Measuring Tree Sway Frequency Using Video Processing

Joseph Ammatelli

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Civil Engineering

University of Washington

2022

Committee:

Jessica Lundquist

Jim Thomson

Program Authorized to Offer Degree:

Civil and Environmental Engineering

©Copyright 2022
Joseph Ammatelli

University of Washington

Abstract

Measuring Tree Sway Frequency Using Video Processing

Joseph Ammatelli

Chair of the Supervisory Committee

Jessica Lundquist

Civil and Environmental Engineering

Measurements of tree properties and fluxes are necessary for improving surface models, monitoring the impact of phenomena such as drought, and mitigating tree related damage. However, collecting tree measurements can be challenging, often requiring invasive, expensive, or cumbersome instruments. A growing body of research has demonstrated that measurements of tree sway frequency can be used to quantify processes, such as drought stress and changes in mass resulting from precipitation interception, that are otherwise difficult to measure. While robust, existing methods for measuring tree sway lack spatial scalability. We investigate whether the virtual vision sensor and multilevel binary thresholding video processing algorithms can be used to accurately extract tree sway frequency at multiple points in a camera field of view, thus enabling scalable measurements of tree properties. Comparing sway frequencies extracted from video and accelerometer data at two sites, we show video processing of tree sway can reproduce

accelerometer sway frequencies with ±0.03 Hz accuracy. The results suggest the video processing algorithms may be suitable for applications where the approximate sway frequency or large changes in sway frequency are associated with the property of interest, for example the identification of trees vulnerable to wind throw and the measurement of snow in trees, respectively.

1. Introduction

Trees help mediate the exchange of water, carbon, oxygen, heat, and momentum between the land and atmosphere. Understanding the timing and magnitude of these processes, including evapotranspiration, precipitation interception, and seasonal changes in biomass, is critical for improving models and monitoring the impact of phenomena such as drought. Meanwhile, predicting mechanical failures in trees is important for mitigating economic losses and infrastructure damage associated with wind throw and wind snap (Moore and Maguire 2004).

Despite its importance to modeling, monitoring, and mitigation efforts, measuring tree properties can be challenging. Conventional instruments for measuring plant physiology, such as dendrometers and sap flow sensors, are invasive and expensive, limiting the number of samples that can be collected. Other methods used to measure tree properties are often cumbersome. For example, previous attempts to quantify canopy interception include weighing severed trees (e.g., Storck et al. 2002) and installing onerous compression-based systems (e.g., Martin et al. 2013).

In recent decades, measurements of natural tree sway frequency have been used to infer difficult-to-measure tree parameters, such as water stress and snow interception, offering a convenient and inexpensive alternative to previous methods. When excited by the wind, trees are compelled into oscillatory motion. The vibration has a dominant frequency defined as the natural frequency, henceforward referred to as the sway frequency. The sway frequency is strongly correlated with architectural parameters such as the diameter at breast height (DBH) and tree height (Bunce et al. 2019, Jackson et al. 2020, Moore and Maguire 2004). It is independent of wind speed in the absence of crown collisions (Rudnicki et al. 2008) and is sensitive to temperature (Granucci et al. 2013). Modeling a tree as a damped harmonic oscillator, the sway frequency can be related to mass and stiffness, enabling other tree processes to be observed and quantified. While many models exist, trees are commonly approximated as cantilever beams (Jackson et al. 2020, Moore and Maguire 2004). Exploiting the relationship between sway frequency, mass, and stiffness, previous studies have used tree sway to classify tree health (Baker 1997), identify changes in tree phenology (Gougherty et al. 2018, Jaeger et al. 2022), monitor plant water content (Ciruzzi and Loheide II 2019, Kooreman 2013), detect snow and rain interception events (Raleigh et al. 2022, Van Emmerick et al. 2017), and estimate changes in canopy mass (Raleigh et al. 2022, Selker et al. 2011).

Sway frequency values have been obtained from contact and contactless sensors including: inclinometers (e.g., Grannucci et al. 2013, Rudnicki et al. 2008), strain gauges (e.g., Jackson et al. 2019), accelerometers (e.g., Jaeger et al. 2022, Peltola 1996), 3d magnetic trackers (e.g., Barbacci et al. 2014), and laser interferometers (e.g., Baker 1997). In general, contact sensors are more economic and give robust long-term measurements but require considerable labor to install. Meanwhile, contactless methods deliver minimally invasive measurements but are expensive and usually unsuitable for unsupervised long-term deployments. In recent studies, the accelerometer has been the most common tool for measuring tree sway, given its low cost and performance in challenging environments. Despite its popularity, using an accelerometer to obtain a vibration signal with a high signal to noise ratio typically requires placing the sensor several meters above the ground (Ciruzzi and Loheide II 2019, Raleigh et al. 2022, Van Emmerick et al. 2017), which introduces installation overhead and safety concerns. Thus, accelerometers can provide useful

data for a small number of trees, but it can be infeasible or expensive to scale the measurements to the community or stand scales, especially when and where there is a significant degree of heterogeneity.

Recent work has demonstrated the potential of using a video camera to simultaneously measure mechanical vibrations at different points in space. Videos encode changes in brightness across space and time for the entire camera field of view (FOV), effectively providing an array of motion sensors. Temporal variations in brightness correspond with local motion so long as changes in lighting, changes to the visual properties of objects in the FOV, camera motion, and camera noise are all negligible. Borrowing terminology from fluid mechanics, vibration signals can be extracted from videos using Lagrangian and Eulerian frameworks. In the former, a feature is tracked across space and time. In the latter, the evolution of a fixed coordinate is analyzed over time. Significant energy has been invested into developing video-based vibration analysis tools for nondestructive testing and noncontact structural health monitoring applications in civil, mechanical, and aerospace engineering. Natural frequencies of cantilever beams (Chen 2016, Schumacher and Shariati 2013), structures (Chen et al. 2017, Ferrer et al. 2013, Schumacher and Shariati 2013, Shang and Shen 2018), and satellite components (Oda et al. 2011) have been accurately captured at multiple points in a camera FOV using both Eulerian and Lagrangian video processing algorithms. Further work has shown that video processing of vibrations can be used to measure the pulse of a human (Jeng and Wu 2012, Poh et al. 2010), perform fruit detachment analysis (Torregrosa et al. 2014), and estimate mechanical properties of plant stems (Nakata et al. 2018).

Several studies have used Lagrangian (feature tracking) video processing algorithms to characterize tree motion and properties. Diener et al. (2006) use the pyramidal implementation of the Kanade-Lucas-Tomasi (KLT) tracking algorithm to extract motion fields of real shrubs for animation purposes. Barbacci et al. (2014) measure tree velocity using two tracking algorithms, including an augmented version of KLT. Tadrist et al. (2018) use KLT video processing to assess the relative motion of branches and leaves at different wind speeds. Applying the Minimum Output Sum of Squared Error Filter (MOSSE), Wang et al. (2022) quantify the sway frequency of a leafless tree. Despite demonstrated success measuring the motion of tree features, Lagrangian vibration analysis can only resolve pixel scale motion for individual features (i.e., it does not provide motion information for the entire FOV) and suffers from issues such as target occlusion.

To analyze the sway of many trees in a stand, we need to move beyond accelerometers and Lagrangian video processing methods. Eulerian video processing methods can resolve sub-pixel motion across large spatial extents (Chen et al. 2017, Ferrer et al. 2013, Schumacher and Shariati 2013), showing promise for measuring the sway frequency of multiple trees in a camera FOV. We investigated whether two simple Eulerian video processing algorithms based on the methods of Schumacher and Shariati (2013) and Ferrer et al. (2013) can be used to capture the frequencies of swaying trees and enable analysis of tree form and function over greater spatial extents. In particular, we had the following objectives:

1. Evaluate the performance of the two algorithms by comparing the video processing output to accelerometer output for multiple trees and conditions

2. Assess the efficacy of the workflow for tree sway studies requiring high spatial and temporal resolution, provide recommendations for use cases and data collection, and suggest directions for future tree sway video processing work
3. Develop and share the vetted Python toolkit for extracting the sway frequency of a tree from a video

2. Materials and Methods

2.1 Video and Accelerometer Tree Sway Datasets

For video processing development and testing, we compiled videos of swaying trees (input data) and collocated accelerometer data (validation data) from several existing studies. Comprehensive sensor configurations and site characteristics for each dataset are summarized in Table 1.

Table 1. Dataset characteristics, camera configuration, and accelerometer configuration by site

	Trout Lake	**Manitou Experimental Forest**
Dataset Characteristics		
Site location	Trout Lake Research Station, WI	Manitou Experimental Forest, CO
Date/Period of samples	2019-08-15	2020-05-16 through 2020-09-03
Trees in FOV equipped with accelerometers	1	6
Tree species	Red Oak *(Quercus rubra)*	Ponderosa Pine *(Pinus ponderosa, PIPO)*
Camera Configuration		
Camera placement	Strapped to an adjacent tree near the ground, slightly upward oblique angle	Installed at the top of the 28-m tower, slightly downward oblique angle
Camera model	Bushnell TrophyCam Aggressor HD	GoPro Hero 4
Framerate (fps)	30	30
Frame resolution	1920x1080	1920x1080
Lens/FOV angle	45°	155°
Video Count/Collection Interval	5 videos	Every 30 min. May - Mid June Every 15 min. Mid-June - Mid August Every 30 min. Mid-August - Early September
Video duration	60s	30s
Accelerometer Configuration		
Accelerometer placement	~8m above ground on trunk	~2-3 m from crown on trunk (~6-8 m above ground)
Accelerometer model	Gulf Coast Data Concepts 2g MEL-X2	Gulf Coast Data Concepts 2g MEL-X2
Sampling Rate	16Hz	16Hz

Video and accelerometer data collected by Ciruzzi and Loheide II (2019) for a red oak tree (*Quercus rubra*) in Trout Lake, Wisconsin were used for a simple single-tree validation. A video camera (Bushnell TrophyCam, 30 fps, 1080p resolution, 45° FOV) was tilted upwards at a slightly oblique angle towards the target tree and fastened with straps to the base of an adjacent tree so camera motion could be assumed negligible. Five, 60-second videos were recorded over a 2-hour period on August 15, 2019. A 3-axis accelerometer (Gulf Coast Data Concepts 2g MEL-X2, 16 Hz continuous sampling) was positioned beneath the main branching of the target tree at ~8m.

