Hand-tremor frequency estimation in videos

Silvia L. Pintea[1], Jian Zheng[1], Xilin Li[1], Paulina J.M. Bank[2],
Jacobus J. van Hilten[2], Jan C. van Gemert[1]

[1]Vision Lab, Delft University of Technology, Netherlands
[2]Department of Neurology, Leiden University Medical Center, Netherlands

Abstract. We focus on the problem of estimating human hand-tremor
frequency from input RGB video data. Estimating tremors from video
is important for non-invasive monitoring, analyzing and diagnosing pa-
tients suffering from motor-disorders such as Parkinson's disease. We
consider two approaches for hand-tremor frequency estimation: (a) a La-
grangian approach where we detect the hand at every frame in the video,
and estimate the tremor frequency along the trajectory; and (b) an Eule-
rian approach where we first localize the hand, we subsequently remove
the large motion along the movement trajectory of the hand, and we
use the video information over time encoded as intensity values or phase
information to estimate the tremor frequency. We estimate hand tremors
on a new human tremor dataset, *TIM-Tremor*, containing static tasks
as well as a multitude of more dynamic tasks, involving larger motion of
the hands. The dataset has 55 tremor patient recordings together with:
associated ground truth accelerometer data from the most affected hand,
RGB video data, and aligned depth data.

Keywords: Video hand-tremor analysis, phase-based tremor frequency detec-
tion, human tremor dataset, Eulerian hand tremors.

1 Introduction

We focus on human hand-tremor frequency estimation from videos captured with
common consumer RGB cameras. The problem has a considerable importance
in medical applications for aiding the medical personnel in the task of motor-
disorder patient monitoring and tremor diagnosing [2,13,25,34]. Traditionally the
clinical practice uses body-worn accelerometers which offer excellent measure-
ments, yet is intrusive, slow to setup, and allows only measuring a single location
per accelerometer. Replacing accelerometers with a common RGB camera brings
forth a non-intrusive method of measuring full-body tremors, offering a strong
advantage in the clinical practice.

In the context of tremor analysis, existing approaches require the use of
specialized sensors [6,8,14,15], which makes it difficult to apply these methods
in practice. Moreover, the targeted application of these approaches are the more
high-level tremor diagnosing problem [14,26] or tremor/no-tremor classification

[28]. We propose to estimate human hand-tremor frequency from RGB videos, and compare against ground truth accelerometer data.

The main challenge, when performing human tremor frequency estimation, is the current lack of openly available realistic datasets. Existing work on human tremor analysis either evaluates using in-house data that is not publicly available [6,14,37], or on simulated tremor data where no ground truth tremor statistics are provided [28]. This limits the assessment of human tremor analysis methods and, thus, its progress. An open evaluation dataset is needed.

In this work: (i) we evaluate the frequency of human hand-tremors from RGB videos and we analyze two possible approaches: (i.a) a Lagrangian approach that focuses on the motion of the hand in the image plane, and estimates tremors over the hand positions; (i.b) an Eulerian approach that aligns the hand position over a temporal window, by tracking it, and subsequently uses the image information over time as extracted from intensity values and phase-images, to perform a windowed Fourier analysis at every hand pixel; (ii) we bring forth the *TIM-Tremor* dataset, containing: 55 RGB patient videos, together with associated ground-truth accelerometer recordings on the most affected hand, as well as aligned depth-data; (iii) we analyze two variants of the Lagrangian approach and two variants of the Eulerian approach and evaluate them numerically on our proposed *TIM-Tremor* human tremor dataset.

2   Related work

2.1   Motion analysis

Periodic motion. The work in [22] performs action recognition by using space-time repetitive motion templates. Similar to using templates, in [5] a self-similarity relying on time-frequency analysis is used for action recognition. The work in [12] performs a spectral decomposition of moving objects to encode periodic motions for object recognition, while [30] performs eigen decomposition and describes periodic motion by the circularity or toroidality of an associated geometric space. Following a similar trend, in [19] complex motion is decomposed into a sequence of simple linear dynamic models for motion categorization. The work in [24] focuses on pedestrian detection through periodic movement analysis. Similar to us, the work in [28] performs tremor analysis, however in [28] videos are classified into tremor /no-tremor using optical flow features and SVM. In the recent work in [29] a CNN is used for discriminating between Parkinson patients and non-Parkinson patients, using wrist-worn senors. In this work we also focus on periodic motion analysis, however our end goal is tremor frequency estimation rather than action recognition, object tracking or recognition.

Differently, in [18,27,33] deep network architectures are trained for counting action repetitions. These actions must be clearly visible and recognizable in the camera view for the deep network architectures to work, while we focus on tremors which are subtle motions.

The most similar work to our work is the work performed in [31,32] where tremor frequency is measure from pixel intensities in the video. However these

methods assumes the location of the body part at which the tremor is measured to be known in advance and moreover, the frequency is estimated over intensity values rather than detected hand location over time, or image phase-information over time, as we propose here. The authors do not publicly provide either code or data, which makes it impossible for us to compare with their approach.

Subtle motion. Small motion, difficult to see with the bare eyes, can be magnified [17,35] through a complex steerable pyramid. In the more realistic case, when the subtle motion is combined with a large motion, follow up work can magnify subtle motions such as tremors in the presence of large object motion such as walking [9,39]. Video frequency analysis has been also employed for estimating the properties of physical materials [7]. We also employ signal analysis in the Fourier domain, however rather than magnifying the subtle motion or estimating material properties, we estimate the frequency of the subtle tremor motion. The works in [14,15] use specialized sensors or a digital light-processing projector, and a high frame-rate camera to detect small vibrations. Unlike [14,15], we do not employ specific cameras or expensive sensors, we estimate the tremor frequency from common RGB videos.

