Uncertainty-Aware Relative Compliance Estimation for a Wind-Excited Plant Branch from Monocular Video

Abhimanyu Bhowmik, Jan Behrens and Robert Babuska

***Abstract*—Vision-based quantification of the mechanical properties of plant branches before manipulation can be valuable for automated horticulture tasks. Efficient manipulation cannot avoid collisions, and hard contact with plants can harm them or require overly cautious motion. Therefore, gaining awareness of a plant's mechanical compliance will be an asset for many downstream robotic plant-handling and monitoring tasks, such as flower picking or selective harvesting of fruits and vegetables. Estimating properties such as stiffness is fundamentally challenging because it requires force measurements, which are unreliable, expensive, and time-consuming to collect at scale. We propose using artificial wind as excitation to estimate relative compliance, which captures how much each branch segment bends compared to its neighbors. We propose a vision-based pipeline to track points on the branch, smooth their trajectories, and estimate per-segment compliance from the resulting bending signals. We estimate the uncertainty of the compliance metric by using a moving-block bootstrap, which quantifies the stability of the result under temporal resampling. This proof of concept is demonstrated using two wind-excited plant videos of 1104 and 590 frames, respectively. The proposed pipeline recovers a proximal-to-distal relative motion gradient and an interpretable relative compliance map. Our code and second experiment are reported in the supplementary material, which can be found at https://abhimanyubhowmik.github.io/rce/.**

I. Introduction

Estimating the mechanical flexibility of plant branches is a key capability for agricultural robots involved in selective harvesting and manipulation. For example, harvesting an orange often requires pushing aside leaves and smaller branches while avoiding stiffer plant parts. In practice, a robot must predict how a branch will deform under contact forces. However, acquiring this information at scale is difficult: contact-based stiffness measurement requires a physical probe. Therefore, wind force is a reasonable proxy that can be produced by an artificial source such as a blower or pressurized air. This paper aims to estimate the relative flexibility of branch segments under wind load using existing vision-based keypoint-tracking approaches.

Previous work on plant motion analysis has focused mainly on automated phenotyping tasks [1], [2], rather than recovering material properties from dynamic video. Davis [3] showed that passive video can be used to recover vibration signals and infer the material properties of visible objects. Still, their analysis does not focus on articulated objects such as plant branches. Tree biomechanics studies [4] identified that branch dynamics under wind are largely controlled by tree morphologies. Few prior works, such as [5], use 3D Gaussian splats and simulation environments to learn physical properties of 3D particles. However, these approaches are intended for physically accurate, photorealistic rendering of plant structures rather than understanding physical parameters for downstream robotic manipulation tasks.

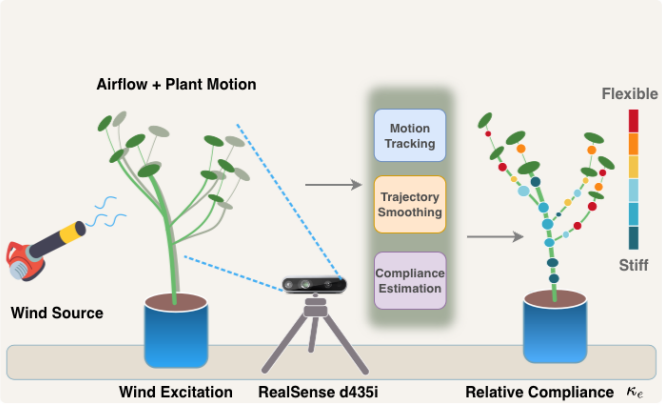


Fig. 1: The monocular video input of a wind-excited plant is processed through TAPIR-based [6] tracking and trajectory smoothing to compute the image-plane relative compliance of the plant stem at various nodes.

In this paper, we investigate whether it is feasible to identify the flexibility of plant branches from non-tactile excitation using existing image-processing and parameter-estimation methods (Fig. 1). Our proposed relative compliance proxy κ_e captures the image-plane-projected component of 3D mechanical compliance. Our factor-graph-inspired smoother encourages temporal smoothing and edge inextensibility while respecting TAPIR [6] tracking. Uncertainty estimates for κ_e are obtained via the moving-block bootstrap, which is appropriate for temporally correlated bending signals. Finally, our interpretable compliance gradient is registered to the initial input video frame, providing relative edge compliance values for downstream motion planning tasks.

