Contactless hand tremor amplitude measurement using smartphones: development and pilot evaluation

James Bungay[1], Osasenaga Emokpae[1], Samuel D. Relton[2], Jane Alty[3], Stefan Williams[4], Hui Fang[5] and David C. Wong[2]

Abstract— Background - Physiological tremor is defined as an involuntary and rhythmic shaking. Tremor of the hand is a key symptom of multiple neurological diseases, and its frequency and amplitude differs according to both disease type and disease progression. In routine clinical practice, tremor frequency and amplitude are assessed by expert rating using a 0 to 4 integer scale. Such ratings are subjective and have poor inter-rater reliability. There is thus a clinical need for a practical and accurate method for objectively assessing hand tremor.

Objective - to develop a proof-of-principle method to measure hand tremor amplitude from smartphone videos.

Methods - We created a computer vision pipeline that automatically extracts salient points on the hand and produces a 1-D time series of movement due to tremor, in pixels. Using the smartphones' depth measurement, we convert this measure into real distance units. We assessed the accuracy of the method using 60 videos of simulated tremor of different amplitudes from two healthy adults. Videos were taken at distances of 50, 75 and 100 cm between hand and camera. The participants had skin tone II and VI on the Fitzpatrick scale. We compared our method to a gold-standard measurement from a slide rule. Bland-Altman methods agreement analysis indicated a bias of 0.04 cm and 95% limits of agreement from -1.27 to 1.20 cm. Furthermore, we qualitatively observed that the method was robust to limited occlusion.

Clinical relevance - We have demonstrated how tremor amplitude can be measured from smartphone videos. In conjunction with tremor frequency, this approach could be used to help diagnose and monitor neurological diseases.

I. Introduction

Hand tremor is a common symptom of multiple diseases, including Parkinson's disease, essential tremor, and multiple sclerosis. Assessment of tremor activity is an important clinical task that can help in diagnosis of disease and evaluating response to treatment.

Tremor is assessed clinically by considering its frequency and amplitude. The standard clinical methods of measuring both tremor amplitude and frequency are subjective. A clinician visually observes a patient tremor and makes an estimate of both measures, categorising it with a severity rating [1], [2]. Such visual estimates of movement disorders are usually performed in face-to-face consultations, and there is large inter-rater variability between expert clinicians such that tremor diagnoses are frequently incorrect [3], [4]

Objective measurement of tremor frequency is possible using an accelerometer strapped to the hand [5]. Tremor amplitude is rarely derived directly, and instead, the amplitude of the acceleration signal is taken as a proxy for the amplitude of the displacement. The practical use of accelerometers is limited in two respects; it requires non-standard specialist equipment, and also adds weight to the hand in a way that may alter tremor characteristics.

Instead, it may be possible to derive tremor frequency and amplitude measurements directly from smartphone videos. Analysis of smartphone videos has been used to assess other biomarkers of neurological conditions [6], [7]. The ubiquity of smartphones means that such computer vision approaches have the potential to be used in multiple contexts. For example, they could be used for remote consultations or for monitoring of disease progression. Recently, we have developed a method for extracting tremor *frequency* directly from smartphone video recordings of hand tremor [8].

Here, we propose and demonstrate proof-of-principle for a method that enables measurement of tremor *amplitude*, with respect to the image plane, from smartphone videos.

II. Method

A. Technical Description

The method consists of four main parts, shown in figure 1. We assume that data have been collected via a modern smartphone that can capture both video and a depth measurement at the centre of the camera frame. In our work, we used an iPhone XR to provide both measurements.

1. Extract hand features - Using the video data, we extracted salient points on the hand using the Mediapipe hand tracker [9]. This two-stage process identifies the palm region using a U-net and then fits salient points that are consistent with pre-specified hand pose model. The hand tracker returns a tuple of $\{x, y, t\}$ corresponding to the x and y pixels, and time. For robustness, we monitored movement over multiple points, corresponding to the base (metacarpal), middle (interphalangeal) and tip of thumb and forefinger.

