investigated distances, collecting a 10 seconds recording for each trial. To coordinate execution and align the number of repetitions inside the time window, subjects were asked to synchronise their movements with a metronome running at 45 bpm. For VF validation, the same 5 subjects were asked to position their hand at 60 cm (middle distance) from the camera and perform again both FT and OC, two trials for each exercise and speed (slow, medium, fast).

For the final automatic classification, other additional 5 subjects were recruited. The 10 subjects performed, two times, three different trials at 60 cm from the camera: a normal (N) execution with natural movement amplitude and speed; a reduced (R) execution, with normal speed but reduced movement amplitude; a slow (S) execution, with normal amplitude but reduced speed. The altered trials (R and S) were requested to reproduce typical features of target population, which could be easily mimicked by healthy subjects. For example, reduced speed and low amplitude are typical signs in PD and elderly people affected by hand arthritis.
Table I summarises the total number of trials collected for validation and automatic classification purposes, for each motor task.

C. Automatic characterisation of altered performance

The last part of our pipeline investigates whether some kinematic parameters computed from GMH-D coordinates could be used to distinguish between altered executions of FT and OC. The kinematic characterisation of the two motor tasks was performed by considering the trajectory over time of the 3D distance between specific pairs of joints for each movement: Thumb Tip (TT) and Index-finger Tip (IT) joints for FT; Wrist (W) and Middle-finger Tip (MT) joints for OC. Feature extraction was automatically performed by custom-written MATLAB scripts; as only form of pre-processing of raw tracking data, a low-pass filter with cutoff frequency at 10 Hz was applied to remove high-frequency noise, prior the computation of TT-IT and W-MT distances. Table II reports the parameters that were estimated for the two motor tasks with a brief description: these features have been considered on the basis of previous works on hand motor evaluation in PD [21].

Finally, a statistical analysis of the estimated features was carried out to identify trends in the data for the three types of executions investigated (i.e., N, S, R).

For automatic classification, Support Vector Machine (SVM) was employed. Features were first normalised using z-score, then two separate classifiers were trained for FT and OC. Nested cross-validation (10 outer train-test splits, 3 internal splits) was employed to evaluate performance: kernel,

TABLE I
TRIALS COLLECTED FOR EACH MOTOR TASK

Type of Trial		**N. of Trials**
DF Validation	***Near (30 cm)***	10
	Middle (60 cm)	10
	Far (1 m)	10
VF Validation	***Slow (45 bpm)***	10
	Medium (80 dpm)	10
	Fast (120 bpm)	10
Classification	***Normal (N)***	20
	Slow (S)	20
	Reduced (R)	20

TABLE II
ESTIMATED KINEMATIC PARAMETERS FOR FT AND OC

Parameter	***Description***	***Unit***
mo_m	Range of motion as max - min (Mean)	mm
mo_cv	Range of motion as max - min (Variation Coefficient)	_
ma_m	Maximum Amplitude of movement (Mean)	mm
ma_cv	Maximum Amplitude of movement (Variation Coefficient)	_
mos_m	Maximum Opening Speed (Mean)	mm/s
mos_cv	Maximum Opening Speed (Variation Coefficient)	_
mcs_m	Maximum Closing Speed (Mean)	mm/s
mcs_cv	Maximum Closing Speed (Variation Coefficient)	_
dur_m	Movement duration (Mean)	s
dur_cv	Movement duration (Variation Coefficient)	_
freq_low	Movement Frequency	Hz

TABLE III
RMSE WITH RESPECT TO DF FOR FT AND OC

DF	Method	FT		OC	
		RMSE (max peaks)	***RMSE (min peaks)***	***RMSE (max peaks)***	***RMSE (min peaks)***
Near	***WCA***	1.54	4.90	0.64	2.78
	KA	1.42	0.36	1.29	0.58
	GMH-D	0.83	0.26	0.36	0.69
Middle	***WCA***	1.84	5.25	0.59	3.08
	KA	0.93	0.33	1.13	0.76
	GMH-D	0.48	0.32	0.40	0.49
Far	***WCA***	2.01	5.23	1.65	3.24
	KA	1.04	0.74	1.08	0.67
	GMH-D	0.55	0.50	0.30	0.58

TABLE IV
RMSE WITH RESPECT TO VF FOR FT AND OC

VF	Method	FT		OC	
		RMSE (max peaks)	***RMSE (min peaks)***	***RMSE (max peaks)***	***RMSE (min peaks)***
Slow	***WCA***	1.54	4.96	1.13	3.08
	KA	1.42	0.35	0.60	0.76
	GMH-D	0.83	0.26	0.40	0.50
Normal	***WCA***	1.84	5.25	1.23	2.53
	KA	0.93	0.33	2.77	3.45
	GMH-D	0.48	0.33	0.67	0.32
Fast	***WCA***	2.00	5.23	1.07	2.59
	KA	1.04	0.74	6.03	2.57
	GMH-D	0.55	0.50	0.60	0.51

Note: RMSE values are in cm

cost of misclassification (C), γ and degree were optimised using Grid Search in the internal cross-validation. Both multi-classes (N, S, R) and binary classifications (Normal vs Altered) were explored, considering either all the estimated features or a subset of the most correlated to class label, with Pearson's correlation $\|\rho\| > 0.6$.