2.2   Human body pose estimation

Works such as [3,21] perform body pose estimation over multiple people, in deep networks. In [3,4,23,36] cascaded prediction or iterative optimizations are used for body pose estimation. We use the method in [36] for estimating where to measure the tremors. We opt for [36] due to its ease of usage and robustness. In this work we use the MPII Human Body Pose dataset [1] for training the human body pose estimation models.
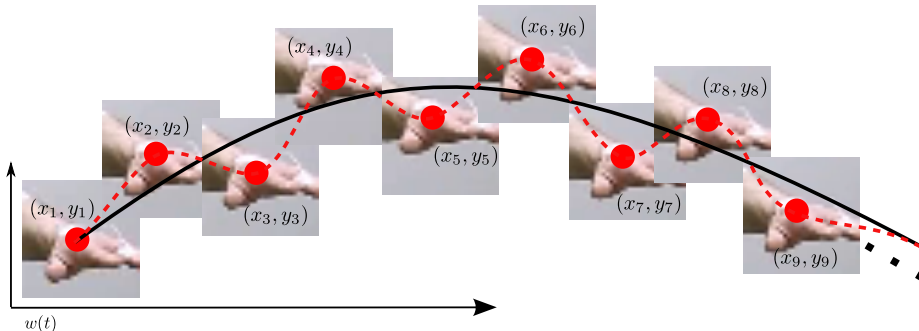
3   Hand-tremor frequency estimation

We start by localizing the affected hand. Subsequently, we consider two methods for hand-tremor frequency estimation: (a) Lagrangian hand-tremor frequency estimation, and (b) Eulerian hand-tremor frequency estimation.

3.1   Hand location estimation

A first step in estimating human hand-tremors, is localizing the affected hand. For this, we use the robust human body pose estimation proposed in [36]. This method provides us a hand location per frame (x_i, y_i). We perform the tremor analysis on shorter temporal windows of the video, $w(t)$.

3.2   (a) Lagrangian hand-tremor frequency estimation

Figure 1 depicts the idea behind the Lagrangian hand-tremor frequency estimation. We start by detecting the hand locations $(x_i, y_i)_{i \in w(t)}$ over the temporal
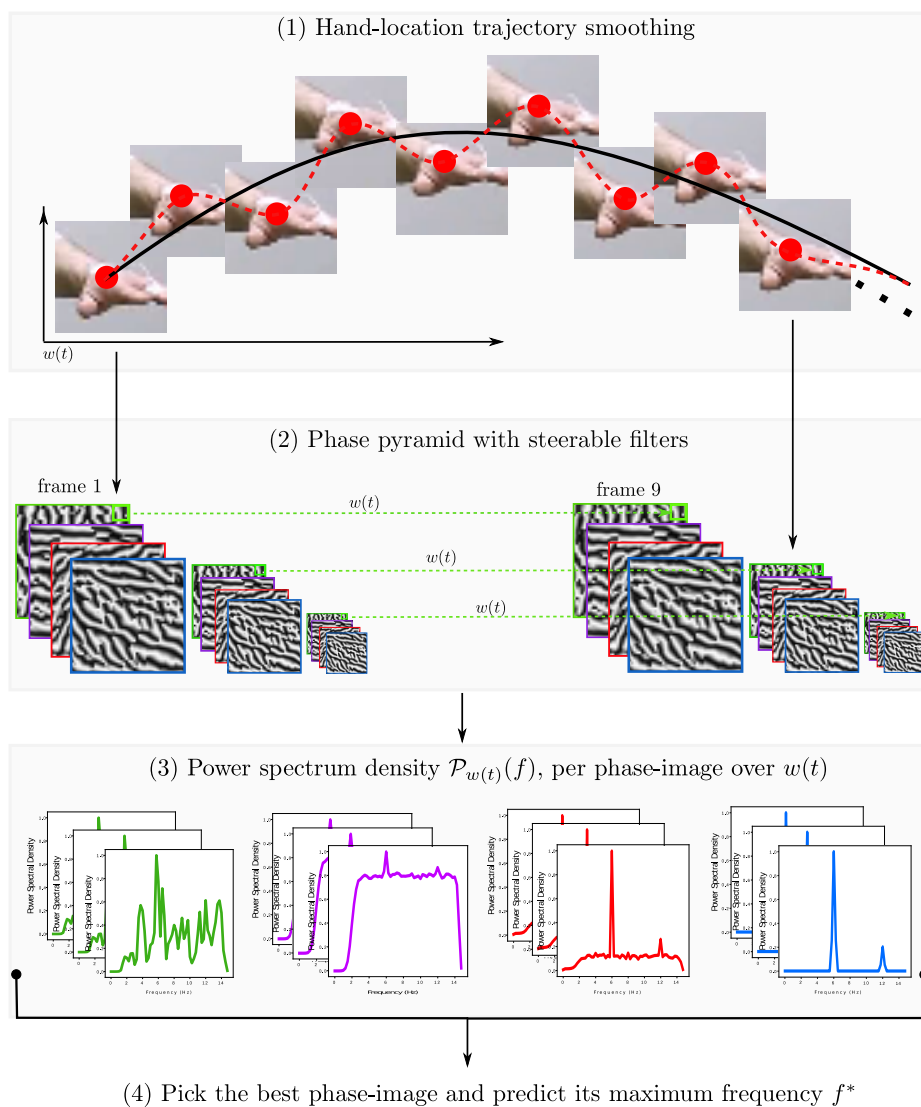
Fig. 1. Lagrangian hand-tremor estimation is based on frequency estimation of (x, y) coordinates. We detect the hand position $(x_i, y_i)_{i \in w(t)}$ at every frame i over a temporal window $w(t)$. The hand motion is characterized by a large motion, depicted by the black line, and a small motion, depicted by the red dotted line. We smooth this information over time, using a Kalman tracker to obtain the smooth coordinates of the hand.