II. Problem Formulation

We model each branch segment as an Euler-Bernoulli beam [8]. Let $s \in [0, L]$ denote the arc length along the branch segment, with $s = 0$ near the parent joint and $s = L$ at the distal tip. Let $v(s, t)$ be the transverse deflection at time t, and let $q(s, t)$ be the distributed transverse load induced by airflow.

A. Bhowmik, J. Behrens, and R. Babuska are with the Czech Institute of Informatics, Robotics and Cybernetics, Czech Technical University in Prague, Prague, Czech Republic.

R. Babuska is also with the Department of Cognitive Robotics, Faculty of Mechanical Engineering, Delft University of Technology, Delft, Netherlands.

This work was supported by the European Union under the project Robotics and advanced industrial production (reg. no. CZ.02.01.01/00/22_008/0004590)

Correspondence:bhowmikabhimanyu@gmail.com

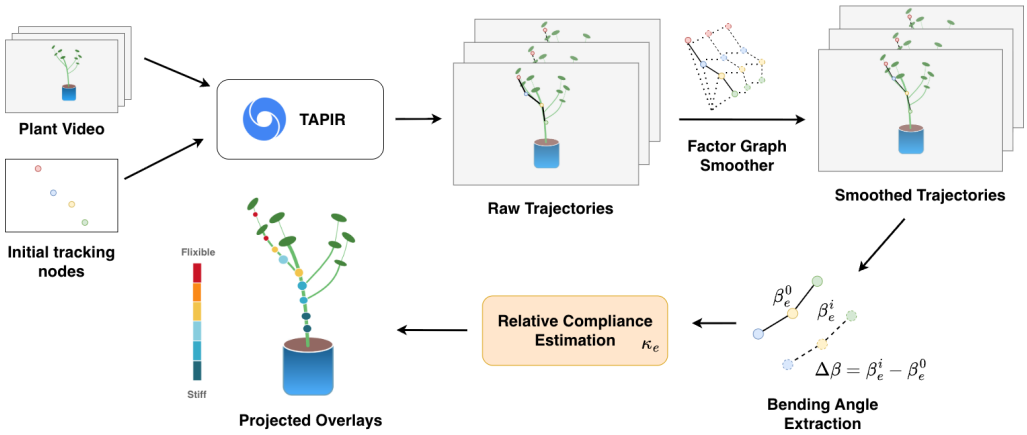


Fig. 2: Overall Methodology for projected (2D) relative compliance estimation. We start with a video of a plant branch and corresponding annotated tracked points, which are processed through TAPIR to generate the raw trajectories. We use a factor graph-style smoothing to recover smooth, noise-free trajectories from the moving average of the raw trajectories. Finally, we extract bending angles and estimate relative compliance using alternating least squares [7], and project the compliance heatmap onto the plant branch.

The standard Euler-Bernoulli equation [8] relates the applied load to the fourth spatial derivative of deflection. In its static form:

$$EI\frac{\partial^4 v(s,t)}{\partial s^4} = q(s,t), \tag{1}$$

where EI is the bending stiffness, with E the Young's modulus and I the second moment of area. We adopt this equation as a quasi-static approximation, treating each frame independently, which is reasonable when the excitation is well below the structure's natural frequency [4].

We express the equation in terms of the bending moment. By integrating twice and applying Cauchy's formula [9] for repeated integrals, we get

$$\frac{\partial^2 v(s,t)}{\partial s^2} = \frac{1}{EI}\int_s^L q(\xi,t)(\xi - s)\,\mathrm{d}\xi. \tag{2}$$

Thus, the observable curvature is proportional to the applied load and inversely proportional to the bending stiffness. This second-order integral form is equivalent to the standard fourth-order Euler-Bernoulli equation [8]. However, it explicitly shows how distal airflow loads generate bending moment through their lever arms.