V. RESULTS

A. GMH-D measures validation

The comparison between GMH-D, WCA, KA and RD was performed only at maximum and minimum peaks of fingers movement, both for FT and OC. In fact, the peaks permit the most accurate evaluation from videos of 3D distances between reference points with respect to the ruler. In addition, they correspond to the key moments in time evolution that must be identified to perform all subsequent motion analyses. The comparison was performed considering TT-IT distance for FT, and the W-MT distance for OC. Fig. 4 contains some examples of the task trajectories. It is important to notice that minimum values for FT are not 0 because finger thickness and positioning of virtual joints produce a residual distance (0.5 cm - 1 cm) even when fingers are touching.

Table III reports the Root Mean Square Error (RMSE) of GMH-D, WCA and KA methods versus RD, for DF validation: the RMSE estimated for maximum and minimum peaks were kept separated to highlight the different behaviour of the three methods. As it can be appreciated, GMH-D distance estimations are the most stable to DF and the closest to real measures, both in maximum and minimum peaks: RMSE values are always smaller than 1 cm and often below 0.5 cm for both tasks. On the contrary, RMSE values for WCA are high for FT maximum peaks (> 1.5 cm), but even more for FT and OC minimum peaks. To investigate this failure, the wrong reconstruction of virtual hand by WCA during a FT trial was analysed (Fig. 5): GMH world model fails to estimate relative depth when the two joints are very close, generating a 3D model in which they are misaligned over the z direction. On the contrary, GMH-D and KA do not present the same phenomenon thanks to the embedding of depth camera information in their estimation procedure.

In addition, WCA tends to worsen as DF increases. WCA problems could depend on the data used to train the machine learning model behind joint estimation.

Regarding KA, it provides RMSE values closer to GMH-D, in particular for mimimum peaks. However, it does not guarantee an overall good reconstruction: RMSE values for maximum peaks are often > 1 cm with respect to RD. As expected, KA seems to be less affected by DF. In general, GMH-D surpasses both methods, considering that KA strongly relies on the post-processing correction procedure described in Section III(C) and it does not always guarantee a perfect reconstruction (Fig. 4b). A distance of 60 cm from the camera seems to be optimal to evaluate FT and OC using GMH-D, hence this value was employed for the remaining steps of the pipeline evaluation.

With respect to VF, Table IV reports RMSE for minimum and maximum peaks of the three investigated methods for both motor tasks. Again, GMH-D proves to be the most stable approach: whether the execution speed, the RMSE for maximum and minimun peaks is always < 1 cm. On the contrary, WCA and KA are characterised by more variable and less accurate performance, with an average worsening as velocity increases. This effect is evident especially for KA, due to the phenomena described in Section II, which are not completely solved by the proposed correction algorithm, especially when the tasks become more challenging for hand tracking.

B. Statistical analysis of kinematic features

For the statistical analysis, FT and OC trials divided in the three previously mentioned groups (i.e., N, S, and R) were considered. Table V shows mean values and standard deviations for the estimated parameters divided by type of execution. As expected, the values are in line with the type of test: in the S category, there is a sharp drop in speed for both tasks; in the R category, there is a marked decrease in amplitude, especially for the FT task. Furthermore, in the R category, there is also a marked decrease in the execution speed. The subsequent analysis will allow highlighting the significant differences between pairs of categories.

On the extracted kinematic features *paired-t-test* was performed, to identify which parameters can discriminate more between pairs of execution categories. In addition, correlation of kinematic parameters to class label by Pearson's correlation was assessed. Results for the two statistical tests are reported, with their significance (as expressed by *p-value*), in Table VI. From the analysis, the difference N(FT) and S(FT) is more significant for parameters related to mean duration (*dur_m*)