2. Calculate tremor amplitude in pixels - In our controlled scenarios, the primary direction of tremor was horizontal. We therefore used only $\{x, t\}$ to represent the motion waveform over time. To filter out low frequency motion due to gross arm movements, we processed the waveform by

[1]J Bungay and O Emokpae are with the Department of Computer Science, University of Manchester, UK,

[2]DC Wong and SD Relton are with the Leeds Institute of Health Sciences, University of Leeds, UK, `corresponding author: d.c.wong@leeds.ac.uk, s.d.relton@leeds.ac.uk`

[3]J Alty is with the University of Tasmania, Australia, University of Leeds, UK, `jane.alty@utas.edu.au`

[3]S Williams is with Leeds Teaching Hospitals Trust, UK, `stefan.williams2@nhs.net`

[3]H Fang is with the Department of Computer Science, Loughborough University, UK, `h.fang@lboro.ac.uk`
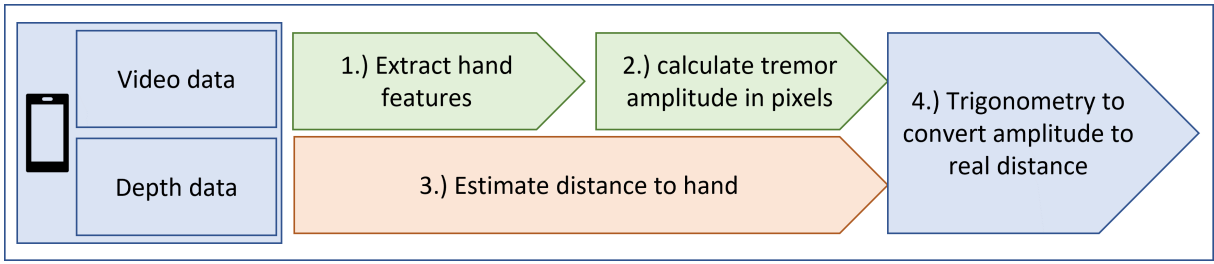
Fig. 1: Overview of hand-tremor amplitude measurement using smartphone video analysis

first extracting peaks and troughs. We used a simple forward-difference to estimate the gradient and located zero crossings. The difference in x between adjacent zero crossings was an instantaneous estimate of the amplitude - this was calculated for all adjacent pairs of zero crossings. From this set, we used the median value to be robust against artificial increase in tremor due to 'ramp-up' of the tremor motion from an at-rest state.

Camera Property	Symbol	Value
Physical Lens Focal Length	f	2.87 mm
35mm Equiv. Lens Focal Length	f_e	32 mm
Sensor Aspect Ratio	a_s	0.75 (i.e. 3:4)
Video Aspect Ratio	a_v	0.5625 (i.e. 9:16)
Horizontal Video Resolution	r_h	1080 px

TABLE I: Specifications of the iPhone XR front facing camera.

3. Convert amplitude from pixels to distance units: Finally, we converted the pixel distance into true distance units. The conversion relies on the distance between the smartphone and the hand. We measured this distance using the Apple TrueDepth sensor, using the front-facing camera on an iPhone XR. In controlled experiments, we assessed the accuracy of the depth sensor by reading the TrueDepth sensor distance to an object at known depths. In these experiments, the camera was affixed to a tripod in good lighting conditions. At a known distance of 40cm, the root mean squared error over six sensor measurements, taken over one minute, was 0.12 cm. At 100 cm, the error increased slightly to 0.38 cm.

The conversion also requires knowledge of physical characteristics of the camera described in Table I. For a video shot in portrait, the width of the physical sensor within the camera, v_w is given by:

$$v_w = \frac{f_e a_v}{f a_s}$$

From this, we can calculate the width of the view in the scene, w, at a given depth, d:

$$w = v_w \frac{f}{d}$$

The conversion between pixels to distance is:

$$dist = pix \frac{w}{r_h}$$

Code for extracting the depth measurement and for computing the amplitude is available at `https://github.com/jamesbungay/cv-tremor-amplitude`. The output of the entire process, for two example waveforms, is shown in Figure 2. In figure 2a, the dominant frequency of the tremor is visible, and the median peak-trough distance is 5.77 cm. In contrast, Figure 2b shows an example in which there is a gross change in x over the 12 second video recording, but there is no high frequency oscillation caused by tremor. In this case, the method correctly determines that there is no meaningful tremor (Median tremor amplitude = 0.09 cm)

B. Method Validation

We undertook a methods agreement analysis to assess the performance of the tremor amplitude algorithm. No participants were recruited; all data were self-recordings containing no personally identifiable information. Given this, the University of Manchester advised that local ethics was not required.

