

The Vision Dexterity Gap: Evaluating Webcam-Based Hand Motor Assessment Across Visual Conditions

Divyanshi Dubey

Alpharetta High School, Alpharetta, Georgia, USA

Abstract— Objective quantification of hand motor performance commonly relies on clinical observation, specialized sensors, or controlled laboratory equipment, limiting accessibility outside traditional assessment environments. This study presents Vision Dexterity, a webcam-based computer vision framework designed to quantify hand movement through three motor tasks: finger tapping, sequential finger opposition, and rapid alternating movement (RAM). The system uses Media Pipe Hand Landmarked to extract 21 three-dimensional hand landmarks from ordinary video and applies task-specific algorithms to quantify movement events. Finger tapping and sequencing were detected using normalized inter-landmark distances and hysteresis thresholds, while RAM orientation was classified using geometric hand features and a participant-specific logistic regression model. Robustness of RAM cycle counting was evaluated by systematically modifying recorded videos using brightness adjustment and Gaussian blur while preserving the underlying movements. An initial reference recording showed sensitivity to these transformations despite consistently high hand-tracking availability. However, across the expanded multi-recording evaluation, eight of nine tested video conditions achieved 100% cycle-count agreement with their corresponding undegraded recordings; only the 150% brightness condition showed reduced agreement at 75%. These findings demonstrate the feasibility of extracting quantitative hand-movement measures from conventional videos while showing that successful hand tracking alone does not guarantee stable downstream measurement. Vision Dexterity provides a low-cost framework for investigating camera-based hand motor assessment and highlights the importance of evaluating robustness across multiple recordings and visual conditions.

Index Terms—Computer vision, hand tracking, motor assessment, Media Pipe, rapid alternating movement, finger tapping, human-computer interaction

I. INTRODUCTION

Hand dexterity is essential for many activities of daily living and depends on the ability to perform rapid, coordinated, and controlled movements of the fingers, hand, and forearm [1]. Dexterity assessments are commonly used to quantify motor performance, characterize functional impairment, and evaluate changes in hand function over time. Established assessments include performance-based tasks that measure characteristics such as movement speed, object manipulation, and fine hand use [1], [2]. However, many conventional assessments require physical objects or equipment, standardized administration procedures, and direct observation, which can make frequent or remote assessment more difficult [1], [2].

Recent advances in computer vision provide a potential alternative to equipment-based assessment. Markerless hand-pose estimation methods can estimate hand position and movement directly from camera images without requiring physical markers or wearable sensing devices [4]. MediaPipe Hands introduced a real-time hand-tracking pipeline capable of predicting a hand skeleton and landmark coordinates from a single RGB camera input [3]. Such landmark representations provide a computational basis for extracting movement characteristics from video and may allow measures such as repetition frequency, movement sequencing, and changes in hand orientation to be calculated automatically.

The application of camera-based motion analysis to hand function is supported by a broader shift toward quantitative motion-tracking technologies in rehabilitation and movement assessment [6]. Nevertheless, translating hand-landmark coordinates into meaningful dexterity measurements presents several challenges. Hand and finger movements are

highly variable, and differences in task execution can reduce the repeatability of kinematic measurements [6]. Hand-pose estimation itself is also challenging because of the complex structure and large number of possible configurations of the human hand [4]. Factors such as hand position, orientation, camera viewpoint, movement speed, and transitions between poses may therefore influence measurements derived from video. An additional challenge is that dexterity is not a single, uniformly defined construct. Performance-based assessments measure different components of hand function and often rely heavily on movement speed, while other aspects of dexterity may require different forms of measurement [1],[2]. A computer-vision framework intended to assess multiple hand movements must therefore use task-specific methods for translating landmark trajectories into meaningful quantitative variables.

To investigate this problem, we developed VisionDexterity, a webcam-based computer vision framework for assessing three hand motor tasks: finger tapping, finger sequencing, and rapid alternating movements (RAM). The system uses markerless hand-landmark tracking and task-specific computational analysis to transform webcam video into quantitative measures of motor performance.