To evaluate the efficacy of resolving the frequency of multiple trees in a video FOV, we processed tree sway data recorded by Bush (Sidney Bush, personal communication, August 12, 2022) at the Manitou Experimental Forest in Colorado. Bush, investigating the potential of using video processing of tree sway to observe drought stress, equipped six ponderosa pine trees with 3-axis accelerometers (Gulf Coast Data Concepts 2g MEL-X2, 16 Hz continuous sampling) and mounted a GoPro camera (30 fps, 1080p resolution, 155° FOV, 30s videos every 15-30 minutes) at the top of a nearby tower to capture the motion of each study tree in the FOV.

2.2 Video Data Processing

To extract tree sway frequency from video data, we developed a framework in Python that supports customizable video processing workflows. We decomposed the video processing logic into three generalized steps: translating a video into vibration signals, estimating the power spectral density of each signal, and aggregating the frequency content across all vibration spectra. For each step, we implemented a collection of functions based on existing methods. We then combined functions from each step to create end-to-end algorithms. We focused our analysis on two end-to-end algorithms based on the methods of Schumacher and Shariati (2013) and Ferrer et al. (2013). In the following subsections, each of which corresponds to a module in the toolkit, we describe in more detail our implementations ofe the three generalized steps and the end-to-end algorithms.

2.2.1 Translating a Video into Vibration Signals

Digital color video cameras encode changes in color and brightness across both space (pixels) and time (frames) for the entire camera FOV. When a RGB color model is used, each pixel is assigned three values representing the brightnesses of red, green, and blue in the space spanning the pixel. If we assume changes in lighting, changes to the visual properties of objects in the FOV, camera motion, and camera noise are all negligible, changes in pixel brightness from one frame to another correspond to the motion of objects in the FOV. Thus, each pixel brightness time series can be interpreted as a virtual motion time series and, in the terminology of Schumacher and Shariati (2013), each pixel can be regarded as a "virtual vision sensor." When an object in the camera FOV vibrates across a pixel, changes in the pixel brightness will be periodic, and we can use spectral analysis techniques to extract information about the vibration frequency content (Appendix A).

We implemented two methods for translating video data into vibration signals, both of which are predicated on the idea that pixel brightness variations correspond to motion across that pixel (Figure 1). The first method, applying the theoretical framework of Schumacher and Shariati (2013), treats each pixel as a "virtual vision sensor," whose time series captures the temporal evolution of local motion. Here, we loaded the region of interest (ROI) of the input video into an array with time, x, and y dimensions. To obtain a single brightness value from the three input color channels, we applied the `cv2.COLOR_RGB2GRAY` grayscale reduction from the OpenCV Python library. Instead of analyzing single pixels as Schumacher and Shariati do, we aggregated frequency information across all pixels in the ROI during the aggregation step. The second method, applying the theoretical framework of Ferrer et al. (2013), generates vibration signals by counting the number of pixels in the ROI whose grayscale brightness is below a particular threshold at each time step. Each threshold emits one vibration signal, so multiple vibration signals are generated using a suite of thresholds. The resultant signals combine information from all pixels in the ROI and represent how much a given object below a particular brightness occupies the ROI at each point in time. The multilevel binary thresholding method is predicated upon the motion of some high contrast boundary. Therefore, unlike the virtual vision sensor, the multilevel binary thresholding method requires a distinct boundary between the vibrating object and the background. Henceforward, the methods will be referred to as the virtual vision sensor (VVS) and multilevel binary thresholding (MBT) methods, respectively.

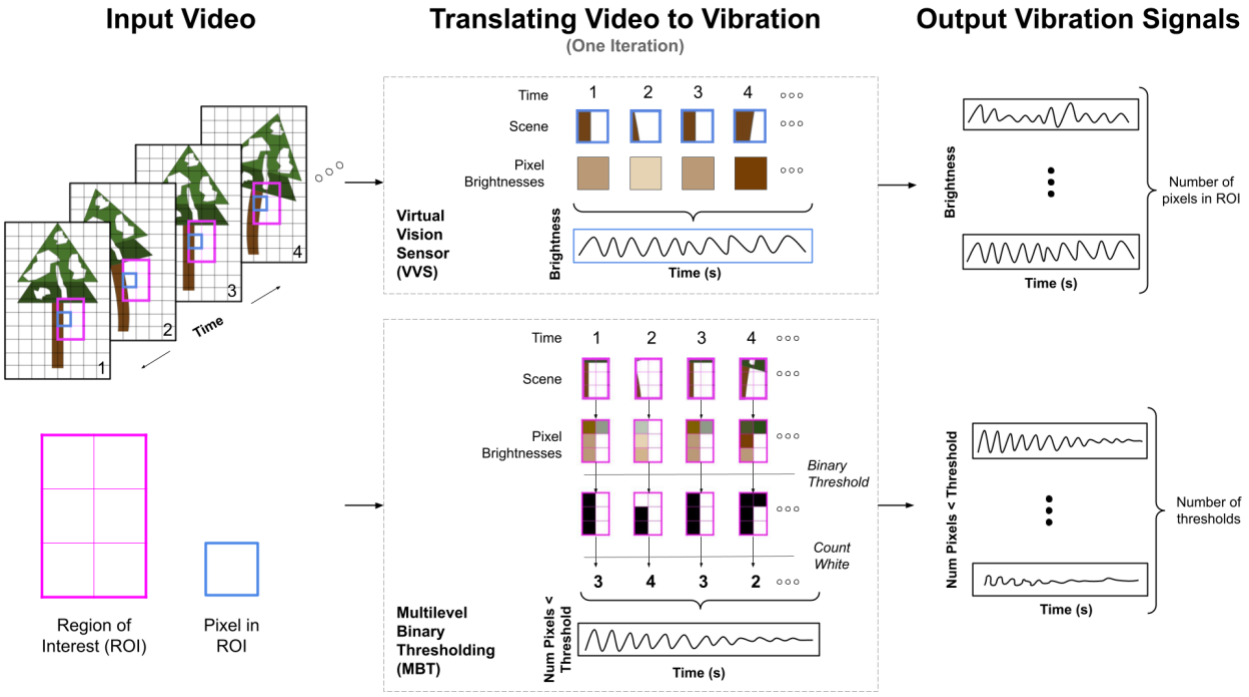


Figure 1. Conceptual illustration of how vibration signals are extracted from a video using two methods. For the virtual vision sensor method (top), pixel brightness time series are treated as vibration signals. One vibration signal is generated for each pixel in the region of interest. For the multilevel binary thresholding method (bottom), the number of pixels in the ROI below a certain threshold is counted at each time step. N vibration signals are generated using n thresholds.

For both methods, the ROI was manually selected using criteria outlined by the authors of both VVS and MBT, who emphasize the importance of evaluating areas of high contrast. To maximize brightness fluctuations and evaluate pixels most representative of the tree sway frequency, we visually inspected the FOV for areas of exposed trunk with high contrast backgrounds (e.g., the sky). The vertical range of the bounding box was chosen such that it spanned the longest vertical portion of the exposed trunk with a high contrast background. The horizontal range was chosen such that it included one edge of the trunk, approximately equal parts of trunk and background, and enough trunk/background pixels to maintain a baseline background signal for MBT. ROIs high on the trunk, where larger sway displacements result in greater changes in pixel brightness, were preferred to ROIs lower on the trunk. When the trunk was difficult to parse or the exposed trunk sway displacement was observed to be insignificant (apparent low signal to noise ratio), ROIs on the crown with high contrast were evaluated.

Formally, suppose we have an input video array given by $V(n, x, y, c)$ where n is the frame number, x is the horizontal coordinate of a pixel, y is the vertical coordinate of a pixel, and c is the RGB color channel. For both methods we extract a rectangular ROI with width w, height h, and corner nearest the origin (r_x, r_y) and combine the three color channels to get the grayscale brightness for each pixel across all frames:

$$V_{ROI}(n, i, j) = 0.299 \cdot V_R(n, i, j) + 0.587 \cdot V_G(n, i, j) + 0.114 \cdot V_B(n, i, j) \quad (1)$$

where

$$V_{channel}(n, i, j) = V\big(n, i + r_x, j + r_y, channel\big) \quad (2)$$
$$i = x - r_x \; for \; x \; in \; [r_x, r_x + w] \quad (3)$$
$$j = y - r_y \; for \; y \; in \; \big[r_y, r_y + h\big]. \quad (4)$$

Then, the vibration signals generated by the VVS method are given by:

$$V_{VVS}(n, i, j) = V_{ROI}(n, i, j). \quad (5)$$

That is, V_{VVS} is simply the input video clipped to the ROI and reduced to grayscale brightness. Meanwhile, for the MBT method, the vibration signals are given by:

$$V_{MBT}(n, l) = \sum_i \sum_j F(n, i, j, l) \quad (6)$$

where

$$F(n, i, j, l) = 1 \; when \; V_{ROI}(n, i, j) < l \quad (7)$$
$$F(n, i, j, l) = 1 \; when \; V_{ROI}(n, i, j) \geq l \quad (8)$$

and l is some brightness threshold.