Data Collection: We recorded a set of videos of two members of the study team (JB, OE) Videos were recorded using an iPhone XR smartphone at 1080p resolution and at 60 fps. The smartphone was attached to a tripod, and we ensured that the videoed area was lit well using a ring light. A ruler was placed directly behind the hand. This allowed us retrospectively to measure the tremor amplitude from the video recording.

We simulated two common types of tremor. Resting tremor was simulated by the subject resting their forearm on a chair arm and rotating their wrist to create side-to-side motion as shown in Figure 3. Postural tremor was simulated by the subject raising an outstretched arm parallel to the floor, and with the thumb closest to the camera. The subject made oscillatory hand movements up and down. The camera was oriented so that the principal direction of tremor was horizontal, with respect to a portrait video frame.

For both resting and postural tremor, we simulated tremor amplitudes according to five categories (No tremor, small ($<$1cm), medium ($\approx$2cm), large ($\approx$5cm) and very large ($>$10cm), which correspond to the tremor categories used within both the Unified Parkinson's Disease Rating Scale and Essential Tremor Rating Assessment Scale [1], [2]. Videos were recorded at three depths 50 cm, 75 cm, 100 cm. The two team members had skin tones of II and VI on the Fitzpatrick scale. In total $2 \times 2 \times 5 \times 3 = 60$ videos were recorded.

Waveform of a Postural Tremor with a Measured Median Amplitude of 5.77cm, Recorded at a Depth of 74.4cm

(a)

Waveform of a Postural Tremor with a Measured Median Amplitude of 0.09cm, Recorded at a Depth of 77.7cm
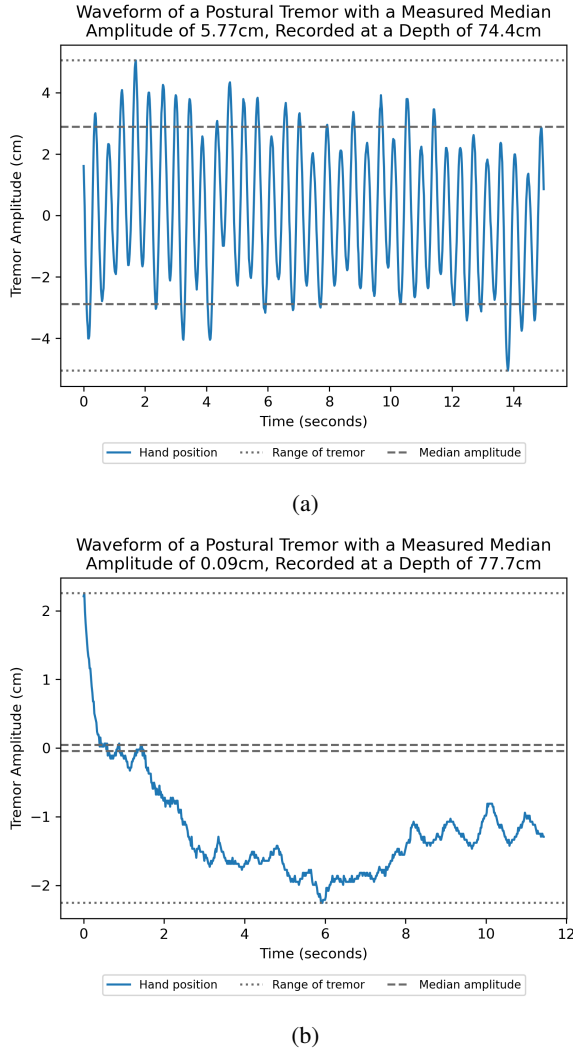
(b)

Fig. 2: (a) shows a 'typical' tremor signal with a consistent dominant frequency. The median peak-trough distance is 5.77 cm. (b) shows gross translational movement over time that does not correspond to tremor, which is correctly ignored when amplitude is calculated.

We recorded an additional set of 8 videos to assess how well hand detection worked under various types of occlusion. For both subjects, we recorded videos of each wearing a ring, a plaster on the dorsum of the hand that simulated having an accelerometer strapped to the hand, and a hand with fewer than five fingers showing.