We hypothesized that hand-landmark data obtained from ordinary webcam video could be transformed into consistent task-specific measures across multiple types of hand movement. By evaluating VisionDexterity across controlled trials, this study examines the feasibility and limitations of webcam-only hand dexterity assessment and provides a foundation for future validation of accessible computer vision-based motor assessment.

II. RESULTS

VisionDexterity successfully processed hand movement across the three implemented motor tasks: finger tapping, sequential finger opposition, and rapid alternating movement (RAM). The finger-based assessments used normalized landmark distances to identify discrete thumb-to-finger contact events, while the RAM pipeline classified changes in hand orientation and converted these state transitions into complete movement cycles. Across the controlled robustness experiment, the unmodified reference recording produced two detected RAM cycles. A re-

encoded control preserved the same two-cycle measurement, whereas each tested brightness- and blur-degraded version produced three detected cycles despite containing the same underlying physical movement.

Table I: RAM Robustness Under Controlled Video Degradation

Condition	Level	Tracking (%)	Detected Cycles	Cycle Error	Error (%)
Reference	—	96.53	2	0	0
Brightness	50	98.26	3	+1	50
Brightness	75	97.88	3	+1	50
Brightness	125	97.04	3	+1	50
Brightness	150	96.97	3	+1	50
Gaussian Blur	3	97.19	3	+1	50
Gaussian Blur	7	97.19	3	+1	50
Gaussian Blur	11	97.19	3	+1	50
Gaussian Blur	21	96.97	3	+1	50
Re-encoded Control	—	96.53	2	0	0

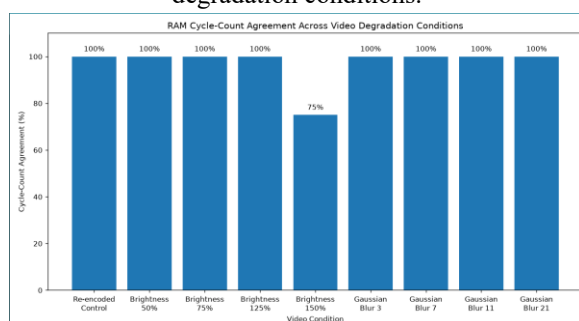
The controlled degradation experiment revealed a consistent cycle-count error despite high hand-tracking availability (Table I). The reference recording produced two detected RAM cycles with 96.53% of processed frames successfully tracked. The re-encoded control also produced two cycles with the same tracking percentage, indicating that re-encoding alone did not alter the measurement. In contrast, all four brightness-modified conditions and all four Gaussian-blur conditions produced three detected cycles. Each degraded condition therefore introduced one additional detected cycle, corresponding to a 50% cycle-count error relative to the two-cycle reference. Notably, the cycle-count error was not accompanied by a substantial loss of hand tracking. Tracking remained between 96.97% and 98.26% across all degraded conditions and was therefore comparable to, or slightly higher than, the 96.53% observed for the reference recording. This indicates that, within this initial reference recording, successful hand detection alone was insufficient to guarantee accurate downstream movement quantification. However, because these findings were obtained from transformations of a single recording, additional recordings were evaluated to determine whether the observed cycle-count instability generalized across movement samples.

RAM cycle-count robustness across multiple recordings

To determine whether the cycle-count instability observed in the initial recording generalized across

different movement samples, the robustness analysis was expanded to four independently recorded RAM videos. Cycle-count agreement was calculated as the percentage of degraded videos for which the detected number of RAM cycles matched that of the corresponding undegraded reference recording. Across the expanded evaluation, eight of the nine tested video conditions achieved 100% cycle-count agreement (Fig. 1). The only reduction occurred under the 150% brightness condition, which achieved 75% agreement. All Gaussian-blur conditions, including the largest tested blur kernel, maintained 100% agreement, as did the re-encoded control and the remaining brightness conditions.