We generated eight MBT vibration signals using eight grayscale thresholds evenly spaced between the minimum and maximum brightnesses in the ROI across all frames. Specifically, if

$$I_{min} = \min\{V_{ROI}(n,i,j)\}_{n,i,j} \quad (9)$$
$$I_{max} = \max\{V_{ROI}(n,i,j)\}_{n,i,j} \quad (10)$$

are the minimum and maximum brightness across all pixels and frames, the set of thresholds, L, used to generate vibration signals is given by

$$L = \left\{ \left\lfloor I_{min} + \frac{i \cdot (I_{max} - I_{min})}{9} \right\rfloor \right\}_{i=1}^{8}. \quad (11)$$

2.2.2 Estimating the Power Spectral Density of Each Vibration Signal

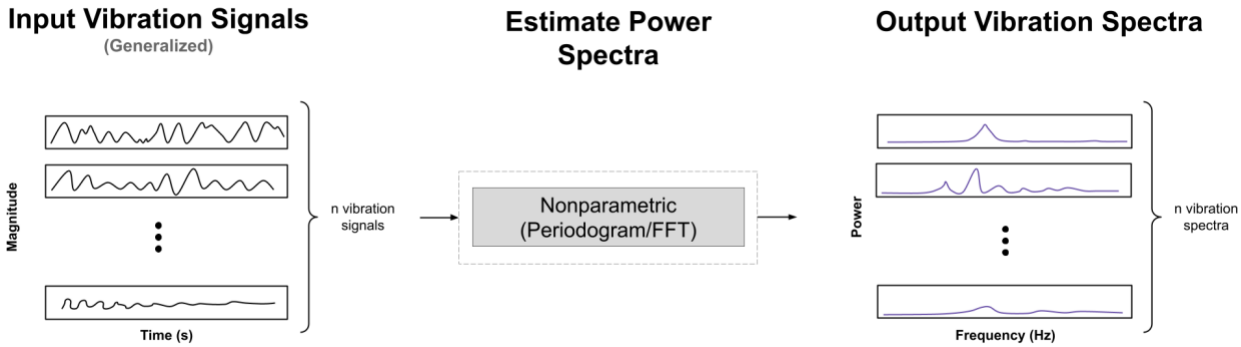


Figure 2. Illustration of how the frequency content of each vibration signal is estimated. Input vibration signals are the output from the previous step (either VVS or MBT vibration signals). For each vibration signal, the power spectrum is estimated using the nonparametric periodogram method.

We analyzed the sway frequency content of each vibration signal by estimating its power spectral density (PSD) (Figure 2). Experimenting with both parametric (autoregressive) and nonparametric (periodogram) methods for estimation, we ultimately decided to use the nonparametric `periodogram` function in the `scipy.signal` package given its comparable output spectrum and improved efficiency (Appendix B). Prior to computing the PSD using the periodogram, we removed any linear trend and subtracted the mean, applied a windowing function, and zero padded each input vibration signal. Window type and zero padded length were left as tunable parameters, which were selected by choosing the pair of parameters that resulted in the smallest difference with accelerometer frequencies for an arbitrarily chosen video in the dataset. We experimented with rectangular and Hann windows and zero padded lengths equal to twice and four times the nearest power of two. The narrow main lobe of the rectangular window in the frequency domain is useful for maximizing frequency resolution, despite its greater spectral leakage. Meanwhile, the Hann window reduces spectral leakage by compromising between main lobe width and side lobe height. The spectra from the 60 second Trout Lake videos had a frequency resolution of 0.016 Hz while the 30 second Manitou Experimental Forest videos had a frequency resolution of 0.033 Hz. Before aggregation, all vibration signal spectra had two degrees of freedom (a measure of statistical quality). To simplify peak selection and

reduce the processing memory requirements, we trimmed the resultant PSDs to a range of plausible candidate frequencies: 0.15-0.5 Hz for both datasets.

2.2.3 Aggregating Frequency Content Across All Vibration Signals

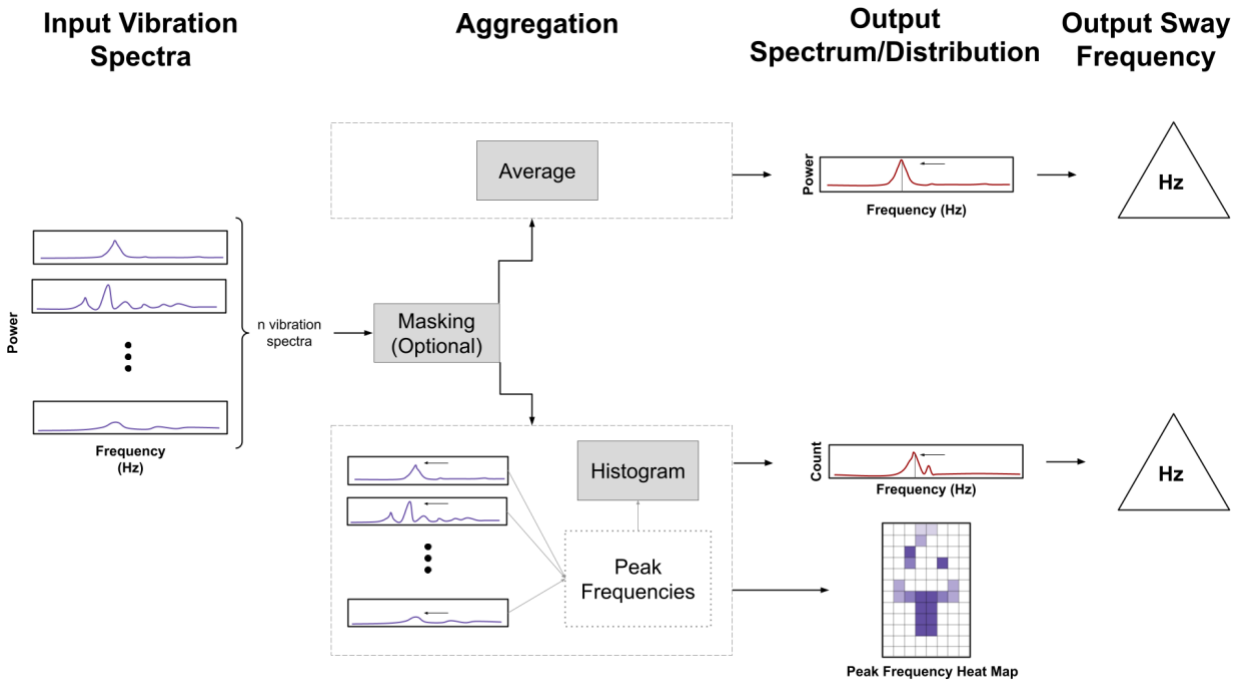


Figure 3. Illustration of the two methods used for aggregating power spectra generated during the previous step. In the first method (top), the input spectra are averaged, and the peak frequency of the average spectrum is outputted. In the second (bottom), the peak frequency of each input spectrum is computed, and the mode peak frequency is outputted. For the second method, when the input spectra have spatial coordinates, a peak frequency heat map can be constructed for spatial analysis.

Once we estimated the power spectral density for each vibration signal, we computed sway frequencies for the ROI using up to two aggregation approaches (Figure 3): (1) finding the peak frequency of the average power spectrum and (2) finding the most commonly occurring peak frequency among the vibration power spectra. The second aggregation method was only applied to VVS spectra since there were relatively few MBT spectra, making the mode potentially misleading. The peak frequency of a spectrum was defined to be the frequency of the local maximum with the greatest magnitude. Such peaks and their corresponding prominences, where prominence is defined as the vertical distance between the peak and its lowest contour line, were found using the `scipy.signal.find_peaks` function. For the first method, we simply computed the unweighted average of the power spectra before finding the peak frequency as described above. Prior to averaging the VVS spectra, members of the input power spectra corresponding to background pixels or vibration signals with low signal to noise ratios were masked based on the prominence of their dominant frequency peak. The average spectrum

provides useful insights into the prominence and magnitude of the entire range of candidate frequencies and improves the degrees of freedom to two times the number of input spectra such that the degrees of freedom are typically > 50 for the VVS average spectrum and 16 for the MBT average spectrum. This increase in degrees of freedom indicates the extracted peak frequency is more statistically robust than selecting individual peak frequencies from the input spectra, which only have two degrees of freedom. For the second method, we computed the peak frequency of each vibration spectrum and then counted the number of times each peak frequency occurred, building a histogram for spectra peak frequencies (where the bin width is equal to the frequency resolution). The mode peak frequency was chosen as the dominant sway frequency for the ROI. As before, pixels or vibration signals with low signal to noise ratios were masked based on the prominence of their dominant frequency peak. Since the VVS spectra were associated with spatial coordinates, the peak frequencies were used to create a peak frequency heat map so the spatial distribution of peak frequencies could be assessed.

2.2.4 End-to-end Video Processing Algorithms

The first video processing algorithm we developed was an extension of the "Virtual Vision Sensor" method proposed by Schumacher and Shariati (2013). We obtained vibration signals from the input video using the VVS framework of treating each grayscale pixel brightness time series in the ROI as a vibration signal. The vibration spectra were estimated using the nonparametric periodogram technique. Whereas Schumacher and Shariati analyzed individual pixels, we aggregated vibration spectra using both methods discussed in 2.2.3 and thus outputted two sway frequencies. For both aggregation techniques, we masked any spectra whose frequency peak prominence was below the 75th percentile of all peak prominences in the ROI. We performed no normalization to the input spectra so that only signals with high absolute prominence were considered (Appendix C for illustration of impact of normalization).