window $w(t)$. The hand motion is typically characterized by a combination of two motions: a large hand trajectory motion, depicted through the continuous black line, and a small motion corresponding to the tremor, depicted in Figure 1 by the dotted red line. We first apply a Kalman-filter tracker [38] to the initial hand locations, detected by the pose estimation algorithm [36]. This step is used for smoothing the hand trajectory, to obtain the large hand motion. We subsequently subtract this smooth trajectory from the original hand locations to retain only the *x and y locations* of the small hand motion, corresponding to the tremor. Thereafter, we apply the windowed Fourier transform over these corrected locations. This provides us a PSD (Power Spectrum Density) function. We use the maximum frequency as the estimated hand-tremor frequency.

3.3   (b) Eulerian hand-tremor frequency estimation

Figure 2 illustrates the Eulerian frequency estimation. The first step is the same as in Figure 1, where the hand locations are detected using the pose estimation method in [36], and subsequently, we smooth the trajectory given by these hand detections using a Kalman tracker. This gives us the smooth trajectory of the hand over time, in the video. We crop image windows around the temporally smoothed locations of the hand in the video — along the black line depicted in Figure 2.(1). For each such image crop, we extract local motion information encoded as phase over different scales and orientations. Thereafter, we compute the frequency of the hand-tremor by using the most informative phase-image. Figure 2 depicts these individual steps.

Phase-image computation. Works such as [10,20,35,39] claim that the phase responses over time contain descriptive information regarding the motion present in the image. In [35] the use of complex steerable filters [11] is proposed for

Fig. 2. Eulerian hand-tremor estimation is based on frequency estimation in images. (1) The fist step is the same as in the Lagrangian illustrated in Figure 1: detecting a Kalman-filtered smoothed hand position at every frame over a temporal window $w(t)$. (2) We crop image windows around the smoothed hand locations. Each such cropped image is transformed into a phase-pyramid with 4 orientations and 3 scales using a steerbale filter bank. (3) For every pixel, in every phase-image over the temporal window $w(t)$ we estimate a PSD (Power Spectrum Density). We accumulate these over the pixels in one phase-image, to obtain one PSD per phase-image. (4) We select the most informative phase-image PSD and use it to estimate the tremor-frequency.

extracting local motion information. Given an input image $I(x, y)$ and a set of complex steerable filters of the form: $G_\sigma^\theta + iH_\sigma^\theta$, where $i = \sqrt{-1}$, σ defines the scale of the filter, and θ the orientation, we obtain a complex steerable pyramid by convolving the image with this set of filters

$$(G_\sigma^\theta + iH_\sigma^\theta) \circledast I(x, y) = A_\sigma^\theta(x, y)e^{i\phi_\sigma^\theta(x, y)}, \tag{1}$$

where $\circledast$ denotes the convolution operations, and $A_\sigma^\theta(x, y)$ is the resulting amplitude for scale σ and orientation θ, and $\phi_\sigma^\theta(x, y)$ is the corresponding phase information. To obtain a phase-image, we set the amplitude to 1 and apply the inverse transformation [11] to reconstruct back the image. Examples of phase-images are depicted in Figure 1.(2). We use 4 orientations: $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$ and 3 scales: $\sigma \in \{1.0, 0.5, 0.25\}$, giving rise to 12 phase-images. In addition to the 12 phase-images, we add the grayscale version of the cropped hand-image. Therefore, we have in total 13 images, which we merge into a single image with 13 channels, over which we estimate the hand-tremor frequency.

Hand-tremor frequency estimation. We filter each one of the 13 input channels over time with a 4th-order Butterworth band-pass filter. This eliminates noisy frequencies that cannot correspond to a natural human tremor.

  To reduce the effect of the considered temporal window, $w(t)$, we use an adjustable Tukey window with the parameter α set to $\frac{f_s}{N-1}$, where f_s is the sampling rate and N is the total number of frames in $w(t)$. This ensures that the video signal over time is processed in a consistent manner while allowing for adjustable temporal window sizes, $w(t)$.

  Within each temporal window, $w(t)$, we estimate a PSD function, over every input channel, at every pixel location. For an input channel, c, we estimate the final PSD, $\mathcal{P}_{w(t)}^c(f)$, by averaging spatially the PSDs over the pixels in that channel. We repeat this process for all 13 channels, giving rise to 13 PSD functions.

  In [5] the power spectrum is considered to be periodic at a certain frequency, f, if the PSD response at that frequency is at least a few standard deviations away from the mean PSD response. This is indicative of how noisy is the PSD function. We use this same criterion to pick the most informative image channel; this is the channel over which we estimate the final hand-tremor frequency. We define for each channel a score, $\mathcal{S}^c(f)$:

$$\mathcal{S}^c(f) = \frac{1}{\mid w(t) \mid} \sum_{w(t)} \left(\mathcal{P}_{w(t)}^c(f) - \mu_{\mathcal{P}_{w(t)}^c} - k\sigma_{\mathcal{P}_{w(t)}^c} \right), \tag{2}$$

where $\mid w(t) \mid$ is the number of temporal windows per video, $\mu_{\mathcal{P}_{w(t)}^c}$ represents the mean of the PSD response, and $\sigma_{\mathcal{P}_{w(t)}}$ denotes the standard deviation, while k is an adjustable parameter. We set $k = 3$ in our experiments.