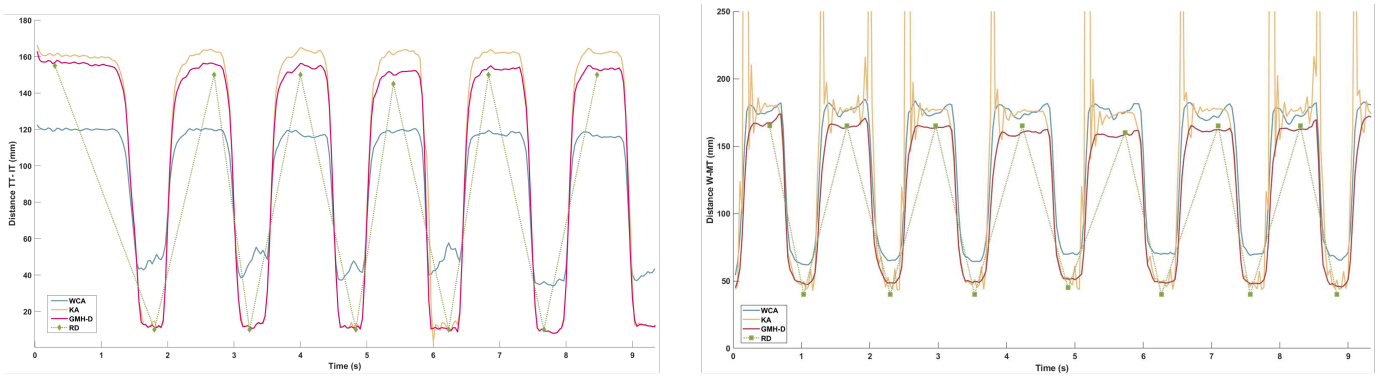

Fig. 4. a) TT-IT distance during an FT trial: GMH-D estimates the closest values to RD peaks, whereas WCA completely fails in minimum peaks; b) W-MT distance in an OC trial: KA method exhibits some spikes in the estimation of distance due to unstable depth tracking

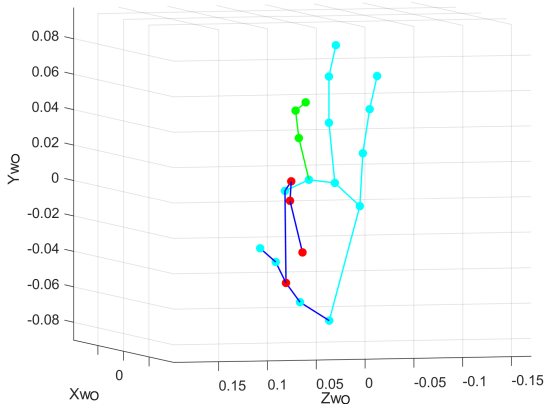

Fig. 5. Wrong reconstruction by WCA when joints TT and IT are in contact

and movement frequency (*freq_low*). In particular, the negative sign for *dur_m* and the positive sign for *freq_low* highlight that N(FT) is characterised by shorter duration and higher frequency with respect to S(FT). The other kinematic parameters are not relevant, hence the two types of execution are similar from that viewpoint, i.e., fingers speed was the same for the two executions but a different cadence was followed, as highlighted by *freq_low*.

In the discrimination between N(FT) and R(FT), all parameters show relevant discrepancy between the two categories, especially for mean amplitude, velocity, duration and frequency: this result seems to support the idea that reducing amplitude has affected indirectly also the other motor components during test execution. Finally, in the comparison between the two types of alterations (i.e., S(FT) and R(FT)) features related to movement mean amplitude and velocity are the most significant, whereas coefficients of variations are less discriminating, but for *ma_cv*: S(FT) likely favours a finer control in the amplitude of the movement performed, whereas R(FT) is characterised by higher variability in the finger opening phase. About OC task, parameters related to mean duration, velocity and frequency are the most discriminating when comparing N(OC) and S(OC). Coefficients of variation for duration (*dur_cv*) and closing speed (*mcs_cv*) are also significant, whereas the ones for opening speed and amplitude are not: this could depend on the intrinsic dynamic of the required movement.

Considering N(OC) and R(OC), there is a significant difference for all parameters except *dur_cv* (as in the case of FT) and the variability parameters related to amplitude. This may depend on the fact that OC movement is more constrained than the movement involving two fingers only, and there is still a tendency to reach more or less always the same maximum and the same minimum, as it happens also in the pathological condition. As for FT, the analysis shows that the reduced execution has indeed led to an overall variation in kinematics, which has impacted not only directly on the amplitude but, indirectly, also on the other motor components.

As regards the comparison between S(OC) and R(OC), the differences in the average values of the parameters related to the amplitude and speed of movement are statistically significant. There is no significant difference for the parameters relating to duration and variability of speed. Concerning the variability in amplitude, the difference is instead significant: certainly the S(OC) performance favours a greater control of the amplitude of the fingers compared to the R(OC) performance that can show more variability during the test.