Data Analysis: We compared the gold-standard ruler-measured amplitude with our computer vision approach using Bland-Altman agreement analysis [10]. The output of this is bias, which is the mean difference between the computer vision and gold standard and 95% limits of agreement (LoA), which may be regarded as the maximum difference between the two methods for 95% of future measurements. In subgroup analysis, we assessed whether there were differences in the bias and LoA for skin tones II and VI using a t-test.

The additional videos with occlusion were analysed qual-

Fig. 3: A research participant in position to measure resting tremor.

itatively in two stages. First, we assessed whether the hand tracker could correctly and reliably identify the salient points. Second, we compared the calculated tremor amplitude against the ruler-based amplitude measurement. Bland-Altman plots were not plotted in these cases, as the small number of videos would render them meaningless.

III. RESULTS

Bland-Altman analysis showed a mean difference of -0.04cm, with 95% LoA of -1.27cm to 1.20cm; the associated Bland-Altman plot is shown in Figure 4. There was no meaningful or statistically significant difference in bias and LoA when the cohort was split according to Fitzpatrick score.

The high limits of agreement can be explained by the manual measurements having a low precision of ± 1 cm. This low precision resulted from two factors. First, the resolution of some of the videos was insufficient to be able to discern the millimetre markings on the ruler. Second, parts of the ruler were occasionally obscured from view by the hand. Measurements thus had to be taken by extrapolating neighbouring ruler markings. Mean difference is not affected by the low precision of manual measurements.

DISCUSSION

This pilot work shows, for the first time, a smartphone video method for estimating amplitude of hand tremor. In our controlled tests on simulated tremor from healthy individuals, our method showed minimal bias and 95% limits of agreement -1.27cm to 1.20cm over a wide range of tremor amplitudes, for two different skin tones. In addition, *ad hoc* tests qualitatively showed this approach to be robust under a range of simulated real world conditions.

Previous research for measuring hand tremor amplitude has used bespoke sensors. The acceleration signal amplitude recorded by wrist-worn accelerometers has been used as a proxy for the true tremor size [11]. In principle, the
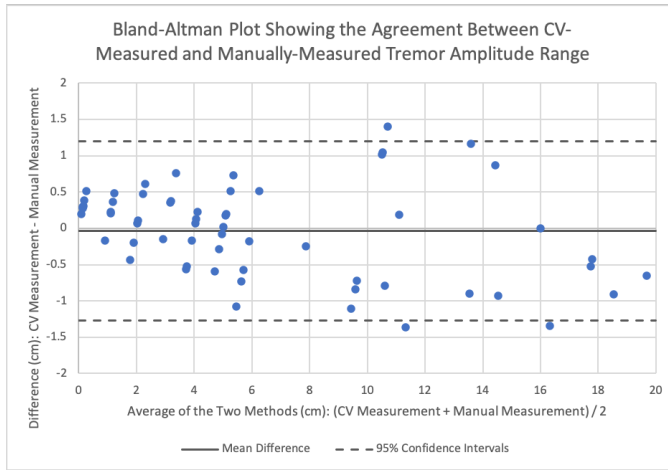
Fig. 4: Bland-Altman plot of CV-measured and manually-measured amplitude range.

acceleration signal can be integrated to provide true distance, but the resulting signal is likely to be noisy. Electromagnetic position sensors have been used, and claim a fidelity of 0.45 mm [12]. Our approach differs by using common sensor modalities that are readily available on most modern smartphones.

While our results are slightly poorer than some existing methods, we believe that these are mainly due to limitations with our experimental setup. Our method contains three potential sources of error. First, from rounding error due to discretization of distance in pixels. For a modern smartphone camera with high resolution, we can assume this to be negligible. Second, from errors in depth measurement. In local tests, we showed an average depth error of 0.38 cm at true distance 100 cm. Using trigonometry, we calculate that this would correspond to a possible error of up to $0.38/100 = 0.38\%$. These sources of error are limitations of the camera technology and their sum can be considered a maximum lower bound on error. In addition, a third error, caused by rounding rounding the visual gold standard to the nearest centimeter, leads to an error of ± 0.5 cm.