  The final predicted frequency over the 13 channels becomes:

$$f^* = \arg\max_f (\max_c \mathcal{S}^c(f)). \tag{3}$$
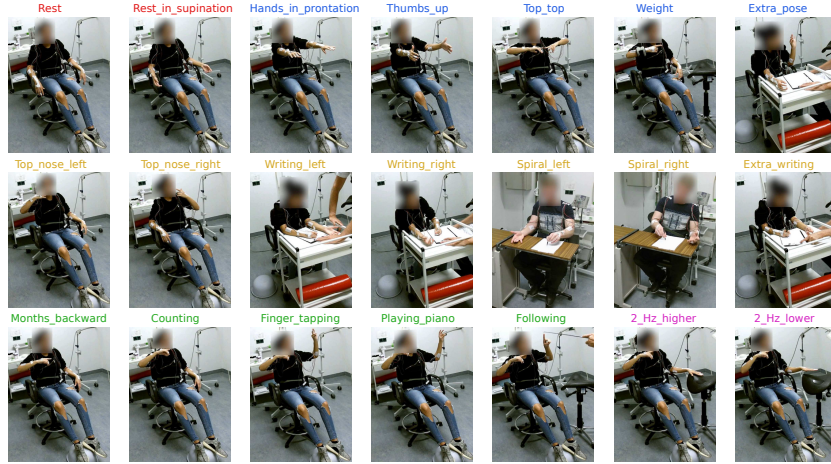
(a) Recorded tasks.

Task	Description
Rest	
Rest	Resting the arms on the chair handles.
Rest_in_supination	Resting the arms on the chair handles, hands in supination position.
Postures	
Hands_in_pronation	Both arms outstretched forward, hands in pronation position.
Thumbs_up	Both arms outstretched forward, thumbs up.
Top_top	Both hands in front of the chest with tips of the index fingers almost touching each other, elbows lifted sideways at approx. 90 degrees angle.
Weight	Affected arm outstretched forward, with a weight (1 kg) attached to the wrist.
Extra_pose	Holding a pose proposed by the medical expert to better visualize the tremor.
Actions	
Top_nose_left	Touching the top of the nose with the left index finger.
Top_nose_right	Touching the top of the nose with the right index finger.
Writing_left	Writing a given sentence with the left hand.
Writing_right	Writing a given sentence with the right hand.
Spiral_left	Drawing a spiral with the left hand.
Spiral_right	Drawing a spiral with the right hand.
Extra_writing	Extra writing task with a special pen, or diverging from the standard writing task with the affected hand.
Distraction	
Months_backward*	Naming the months backwards.
Counting*	Counting backwards (100 minus 7).
Finger_tapping*	Tapping with the index finger and thumb of the contralateral hand.
Playing_piano*	Moving the thumb of the contralateral hand across all fingers, from the index to the pinky finger and back.
Following*	Following a moving pointer with the index finger of the contralateral hand.
Entrainment	
2_Hz_higher*	Tapping with the contralateral hand in the rhythm of a flashing light, 2 Hz higher than the estimated tremor frequency of the affected hand.
2_Hz_lower*	Tapping with the contralateral hand in the rhythm of a flashing light, 2 Hz lower than the estimated tremor frequency of the affected hand.

*During these tasks, the affected hand was kept in the posture in which the tremor was most pronounced (i.e. arm on chair handle, arm outstretched with hand in pronation or thumbs up, or in front of the chest).

(b) Explanation.

Fig. 3. (a) We record motor-disorder patients in 21 tasks. Each task may elicit a tremor. (b) Short explanation of what each task involves.

Fig. 4. Examples from the recording setup together with the predicted body joint locations using [36]. We use this to obtain the location of the hand where we estimate the tremor frequency.

4   Experiments

We test the considered frequency estimation approaches on our tremor patient dataset, *TIM-Tremor*, containing a multitude of tasks. The anonymized *TIM-Tremor* patient data can be found at https://doi.org/10.4121/uuid:522d14ed-3019-4206-b49e-a4e674b6440a.

4.1   Patient data evaluation

Data description. We recorded the *TIM-Tremor* dataset, in which 55 patients are videotaped sitting in a chair and performing a multitude of tasks. The data is recorded with a Kinect^{TM} v2 device, and it consists of short RGB videos of resolution 1920×1080 px, and associated depth video recordings of 512×424 px using a 16-bit encoding, as well as depth videos aligned with the RGB videos following the method in [16]. To reduce the storage requirements, we rescale the video resolution to 960×540 px. The ground truth tremor frequency is measured on the wrist of the most affected hand: left/right. On this hand, during the recording, we position an accelerometer. The accelerometer recordings are included in the dataset. The hand on which the accelerometer is positioned, is annotated in the dataset for each patient. Thus, for each patient and each performed task, we provide a set of recorded videos of approximately 1 minute each, together with a corresponding aligned depth map video, and the ground truth accelerometer recording from the most affected hand.