Regarding the results of Pearson's correlation, in FT it appears that all parameters are significant in identifying the test type, with the exception of *dur_cv* which is compliant with *paired-t-test* outcome. The analysis also confirms that the parameters most correlated with altered performance are indeed the mean values of amplitude, speed, duration, and the frequency of movement ($\|\rho\| > 0.5$). In general, these features have an inverse relationship (negative sign) with respect to the type of test (normal or with abnormality) denoting a reduction in their value for altered performance, with the exception of *dur_m* which has a direct relationship (positive sign), denoting an increase of the parameter value for altered performance. Vice versa, the parameters relating to variability are less significant

TABLE V
MEAN VALUES AND STANDARD DEVIATIONS OF PARAMETERS FOR FT AND OC TASKS

Parameter	Mean ± std					
	FT			OC		
	Normal (N)	*Slow (S)*	*Reduced (R)*	*Normal (N)*	*Slow (S)*	*Reduced (R)*
mo_m (mm)	110.11 ± 18.65	118.60 ± 13.96	60.69 ± 14.84	123.04 ± 7.02	126.23 ± 8.47	105.05 ± 9.72
mo_cv (-)	0.06 ± 0.06	0.03 ± 0.02	0.15 ± 0.09	0.03 ± 0.01	0.02 ± 0.01	0.05 ± 0.03
ma_m (mm)	117.88 ± 18.21	126.00 ± 14.33	71.99 ± 15.12	164.84 ± 5.86	167.95 ± 5.78	156.76 ± 7.19
ma_cv (-)	0.05 ± 0.04	0.03 ± 0.02	0.11 ± 0.08	0.02 ± 0.01	0.01 ± 0.01	0.03 ± 0.02
mos_m (mm/s)	969.11 ± 213.09	856.12 ± 387.73	250.40 ± 84.36	1651.44 ± 150.45	1279.71 ± 271.82	581.78 ± 228.36
mos_cv (-)	0.13 ± 0.05	0.19 ± 0.11	0.25 ± 0.15	0.09 ± 0.03	0.12 ± 0.08	0.19 ± 0.10
mcs_m (mm/s)	1072.20 ± 309.05	1057.34 ± 496.93	303.61 ± 138.92	1764.80 ± 258.96	1217.33 ± 403.14	469.50 ± 222.07
mcs_cv (-)	0.17 ± 0.08	0.21 ± 0.14	0.33 ± 0.21	0.10 ± 0.03	0.16 ± 0.06	0.18 ± 0.09
dur_m (s)	0.78 ± 0.23	1.35 ± 0.13	1.40 ± 0.36	0.73 ± 0.02	1.41 ± 0.20	1.57 ± 0.35
dur_cv (-)	0.13 ± 0.09	0.12 ± 0.08	0.11 ± 0.04	0.13 ± 0.04	0.08 ± 0.03	0.12 ± 0.10
freq_low (hz)	1.33 ± 0.05	0.76 ± 0.12	0.98 ± 0.28	1.32 ± 0.05	0.76 ± 0.16	1.06 ± 0.31

TABLE VI
PAIRED-T-TEST AND PEARSON'S CORRELATION TO CLASS LABEL FOR FT AND OC

Parameter	Paired-T-Test						Pearson's coefficient	
	FT			OC			FT	OC
	N vs S	*N vs R*	*S vs R*	*N vs S*	*N vs R*	*S vs R*		
mo_m (mm)	-1.34	7.39 ***	11.70 ***	-1.03	6.69 ***	6.01 ***	-0.68 ***	-0.59 ***
mo_cv (-)	1.45	-4.59 ***	-2.99 *	1.74	-2.76 *	-3.32 **	0.48 ***	0.38 *
ma_m (mm)	-1.29	7.07 ***	10.80 ***	-1.30	3.40 **	5.28 ***	-0.66 ***	-0.43 **
ma_cv (-)	1.32	-2.54 *	-3.84 **	2.11	-1.77	-2.69 **	0.43 **	0.27
mos_m (mm/s)	0.94	11.7 ***	6.60 ***	4.25 ***	13.6 ***	12 ***	-0.73 ***	-0.89 ***
mos_cv (-)	-1.69	-2.93 *	-1.50	-1.62	-3.55 **	-2.09	0.44 **	0.51***
mcs_m (mm/s)	0.12	8.32 ***	6.21 ***	4.63 ***	13.0 ***	6.27 ***	-0.64 ***	-0.87 ***
mcs_cv (-)	-1.09	-2.90 *	-1.84	-4.23 ***	-3.42 **	-0.617	0.39 **	0.46 **
dur_m (s)	-8.17 ***	-5.71 ***	-0.54	-13.20 ***	-9.00 ***	-1.56	0.67 ***	0.80 ***
dur_cv (-)	0.44	0.77	0.39	4.87 ***	0.457	-1.56	-0.13	-0.08
freq_low (hz)	16.8 ***	5.04 ***	-2.46 *	12.4 ***	3.04 **	-3.37 **	-0.50***	-0.34 *