Figure 1: RAM cycle-count agreement across video degradation conditions.



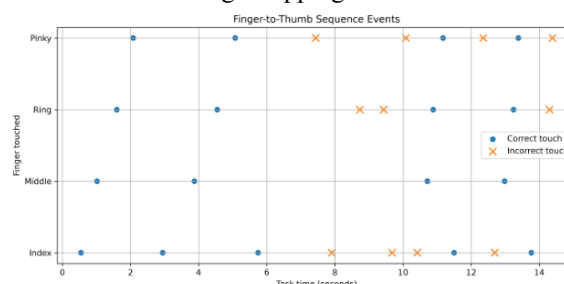
These results indicate that the cycle-count error observed in the initial controlled experiment was not consistently reproduced across the expanded set of recordings. Instead, RAM cycle counting remained stable under most tested transformations, while the highest brightness condition produced reduced agreement. The difference between the initial and expanded experiments also demonstrates the importance of evaluating robustness across multiple recordings rather than drawing conclusions from a single reference video.

Finger-tapping events were detected from normalized landmark distance

For the finger-tapping task, movement was quantified using the normalized distance between the thumb and index fingertips over time. Repeated decreases in thumb-index distance produced distinct tapping events, while separate tap and release thresholds were used to distinguish completed movements and reduce repeated detections during a single contact (Fig. 2). The resulting signal demonstrates that repeated finger-

tapping movements could be extracted directly from webcam-derived hand landmarks without requiring wearable sensors or specialized motion-capture equipment.

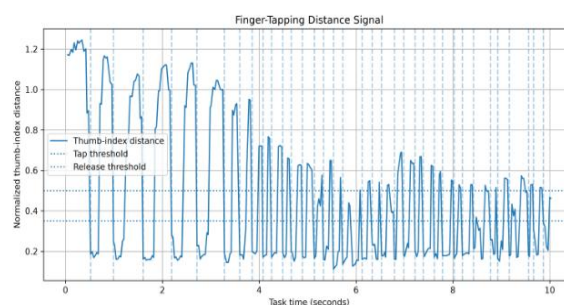
Figure 2. Normalized thumb-index distance signal and detected finger-tapping events over time.



Finger-sequence performance captured correct and incorrect touches

The finger-to-thumb sequencing task was evaluated by identifying individual finger-contact events and comparing the detected finger with the expected sequence. Correct and incorrect touches were mapped over task time, allowing both the temporal pattern and ordering accuracy of the sequence to be visualized (Fig. 3). Correct touches were observed throughout the recording, while incorrect touches were also identified during portions of the sequence, demonstrating the framework's ability to distinguish successful contacts from sequence errors rather than relying solely on total movement count.

Figure 3. Detected finger-to-thumb sequence events over time, showing correct and incorrect finger contacts.



III. DISCUSSION

The findings demonstrate that VisionDexterity can extract multiple measures of hand motor performance from standard webcam video using a single landmark-based computational framework. Across the three

implemented tasks, the system captured distinct aspects of movement: repetitive thumb–index motion during finger tapping, ordered finger contacts during finger sequencing, and alternating hand orientation during rapid alternating movement (RAM). Together, these results suggest that webcam-derived hand landmarks can support more detailed movement analysis than simple hand detection alone.

The finger-tapping and finger-sequencing results illustrate the framework's ability to convert continuous landmark data into task-specific movement events. In finger tapping, changes in normalized thumb–index distance produced a repeated movement signal from which individual taps could be detected. Finger sequencing required a different analysis approach: instead of measuring repetition alone, the system identified which finger contacted the thumb and classified each contact as correct or incorrect relative to the expected sequence. Together, these findings show that the same hand-landmark framework can support different measures of motor performance through task-specific analysis rules.