Data collection occurred in parallel to the standard tremor clinical evaluation. The standard tremor evaluation consists of a set of 21 tasks, which are illustrated in Figure 3.(a) and described in Figure 3.(b). The tasks vary with respect to the adopted posture: e.g. arm supported by the arm rest, or held outstretched in front of the patient, the amount of motion involved: e.g. rest – no motion, or touching the top of the nose – intention-oriented motion, as well as the focus of attention:
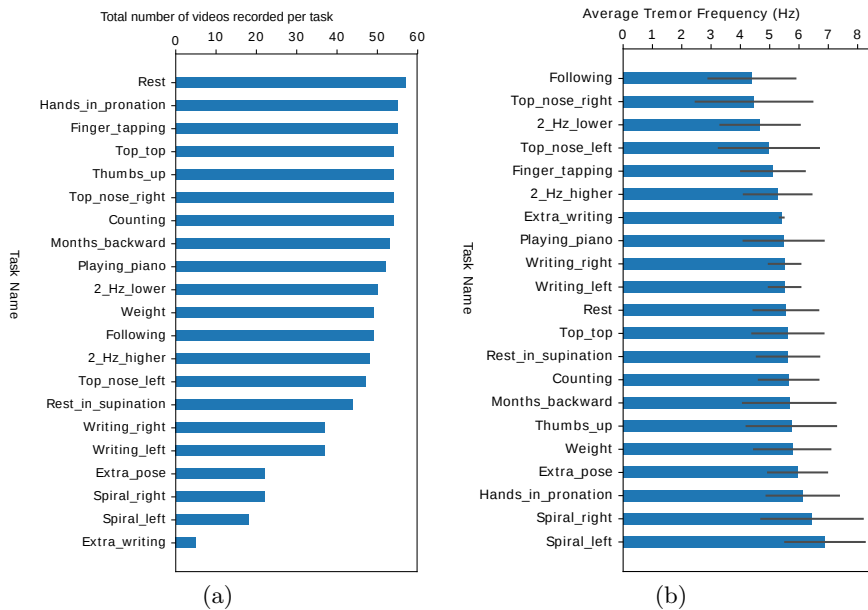
Fig. 5. (a) Total number of videos recorded per task. (b) Average frequency and standard deviation for all tasks across all 55 patient recordings.

e.g. distraction by mental task. Changes in tremor frequency between these tasks are analyzed by the medical expert to classify the tremor. For example, certain types of tremor are present across most or all tasks (e.g. "Parkinsonian tremor"), while other types of tremor may only occur when performing a specific task (e.g. "postural tremor" occurs only when a patient maintains a specific posture such as *Thumbs_up*), while other tremors may show considerable variation in tremor frequency between tasks (e.g. "functional tremor").

Figure 4 displays a few examples of the recording setup together with the estimated joint locations using [36]. In Figure 5 we show the total number of videos recorded for each task, and the average hand tremor frequency, as estimated by the accelerometer, together with the standard deviation, computed across all 55 patients. The average tremor frequency is around 5 Hz, which is a common in tremor affections such as Parkinson and Dystonia.

Experimental evaluation. We estimate the body pose in the videos using the method in [36], pretrained on the MPII dataset [1]. We apply the method a every frame. We use a temporal window, $w(t)$, of 60 frames for frequency estimation. Unless stated otherwise, we evaluate our method in terms of MAE (Mean Absolute Error) with respect to the ground truth frequency detected by the accelerometer. We only evaluate on video segments in which a periodic tremor has been detected, using the accelerometer data.

Task	Lag_no_smooth (Hz)	Lag_with_smooth (Hz)
2_Hz_higher	1.917 (± 2.395)	**1.879** (± 2.127)
2_Hz_lower	2.248 (± 2.770)	**1.721** (± 2.266)
Counting	1.731 (± 2.336)	**1.377** (± 2.246)
Extra_pose	3.590 (± 2.369)	**1.918** (± 1.328)
Extra_writing	**1.968** (± 0.000)	**1.968** (± 0.000)
Finger_tapping	1.989 (± 2.783)	**1.326** (± 1.974)
Following	1.607 (± 1.745)	**1.312** (± 1.728)
Hands_in_pronation	2.582 (± 2.154)	**2.398** (± 2.024)
Months_backward	2.544 (± 2.703)	**2.031** (± 2.500)
Playing_piano	2.443 (± 2.826)	**2.033** (± 2.516)
Rest	**3.300** (± 3.271)	3.395 (± 3.226)
Rest_in_supination	2.889 (± 3.228)	**2.059** (± 2.248)
Spiral_left	**6.721** (± 1.896)	**6.721** (± 1.896)
Spiral_right	3.246 (± 1.762)	**3.148** (± 1.803)
Top_nose_left	3.743 (± 3.262)	**3.688** (± 3.242)
Top_nose_right	1.928 (± 2.323)	**1.771** (± 2.204)
Top_top	**1.388** (± 1.797)	1.669 (± 1.888)
Thumbs_up	**1.694** (± 1.807)	1.694 (± 1.836)
Weight	**2.660** (± 2.667)	2.795 (± 2.569)
Writing_left	**2.557** (± 1.139)	**2.557** (± 1.139)
Writing_right	**2.557** (± 1.139)	**2.557** (± 1.139)
Average MAE	2.633 (± 2.208)	**2.382** (± 1.995)

Table 1. Exp 1.1: MAE when comparing the Lagrangian method with trajectory smoothing by using the Kalman tracker — *Lag_with_smooth*, versus not using trajectory smoothing, *Lag_no_smooth*. *Lag_with_smooth* performs slightly better than the default Lagrangian method, *Lag_no_smooth*. We highlight in bold the better performing method (lower is better).

4.2   Exp. 1: Design choices

In this experiment we test individual choices in the considered Lagrangian and Eulerian approaches. For the Lagrangian approach we test in **Exp 1.1** if removing the smooth trajectory, corresponding to the large motion of the hand, helps the frequency estimation. For the Eulerian approach, in **Exp 1.2** we test the added value of computing hand-tremor frequency over the phase information, rather than using only the intensity values of the image.