Note: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

and with a direct relationship (positive sign of the coefficient), demonstrating an increase in variability for altered trials. In OC, it appears that all features are significant in identifying the category, with the exception of the variability of duration and amplitude: this result is in agreement with the *paired-t-test*, from which it emerges that the *dur_cv* parameter is never significant in the comparison between type of trials, while the *mo_cv* and *ma_cv* are only significant in the comparison between S(OC) and R(OC) types. The analysis also confirms that the parameters most correlated with altered performance (whether in terms of speed or amplitude) are indeed the mean values of the amplitude, speed, and duration parameters and the variability of the speed parameters ($\|\rho\| > 0.5$): in general, these parameters have an inverse relationship (negative sign of the coefficient) with respect to the type of test, denoting a reduction of the feature for altered performance, with the exception of the *dur_m* which instead has a direct relationship (positive sign of the coefficient). Vice versa, the coefficient of variation of amplitude and duration are less significant or not significant, with a direct relationship denoting an increase in variability for altered performance. Fig. 6 shows scatter plots of the data for FT and OC with respect to the top three most correlated features: we can observe that N points are well separated (with the exception of few outliers in FT) from S and R points, which can be seen as an unique "altered" (A) class.

C. Results of the automatic classification

Accuracy performance of SVM classifiers from nested-cross validation are reported in Table VII. Although complex combinations of hyperparameters were considered, due to the high separation among categories (Fig. 6), linear kernel and C=1 proved to be the most recurrent (and sufficient) configuration to classify FT and OC, both in the multi-classes and binary problem. The neat separation between N and A classes (i.e., grouping R and S) also explains the very high accuracy achieved by the binary classifiers. Furthermore, as expected, performance increased when reducing the feature set to the most correlated parameters, thus removing irrelevant or confusing features.

Even if these results are high, their statistical meaningfulness is limited because they were achieved on data from a small

TABLE VII
ACCURACY OF SVM CLASSIFIERS

Feature set	FT		OC	
	Multiclass (%)	*Binary (%)*	*Multiclass (%)*	*Binary(%)*
Full	91.5 ± 10.5	95.0 ± 10.0	95.5 ± 9.1	97.5 ± 7.5
$\rho > \mathbf{0.6}$	94.0 ± 9.2	97.5 ± 7.5	98.0 ± 6.0	100.0 ± 0.0

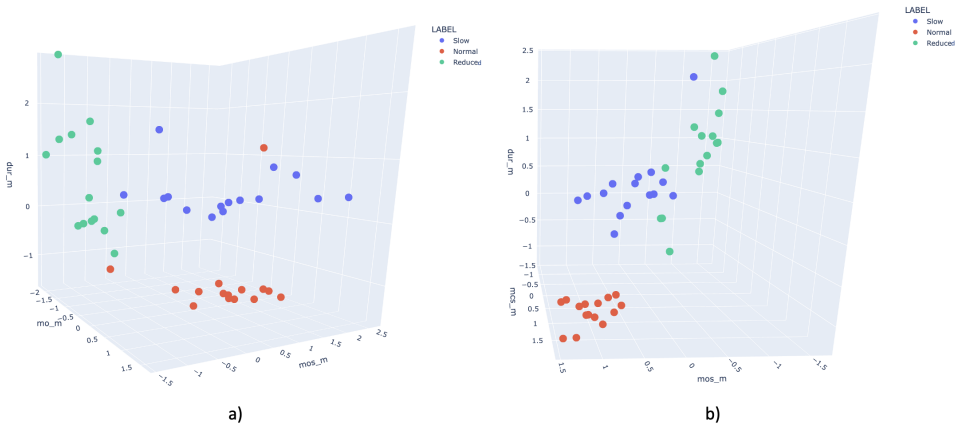

Fig. 6. Scatter plots of data for classification with respect to the three features most correlated to the class label: a) Scatter plot of FT data by type of trials performed, Normal class is separated neatly from Reduced and Slow but for few outliers; b) Scatter plot of OC data by type of trials performed, Normal class is separated neatly from Reduced and Slow

cohort of healthy subjects, performing simulations. For these reasons, the two classes N and A can be considered as extreme cases in a real, more variegated spectrum. It is expected that real elderly and pathological subjects will span in the middle between these two extremes, making the classification task more challenging, but still feasible for the presented system.