The RAM experiments additionally revealed an important distinction between hand-tracking availability and measurement reliability. In the initial controlled-degradation experiment, landmark tracking remained high even when the derived cycle count changed. Thus, successful detection of a hand in most video frames did not necessarily guarantee that the downstream movement metric would remain unchanged. This distinction is important for computer-vision-based assessment because previous research on markerless motion-capture systems has similarly shown that measurement validity and reliability can vary across particular movement variables and tasks [5], [6].

However, the expanded multi-recording experiment provided a more nuanced picture of this vulnerability. Cycle-count agreement was preserved across nearly all tested degradation conditions, with reduced agreement occurring only under the 150% brightness condition. The widespread errors observed in the initial reference-video experiment therefore did not generalize across the broader recording set. Rather than indicating that RAM measurement is inherently unstable under brightness or blur, these findings suggest that robustness may depend on interactions between individual recordings and particular visual transformations. This interpretation is consistent with

evidence that the performance of markerless motion-capture systems can vary according to the movement task and measurement context [5], [6]. More recent experimental evidence has also demonstrated that lighting conditions and camera positioning can significantly affect kinematic measurements obtained from markerless motion capture [7].

Several limitations should be considered when interpreting these findings. The present study was designed primarily as a proof-of-concept computational evaluation rather than a clinical validation study. The number and diversity of recordings were limited, and the controlled brightness and Gaussian-blur transformations represent only a subset of the conditions encountered during real-world webcam use. Factors such as camera angle, hand-to-camera distance, background complexity, illumination, motion blur, occlusion, and differences in camera hardware may further influence landmark estimation and downstream measurements. Camera positioning and lighting have been experimentally shown to influence measurements produced by markerless motion-capture systems [7], while suboptimal camera placement and field of view have also been associated with reduced reliability [8]. In addition, although VisionDexterity can quantify movement features and task performance, the present results do not establish diagnostic accuracy or correspondence with clinician-administered motor assessments.

Future work should therefore evaluate VisionDexterity in larger and more diverse participant samples and compare its measurements with established clinical or instrumented reference measures. Validation against reference systems is particularly important because systematic reviews of markerless motion capture have found promising validity and reliability for some movement measures while also identifying substantial variation across joints, tasks, systems, and measurement conditions [5], [6]. Additional robustness testing could systematically vary camera position, resolution, frame rate, lighting, and occlusion while examining not only whether landmarks remain detectable but whether derived measures such as tap frequency, sequencing accuracy, and RAM cycle count remain stable. Such validation would help determine which webcam-derived metrics are sufficiently reliable for remote

motor assessment and where additional calibration or quality-control procedures are required.

Overall, VisionDexterity demonstrates the feasibility of transforming ordinary webcam video into quantitative measures across multiple hand motor tasks without requiring specialized sensing hardware. The results also emphasize that accessibility alone is not sufficient: a useful webcam-based assessment system must preserve the reliability of its downstream movement measurements across realistic recording conditions. By combining accessible video capture with task-specific movement analysis and explicit robustness testing, VisionDexterity provides a foundation for further investigation of scalable computer-vision-based hand motor assessment.

IV. MATERIALS AND METHODS

VisionDexterity framework

VisionDexterity was developed in Python as a webcam-based framework for quantifying hand movement during three motor tasks: finger tapping, finger-to-thumb sequencing, and rapid alternating movement (RAM). Video acquisition and frame processing were performed using OpenCV [9], while hand localization was performed using the MediaPipe Hand Landmarker [3]. The framework tracked a single hand and extracted three-dimensional landmark coordinates from each successfully processed frame. These landmarks were subsequently converted into task-specific geometric features and movement events. The same landmark representation was therefore used across all three tasks, while separate analysis procedures were implemented according to the movement being assessed.

Finger-tapping task

The finger-tapping task was recorded for 10 s following a 3-s countdown. Finger tapping is an established approach for evaluating aspects of motor performance, including movement speed and rhythm [10]. Movement was quantified using the three-dimensional Euclidean distance between the thumb tip and index fingertip. Landmark distance was normalized by the three-dimensional distance between the wrist and middle-finger metacarpophalangeal (MCP) landmark before event detection. A hysteresis-based procedure was used to distinguish individual taps: a normalized thumb-index distance below 0.35

indicated contact, while the distance was required to exceed 0.50 before another tap could be registered. The use of separate contact and release thresholds reduced repeated detections during a single sustained contact.