A. Why absolute stiffness is not identifiable

From monocular video, we observe branch motion, which we use to estimate a discrete bending signal. However, the airflow load $q(s,t)$ is not measured. This creates a fundamental scale ambiguity. For any positive scalar c,

$$EI \mapsto c\,EI, \qquad q(s,t) \mapsto c\,q(s,t) \tag{3}$$

leaves (2) unchanged, as both the numerator and denominator are scaled by the same factor. Therefore, the same observed motion can be explained either by a stiff branch under a strong load or by a compliant branch under a weak load. So, without any knowledge about external wind load, absolute bending stiffness EI is not identifiable.

B. What is identifiable: projected relative compliance

Although absolute stiffness cannot be recovered, image-plane-projected relative compliance between branch segments can be estimated under a shared-excitation assumption, i.e., the whole branch experiences the wind force simultaneously at any given time instant. In our setting, the branch is excited by a common airflow source in a constant direction throughout the estimation period. Therefore, we can assume that a single scalar can describe the instantaneous airflow intensity experienced by the whole branch.

We model the load on edge e as separable into a temporal excitation amplitude and a spatial load profile:

$$q_e(\xi,t) \approx a_t\,\bar{q}_e(\xi), \tag{4}$$

where $a_t \geq 0$ captures the airflow intensity at frame t, and $\bar{q}_e(\xi)$ captures the spatial distribution of load along edge e. Substituting (4) into (2) gives

$$\frac{\partial^2 v_e(s,t)}{\partial s^2} = a_t\frac{1}{EI_e}\int_s^{L_e} \bar{q}_e(\xi)(\xi - s)\,\mathrm{d}\xi. \tag{5}$$

In the video, we do not directly measure the continuous curvature. Instead, for each edge e and frame t, we compute a bending observation $y_{e,t}$ discretized in both space and time. It is the change in local angle between tracked skeleton points. This quantity is a finite-dimensional proxy for the curvature term in (5). Applying this discrete measurement operation to (5) yields

$$y_{e,t} \approx a_t \underbrace{\left[\frac{1}{EI_e}C_e\right]}_{\text{edge-dependent compliance}}, \tag{6}$$

where C_e absorbs the edge-specific geometric and load-shape terms induced by the integral and the discrete angle measurement. We therefore define an effective relative compliance proxy of edge e as $\kappa_e \propto C_e/EI_e$ and define a shared frame-wise excitation proxy z_t proportional to the airflow amplitude a_t. This gives the rank-1 observation model

$$y_{e,t} \approx \kappa_e z_t. \tag{7}$$

Here, $y_{e,t}$ is the observed bending signal for edge e at frame t, $\kappa_e > 0$ estimates the relative projected compliance of that edge, and $z_t \geq 0$ is the shared excitation magnitude at frame t. Edges with $\kappa_e > 1$ bend more than average under the same excitation and are therefore relatively compliant. Edges with $\kappa_e < 1$ bend less than average and are therefore relatively stiff.

III. Methodology

The full pipeline has three stages as shown in Fig. 2: (1) trajectory smoothing via factor-graph optimization; (2) bending angle extraction with outlier rejection; (3) relative projected compliance estimation with bootstrap uncertainty.

A. TAPIR point tracking

We use TAPIR (Tracking Any Point with per-frame Initialization and temporal Refinement) by Doersch et al. [6], to track the movement of all annotated points or nodes on the plant stem under wind load. TAPIR uses a set of initialized query points on visible plant structures to predict their corresponding image locations in subsequent frames. It also estimates whether each point is visible or occluded. TAPIR uses a 2-stage framework: (1) A per-frame matching system independently proposes potential trackpoints at each timeframe. (2) In the next stage, it uses temporal context and local correspondence to refine the proposed trajectories. This approach makes TAPIR suitable for tracking points on deforming objects such as plant branches, where reliable frame-to-frame correspondence is required despite motion, appearance changes, and temporary occlusions. However, these raw TAPIR tracks exhibit very high noise along the branch axis due to disturbances such as pixel noise and lighting changes. This is expected, as the branch surface is visually uniform in color and texture along its length. On the other hand, tracking accuracy perpendicular to the branch is more reliable because the stem-background boundary creates a clear intensity contrast. This creates unwanted high variations in the edge lengths of the tracked points across frames. Since we are not processing the raw TAPIR tracks in real time, we can extract clean, consistent trajectories.