VI. Conclusions

This work proposes a pipeline suitable for automatic and non-invasive remote assessment of hand motor skills in elderly and pathological subjects. This pipeline is based on GMH-D, an innovative mixed approach to track and reconstruct a 3D, 21-joints, virtual hand combining Google MediaPipe Hands (GMH) and video recordings from a RGB-Depth camera (Azure Kinect). GMH-D fuses together 2D coordinates and relative depth estimations provided by GMH with depth data from the Kinect, exploting then predefined 2D-to-3D projection routines to infer 3D joints trajectories over time.
The validation results demonstrated that this method provides an accurate and stable estimation of 3D coordinates at 60 cm from the camera, also in conditions where the hand is involved in highly dynamic and fast tasks as finger tapping (FT) or repeated opening and closing of the hand (OC).
GMH-D is able to correctly estimate distances among joints of interest at key time instants (minimum and maximum peaks) with an error smaller than 1 cm, surpassing accuracy and reliability of MediaPipe native 3D world coordinates (WCA) and another method exploiting depth information from the camera for all points estimation (KA). This result configures the proposed approach as consistent for clinically-relevant kinematic analysis.
Moreover, the method allows extracting kinematic parameters that proved significant to identify simulated alterations of hand movements. Indeed, supervised classifiers (SVM) trained on these features were able to achieve a mean accuracy higher than 90% in the recognition of altered FT and OC trials.

The proposed approach is clearly not exempted from limitations: the method can currently only evaluate one hand at a time; only motor tasks performed with the hand in front of the camera and visible wrist were evaluated; the automatic classification results were obtained on a small cohort of subjects performing simulated alterations. However, future works will focus on overcoming these limitations: for example, exploring alternatives to MOCAP systems, which could alter the quality of the proposed method, to evaluate reconstructed distances in more complex camera setup and more challenging hand motor tasks; testing the pipeline on subjects with real impaired hand mobility; migrating the solution on other devices, like smartphones or tablets with an embedded RGB-D camera, to make the acquisition system even more portable, hence suitable for remote assessment scenarios.
Nevertheless, considering the preliminary results obtained by this study, it is reasonable to expect that the proposed approach could find a range of applications in clinical and non-clinical scenarios where the computation of accurate kinematic parameters is required to estimate hand motor performance.

References

[1] E. Carmeli, H. Patish, and R. Coleman, "The aging hand," *J. Gerontol. A Biol. Sci. Med. Sci.*, vol. 58, no. 2, pp. 146–152, 2003.
[2] V. K. Ranganathan, V. Siemionow, V. Sahgal, and G. H. Yue, "Effects of aging on hand function," *J. Am. Geriatr. Soc.*, vol. 49, no. 11, pp. 1478–1484, 2001.
[3] J. L. Bowden and P. A. McNulty, "The magnitude and rate of reduction in strength, dexterity and sensation in the human hand vary with ageing," *Exp. Gerontol.*, vol. 48, no. 8, pp. 756–765, 2013.
[4] J. R. Lukos, H. Poizner, and J. I. Sage, "Hand function in parkinson's disease," in *Hand Function*, pp. 133–149, New York, NY: Springer New York, 2014.
[5] E. L. Proud, K. J. Miller, B. Bilney, S. Balachandran, J. L. McGinley, and M. E. Morris, "Evaluation of measures of upper limb functioning and disability in people with parkinson disease: a systematic review," *Arch. Phys. Med. Rehabil.*, vol. 96, no. 3, pp. 540–551.e1, 2015.
[6] W. Muellbacher, C. Richards, U. Ziemann, G. Wittenberg, D. Weltz, B. Boroojerdi, L. Cohen, and M. Hallett, "Improving hand function in chronic stroke," *Arch. Neurol.*, vol. 59, no. 8, pp. 1278–1282, 2002.
[7] O. H. Gündüz and C. Ş. Toprak, "Hand function in stroke," in *Hand Function*, pp. 125–135, Cham: Springer International Publishing, 2019.

[8] C. Barut, E. Kızıltan, E. Gelir, and F. Köktürk, "Advanced analysis of finger-tapping performance: a preliminary study," *Balkan Med. J.*, vol. 30, no. 2, pp. 167–171, 2013.

[9] L. Carment, A. Abdellatif, C. Lafuente-Lafuente, S. Pariel, M. A. Maier, J. Belmin, and P. G. Lindberg, "Manual dexterity and aging: A pilot study disentangling sensorimotor from cognitive decline," *Front. Neurol.*, vol. 9, p. 910, 2018.

[10] G. M. Johansson and C. K. Häger, "Measurement properties of the motor evaluation scale for upper extremity in stroke patients (MESUPES)," *Disabil. Rehabil.*, vol. 34, no. 4, pp. 288–294, 2012.