Exp 1.1: The need of trajectory smoothing. We experimentally compare two variants of the Lagrangian frequency estimation. The *Lag_no_smooth* variant uses raw hand trajectory points as computed by the pose estimation algorithm. The *Lag_with_smooth* variant removes the large motion of the hand obtained by subtracting the output of a Kalman tracker, which in effect retains only the small motions. The MAE numbers in Table 1 show that removing the large motion by using the Kalman tracker is beneficial to the overall performance. This is explained by the fact that subtracting the trajectory returned by the Kalman

Task	Euler_gray (Hz)	Euler_phase (Hz)
2_Hz_higher	0.882 (± 1.142)	**0.857** (± 1.533)
2_Hz_lower	1.335 (± 2.022)	**0.984** (± 1.333)
Counting	0.767 (± 1.252)	**0.472** (± 0.780)
Extra_pose	**1.180** (± 2.006)	1.623 (± 1.185)
Extra_writing	**0.984** (± 0.000)	**0.984** (± 0.000)
Finger_tapping	0.492 (± 0.893)	**0.385** (± 0.647)
Following	0.820 (± 1.327)	**0.557** (± 0.503)
Hands_in_pronation	1.271 (± 1.755)	**1.066** (± 1.506)
Months_backward	1.133 (± 1.848)	**1.219** (± 1.933)
Playing_piano	1.148 (± 1.832)	**1.148** (± 1.714)
Rest	1.459 (± 1.759)	**1.253** (± 1.770)
Rest_in_supination	**1.475** (± 1.919)	1.537 (± 1.728)
Spiral_left	3.278 (± 1.671)	**2.951** (± 1.391)
Spiral_right	3.246 (± 2.936)	**2.509** (± 2.021)
Top_nose_left	2.595 (± 2.216)	**1.776** (± 2.008)
Top_nose_right	3.108 (± 2.739)	**2.164** (± 2.015)
Top_top	0.860 (± 1.311)	**0.720** (± 1.200)
Thumbs_up	1.002 (± 1.419)	**1.002** (± 1.273)
Weight	1.207 (± 1.695)	**0.961** (± 1.226)
Writing_left	**0.394** (± 0.573)	0.492 (± 0.538)
Writing_right	**0.394** (± 0.573)	0.492 (± 0.538)
Average MAE	1.382 (± 1.566)	**1.198** (± 1.278)

Table 2. Exp 1.2: MAE showing the added value of the phase information. We compare the *Euler_gray* — Eulerian frequency estimation over grayscale hand-images, with *Euler_phase* – Eulerian frequency estimation over 12 phase-images and 1 grayscale image. Adding the 12 extra phase-images is beneficial for the frequency estimation. We highlight in bold the better performing method (lower is better).

tracker from the original hand trajectory works as a data detrending step. This allows for the frequency to be estimated only over the small tremor motion.

Exp 1.2: The added value of using phase-images. For both considered Eulerian approaches we stabilize the trajectory along which we measure the tremor by using the Kalman tracker, and subsequently perform the frequency estimation over the complete hand window. In Table 2 we test the added value of using phase information for frequency estimation. We compare two variants. The *Euler_gray* variant estimates the frequency over gray-scale pixels over gray-scale hand-images, obtained by cropping the hand location along the smoothed trajectory of the hand. The *Euler_phase* variant adds the 12 phase channels as detailed in section 3.3. The phase channels allow the *Euler_phase* to more precisely capture the small motion corresponding to the tremor, because the phase is effective for describing motion. The MAE numbers in Table 2 validate that adding the phase information is beneficial for the hand-tremor frequency estimation.
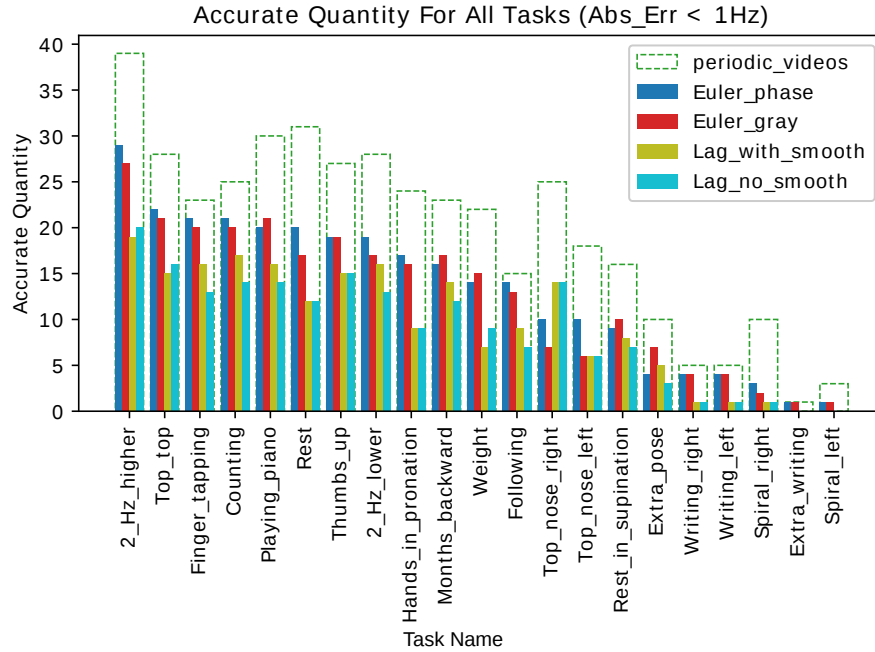
Fig. 6. Exp 2: We report accuracy on all recorded tasks, over the 55 patient recordings (higher is better). We consider the hand-tremor frequency to be correctly estimated for a task if the MAE (Mean Absolute Error) for that task is lower than 1 Hz. We plot in dotted green line the total number of videos recorded for each task, on which we have detected a periodic tremor. For each of our considered methods we show the number of videos for which the frequency was correctly estimated. On average the Eulerian methods perform better than the Lagrangian methods.

4.3   Exp 2: Eulerian versus Lagrangian tremor frequency estimation

In Figure 6 we display the accuracy of our proposed frequency estimation methods over the complete set of 55 patient recordings, for all tasks. We show in dotted green line the number of videos per task where a periodic tremor was detected, according to the accelerometer data. In corresponding color, we show the number of videos in which we have correctly estimated the hand-tremor frequency, for each frequency estimation method: *Euler_phase* is the Eulerian method using 12 phase-channels and 1 grayscale channel; *Euler_gray* is the Eulerian method on image intensity information only; *Lag_no_smooth* is the Lagrangian method without Kalman trajectory smoothing; *Lag_with_smooth* is the Lagrangian method with Kalman trajectory smoothing. We consider an estimated tremor frequency to be correct if the MAE between the accelerometer frequency and the one estimated by the method is lower than 1 Hz.