B. Factor-graph style trajectory smoother

After extracting the raw trajectories from TAPIR, we first apply a simple moving average and then use factor-graph style optimization [10] to reduce the noise and smooth the trajectories. Let $\boldsymbol{x}_{i,t} \in \mathbb{R}^2$ be the latent smoothed 2-D position of node i at frame t, and let $\tilde{\boldsymbol{m}}_{i,t}$ be the corresponding raw inputs. The branch is represented as a rooted tree with node set $\mathcal{V}$ and edge set $\mathcal{E}$. We minimize the following quadratic cost function

$$\begin{aligned}\mathcal{L}(\boldsymbol{X}) = \lambda_m \sum_{i,t} \|\boldsymbol{x}_{i,t} - \tilde{\boldsymbol{m}}_{i,t}\|^2 \\ + \lambda_v \sum_{i,t} \|\boldsymbol{x}_{i,t+1} - \boldsymbol{x}_{i,t}\|^2 \\ + \lambda_a \sum_{i,t} \|\boldsymbol{x}_{i,t+1} - 2\boldsymbol{x}_{i,t} + \boldsymbol{x}_{i,t-1}\|^2 \\ + \lambda_e \sum_{(p,c)\in\mathcal{E},t} \|(\boldsymbol{x}_{c,t} - \boldsymbol{x}_{p,t}) - \boldsymbol{r}_{pc}^0\|^2\end{aligned} \tag{8}$$

where $\boldsymbol{r}_{pc}^0$ is the initial edge vector (encoding approximate inextensibility and organic deformation of branch), and $\lambda_m, \lambda_v, \lambda_a, \lambda_e$ balance measurement fidelity against physical priors.

The first term, λ_m, is the measurement fidelity, which prevents the smoothed tracks from deviating too far from the actual visual data seen by TAPIR. The terms λ_v and λ_a account for velocity and acceleration smoothness, respectively. They penalize sudden changes in position or velocity and encourage a steady flow of motion. The last term λ_e maintains structural integrity, so that the branch segments (edges) preserve their original length even as they bend. It also acts like a regularization and preserves the organic shape of the branch by retaining the initial direction of the vector $\boldsymbol{r}_{pc}^0$ as well as its length. By framing this way, we tried to maintain physical consistency, as the method aims to smooth the motion of each point individually and also of the entire skeleton as a single connected object. Since Eq. (8) is a quadratic in $\boldsymbol{X}$, the global minimum is found by solving a sparse linear system, which is efficient even for $\sim$20 nodes over $\sim$1000 frames.

C. Bending angle extraction

For each ordered triplet (grandparent g, parent p, child c) with $e = (g, j, c)$, the instantaneous bending angle at frame t is:

$$\beta_{e,t} = \arccos\left(\frac{\boldsymbol{v}_{gp,t} \cdot \boldsymbol{v}_{pc,t}}{\|\boldsymbol{v}_{gp,t}\| \, \|\boldsymbol{v}_{pc,t}\|}\right) \tag{9}$$

where $\boldsymbol{v}_{gp,t} = \boldsymbol{x}_{g,t} - \boldsymbol{x}_{p,t}$ and $\boldsymbol{v}_{pc,t} = \boldsymbol{x}_{c,t} - \boldsymbol{x}_{p,t}$. The initial angle β_e^0 is taken from the first frame without any excitation. The signed deviation $\Delta\beta_{e,t} = \beta_{e,t} - \beta_e^0$ is then computed for each frame. The outliers are rejected via a median absolute deviation (MAD) mask [11]. The bending signal described in our rank 1 observation model (7) is $y_{e,t}$, which is infinitesimal in nature. Here, we use the $|\Delta\beta_{e,t}|$ as a finite and discrete proxy of the observation signal. In other words, the signal entering the compliance model is $y_{e,t} = |\Delta\beta_{e,t}|$ for all inlier frames.

D. Relative compliance estimation

1) Alternating least squares: Given bending signals $y_{e,t}$, we estimate the image-plane projected relative compliance (κ_e) and the wind-excitation magnitude at time t (z_t) by alternating:

Initialisation: $z_t^{(0)} = \mathrm{median}_e\, y_{e,t}$.