[11] C. G. Goetz, B. C. Tilley, S. R. Shaftman, Stebbins, *et al.*, "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (MDS-UPDRS): scale presentation and clinimetric testing results: MDS-UPDRS: Clinimetric assessment," *Mov. Disord.*, vol. 23, no. 15, pp. 2129–2170, 2008.

[12] A. M. Valevicius, Q. A. Boser, E. B. Lavoie, G. S. Murgatroyd, P. M. Pilarski, C. S. Chapman, A. H. Vette, and J. S. Hebert, "Characterization of normative hand movements during two functional upper limb tasks," *PLoS One*, vol. 13, no. 6, p. e0199549, 2018.

[13] J. Connolly, J. Condell, B. O'Flynn, J. T. Sanchez, and P. Gardiner, "IMU sensor-based electronic goniometric glove (iSEG-Glove) for clinical finger movement analysis," *IEEE Sens. J.*, pp. 1–1, 2017.

[14] L. di Biase, S. Summa, J. Tosi, F. Taffoni, M. Marano, A. Cascio Rizzo, F. Vecchio, D. Formica, V. Di Lazzaro, G. Di Pino, and M. Tombini, "Quantitative analysis of bradykinesia and rigidity in parkinson's disease," *Front. Neurol.*, vol. 9, 2018.

[15] F. Cavallo, A. Moschetti, D. Esposito, C. Maremmani, and E. Rovini, "Upper limb motor pre-clinical assessment in parkinson's disease using machine learning," *Parkinsonism Relat. Disord.*, vol. 63, pp. 111–116, 2019.

[16] V. Bobić, M. Djurić-Jovičić, N. Dragašević, M. B. Popović, V. S. Kostić, and G. Kvaščev, "An expert system for quantification of bradykinesia based on wearable inertial sensors," *Sensors (Basel)*, vol. 19, no. 11, p. 2644, 2019.

[17] A. Schwarz, M. M. C. Bhagubai, G. Wolterink, J. P. O. Held, A. R. Luft, and P. H. Veltink, "Assessment of upper limb movement impairments after stroke using wearable inertial sensing," *Sensors (Basel)*, vol. 20, no. 17, p. 4770, 2020.

[18] Q. Sanders, V. Chan, R. Augsburger, S. C. Cramer, D. J. Reinkensmeyer, and A. H. Do, "Feasibility of wearable sensing for in-home finger rehabilitation early after stroke," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1363–1372, 2020.

[19] C. Yin, Q. Liu, W. Meng, and Q. Ai, "Quantitative evaluation of hand functions using a wearable glove with multiple sensors," in *2021 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, IEEE, 2021.

[20] P. J. M. Bank, J. Marinus, C. G. M. Meskers, J. H. de Groot, and J. J. van Hilten, "Optical hand tracking: A novel technique for the assessment of bradykinesia in parkinson's disease," *Mov. Disord. Clin. Pract.*, vol. 4, no. 6, pp. 875–883, 2017.

[21] C. Ferraris, R. Nerino, A. Chimienti, G. Pettiti, N. Cau, V. Cimolin, C. Azzaro, G. Albani, L. Priano, and A. Mauro, "A self-managed system for automated assessment of UPDRS upper limb tasks in parkinson's disease," *Sensors (Basel)*, vol. 18, no. 10, p. 3523, 2018.

[22] D. Simonsen, E. G. Spaich, J. Hansen, and O. K. Andersen, "Design and test of a closed-loop FES system for supporting function of the hemiparetic hand based on automatic detection using the microsoft kinect sensor," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 8, pp. 1249–1256, 2017.

[23] S. Cho, W.-S. Kim, N.-J. Paik, and H. Bang, "Upper-limb function assessment using VBBTs for stroke patients," *IEEE Comput. Graph. Appl.*, vol. 36, no. 1, pp. 70–78, 2016.

[24] "Leap motion tracking." https://www.ultraleap.com/product/leap-motion-controller/. Accessed: 2022-3-30.

[25] A. H. Butt, E. Rovini, C. Dolciotti, G. De Petris, P. Bongioanni, M. C. Carboncini, and F. Cavallo, "Objective and automatic classification of parkinson disease with leap motion controller," *Biomed. Eng. Online*, vol. 17, no. 1, p. 168, 2018.

[26] E. D. Oña, C. Balaguer, R. Cano-de la Cuerda, S. Collado-Vázquez, and A. Jardón, "Effectiveness of serious games for leap motion on the functionality of the upper limb in parkinson's disease: A feasibility study," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–17, 2018.