(a) *Hands_in_prontation* task.   (b) *Weight* task.
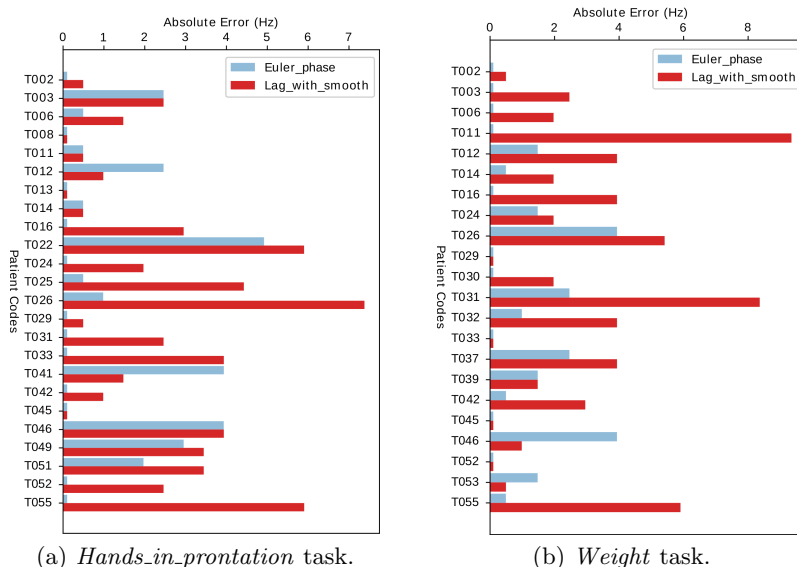
Fig. 7. Exp 2: The MAE per patient, for the two tasks where the Eulerian methods performed better than the Lagrangian methods (lower is better). To avoid over-cluttering the image, we plot only the best performing Lagrangian method: *Lag_with_smooth* — Lagrangian method with Kalman trajectory smoothing, and the best performing Eulerian method: *Euler_phase* — Eulerian method using 12 phase channels and 1 grayscale channel. The Lagrangian method makes large frequency estimation mistakes on a few patient videos, while the Eulerian method is more precise on some of the patient videos. (Note: for certain patients the task has not been recorded or no stable frequency, according to the accelerometer, has been found.)

Figure 6 shows that on average the Eulerian frequency estimation methods are more precise than the Lagrangian methods. The gain of using the Eulerian approaches is especially clear for the *Weight* task and the *Hands_in_pronation* task. Figure 7 displays the MAE scores per patient for these two tasks. To avoid over-cluttering the image, we only show the best Lagrangian method: *Lag_with_smooth*, Lagrangian with Kalman trajectory smoothing, and the best Eulerian method: *Euler_phase*, Eulerian over 12 phase channels and 1 grayscale channel. The Eulerian method gives more precise frequency estimates for some of the patient recordings, while for others it performs similar to the Lagrangian method. The tasks are not characterized by large hand motion. The gain of the Eulerian method over the Lagrangian is explained by the Eulerian method better describing the subtle changes in image information over time at the hand location. Therefore, the Eulerian method more accurately captures the tremor in tasks that do not involve large hand motion, but exhibit small motion.

5   Conclusions

We consider the task of hand-tremor frequency estimation from RGB videos. We propose two different approaches for measuring human hand-tremor frequencies: (a) Lagrangian hand-tremor frequency estimation, using the trajectory of the hand motion in the image plane throughout the video, to assess the hand-tremor frequency; and (b) Eulerian hand-tremor frequency estimation, which measures the change in the image information over time, at the location of hand in the image plane. We experimentally evaluate two variants of each approach on our proposed *TIM-Tremor* dataset containing 55 patient recordings performing a multitude of tasks. From our experimental analysis we learned that the Eulerian approaches are more accurate on average than the Lagrangian methods, with the difference being substantial on tasks on which there is a limited amount of large hand motion, but where there is a small hand-tremor motion present.

References

1. Andriluka, M., Pishchulin, L., Gehler, P., Schiele, B.: 2d human pose estimation: New benchmark and state of the art analysis. In: CVPR. pp. 3686–3693 (2014)
2. di Biase, L., Brittain, J.S., Shah, S.A., Pedrosa, D.J., Cagnan, H., Mathy, A., Chen, C.C., Martín-Rodríguez, J.F., Mir, P., Timmerman, L., et al.: Tremor stability index: a new tool for differential diagnosis in tremor syndromes. Brain p. awx104 (2017)
3. Bulat, A., Tzimiropoulos, G.: Human pose estimation via convolutional part heatmap regression. In: ECCV. pp. 717–732. Springer (2016)
4. Carreira, J., Agrawal, P., Fragkiadaki, K., Malik, J.: Human pose estimation with iterative error feedback. In: CVPR. pp. 4733–4742 (2016)
5. Cutler, R., Davis, L.S.: Robust real-time periodic motion detection, analysis, and applications. PAMI **22**(8), 781–796 (2000)
6. Dai, H., Zhang, P., Lueth, T.C.: Quantitative assessment of parkinsonian tremor based on an inertial measurement unit. Sensors **15**(10), 25055–25071 (2015)
7. Davis, A., Bouman, K.L., Chen, J.G., Rubinstein, M., Durand, F., Freeman, W.T.: Visual vibrometry: Estimating material properties from small motion in video. In: CVPR. pp. 5335–5343 (2015)
8. Elble, R.J., McNames, J.: Using portable transducers to measure tremor severity. Tremor and Other Hyperkinetic Movements **6** (2016)
9. Elgharib, M., Hefeeda, M., Durand, F., Freeman, W.T.: Video magnification in presence of large motions. In: CVPR. pp. 4119–4127 (2015)
10. Fleet, D.J., Jepson, A.D.: Computation of component image velocity from local phase information. IJCV **5**(1), 77–104 (1990)
11. Freeman, W.T., Adelson, E.H., et al.: The design and use of steerable filters. TPAMI **13**(9), 891–906 (1991)