Recorded data were stored on a frame-by-frame basis and included normalized pinch distance and cumulative tap count. Post-processing calculated the total number of detected taps, tapping frequency, inter-tap intervals, and variability in the time between successive taps. The normalized distance signal and detected events were subsequently visualized as a function of task time.

Finger-to-thumb sequencing task

The finger-sequencing task was recorded for 15 s following a 3-s countdown. Participants performed the ordered sequence index $\rightarrow$ middle $\rightarrow$ ring $\rightarrow$ little finger, with each finger contacting the thumb. Sequential finger-to-thumb opposition has been used in previous research as a task for evaluating sequential motor performance [11]. The MediaPipe fingertip landmarks corresponding to the index, middle, ring, and little fingers were compared with the thumb-tip landmark using three-dimensional Euclidean distance. As in finger tapping, fingertip distances were normalized by the three-dimensional distance between the wrist and middle-finger metacarpophalangeal (MCP) landmark and evaluated using separate touch and release thresholds of 0.35 and 0.50, respectively. Each detected contact was assigned to the finger closest to the thumb and compared with the finger expected at that position in the predefined sequence. Events could therefore be categorized according to both finger identity and sequencing accuracy. This allowed the framework to measure ordered motor performance rather than relying solely on the total number of movements.

Rapid alternating movement analysis

RAM was assessed by classifying hand orientation throughout alternating palm and back-of-hand movements. For the RAM pipeline, MediaPipe Hand Landmarker [3] operated in video mode with one hand and minimum detection, hand-presence, and tracking confidence values of 0.60. Geometric features derived from the detected landmarks were provided to a logistic-regression classifier trained using labeled palm and back-of-hand calibration data. The classifier

generated frame-level probabilities representing hand orientation, which were subsequently smoothed using a three-frame moving window before state assignment.

Rather than directly counting changes in classifier output, RAM cycles were identified using a state-machine procedure. A valid cycle required the ordered progression PALM → TRANSITION → BACK → TRANSITION → PALM. Candidate states were required to persist for at least two consecutive frames before being accepted, and movements occurring in an invalid order reset partial cycle progression. This procedure was designed to reduce the influence of brief frame-level classification fluctuations on the final movement count.

Controlled video-degradation experiment

Robustness of the RAM pipeline was evaluated by applying controlled transformations to recorded video while maintaining the same downstream analysis procedure. Four brightness conditions were evaluated at 50%, 75%, 125%, and 150% of the reference intensity. Gaussian blur was separately evaluated using kernel sizes of 3, 7, 11, and 21. A re-encoded version of the original recording was included as a control to determine whether video re-encoding alone altered the calculated movement output.

Each transformed recording was processed through the same RAM pipeline. For each condition, the analysis recorded the number of processed frames, successfully tracked frames, frames containing valid hand geometry, and detected RAM cycles. Tracking availability was calculated as the percentage of processed frames in which the hand was successfully tracked. Absolute cycle-count error was calculated relative to the corresponding undegraded reference recording.

Multi-recording robustness evaluation

To determine whether observations from a single reference video generalized across recordings, the degradation analysis was extended to four independently recorded RAM trials. Each recording served as its own reference, and transformed versions were compared with the cycle count obtained from the corresponding undegraded recording. Cycle-count preservation was defined as equality between the detected cycle count of a transformed recording and that of its reference.

For each degradation condition, agreement was summarized as the proportion of transformed recordings in which the original RAM cycle count was preserved. This analysis complemented frame-level tracking measurements by directly evaluating whether the final task-specific movement output remained stable under controlled changes in video quality.