[27] A. Moshkova, A. Samorodov, N. Voinova, A. Volkov, E. Ivanova, and E. Fedotova, "Parkinson's disease detection by using machine learning algorithms and hand movement signal from LeapMotion sensor," in *2020 26th Conference of Open Innovations Association (FRUCT)*, IEEE, 2020.

[28] T. Zhang, Z.-R. Wang, P. Wang, L. Xing, L.-P. Mei, and J. Zhao, "Leap motion-based virtual reality training for improving motor functional recovery of upper limbs and neural reorganization in subacute stroke patients," *Neural Regen. Res.*, vol. 12, no. 11, p. 1823, 2017.

[29] M. R. S. B. de Souza, R. S. Gonçalves, and G. Carbone, "Feasibility and performance validation of a leap motion controller for upper limb rehabilitation," *Robotics*, vol. 10, no. 4, p. 130, 2021.

[30] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, and M. Grundmann, "MediaPipe: A framework for building perception pipelines," *arXiv:1906.08172*, 2019.

[31] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C. L. Chang, and M. Grundmann, "Mediapipe hands: On-device real-time hand tracking," *arXiv:1906.08172*, 2020.

[32] V. Chunduru, M. Roy, D. R. N., and R. G. Chittawadigi, "Hand tracking in 3D space using MediaPipe and PnP method for intuitive control of virtual globe," in *2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC)*, IEEE, 2021.

[33] W. A. A. Praditasari, R. Aprilliyani, and I. Kholis, "Design and implementation of interactive virtual museum based on hand tracking OpenCV in indonesia," in *2021 8th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, IEEE, 2021.

[34] B. J. Boruah, A. K. Talukdar, and K. K. Sarma, "Development of a learning-aid tool using hand gesture based human computer interaction system," in *2021 Advanced Communication Technologies and Signal Processing (ACTS)*, IEEE, 2021.

[35] R. Damindarov, C. A. Fam, R. A. Boby, M. Fahim, A. Klimchik, and T. Matsumaru, "A depth camera-based system to enable touchless interaction using hand gestures," in *2021 International Conference "Nonlinearity, Information and Robotics" (NIR)*, IEEE, 2021.

[36] Halder, Arpita, and A. Tayade, "Real-time vernacular sign language recognition using mediapipe and machine learning," *International Journal of Research Publication and Reviews*, vol. 2, no. 5, pp. 9–17, 2021.

[37] A. Chaikaew, K. Somkuan, and T. Yuyen, "Thai sign language recognition: An application of deep neural network," in *2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering*, IEEE, 2021.

[38] J. Shin, A. Matsuoka, M. A. M. Hasan, and A. Y. Srizon, "American sign language alphabet recognition by extracting feature from hand pose estimation," *Sensors (Basel)*, vol. 21, no. 17, 2021.

[39] E. D'Antonio, J. Taborri, I. Mileti, S. Rossi, and F. Patané, "Validation of a 3d markerless system for gait analysis based on openpose and two rgb webcams," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17064–17075, 2021.

[40] V. Cimolin, L. Vismara, C. Ferraris, G. Amprimo, G. Pettiti, R. Lopez, M. Galli, R. Cremascoli, S. Sinagra, A. Mauro, and L. Priano, "Computation of gait parameters in post stroke and parkinson's disease: A comparative study using RGB-D sensors and optoelectronic systems," *Sensors (Basel)*, vol. 22, no. 3, p. 824, 2022.

[41] P. Sturm, "Pinhole camera model," in *Computer Vision*, pp. 610–613, Boston, MA: Springer US, 2014.

[42] J. Volak, J. Bajzik, S. Janisova, D. Koniar, and L. Hargas, "Real-time interference artifacts suppression in array of ToF sensors," *Sensors (Basel)*, vol. 20, no. 13, p. 3701, 2020.

[43] M. Naeemabadi, B. Dinesen, O. K. Andersen, and J. Hansen, "Investigating the impact of a motion capture system on microsoft kinect v2 recordings: A caution for using the technologies together," *PLoS One*, vol. 13, no. 9, p. e0204052, 2018.

[44] M. Djurić-Jovičić, I. Petrović, Ječmenica-Lukić, *et al.*, "Finger tapping analysis in patients with parkinson's disease and atypical parkinsonism," *J. Clin. Neurosci.*, vol. 30, pp. 49–55, 2016.

[45] N. S. Frolov, E. N. Pitsik, V. A. Maksimenko, V. V. Grubov, A. R. Kiselev, Z. Wang, and A. E. Hramov, "Age-related slowing down in the motor initiation in elderly adults," *PLoS One*, vol. 15, no. 9, p. e0233942, 2020.