12. Goldenberg, R., Kimmel, R., Rivlin, E., Rudzsky, M.: Behavior classification by eigendecomposition of periodic motions. Pattern Recognition **38**(7), 1033–1043 (2005)
13. Hssayeni, M.D., Burack, M.A., Ghoraani, B.: Automatic assessment of medication states of patients with parkinson's disease using wearable sensors. In: EMBC. pp. 6082–6085. IEEE (2016)
14. Jeon, H., Lee, W., Park, H., Lee, H.J., Kim, S.K., Kim, H.B., Jeon, B., Park, K.S.: Automatic classification of tremor severity in parkinsons disease using a wearable device. Sensors **17**(9), 2067 (2017)
15. Kayaba, H., Kokumai, Y.: Non-contact full field vibration measurement based on phase-shifting. In: CVPR. pp. 3655–3663 (2017)
16. Kooij, J.: Sensecap: synchronized data collection with microsoft kinect2 and leap-motion. In: ACMMM. pp. 1218–1221. ACM (2016)
17. Kooij, J., van Gemert, J.: Depth-aware motion magnification. In: European Conference on Computer Vision. pp. 467–482. Springer (2016)
18. Levy, O., Wolf, L.: Live repetition counting. In: CVPR. pp. 3020–3028 (2015)
19. Lu, C., Ferrier, N.J.: Repetitive motion analysis: Segmentation and event classification. PAMI **26**(2), 258–263 (2004)
20. Pintea, S.L., van Gemert, J.C.: Making a case for learning motion representations with phase. In: ECCV workshop. pp. 55–64 (2016)
21. Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P.V., Schiele, B.: Deepcut: Joint subset partition and labeling for multi person pose estimation. In: CVPR. pp. 4929–4937 (2016)
22. Polana, R., Nelson, R.C.: Detection and recognition of periodic, nonrigid motion. IJCV **23**(3), 261–282 (1997)
23. Ramakrishna, V., Munoz, D., Hebert, M., Bagnell, A.J., Sheikh, Y.: Pose machines: Articulated pose estimation via inference machines. In: ECCV (2014)
24. Ran, Y., Weiss, I., Zheng, Q., Davis, L.S.: Pedestrian detection via periodic motion analysis. IJCV **71**(2), 143–160 (2007)
25. Ripin, Z.M., Chan, P.Y.: Pathological hand tremor measurementchallenges and advances. In: ICIBELS. pp. 3–8 (2017)
26. Roy, K., Rao, G.S., Anouncia, S.M.: A learning based approach for tremor detection from videos. In: ICOS. pp. 71–76. IEEE (2013)
27. Runia, T.F., Snoek, C.G., Smeulders, A.W.: Real-world repetition estimation by div, grad and curl. In: CVPR. pp. 9009–9017 (2018)
28. Soran, B., Hwang, J., Lee, S., Shapiro, L.: Tremor detection using motion filtering and svm. In: ICPR. pp. 178–181 (2012)
29. Taewoong Um, T., Pfister, F.M.J., Pichler, D.C., Endo, S., Lang, M., Hirche, S., Fietzek, U., Kulić, D.: Parkinson's Disease Assessment from a Wrist-Worn Wearable Sensor in Free-Living Conditions: Deep Ensemble Learning and Visualization. CoRR (2018)
30. Tralie, C.J., Perea, J.A.: (quasi) periodicity quantification in video data, using topology. CoRR (2017)
31. Uhríková, Z., Růžička, E., Hlaváč, V., Nugent, C.D.: Treman: a tool for measuring tremor frequency from video sequences. Movement Disorders **25**(4), 504–506 (2010)
32. Uhríková, Z., Šprdlík, O., Hoskovcová, M., Komárek, A., Ulmanová, O., Hlaváč, V., Nugent, C.D., Růžička, E.: Validation of a new tool for automatic assessment of tremor frequency from video recordings. Journal of neuroscience methods **198**(1), 110–113 (2011)
33. Victor, B., He, Z., Morgan, S., Miniutti, D.: Continuous video to simple signals for swimming stroke detection with convolutional neural networks. CVPR (2017)

34. Vidailhet, M., Roze, E., Jinnah, H.A.: A simple way to distinguish essential tremor from tremulous parkinsons disease. Brain **140**(7), 1820–1822 (2017)
35. Wadhwa, N., Rubinstein, M., Durand, F., Freeman, W.T.: Phase-based video motion processing. SIGRAPH **32**(4), 80 (2013)
36. Wei, S., Ramakrishna, V., Kanade, T., Sheikh, Y.: Convolutional pose machines. In: CVPR. pp. 4724–4732 (2016)
37. Xia, L., Zou, B., Liu, H., Su, H., Qianghui, H.: A new method for evaluating postural hand tremor based on cmos camera. Optik-International Journal for Light and Electron Optics **126**(5), 507–512 (2015)
38. Zarchan, P., Musoff, H.: Fundamentals of kalman filtering: A practical approach (2013)
39. Zhang, Y., Pintea, S.L., van Gemert, J.C.: Video acceleration magnification. CVPR (2017)