For the second video processing algorithm, we implemented the multilevel binary thresholding scheme developed by Ferrer et al. (2013). Eight vibration signals were generated using the MBT video to vibration translation with eight binary thresholds. As before, vibration spectra were computed using the nonparametric periodogram PSD estimation. We combined the signals in the frequency domain by finding the peak frequency of the average spectrum. All spectra were included in the average. Since we only generated eight vibration signals, each of which had no spatial coordinate, we did not compute the mode peak frequency or output a heat map.

2.3 Accelerometer Data Processing for Video Processing Validation

The dominant sway frequencies generated by the video processing algorithms were compared against the dominant sway frequencies extracted from acceleration data with two durations/start times. The first accelerometer signal, having the same start time and duration as the video vibration signals, was used to assess how well the video processing performed against an accelerometer signal with the same number of vibration periods. Here, the power spectra across all horizontal accelerometer axes were computed using periodograms of size 4096, rectangular windowing functions, and input detrending. The spectra were then averaged together, and the

peak frequency was identified. The resultant spectrum had the same frequency resolution as the test video (0.016 Hz for Trout Lake and 0.033 Hz for Manitou Experimental Forest) and 4 degrees of freedom. To compare the video processing output to sway frequency values more representative of the average sway over a longer interval, we also extracted peak frequencies from 30-minute acceleration signals centered on the video start times. Here, the spectrum for each accelerometer axis was estimated using Welch's method (`scipy.signal.welch`) with 50% overlapping segments of length 4096 and rectangular windows. Welch's method averages the spectra of segments of the input signal, reducing the spectrum noise and offering a more accurate indication of average frequency content. The 30-minute accelerometer spectra had a frequency resolution of 0.004 Hz and approximately 50 degrees of freedom.

We validated the video processing methods using five videos from the Trout Lake dataset and three videos from the Manitou Experimental Forest dataset. For the Trout Lake dataset, we focused our analysis on a small ROI spanning the sky-trunk boundary and applied three video processing methods: (1) VVS with average spectrum aggregation, (2) VVS with modal aggregation of spectra peaks, and (3) MBT with average spectrum aggregation. We calibrated the processing using the video at 18:09 and estimated all spectra using the Hann window and a zero-padded length of 8192. For the Manitou Experimental Forest dataset, we extracted sway frequencies for the two target trees with the greatest visibility. Other trees equipped with accelerometers were obscured by trees in the foreground, difficult to identify, and lacked worthy ROIs. Since the ROIs enclosing the target trees included mostly tree crown features and lacked distinct, high contrast boundaries, we only applied the two VVS video processing methods, which, unlike the MBT method, are not predicated on measuring the motion of a boundary. We calibrated the processing using the video at 2020-08-20 17:28 and estimated all spectra using the rectangular window and a zero-padded length of 4096.

3. Results

Sway frequency values extracted from video and accelerometer data are reported in Table 2. When computing the sway frequency values, we made use of zero padding during periodogram estimation, which enabled high quality interpolation between the frequency samples (bins) associated with the original length vibration signals. However, zero padding does not improve the frequency resolution (bin width) of the spectrum, which is equal to the inverse of the unpadded signal duration in seconds. Peaks within a bin width of each other can be mapped to the same frequency bin, meaning we can only reliably estimate peaks to the nearest bin. Therefore, the values reported in Table 2 are rounded to two decimal places, the precision of the coarsest frequency resolution in both datasets.

In general, sway frequencies extracted from videos showed strong agreement with sway frequencies extracted from accelerometer data. Nine of 11 trials had video processing output within 0.01 Hz of accelerometer output. Accelerometer frequencies were exactly reproduced to two decimals by at least one video processing method for six of 11 videos. The maximum observed difference between video and accelerometer output was 0.26 Hz (Trout Lake 17:32 sample). Excluding accelerometer frequency peaks with poor prominence, the maximum observed difference was 0.03 Hz (08-20 trial at Manitou Experimental Forest, Tree 2).

Table 2. Output video processing and accelerometer sway frequencies for each site

	Video Processing Peak Frequency (Hz)			Accelerometer Peak Frequency (Hz)	
Video Timestamp (local time)	**VVS Average Spectrum**	**VVS Peak Frequency Histogram**	**MBT Average Spectrum**	**Short Window (same duration as video)**	**30-minute Window**
Trout Lake					
2019-8-15 17:07:18	0.21	0.21	0.21	0.26	0.23
2019-8-15 17:30:01	0.24	0.23	0.25	0.24	0.24
2019-08-15 17:32:34	0.23	0.23	0.23	0.49	0.25
2019-08-15 18:00:01	0.24	0.24	0.24	0.24	0.24
2019-08-15 18:09:06	0.24	0.24	0.24	0.40	0.24
Manitou Experimental Forest, Tree 1					
2020-08-15 12:06:00	0.28	0.28	-	0.29	0.29
2020-08-20 17:28:00	0.29	0.30	-	0.29	0.29
2020-08-31 11:50:00	0.30	0.30	-	0.31	0.28
Manitou Experimental Forest, Tree 2					
2020-08-15 12:06:00	0.37	0.38	-	0.38	0.37
2020-08-20 17:28:00	0.36	0.36	-	0.39	0.37
2020-08-31 11:50:00	0.37	0.36	-	0.36	0.38

3.1 Single Tree Validation – Trout Lake Dataset

Analysis of a ROI spanning the trunk-sky boundary for five test videos demonstrated strong agreement between the sway frequencies returned from video and accelerometer processing. For three of five videos, at least one of the video processing output frequencies was within 0.01 Hz of at least one of the accelerometer outputs, and at least one of the video processing output frequencies was the same as at least one of the output accelerometer frequencies. Across all five videos, the maximum difference between a video processing and accelerometer output frequency

was 0.26 Hz (video at 17:32). Excluding frequencies extracted from the short (i.e., same length as video) accelerometer windows, the maximum difference across all five videos was 0.02 Hz (videos at 17:07 and 17:32). The video processing methods tended to agree, with all three outputs being the same for four of the five videos. The video output tended to underestimate the accelerometer output. Apart from the 1-minute accelerometer spectra, all spectra and the VVS frequency distributions had relatively prominent peaks (i.e., distinct sway frequency signals). Compared to the accelerometer spectra, the video processing output exhibited more prominent peaks.

Considering the video collected at 18:00 as an example (Figure 4), the VVS peak frequency heat maps show that pixels with the most prominent frequency peaks were concentrated around the trunk-sky boundary. Masking removed pixels overlapping with the sky and regions of the trunk with poor contrast. The VVS average spectrum, VVS frequency distribution, and MBT average spectrum all yielded the same peak frequency of 0.24 Hz. Meanwhile, the accelerometer spectra for 1- and 30-minute windows also yielded peak frequencies of 0.24 Hz.

3.2 Multi-Tree Validation – Manitou Experimental Forest Dataset

For the Manitou Experimental Forest dataset, we examined video output for a stand of trees, as illustrated for the 08-15 video sample (Figure 5). Sway frequencies extracted for two trees (marked in Figure 5) from three video samples were similar to resolved accelerometer frequencies and exhibited agreement comparable to the Trout Lake output. Unlike the Trout Lake data, however, the accelerometer spectra had pronounced peaks for all window lengths. In some cases, the accelerometer spectra had adjacent peaks or energy concentrated in a range of frequencies (e.g., Figure 5d). The two focus trees had similar maximum differences between the video processing and accelerometer data.

3.2.1 Tree 1

For all three test videos, the video processing output was within 0.02 Hz of the accelerometer output. In one of the three trials, one of the video processing methods exactly reproduced an accelerometer sway frequency. The outputs of the two video processing methods were the same for two out of the three trials, with the average spectrum result less than the histogram peak frequency result for the remaining trial. Neither video processing method consistently outperformed the other; however, the VVS average spectra peaks tended to be weaker (broader with low prominence) than the VVS histogram peaks (e.g., Figure 5b).