Update κ_e given $\boldsymbol{z}$: The raw least-squares slope is

$$\hat{\kappa}_e^{\text{raw}} = \frac{\sum_t z_t \, y_{e,t}}{\sum_t z_t^2} \tag{10}$$

However, using only alternating least squares [7] has a major limitation. In the video, some parts of the branch may be occluded by leaves for extended periods. If an edge is only seen for a few frames, a standard least-squares ($\hat{\kappa}_e^{raw}$) might produce a noisy estimate because it doesn't have enough evidence to form a stable estimate. We use shrinkage toward neutral as a regularization method. Where neutral compliance is $\kappa = 1$, i.e., a segment with $\kappa_e = 1$ bends exactly as much as the branch average under a given wind load. In practice, we implement shrinkage toward neutral in log space:

$$\kappa_e = \exp\!\big(\alpha_e \log \hat{\kappa}_e^{\text{raw}}\big), \quad \alpha_e = \frac{n_e}{n_e + \lambda} \tag{11}$$

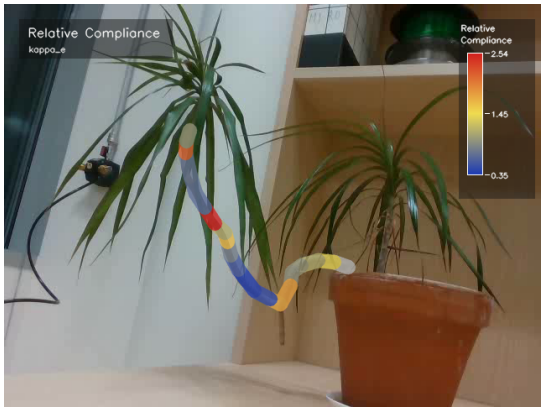
where n_e is the number of inlier frames for edge e and λ is a prior strength hyperparameter. Equation (11) implements a log-normal shrinkage prior: edges with few observations are pulled toward $\kappa_e = 1$ (neutral compliance).

Update z_t given $\boldsymbol{\kappa}$: $z_t = \big(\sum_e \kappa_e \, y_{e,t}\big) / \big(\sum_e \kappa_e^2\big)$.

After each iteration, the geometric mean of $\{\kappa_e\}$ is renormalized to unity to resolve scale ambiguity.

2) Uncertainty: moving-block bootstrap.: Because the bending time series are temporally autocorrelated, the standard iid bootstrap underestimates uncertainty. To avoid it, we use a moving-block bootstrap [12]. Instead of picking single frames, it picks contiguous blocks of time B, which are resampled with replacement to form synthetic sequences of length T. Then the full compliance estimation model is refitted on each. The 95% confidence interval for κ_e is the 2.5th-97.5th percentile of the bootstrap distribution $\{\kappa_e^{(b)}\}$.

The relative Confidence Interval width $U_e = (CI_{\text{hi},e} - CI_{\text{lo},e})/\kappa_e$ serves as a scale-normalised uncertainty index. A value of $U_e < 0.2$ indicates well-constrained estimates; $U_e > 0.5$ flags unreliable edges.



(a)

(b)

Fig. 3: Projection of estimated κ_e and corresponding uncertainty as a heatmap in the Image frame for the 1st plant. Figure (a) shows the relative projected compliance of the stem, where blue indicates the stiff sections and red indicates the more compliant sections. In Figure (b), the second image, the relative uncertainty of the stem indicates the measurement uncertainty of κ_e. Here, green indicates greater confidence in the compliance estimation, and red indicates greater uncertainty.

IV. Experiments and Results

A. Experimental setup

A potted plant with long branches was placed in front of an Intel RealSense D435i camera on a tripod. The plant is excited by a constant-velocity air source. The camera was mounted in a fixed position to avoid background motion, and we used only the monocular RGB video stream as input. Additionally, we collected data from a second plant using the same setup. Results for that experiment are reported in the supplementary material. The initial anchor points on the branch were manually annotated in the first frame (out of convenience; this step is relatively straightforward to automate, too). The movement of these points under air excitation is tracked through all frames using TAPIR [6]. For the main experiment, the branch was represented as a chain of 20 nodes with 19 edges, which remained fixed throughout the video. Plant 2 is annotated as a chain of 25 nodes / 24 edges using the same procedure. The tree is traversed from the root to the tip node, with the root distance increasing monotonically from edge 1 to edge 19. The edges are assigned to three anatomical zones based on their distance from the root: proximal (the first 0-5 edges), mid (edges 6-12), and distal(last 7 edges (edges 13-19)).