While this pilot work provides proof of principle, it is limited in a few key respects. First, the tremor amplitude is only calculated in the plane of the camera image. While this is sufficient for some clinical scenarios, we know that tremor can be very heterogeneous, depending on clinical condition. For instance, tremor associated with Parkinson's disease is commonly described as a 'pill-rolling' tremor, which is characterised by rotation of the wrist. Second, the ruler used to provide a reference amplitude measurement was often occluded by the hand. This meant that we were unable to make accurate reference measurements, instead rounding to the nearest centimeter. This in turn led to unreliable estimates of the true agreement with the video-based method. Third, data collection was undertaken using an iPhone only. We note that most modern smartphones contain at least one method for measuring depth data, and that our approach should therefore be generalisable to other devices. An alternative,

where there is no direct method to measure depth, would be to estimate depth via depth-from-motion methods [13].

To address these issues, we are currently conducting a larger validation study using data from patients with multiple types and acuity of tremor. In this study, we will also investigate whether full 3D depth map videos can improve estimation of tremor amplitude.

CONCLUSION

We have demonstrated a smartphone video approach for measuring tremor amplitude. In conjunction with a method for measuring frequency (see [8]), this method could objectively and contactlessly measure the key clinical components of tremor in near real-time. The method has potential uses for diagnosis, and remote monitoring of disease progression or drug response.

REFERENCES

[1] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel *et al.*, "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008.

[2] R. J. Elble, "The essential tremor rating assessment scale," *Journal of Neurology & Neuromedicine*, vol. 1, no. 4, 2016.

[3] S. Jain, S. E. Lo, and E. D. Louis, "Common misdiagnosis of a common neurological disorder: how are we misdiagnosing essential tremor?" *Archives of neurology*, vol. 63, no. 8, pp. 1100–1104, 2006.

[4] N. P. Bajaj, V. Gontu, J. Birchall, J. Patterson, D. G. Grosset, and A. J. Lees, "Accuracy of clinical diagnosis in tremulous parkinsonian patients: a blinded video study," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 81, no. 11, pp. 1223–1228, 2010.

[5] F. Vial, P. Kassavetis, S. Merchant, D. Haubenberger, and M. Hallett, "How to do an electrophysiological study of tremor," *Clinical Neurophysiology Practice*, vol. 4, pp. 134–142, 2019.

[6] S. Williams, S. D. Relton, H. Fang, J. Alty, R. Qahwaji, C. D. Graham, and D. C. Wong, "Supervised classification of bradykinesia in parkinson's disease from smartphone videos," *Artificial Intelligence in Medicine*, vol. 110, p. 101966, 2020.

[7] Z. Zhao, H. Fang, S. Williams, S. D. Relton, J. Alty, A. J. Casson, and D. C. Wong, "Time series clustering to examine presence of decrement in parkinson's finger-tapping bradykinesia," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 780–783.

[8] S. Williams, H. Fang, S. D. Relton, D. C. Wong, T. Alam, and J. E. Alty, "Accuracy of smartphone video for contactless measurement of hand tremor frequency," *Movement Disorders Clinical Practice*, vol. 8, no. 1, pp. 69–75, 2021.

[9] 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 preprint arXiv:2006.10214*, 2020.

[10] J. M. Bland and D. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The lancet*, vol. 327, no. 8476, pp. 307–310, 1986.

[11] L. Aljihmani, Y. Zhuz, H. T. Abbas, R. Mehta, F. Sasangohar, M. Erraguntla, Q. H. Abbasi, and K. Qaraqe, "Spectral analysis of hand tremors induced during a fatigue test," in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2019, pp. 658–663.

[12] T. Perera, W.-L. Lee, S. A. Yohanandan, A.-L. Nguyen, B. Cruse, F. M. Boonstra, G. Noffs, A. P. Vogel, S. C. Kolbe, H. Butzkueven *et al.*, "Validation of a precision tremor measurement system for multiple sclerosis," *Journal of Neuroscience Methods*, vol. 311, pp. 377–384, 2019.

[13] J. Valentin, A. Kowdle, J. T. Barron, N. Wadhwa, M. Dzitsiuk, M. Schoenberg, V. Verma, A. Csaszar, E. Turner, I. Dryanovski *et al.*, "Depth from motion for smartphone ar," *ACM Transactions on Graphics (ToG)*, vol. 37, no. 6, pp. 1–19, 2018.