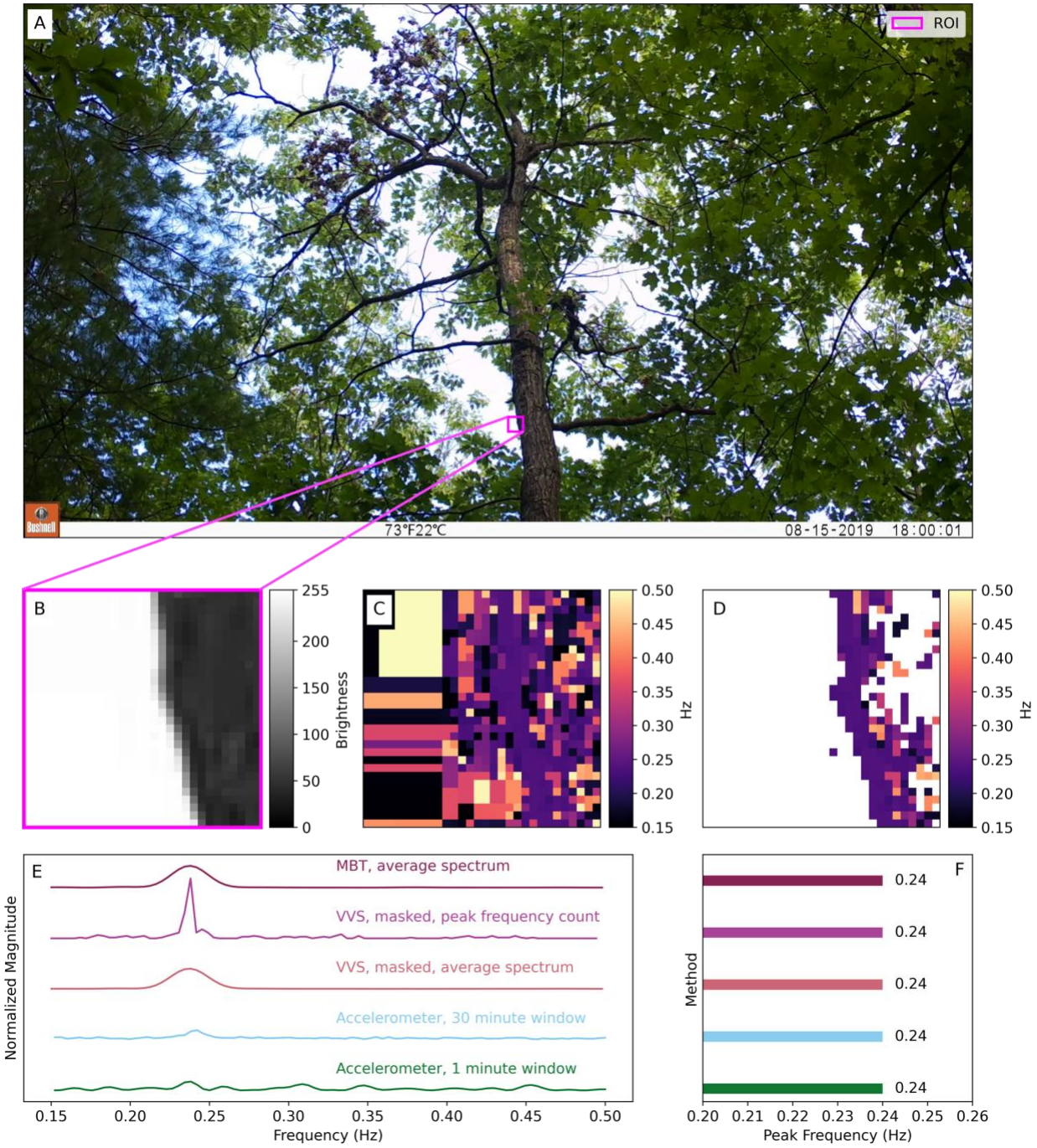


Figure 4. Experimental results from Trout Lake video sample collected at 18:00. (A) Camera FOV. (B) Grayscale region of interest. (C) Unmasked frequency heat mask. (D) Masked frequency heat map where only pixels whose spectrum's peak prominence are above the 75th percentile of prominences are shown. (E) Video processing and accelerometer spectra. All spectra/distributions have been normalized for comparison. (F) Peak frequencies extracted from each signal.

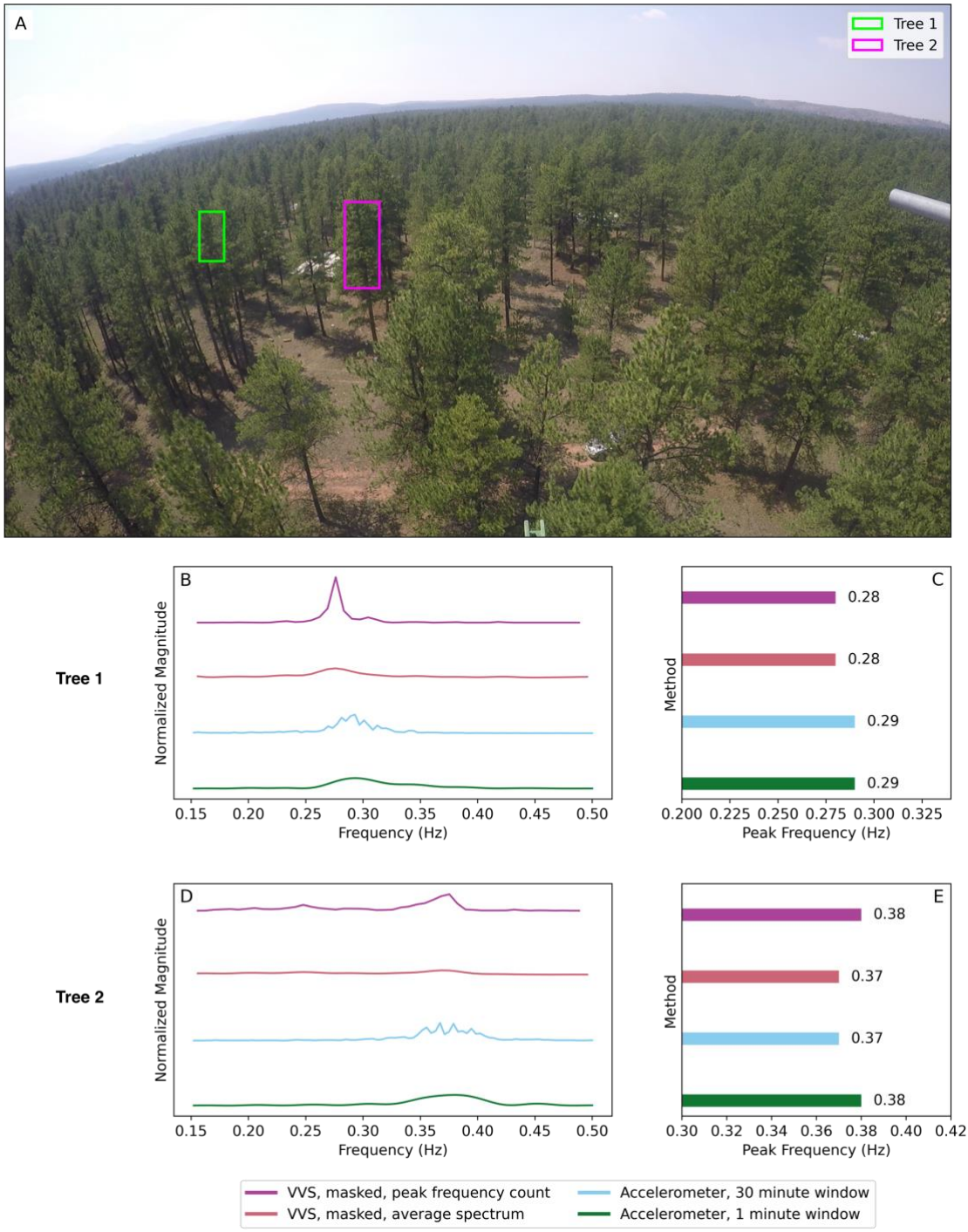


Figure 5. Sample output for the 8-15 video from the Manitou Experimental Forest dataset. Note that only VVS methods are reported since MBT was not suitable for the ROIs enclosing the crowns. All spectra/distributions have been normalized for comparison. (A) Camera FOV and ROIs. (B) Output video processing and accelerometer spectra for Tree 1. (C) Tree 1 peak frequency by method. (D) Output video processing and accelerometer spectra for Tree 2. (E) Tree 2 peak frequency by method.

3.2.2 Tree 2

For all three test videos, the video processing output was within 0.03 Hz of the accelerometer output. In two of the three trials, video processing reproduced at least one of the accelerometer outputs to the nearest hundredth of a Hz. The outputs of the two video processing methods were the same for one of the three trials, and, like the Tree 1 output, the VVS average spectra peaks were broader and less prominent than the peak frequency histogram peaks.

4. Discussion

4.1 Assessing Video Processing Performance

4.1.1 How Accurate are the Video Processing Sway Frequency Measurements?

We demonstrate that the two Eulerian video processing algorithms can be used to measure tree sway frequency to within approximately ±0.03 Hz of analogous frequencies extracted from multiple lengths of accelerometer data. The worst observed video processing-accelerometer difference of 0.26 Hz involved an accelerometer sway frequency extracted from a short (60s) window whose power spectrum had no clear peak and only 2 degrees of freedom (poor statistical quality). Excluding short window accelerometer spectra with poor peak prominences, all differences were less than 0.03 Hz. Agreement with accelerometer output to within ±0.01 Hz was achieved in nine of 11 trials, indicating the video processing measurements commonly have an even higher degree of agreement with accelerometer sway frequencies. When accelerometer spectra for both lengths of data had prominent peaks, the video processing output did not favor sway frequencies from one accelerometer signal length (i.e., the video processing showed similar performance measuring the short-window and longer-term average sway frequencies).

4.1.2 Why do the Video and Accelerometer Sway Frequencies Disagree?

We were surprised to observe that the video processing output did not show greater agreement with the short window accelerometer sway frequencies, which were generated from vibration signals with the same start time and duration as the test videos. The discrepancies leave room for several possibilities. First, it is possible that the video and accelerometer vibration signals encoded different vibrations. Deviations between the vibration signals may have been due to differences in ROI and accelerometer positions, camera motion, camera quantization noise, changes in lighting, or lens distortion. Furthermore, trees have been observed to have sway motion along two primary axes, each of which have their own sway frequency (Kovacic et al. 2018). While existing measurements indicate these sway frequencies are nearly identical, it is possible the video and accelerometer vibration signals were describing the motion of different axes. In the case of the Manitou Experimental Forest trees, since the ROIs enclosed part of the tree crowns, branch motion may have also biased the video processing measurements. Second, the accelerometer spectral analysis parameters were fixed, while the analogous video processing parameters (the zero padded length and window type) were tunable. Thus, differences in how the vibration signals were processed may have resulted in small frequency deviations. Third, in some cases, the small windows of acceleration data had poor signal to noise ratios, leading to the

extraction of potentially misleading sway frequencies. In contrast to the 30-minute accelerometer spectra, the short length accelerometer spectra, which only have two degrees of freedom, had peaks with considerable magnitude uncertainty. For the Trout Lake dataset, the spectra resulting from short accelerometer signals had multiple peaks with similar magnitudes, but the 30-minute spectra had a single distinct peak or a cluster of adjacent peaks (Figure 4). In general, this suggests the short accelerometer vibration signals do not contain enough vibration cycles and/or are disproportionately impacted by inherent noise.