V. CONCLUSION

VisionDexterity demonstrates the technical feasibility of using standard webcam video and hand-landmark analysis to quantify multiple aspects of hand motor performance. The framework successfully extracted finger-tapping events, finger-sequencing accuracy, and RAM cycles, while robustness testing showed that cycle counts were preserved across most tested video-degradation conditions. These findings support further investigation of accessible, camera-based approaches to quantitative hand motor assessment. Future work should validate the framework across larger and more diverse participant and recording sets and compare its measurements with established clinical or instrumented assessments.

REFERENCES

- [1] K. E. Yancosek and D. Howell, "A narrative review of dexterity assessments," *Journal of Hand Therapy*, vol. 22, no. 3, pp. 258–270, 2009, doi: 10.1016/j.jht.2008.11.004.
- [2] J. Yong, J. C. MacDermid, T. Packham, P. Bobos, J. Richardson, and S. Moll, "Performance-based outcome measures of dexterity and hand function in person with hands and wrist injuries: A scoping review of measured constructs," *Journal of Hand Therapy*, vol. 35, no. 2, pp. 200–214, 2022, doi: 10.1016/j.jht.2021.04.017.
- [3] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "MediaPipe Hands: On-device real-time hand tracking," in *Proc. CV4ARVR*, 2020.
- [4] L. Reissner, G. Fischer, R. List, P. Giovanoli, and M. Calcagni, "Assessment of hand function during activities of daily living using motion tracking cameras: A systematic review," *Proceedings of the Institution of*

- Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 233, no. 8, pp. 764–783, 2019, doi: 10.1177/0954411919851302.
- [5] S. Scataglini, E. Abts, C. Van Bocxlaer, M. Van den Bussche, S. Meletani, and S. Truijen, “Accuracy, validity, and reliability of markerless camera-based 3D motion capture systems versus marker-based 3D motion capture systems in gait analysis: A systematic review and meta-analysis,” *Sensors*, vol. 24, no. 11, Art. no. 3686, 2024, doi: 10.3390/s24113686.
- [6] F. Varcin and M. G. Boocock, “The accuracy, validity and reliability of Theia3D markerless motion capture for studying the biomechanics of human movement: A systematic review,” *Artificial Intelligence in Medicine*, vol. 173, Art. no. 103332, 2026, doi: 10.1016/j.artmed.2025.103332.
- [7] Thomas, C. B. Tucker, D. E. Lunn, M. J. Cooke, G. Nicholson, and J. Walker, “Lights, cameras, action—The influence of lighting and camera position on walking and running kinematic measurements using Theia3D markerless motion capture,” *Journal of Biomechanics*, vol. 203, Art. no. 113343, 2026, doi: 10.1016/j.jbiomech.2026.113343.
- [8] M. W. Lenhoff, S. Zanini, G. Chingas, S. Patel, R. DiGiacomo, and H. J. Hillstrom, “Intra and inter-system reliability of a markerless motion capture system in different configurations during functional tasks: Implications for utility,” *Medical Engineering & Physics*, vol. 142, Art. no. 104375, 2025, doi: 10.1016/j.medengphy.2025.104375.
- [9] G. Bradski, “The OpenCV library,” *Dr. Dobb’s Journal of Software Tools*, vol. 25, no. 11, pp. 120–125, 2000.
- [10] J. G. V. Habets, R. K. Spooner, V. Mathiopoulou, L. K. Feldmann, J. L. Busch, J. Roediger, B. H. Bahnners, A. Schnitzler, E. Florin, and A. A. Kühn, “A first methodological development and validation of ReTap: An open-source UPDRS finger tapping assessment tool based on accelerometer-data,” *Sensors*, vol. 23, no. 11, Art. no. 5238, 2023, doi: 10.3390/s23115238.
- [11] Karni, G. Meyer, C. Rey-Hipolito, P. Jezzard, M. M. Adams, R. Turner, and L. G. Ungerleider, “The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 3, pp. 861–868, 1998, doi: 10.1073/pnas.95.3.861.