The raw trajectories produced by TAPIR are first processed with a moving-average filter of window size 5. They are then smoothed via factor-graph style optimization with the following hyperparameters: $\lambda_m = 1.0$, $\lambda_v = 0.35$, $\lambda_a = 0.2$, $\lambda_e = 0.45$. Next, we estimated the relative compliance with Alternating Least Squares [7] and Uncertainty with the moving-block bootstrap with the following hyperparameters: Prior strength $\lambda = 12$; block bootstrap with block size $B = 25$ frames; 250 bootstrap repetitions. The same hyperparameters are used for both plants.

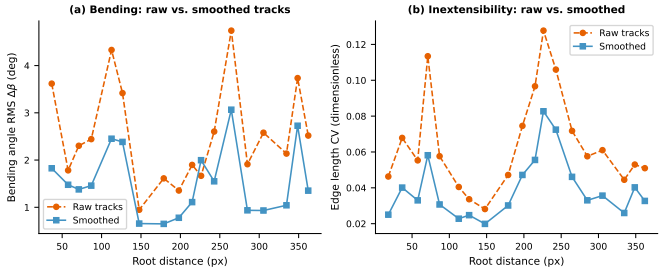


Fig. 4: Effect of factor-graph smoothing on the first plant. Left: bending angle RMS across the branch. Right: edge-length coefficient of variation. Smoothed tracks (blue) show ∼39% lower noise than raw TAPIR tracks with just a moving average filter (orange) while preserving the motion signal.

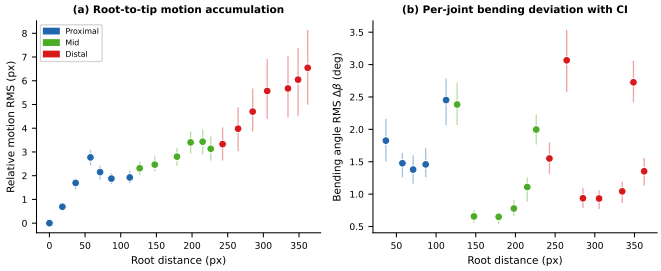


Fig. 5: Branch motion analysis for plant 1. (a) Node relative motion RMS as a function of root distance, showing monotonic amplification toward tips. (b) Per-joint bending angle RMS $\Delta\beta$ with 95% Confidence Intervals (vertical bars).

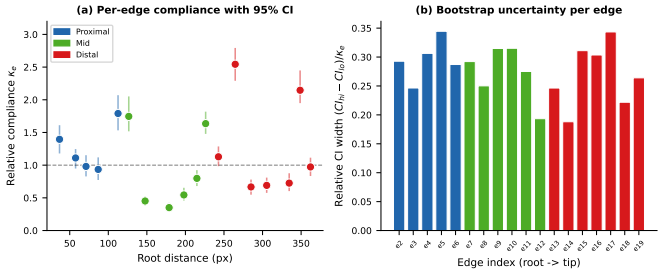


Fig. 6: (a) κ_e per edge as a function of root distance; vertical bars are 95% moving-block bootstrap CIs; dashed line marks neutral compliance ($\kappa = 1$). Colors indicate anatomical zone. (b) Bootstrap uncertainty (relative CI width) per edge; edges with poor support show higher uncertainty.

B. Impact of smoothing on noise reduction

The factor-graph based smoother substantially reduces the track's geometric noise while preserving the bending signal (Fig. 4). We used the Root Mean Square (RMS) of the bending signal ($y_{e,t}$), as it effectively accounts for both positive and negative bending of the stem from its resting position across the entire video and reduces it to a single intensity score. For all the bends (leaving the root node), the mean bending RMS decreases from 2.53° (raw moving average) to 1.54° (smoothed) for plant 1, which is a 39.1% reduction. For plant 2 experiment RMS decreases from 4.42 ° to 2.11°, a 52.2% reduction. We used the edge length coefficient of variation $CV = \frac{\sigma}{\mu}$ to measure the relative variations of the branch's edge length across the frames, where the σ is the standard deviation and μ is the mean of edge lengths. The CV decreases from 0.065 to 0.040 (38.4%) for plant 1 and 0.077 to 0.044 (42.8%) for plant 2 after smoothing, confirming that smoothed tracks better respect the inextensibility prior. The reduction is consistent across all zones.