Differences between the video processing output and peak frequencies extracted from 30-minute accelerometer windows were expected given the difference in vibration signal lengths. The long acceleration windows contained more tree sway cycles and therefore reflected the sway over a longer period of time. Moreover, using Welch's method for spectral estimation involved averaging spectra from segments of the acceleration signals and therefore yielded the average sway frequency over the interval. In some cases, after the averaging, two dominant frequency peaks were observed in the accelerometer spectra, suggesting different segments of the accelerometer signal yielded subtly different frequency peaks or segments tended to have two dominant modes (e.g., Figure 5, Control East). This could be due to small natural changes in sway frequency or possibly changes in wind direction resulting in sway along a different primary axis.

4.1.3 Comparing the Video Processing Algorithms

All three algorithms (VVS with average spectrum aggregation, VVS with modal aggregation of spectra peaks, and MBT with average spectrum aggregation) performed similarly when applied to suitable ROIs. Each method overestimated, underestimated, and exactly resolved accelerometer frequencies on different trials. In many cases, all three methods returned the same frequency, increasing the confidence in the video processing output. Both the average spectrum and peak frequency histogram aggregation techniques applied to the VVS vibration spectra proved to be comparable but were not always the same, suggesting both are worth examining. However, in general, the average spectrum aggregation technique should be preferred given its larger number of degrees of freedom. When possible, we recommend applying multiple video processing methods to increase the confidence in the output sway frequencies.

Despite possessing similar degrees of accuracy, the different algorithms, specifically the different methods for translating a video into vibration signals, proved to have distinct practical tradeoffs. Unlike the MBT algorithm, the VVS algorithms can successfully resolve frequencies from ROIs with no distinct boundary motion (e.g., the tree crowns in the Manitou Experimental Forest dataset). However, because the MBT classifies pixels based on some brightness threshold, it is more robust against ambient lighting changes. The VVS algorithm can output pixel wise peak frequencies, which is useful for visualizing frequency content over large spatial extents. However, since the VVS method requires loading and operating on each pixel time series, it is more computationally and memory intensive compared to the MBT method, which only operates on a small number of vibration signals.

4.2 Viability of Measuring Tree Sway Frequency Across Space and Time Using Videos

We show sway frequencies for multiple trees in a camera FOV can be resolved to a reasonable degree of accuracy. Despite promising results for individual videos sampled from small time windows, scaling analysis to video time series with multiple target trees presents several challenges. First, the results demonstrate that agreement between video processing and accelerometer sway frequency output is sensitive to the choice of frequency range, ROI, spectral estimation parameters (zero padded length, window type), and aggregation approach. Choosing the optimal combination of parameters for many videos and conditions would likely be challenging. Second, it remains to be seen how well the video processing performs when sway displacements are relatively small. In most cases, the videos and trees we analyzed had strong sway signals with visually apparent displacements and distinct accelerometer peaks. Tree 2 in the Manitou Experimental Forest had the weakest visually apparent sway signal and exhibited the least degree agreement between video and accelerometer output, foreshadowing potential decreases in video processing performance when the sway displacement is small. Third, it can be challenging to resolve the sway of multiple trees distributed across a forest stand. When the camera is at a high point looking down, as can be necessary to observe multiple trees, and/or when there is a dense forest canopy, it can be difficult to parse individual trees and identify ROIs with suitable contrast (e.g., with the sky). In many cases, trees in the foreground block trees of interest and desirable ROIs. Moreover, as the distance between the camera and target trees increases, the apparent sway displacement decreases, leading to potentially weaker video processing vibration signals.

While difficult to scale across space, accelerometers provide several distinct advantages when collecting data over long periods of time. Accelerometers log continuously throughout the night and day, providing sway signals with high temporal resolution. Given the memory overhead of collecting video data, achieving similar results with video processing would require considerable real time processing. When placed appropriately, accelerometers deliver strong sway signals for two perpendicular axes. As a result, dominant sway frequencies tend to be more prominent and sway motion for different axes can be quantified and compared. Accelerometers are reliable in a wide variety of conditions, simplifying data processing.

4.3 Tree Sway Video Processing Use Cases and Deployment Recommendations

4.3.1 Tree Sway Video Processing Use Cases

While a typical accuracy of ±0.03 Hz is acceptable for many applications, it may not be sufficient for studies requiring 0.01 Hz accuracy or finer, which has been the case for several existing tree sway studies. Ciruzzi and Loheide II (2019), Gougherty et al. (2018), and Van Emmerick et al. (2017) attribute changes in tree sway on the order of 0.01 Hz to diurnal variations in water content, phenology changes, and rain interception, respectively. Meanwhile, Ciruzzi and Loheide II (2019), Jaeger et al. (2022) and Raleigh et al. (2022) use changes in tree sway an order of magnitude larger to observe drought stress, estimate the timing of major phenological changes, and quantify the timing and magnitude of snow interception, respectively. Thus, the applicability of the video processing methods to tree sway studies depends on the

expected magnitude of sway frequency changes resulting from the phenomena of interest and the periodogram frequency resolution that can be achieved with the given length of video data. With the current degree of accuracy, the video processing may be useful for observing severe drought stress and quantifying snow interception. Additionally, the VVS method and resultant frequency heat map may be useful for applications where measuring the approximate sway frequency is desirable, such as for identifying trees at greater risk of wind throw.

We recommend limiting video processing analysis to tall, slender trees that are well represented by the cantilever model and whose stem sway dominates branch sway. We observed the strongest sway signals when and where there was a large sway displacement and a distinct boundary between the tree and its background. Trees with small sway displacements and low contrast boundaries tended to have noisier vibration signals with less obvious dominant sway frequencies. We also recommend choosing trees with significant portions of exposed trunk. Unlike sway signals from the crown, sway signals from the trunk-background boundary are straightforward to interpret and usually less noisy.

4.3.2 Deployment Recommendations

The capacity to extract meaningful sway frequency data from video data depends heavily on how the video data are collected. Simultaneously, field site constraints and the memory burden associated with video data impose considerable practical limitations. Here, we provide guidelines for collecting video data such that the memory overhead is reduced and desired signals processing properties are preserved.

The Nyquist-Shannon sampling theorem dictates that the maximum resolvable frequency in a discrete time series is equal to half of the sampling rate. Therefore, we recommend choosing the smallest available video frame rate greater than twice the highest expected sway frequency, plus some padding. Most trees vibrate at low frequencies, usually well below 5 Hz. Hence, when possible, we advise choosing a framerate around 8-10 Hz.

The frequency resolution of a Discrete Fourier Transform (DFT), which underlies the periodogram, indicates how close two peaks in the spectrum can be such that both are resolved. When the frequency resolution is too large, the energy from adjacent peaks can be contained within the same frequency bin, yielding a single peak. We can thus interpret the frequency resolution as the frequency bin width or a measure of the size of frequency changes that can be detected by the DFT. The frequency resolution of the DFT is equal to the sampling frequency divided by the number of samples in the time series. In our case, the latter is simply the sampling frequency multiplied by the duration of the signal, so the frequency resolution is equal to the inverse of the duration of the video in seconds. Thus, if we want to reliably detect 0.01 Hz changes, which is necessary for many tree sway applications, we need to have a 100 second (or longer) video. To enhance the statistical quality of the vibration spectra by, e.g., applying Welch's method, the length of the signal would need to be multiple times larger (1800 seconds is a good target). However, for studies requiring fine temporal resolution over long periods of time, this would result in hundreds of gigabytes of video data. We therefore recommend computing the necessary video length for the desired frequency resolution and recording videos for as long as

memory/data collection resources allow. Long video records should be preferred to more frequent data collection.

The frame resolution, lens type, and distance between the camera and target trees control the size of the observed sway and corresponding signal strengths. We suggest placing the camera as close to the target trees as possible and using the largest resolution that memory constraints allow. Presently, it is unclear how the lens FOV angle and corresponding distortion of motion affects sway measurements. We therefore encourage avoiding wide angle lenses. For best results, we also recommend augmenting the field site with complementary sensors. We suggest equipping the feature the camera is mounted to with an accelerometer so camera motion can be quantified and, if necessary, removed from vibration spectra (Chen 2016). Stationary objects in the frame can also be used for similar purposes. Finally, wind speed data collected near the site may be useful for identifying periods when there are likely to be strong sway signals.

4.4 Directions for Future Work

Our work validates two video processing methods using a small collection of videos. Further analysis, particularly of video time series over long periods (days, weeks, months), is needed to clarify when and where video processing can infer tree properties. Future work should begin by comparing sway frequency time series extracted from video and accelerometer data when there are expected large changes in sway frequency.

Numerous opportunities exist for improving the video processing algorithms. The phase based optical flow video vibration analysis technique (Chen et al. 2017, Wadhwa et al. 2013) has shown promise measuring both the frequency and magnitude of mechanical vibrations. The authors demonstrate the phase-based approach enables robust analysis of signals with small displacements, making it well suited for tree sway analysis. Additionally, future work could experiment with different approaches for estimating the periodicity of the vibration signals. For improved estimation in the frequency domain, the multitaper method could be used. Furthermore, time domain methods for computing periodicity, such as autocorrelation, may improve the frequency resolution and hence bypass some of the limitations of PSD estimation.

Experimentation with cameras recording different wavelengths could also be studied. Videos recorded with a near infrared camera may extend video data collection into hours of darkness and provide desirable high contrast ROIs through the use of the Normalized Difference Vegetation Index (NDVI).

