Hand-tremor frequency estimation in videos

S.L. Pintea[1], J. Zheng[1], X. Li[1], P.J.M. Bank[2], J.J. van Hilten[2], J.C. van Gemert[1]

Delft University of Technology

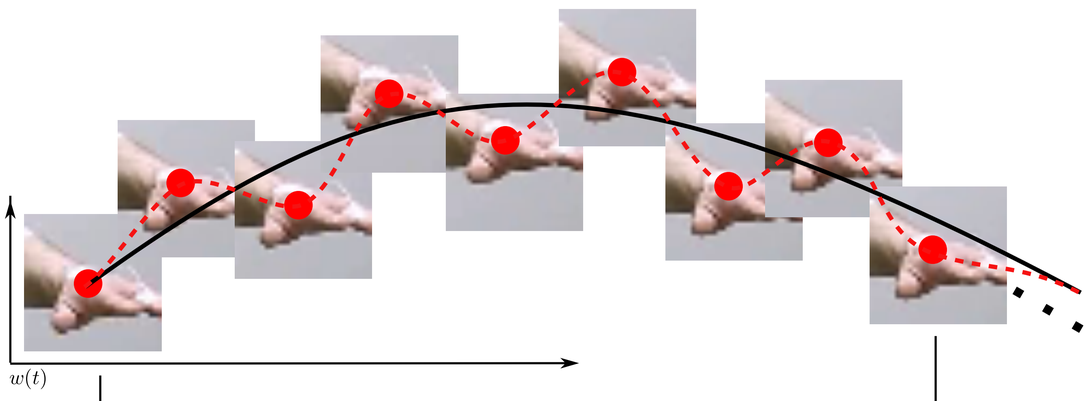
[1]Vision Lab, Delft University of Technology, Netherlands

[2]Department of Neurology, Leiden University Medical Center, Netherlands

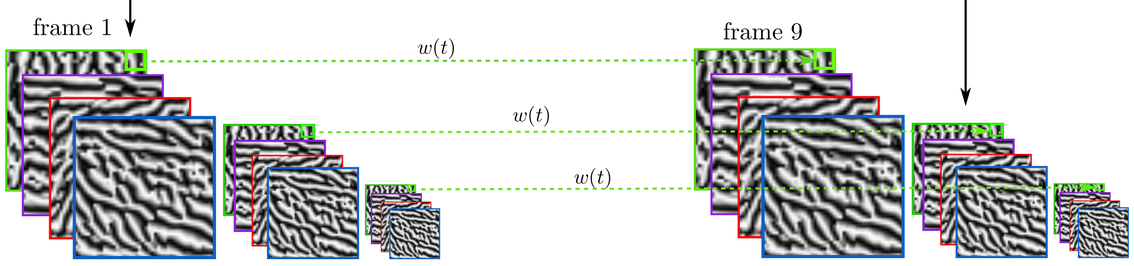
(I) Contributions:

- Two hand-tremor frequency estimation methods from RGB videos;
  - Lagrangian using the motion of the hand in the image plane;
  - Eulerian using image information over time at the hand location, as extracted from intensity values and phase-images.
- TIM-Tremor dataset containing:
  - 55 RGB patient videos,
  - ground-truth accelerometer recordings on the most affected hand,
  - depth video-data.

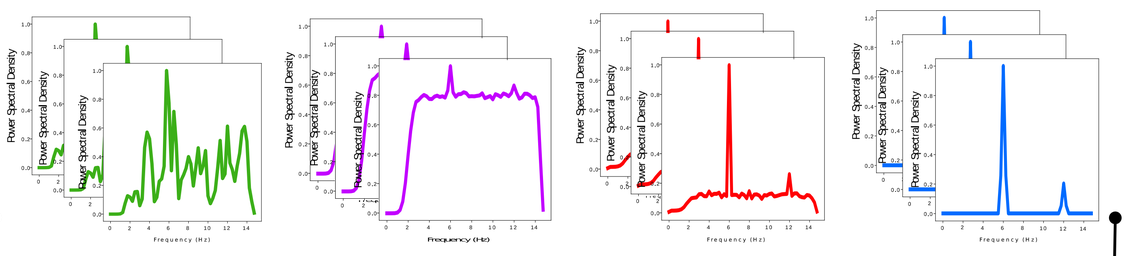
(1) Hand-location trajectory smoothing



(2) Phase pyramid with steerable filters



(3) Power spectrum density $\mathcal{P}_{w(t)}(f)$, per phase-image over $w(t)$



(4) Pick the best phase-image and predict its maximum frequency f^*

(II) Prior work:

- We use the pose estimation in [1] to detect the hand location.
- We use the steerable pyramid from [2] for extracting image phase.