C. Motion amplification along the branch

Since we modeled our branch segments as an Euler-Bernoulli Beam [8], the motion is always monotonically increasing toward the tip because of the lever arm effect. Fig. 5(a) shows node relative motion RMS increasing almost monotonically from the root to the distal tip. This root-to-tip amplification is consistent with the cantilever beam prediction. On the other hand, Fig. 5(b) shows bending angle deviations per ordered triplets. Several bends show elevated RMS ($> 2°$) while others are near zero, reflecting heterogeneous compliance along the branch and not merely as a function of distance from the root.

D. Relative projected compliance map

Fig. 6(a) shows κ_e with 95% bootstrap confidence intervals (CIs) for plant 1. The root edge (edge 1) is not considered because it has zero compliance and can only be a grandparent node in any ordered triplet. The highest compliance is observed at edge 14 (distal zone) with $\kappa_e = 2.54$ and 3.07° deviation angle as shown in the plot 6(a) and 5(b) respectively. Other highly compliant edges are edge 6 (proximal zone), edges 7 and 12 (mid zones), and edge 18 (distal zone). This observation anatomically corresponds to Image 3 (a), where these compliant edges are marked with red, yellow, and orange overlays. However, most edges are not as compliant, and we can clearly observe two distinct, relatively stiff branch sections in the mid and distal regions. It corresponds to edges (8-11) and (15-17) in Figure 6 and blue overlays in the figure 3 (a).

For robotic manipulation, this compliance map has direct practical utility. For example, if a harvesting robot needs to displace this branch, the contact point matters. Gripping above the most compliant sections (marked in red, orange, and yellow in Fig.3(a)) is preferable, because the highly compliant branch segments then lie between the contact point and the root. Under the applied manipulation force, sections with high relative compliance can act as springs, gradually absorbing and distributing the load rather than transmitting it rigidly through the stem. Gripping below these sections, by contrast, would bypass this natural compliance and risk applying concentrated stress to a stiffer part of the branch.

E. Uncertainty analysis

Fig. 6(b) shows the relative Confidence Interval width U_e for each edge for plant 1. The median is 0.29 (29%), indicating that the estimates are well constrained for most edges. However, plant 2 has a noticeably higher median at 0.43 (43%) due to

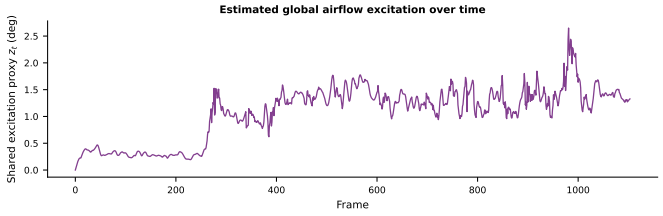


Fig. 7: Estimated shared excitation proxy for plant 1 z_t over entire video sequence, capturing periodic airflow from the wind source.

significantly fewer recorded frames (590). Furthermore, several mid-zone edges lose tracking for extended periods due to occlusion by leaves. Edges with $U_e > 0.4$ tend to be in the distal or mid zones. The distal region has fewer, shorter inlier frames due to repeated leaf occlusions. On the other hand, the mid zone exhibits low-motion signals, making it difficult to distinguish actual signals from pure noise. Here, we can also observe the importance of regularization: the shrinkage prior (11) pulls weakly supported edges toward $\kappa_e = 1$, and their bootstrap Confidence Intervals widen accordingly.

F. Shared excitation proxy

Fig. 7 shows the estimated shared excitation z_t over the full sequence. This result is physically consistent with the video observations: we observe that, for approximately 8 seconds, there was no wind, and the first wind gust occurred around the 264th frame. The signal also exhibits structured oscillations consistent with periodic wind gusts, with an RMS of $1.18°$. The temporal structure of the excitation z_t supports the shared-excitation assumption as well. If different edges had unrelated excitation patterns, the alternating optimization would not converge to a consistent z_t.