5. Conclusions

We demonstrate that the virtual vision sensor (VVS) and multilevel binary thresholding (MBT) Eulerian video processing algorithms can measure tree sway frequency to within 0.03 Hz of accelerometer values at multiple points in a camera field of view when appropriate processing parameters are chosen. Our analysis suggests that the technology is suitable for detecting changes in sway frequency greater than 0.03 Hz, which is sufficient to resolve changes due to

snow interception and phenology variations but not diurnal drought stress, as based on currently published values (Ciruzzi and Loheide II 2019, Jaeger et al. 2022, Raleigh et al. 2022). When high temporal resolution and highly accurate measurements are required, we advise using well-placed accelerometers. While the video processing methods show promise for improving the spatial resolution of tree sway analysis, challenges remain when scaling to a large number of videos and trees. Further work examining when and where video processing can infer tree properties is needed.

Acknowledgements

We gratefully acknowledge funding support from NASA Grant NNX17AL59G.

This work received invaluable contributions from collaborators and peers. We would like to thank Ethan Gutmann (National Center for Atmospheric Research) for spearheading the initial development of the tree sway codebase, contributing keen insights throughout the development and testing phases, and offering additional computational resources. We would like to express our gratitude to Dominick Ciruzzi (University of Wisconsin-Madison/ William and Mary) and Steven Loheide II (University of Wisconsin-Madison), Sidney Bush and Holly Barnard (University of Colorado), and Mark Raleigh (National Center for Atmospheric Research/Oregon State University) who graciously shared the video and accelerometer data that made our analysis possible. We would also like to thank Sanna Sevanto (Los Alamos National Lab) and John Stimberis (Washington Department of Transportation), who kindly allowed us to film trees subjected to drought stress and snow, respectively, on their premises during early data collection efforts. We also thank all past and current members of the Mountain Hydrology Research Group. Finally, the author would like to give sincere thanks to Jessica Lundquist, his advisor, for her guidance, expertise, and support.

Code and Data

All code used for analysis is published on GitHub in the `swayfreq` repository: https://github.com/j-amma/swayfreq.

To access the data, contact the author at jamma@uw.edu.

References

Baker, C. J. (1997). Measurements of the natural frequencies of trees. Journal of Experimental Botany, 48(5), 1125–1132. https://doi.org/10.1093/jxb/48.5.1125

Barbacci, A., Diener, J., Hémon, P., Adam, B., Donès, N., Reveret, L., & Moulia, B. (2014). A robust videogrametric method for the velocimetry of wind-induced motion in trees. Agricultural and Forest Meteorology, 184, 220–229. https://doi.org/10.1016/j.agrformet.2013.10.003

Bunce, A., Volin, J. C., Miller, D. R., Parent, J., & Rudnicki, M. (2019). Determinants of tree sway frequency in temperate deciduous forests of the Northeast United States. Agricultural and Forest Meteorology, 266–267, 87–96. https://doi.org/10.1016/j.agrformet.2018.11.020

Chen, J. G. (2016). Video camera-based vibration measurement of infrastructure [Thesis, Massachusetts Institute of Technology]. https://dspace.mit.edu/handle/1721.1/104119

Chen, J. G., Davis, A., Wadhwa, N., Durand, F., Freeman, W. T., & Büyüköztürk, O. (2017). Video Camera–Based Vibration Measurement for Civil Infrastructure Applications. Journal of Infrastructure Systems, 23(3), B4016013. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000348

Ciruzzi, D. M., & Loheide II, S. P. (2019). Monitoring Tree Sway as an Indicator of Water Stress. Geophysical Research Letters, 46(21), 12021–12029. https://doi.org/10.1029/2019GL084122

Diener, J., Reveret, L., & Fiume, E. (2006). Hierarchical Retargetting of 2D Motion Fields to the Animation of 3D Plant Models. The Eurographics Association. https://doi.org/10.2312/SCA/SCA06/187-194

Ferrer, B., Espinosa, J., Roig, A. B., Perez, J., & Mas, D. (2013). Vibration frequency measurement using a local multithreshold technique. Optics Express, 21(22), 26198–26208. https://doi.org/10.1364/OE.21.026198

Gougherty, A. V., Keller, S. R., Kruger, A., Stylinski, C. D., Elmore, A. J., & Fitzpatrick, M. C. (2018). Estimating tree phenology from high frequency tree movement data. Agricultural and Forest Meteorology, 263, 217–224. https://doi.org/10.1016/j.agrformet.2018.08.020

Granucci, D., Rudnicki, M., Hiscox, A., Miller, D., & Su, H.-B. (2013). Quantifying the effects of freezing on tree sway frequencies. Agricultural and Forest Meteorology, 168, 10–14. https://doi.org/10.1016/j.agrformet.2012.07.016

Jackson, T. D., Sethi, S., Dellwik, E., Angelou, N., Bunce, A., van Emmerik, T., Duperat, M., Ruel, J.-C., Wellpott, A., Van Bloem, S., Achim, A., Kane, B., Ciruzzi, D. M., Loheide II, S. P., James, K., Burcham, D., Moore, J., Schindler, D., Kolbe, S., … Gardiner, B. (2020). The motion of trees in the wind: A data synthesis [Preprint]. Biodiversity and Ecosystem Function: Terrestrial. https://doi.org/10.5194/bg-2020-427

Jackson, T., Shenkin, A., Moore, J., Bunce, A., van Emmerik, T., Kane, B., Burcham, D., James, K., Selker, J., Calders, K., Origo, N., Disney, M., Burt, A., Wilkes, P., Raumonen, P., Gonzalez de Tanago Menaca, J., Lau, A., Herold, M., Goodman, R. C., … Malhi, Y. (2019). An architectural understanding of natural sway frequencies in trees. Journal of The Royal Society Interface, 16(155), 20190116. https://doi.org/10.1098/rsif.2019.0116

Jaeger, D. M., Looze, A. C. M., Raleigh, M. S., Miller, B. W., Friedman, J. M., & Wessman, C. A. (2022). From flowering to foliage: Accelerometers track tree sway to provide high-resolution insights into tree phenology. Agricultural and Forest Meteorology, 318, 108900. https://doi.org/10.1016/j.agrformet.2022.108900

Jeng, Y.-N., & Wu, C.-H. (2012). Frequency Identification of Vibration Signals Using Video Camera Image Data. Sensors, 12(10), Article 10. https://doi.org/10.3390/s121013871

Kooreman, B. D. (2013). Measuring weight fluctuations in trees based on natural frequency. https://repository.tudelft.nl/islandora/object/uuid%3Ab5c5466b-8165-4871-81a3-db8574cf1fd1

Kovacic, I., Radomirovic, D., Zukovic, M., Pavel, B., & Nikolic, M. (2018). Characterisation of tree vibrations based on the model of orthogonal oscillations. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-26726-5

Martin, K. A., Van Stan II, J. T., Dickerson-Lange, S. E., Lutz, J. A., Berman, J. W., Gersonde, R., & Lundquist, J. D. (2013). Development and testing of a snow interceptometer to quantify canopy water storage and interception processes in the rain/snow transition zone of the North Cascades, Washington, USA. Water Resources Research, 49(6), 3243–3256. https://doi.org/10.1002/wrcr.20271

Moore, J. R., & Maguire, D. A. (2004). Natural sway frequencies and damping ratios of trees: Concepts, review and synthesis of previous studies. Trees, 18(2), 195–203. https://doi.org/10.1007/s00468-003-0295-6

Nakata, M. T., Takahara, M., Sakamoto, S., Yoshida, K., & Mitsuda, N. (2018). High-Throughput Analysis of Arabidopsis Stem Vibrations to Identify Mutants With Altered Mechanical Properties. Frontiers in Plant Science, 9. https://www.frontiersin.org/articles/10.3389/fpls.2018.00780

Oda, M., Hagiwara, Y., Suzuki, S., Nakamura, T., Inaba, N., Sawada, H., Yoshii, M., & Goto, N. (2011). Measurement of Satellite Solar Array Panel Vibrations Caused by Thermal Snap and Gas Jet Thruster Firing. In N. Baddour (Ed.), Recent Advances in Vibrations Analysis. InTech. https://doi.org/10.5772/22034

Peltola, H. (1996). Swaying of trees in response to wind and thinning in a stand of Scots pine. Boundary-Layer Meteorology, 77(3), 285–304. https://doi.org/10.1007/BF00123529

Poh, M.-Z., McDuff, D. J., & Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Optics Express, 18(10), 10762–10774. https://doi.org/10.1364/OE.18.010762

Raleigh, M. S., Gutmann, E. D., Van Stan II, J. T., Burns, S. P., Blanken, P. D., & Small, E. E. (2022). Challenges and Capabilities in Estimating Snow Mass Intercepted in Conifer

Canopies With Tree Sway Monitoring. Water Resources Research, 58(3), e2021WR030972. https://doi.org/10.1029/2021WR030972

Rudnicki, M., Meyer, T. H., Lieffers, V. J., Silins, U., & Webb, V. A. (2008). The periodic motion of lodgepole pine trees as affected by collisions with neighbors. Trees, 22(4), 475–482. https://doi.org/10.1007/s00468-007-0207-2

Schumacher, T., & Shariati, A. (2013). Monitoring of Structures and Mechanical Systems Using Virtual Visual Sensors for Video Analysis: Fundamental Concept and Proof of Feasibility. Sensors, 13(12), Article 12. https://doi.org/10.3390/s131216551

Selker, J. S., Lane, J. W., Rupp, D. E., Hut, R., Abou Najm, M. R., Stewart, R. D., Van De Giesen, N., & Selker, F. (2011). The answer is blowing in the wind: Using wind induced resonance of trees to measure time varying canopy mass, including interception. 2011, H11G-1155.