(III) Lagrangian & Eulerian frequency estimation:

Lagrangian frequency estimation (1):

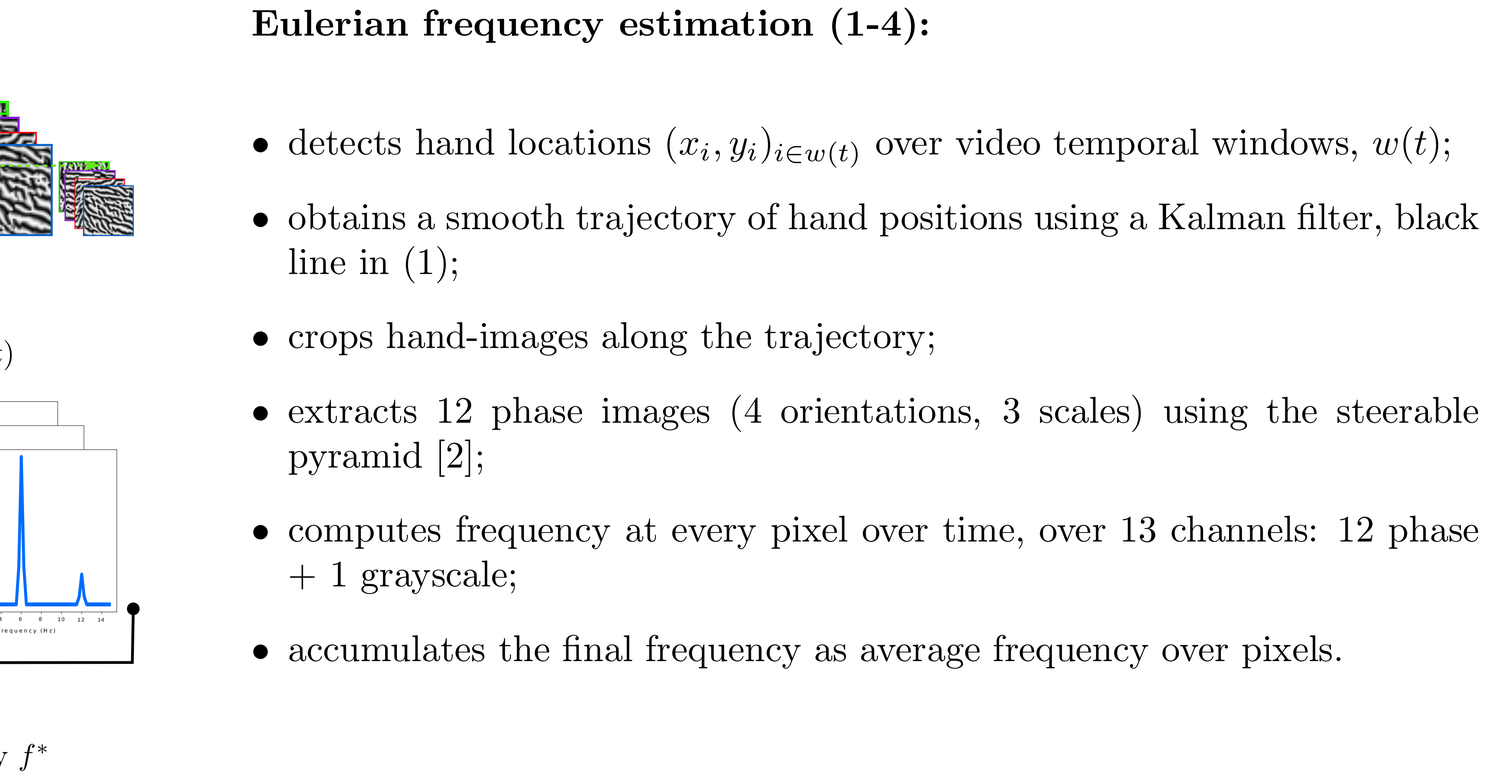
- detects hand locations $(x_i, y_i)_{i \in w(t)}$ over video temporal windows, $w(t)$;
- uses a Kalman filter to subtract the large motion — black line, and retains the small motion — red line;
- windowed Fourier transform is applied to obtain the frequency.

Eulerian frequency estimation (1-4):

- detects hand locations $(x_i, y_i)_{i \in w(t)}$ over video temporal windows, $w(t)$;
- obtains a smooth trajectory of hand positions using a Kalman filter, black line in (1);
- crops hand-images along the trajectory;
- extracts 12 phase images (4 orientations, 3 scales) using the steerable pyramid [2];
- computes frequency at every pixel over time, over 13 channels: 12 phase + 1 grayscale;
- accumulates the final frequency as average frequency over pixels.

(IV) Dataset & Results:

- We evaluate on 55 patients of our TIM-Tremor dataset.
- Tasks vary in: posture, action performed, cognitive distraction, and entrainment.
- We report accuracy as the number of videos where a tremor was present, and it was correctly detected: i.e. the absolute error $< 1Hz$.



(V) Dataset DOI & References:

Dataset DOI: 10.4121/uuid:522d14ed-3019-4206-b49e-a4e674b6440a.

[1] S. Wei et al: "Convolutional pose machines". In: CVPR (2016).

[2] D.J. Fleet, A.D. Jepson: "Computation of component image velocity from local phase information". In IJCV 5(1), 77-104 (1990).