V. Conclusions and Future Work

We proposed a plausible solution to estimate the relative projected compliance between branch segments using a monocular video feed of a plant under wind load. Our approach doesn't require ground truth for the applied force, as its predictions are relative. The key insight is that the rank-1 observation model can provide a basis for estimating both the camera-plane-projected relative compliance and the shared temporal excitation for every edge. These estimations qualitatively agree with the plant's geometry and its video observation data. To constrain the plant's physical properties, we used a factor graph-style optimization. This reduced the geometric noise produced by TAPIR. We measured the uncertainty of our estimates using the moving-block bootstrap, yielding a median relative confidence interval of 29% for the main experiment. Finally, the method is applied over two annotated branch graphs to produce interpretable compliance maps. This work serves as a proof-of-concept for the careful harvesting of produce based on plant stiffness. The method can estimate stem compliance without touching it, enabling the robot to manipulate plants safely.

Limitations and future work

The results are based on two plants recorded with the same wind source and camera setup. Several species, wind sources, and sessions are required to fully confirm the stability of the relative compliance estimation. Furthermore, a crucial assumption in this paper is that the experience of the wind force is simultaneous throughout the branch at any time instant, i.e., wind force should be spatially coherent and quasi-static. So long branches, and strong/turbulent wind might break these assumptions. Moreover, future work should integrate automatic anchor-point detection from segmented plant branches using methods such as SAM2 [13]. This automated initialization will allow the estimation of bootstrap uncertainties across topological variants of the initial skeleton. Lastly, our model assumes that the movement induced by the wind excitation is planar and parallel to the image plane. For non-planar movements induced by wind load that are unobservable by a single monocular camera, we need multiview cameras and updated methods.

References

[1] N. Wagner and G. Cielniak, "Vision-based monitoring of the short-term dynamic behaviour of plants for automated phenotyping," in *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2023, pp. 624–633.

[2] E. de Langre, O. Penalver, P. Hémon, J.-M. Frachisse, M.-B. Bogeat-Triboulot, B. Niez, E. Badel, and B. Moulia, "Nondestructive and fast vibration phenotyping of plants," *Plant Phenomics*, vol. 2019, p. 6379693, 2019.

[3] M. Davis, "Visual vibration analysis," Ph.D. dissertation, 01 2016.

[4] K. James, G. Dahle, J. Grabosky, B. Kane, and A. Detter, "Tree biomechanics literature review: Dynamics," *Arboriculture Urban Forestry*, vol. 40, pp. 1–15, 01 2014.

[5] N. Wada, K. Yamashita, R. Kawahara, and K. Nishino, "M-phygs: Multi-material object dynamics from video," *arXiv preprint arXiv:2512.16885*, 2025.

[6] C. Doersch, Y. Yang, M. Vecerik, D. Gokay, A. Gupta, Y. Aytar, J. Carreira, and A. Zisserman, "TAPIR: Tracking any point with per-frame initialization and temporal refinement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 10 061–10 072.

[7] T. Hastie, R. Mazumder, J. D. Lee, and R. Zadeh, "Matrix completion and low-rank svd via fast alternating least squares," *The Journal of Machine Learning Research*, vol. 16, no. 1, pp. 3367–3402, 2015.

[8] J. M. Gere and B. J. Goodno, *Mechanics of Materials*, 7th ed. Stamford, CT: Cengage Learning, 2009.

[9] G. B. Folland, *Advanced Calculus*. Upper Saddle River, NJ: Prentice Hall, 2002.

[10] M. Xie, A. Escontrela, and F. Dellaert, "A factor-graph approach for optimization problems with dynamics constraints," *arXiv preprint arXiv:2011.06194*, 2020.

[11] P. Rousseeuw and C. Croux, "Alternatives to median absolute deviation," *Journal of the American Statistical Association*, vol. 88, pp. 1273 – 1283, 12 1993.

[12] P. Bühlmann and H. R. Künsch, "Block length selection in the bootstrap for time series," *Computational Statistics & Data Analysis*, vol. 31, no. 3, pp. 295–310, 1999.

[13] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson *et al.*, "SAM 2: Segment anything in images and videos," in *arXiv preprint arXiv:2408.00714*, 2024.