Shang, Z., & Shen, Z. (2018). Multi-point vibration measurement and mode magnification of civil structures using video-based motion processing. Automation in Construction, 93, 231–240. https://doi.org/10.1016/j.autcon.2018.05.025

Storck, P., Lettenmaier, D. P., & Bolton, S. M. (2002). Measurement of snow interception and canopy effects on snow accumulation and melt in a mountainous maritime climate, Oregon, United States. Water Resources Research, 38(11), 5-1-5–16. https://doi.org/10.1029/2002WR001281

Tadrist, L., Saudreau, M., Hémon, P., Amandolese, X., Marquier, A., Leclercq, T., & de Langre, E. (2018). Foliage motion under wind, from leaf flutter to branch buffeting. Journal of The Royal Society Interface, 15(142), 20180010. https://doi.org/10.1098/rsif.2018.0010

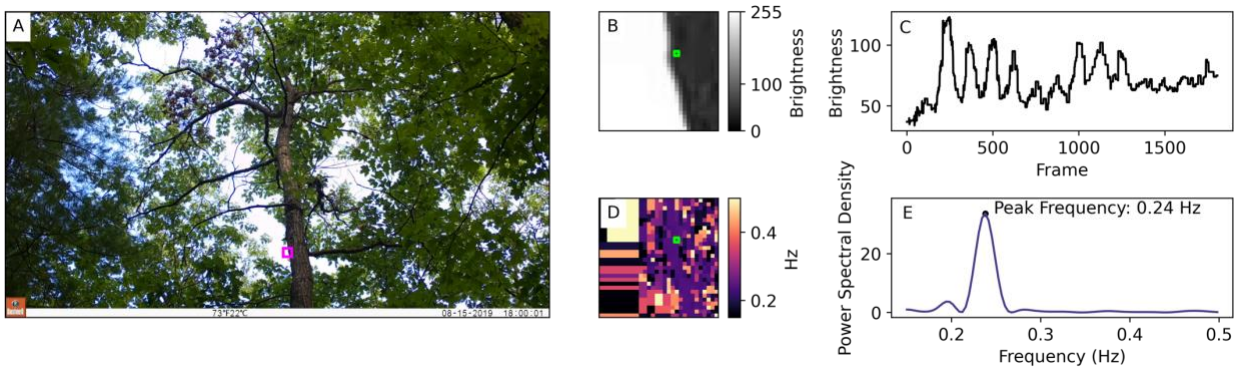
Torregrosa, A., Albert, F., Aleixos, N., Ortiz, C., & Blasco, J. (2014). Analysis of the detachment of citrus fruits by vibration using artificial vision. Biosystems Engineering, 119, 1–12. https://doi.org/10.1016/j.biosystemseng.2013.12.010

Van Emmerik, T., Steele-Dunne, S., Hut, R., Gentine, P., Guerin, M., Oliveira, R. S., Wagner, J., Selker, J., & Van de Giesen, N. (2017). Measuring Tree Properties and Responses Using Low-Cost Accelerometers. Sensors, 17(5), Article 5. https://doi.org/10.3390/s17051098

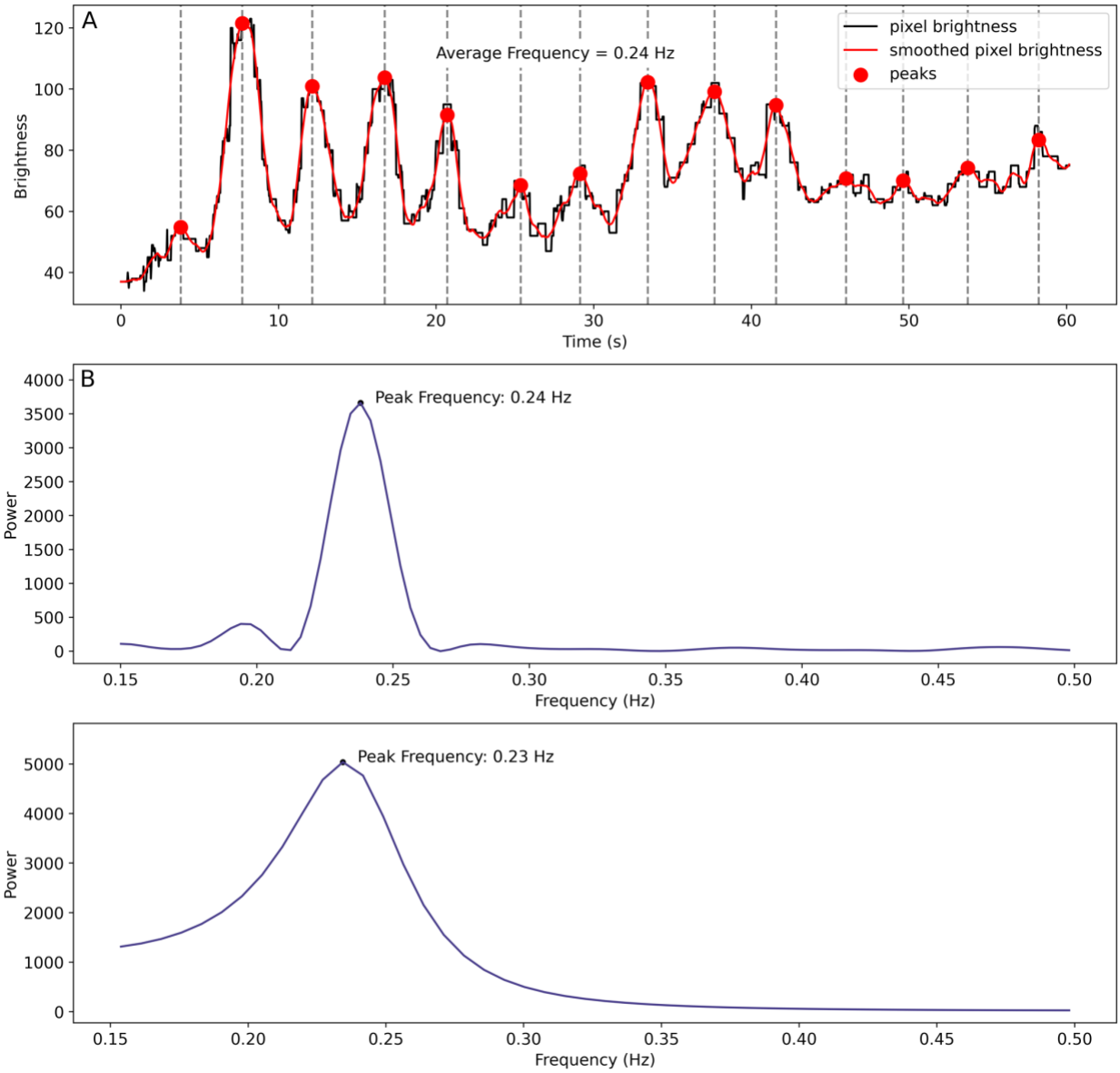
Wadhwa, N., Rubinstein, M., Durand, F., & Freeman, W. T. (2013). Phase-based video motion processing. ACM Transactions on Graphics, 32(4), 1–10. https://doi.org/10.1145/2461912.2461966

Wang, A., Yang, X., & Xin, D. (2022). The Tracking and Frequency Measurement of the Sway of Leafless Deciduous Trees by Adaptive Tracking Window Based on MOSSE. Forests, 13(1), Article 1. https://doi.org/10.3390/f13010081

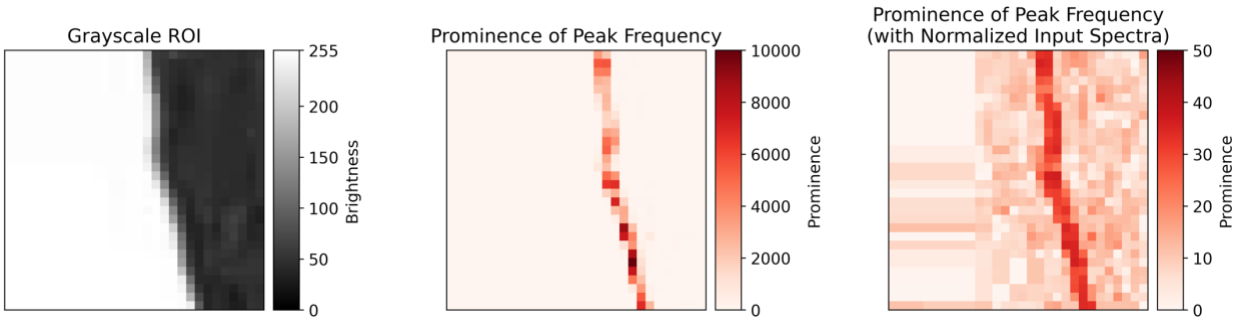
Appendix



Appendix A. Illustration of a pixel vibration signal in the time and frequency domains for the Trout Lake video and ROI. (A) Camera FOV and ROI (B) Grayscale ROI with highlighted pixel in lime (C) Pixel Brightness time series (D) Frequency heat map for ROI with highlighted pixel in lime (E) Power spectrum for pixel brightness time series.



Appendix B. Comparison of time and frequency domain methods for computing the peak frequency. (A) Time domain method: smooth vibration signal, find peaks, compute inverse of average distance between peaks (B) Nonparametric estimation - periodogram (rectangular window, nfft=8192) (C) Parametric estimation - autoregressive model using the Burg method (nfft=8192).



Appendix C. Effect of normalizing pixel spectra. (Left) Grayscale ROI (Middle) Prominence of the unnormalized pixel spectra frequency peaks (Right) Prominence of the normalized pixel spectra frequency peaks.