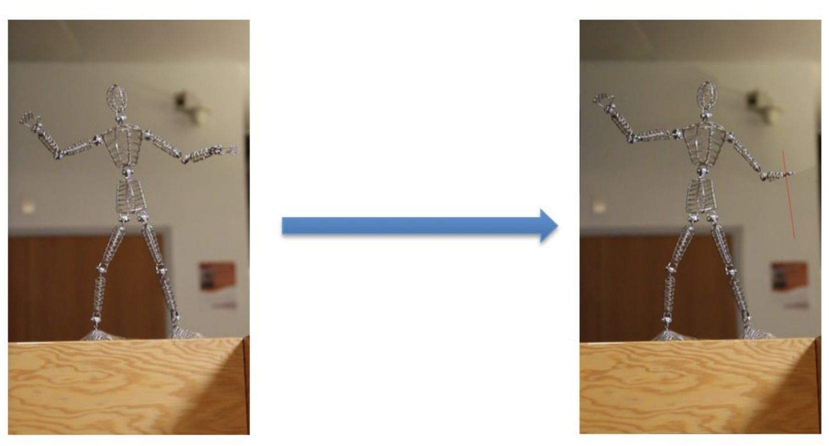
Interactive Video Synthesis using Vibrational Modal Analysis

Navtegh Singh Gill, Stuti Wadhwa, Sehajpreet Kaur
Department of Computer Science, University of Toronto

Motivation

- This project aims to learn vibration modes from videos and synthesize motion in response to unseen forces without any knowledge of scene geometry or material properties.
- We use this information to build image-space models of object dynamics around a rest state, letting us turn short video clips into physically-plausible, interactive animations.
- Recent research Berthy et al. [2] suggests that optical flow is not robust to small motion and that a phase-shift-based approach relying on Steerable Pyramids can be more effective for feature tracking.
- Hence, we comprehensively analyze the strengths and weaknesses of both methods. Additionally, we propose a novel approach that combines the two, offering a more generalized, robust, and accurate simulations of deformable objects.



Related Work

The concept of image-space modal bases for plausible manipulation of objects in video was first introduced by Davis et al. [1]. We implemented our approaches based on the paper's description, with the type of modal analysis being the distinguishing feature between them:

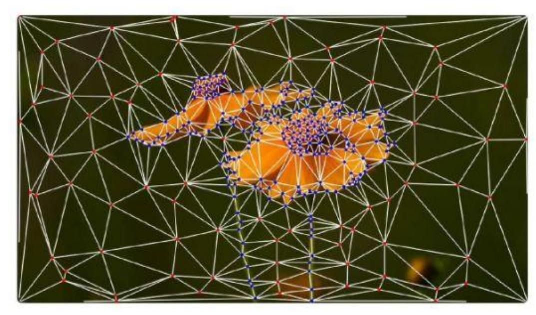
- **Lucas-Kanade Optical Flow [2]:**
  Works on local motion estimation for the given feature points but fails to capture motion of independent components.
- **Steerable Pyramids [3]:**
  Its ability to capture fine-grained motion details makes it effective for small-scale movements but causes it to struggle under high-force scenarios, resulting in excessive deformation or unrealistic outcomes.
- **RAFT Optical Flow [4]:**
  Deep learning based optical flow technique, producing similar results to L-K [2] but computationally expensive.

References

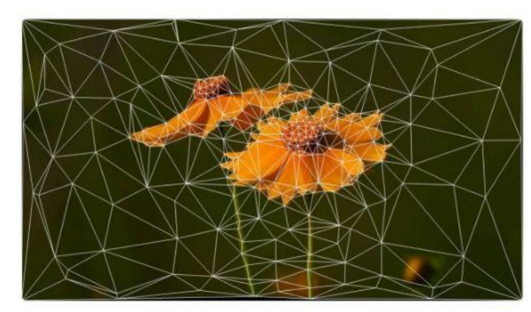
[1] A. Davis, J. G. Chen, and F. Durand, “Image-space modal bases for plausible manipulation of objects in video,” ACM Transactions on Graphics, vol. 34, no. 6, pp. 1–7, Nov. 2015
[2] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI), Vancouver, BC, Canada, 1981
[3] Feng, Berthy T., et al. "Visual vibration tomography: Estimating interior material properties from monocular video." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
[4] Z. Teed and J. Deng, “RAFT: Recurrent All-Pairs Field Transforms for Optical Flow,” Computer Vision – ECCV 2020, pp. 402–419, 2020

New Technique

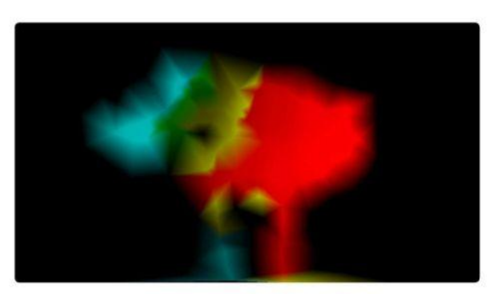
- **Key features** are **detected** in the video frames and **triangulation** is applied to establish spatial relationships between these points.
- **Feature tracking** is done using different motion analysis methods– Lucas-Kanade, Steerable Pyramids, RAFT and a combination of Lucas-Kanade and Steerable Pyramids**.**
  - We introduce this fourth technique which combines and normalizes the FFT results of Lucas-Kanade and Steerable Pyramids, allowing us to **capture fine-grained details** while preserving the **natural motion** of object when large forces are applied**.**
- Displacement data is **analyzed in the frequency domain** using Discrete Fourier Transform and users interactively select the **desired vibration modes.**
- These selected modes are loaded into the **simulation** environment, allowing users to interact with the object.



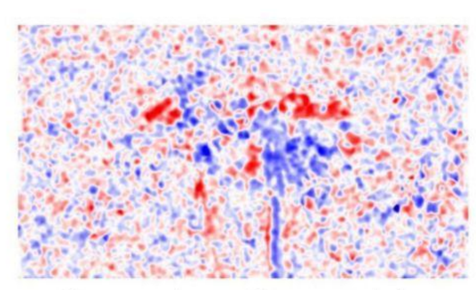
Feature Detection



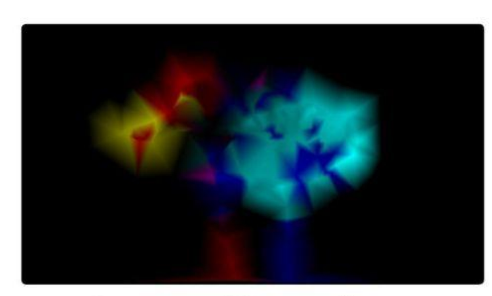
Feature Tracking



Lucas Kanade

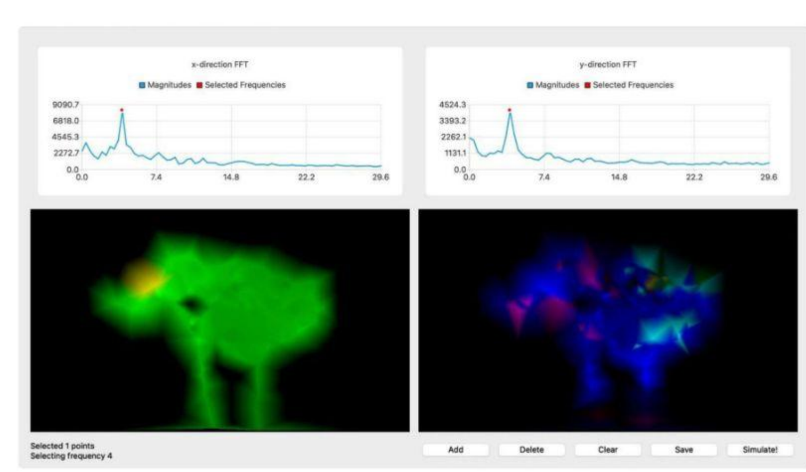


Complex Steerable pyramids



Lucas Kanade + Complex Steerable pyramids

RAFT



Mode Analysis and Selection

Experimental Results

Reference Frame	Lucas Kanade (LK)	Complex Steerable Pyramid (CSP)	LK + CSP	RAFT Model
				
				

Simulation results across various methods

	Lucas Kanade	Complex Steerable Pyramid	LK + CSP	RAFT
1st preference	6.7%	0%	93.3%	0%
2nd preference	66.7%	13.3%	6.7%	13.3%
3rd preference	26.7%	46.7%	0%	26.7%
4th preference	0%	40%	0%	60%

Preference distribution of 30 participants: Ranking methods based on performance in capturing object motion and fine details.