Topological Eulerian Synthesis of Slow Motion Periodic Videos[*][†]

Christopher J. Tralie

Duke University, Department of Mathematics

`ctralie@alumni.princeton.edu`

Matthew Berger

University of Arizona, Department of Computer Science

`matthew.sh.berger@gmail.com`

Abstract

We consider the problem of taking a video that is comprised of multiple periods of repetitive motion, and reordering the frames of the video into a single period, producing a detailed, single cycle video of motion. This problem is challenging, as such videos often contain noise, drift due to camera motion and from cycle to cycle, and irrelevant background motion/occlusions, and these factors can confound the relevant periodic motion we seek in the video. To address these issues in a simple and efficient manner, we introduce a tracking free Eulerian approach for synthesizing a single cycle of motion. Our approach is geometric: we treat each frame as a point in high-dimensional Euclidean space, and analyze the sliding window embedding formed by this sequence of points, which yields samples along a topological loop regardless of the type of periodic motion. We combine tools from topological data analysis and spectral geometric analysis to estimate the phase of each window, and we exploit the sliding window structure to robustly reorder frames. We show quantitative results that highlight the robustness of our technique to camera shake, noise, and occlusions, and qualitative results of single-cycle motion synthesis across a variety of scenarios.

1 Introduction

Repetitive, periodic behavior is ubiquitous in our world. Animal locomotion, mechanical motion, biological rhythms, and musical rhythms can all be characterized as periodic phenomena. In this work, we consider the problem of taking a video that contains such repetitive motion of many periods, and synthesizing a single fine detail, slow motion template cycle. This fine detailed analysis can be applied, for example, to characterize subtle progressions of blood flow in the face during a heartbeat [17]. Having access to a consensus template can also be used to visualize variations from cycle to cycle. This could be used to indicate the onset of failure in a repetitive automated action on an assembly line, to assess the stress of motions in repetitive human actions [13], and to optimize performance in athletic activities.

Our approach to slow motion templates is *Eulerian*; that is, we process the video pixel by pixel with no tracking. Eulerian approaches for video synthesis are attractive due to their simplicity and ease of implementation. For instance, a Fourier bandpass filter can successfully elucidate subtle periodic motions in videos [37, 36]. Time-domain Eulerian approaches can synthesize infinitely playing stochastic "video textures" [25], as well as seamlessly loop video templates from a single cycle of periodic motion [20, 19], though such approaches are not applicable to our problem since we fuse multiple cycles. There are some pitfalls with Eulerian techniques, however. They are known to fail for large motions, and also require the user to manually specify frequencies of interest [37, 36]. Furthermore, period-to-period drift due to a moving camera or small motion variations can be problematic, as can unrelated motion in the background or occlusions [26].

---

[*]Supplementary videos can be found at `http://www.ctralie.com/Research/SloMoLoops/`

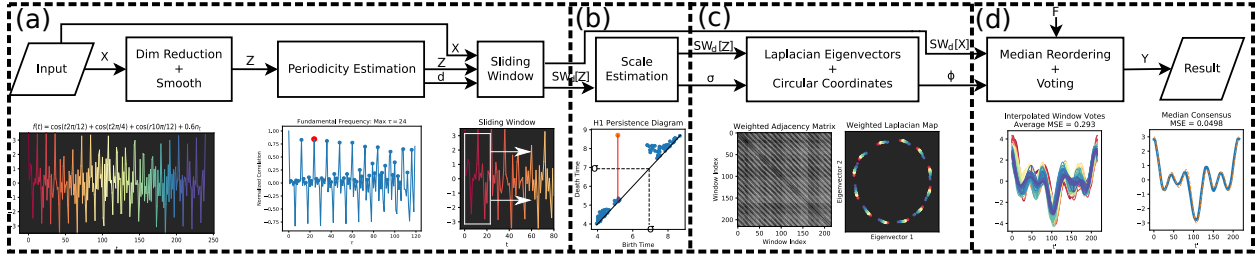[†]Code to replicate results can be found at `https://github.com/ctralie/SloMoLoops`

Figure 1: Overview of our technique for a 1D periodic time series with additive Gaussian noise ($n_t \in \mathcal{N}(0, 1)$). We first estimate the signal's period and extract a sliding window embedding (a), followed by estimating the spatial scale (b) via TDA, which is used to compute the phase ϕ via the graph Laplacian (c). The single-period motion template is then synthesized by having reordered windows vote on individual pixels (d). On the right we show the result superimposed with ground truth (in orange).

To address the problems with an Eulerian representation, we devise a geometric and topological Eulerian framework by constructing *sliding window embeddings* of videos. Sliding window embeddings, or delay reconstructions, have found a diverse array of applications in activity recognition [10, 34], video analysis [31], and motion capture analysis [34]. Sliding window embeddings of periodic time series, in particular, form samples of a topological loop [24], as do generalized multivariate sliding windows on periodic video data, regardless of the type of motion present [29, 31]. Furthermore, the sliding window embedding of an Eulerian video provides a form of regularization that mitigates the effects of drift [31]. Our geometric framework also allows us to use topological data analysis (TDA) [9, 7, 8, 4, 11] and fundamental frequency estimation [22] to autotune the spatial and temporal scales of the periodic motion, respectively, so that no user intervention is required. We use Laplacian Eigenmaps [2] computed on the sliding window embedding to obtain a circular phase $\phi \in [0, 2\pi]$ for each window. We then reorder the windows by ϕ, and we use windows that overlap in time vote on final pixels in each frame of the template, further mitigating drift. Figure 1 highlights the steps of our technique for a sampled 1D periodic time series. There are only 12 samples per period in the original signal, so the details of each period are coarse and noisy. However, once resorted, we get a fine-detailed representation of one period.

Many prior works use a combination of sliding window embeddings and TDA, for instance to detect chatter in mechanical systems [16] and quantify periodic activities in motion capture data [33, 35]. However, our approach in using a sliding window jointly with TDA to *synthesize* a new result is novel. Related works use spectral geometry to rearrange unstructured collections of images around a loop as a pre-processing step for structure from motion [1] or along a line to order the microscopic images of a developing embryo [6]. By contrast, we seek a cyclic reordering given input frames ordered in time, enabling us to construct a sliding window.

2 Slow Motion Templates

Figure 1 shows the different parts of our method for a 1D periodic signal for illustration purposes. The actual input to our method is a 3-channel video of N frames, each frame of resolution $W \times H$. We treat each frame as a point in Euclidean space of dimension $\mathbb{R}^{3 \cdot W \cdot H}$, and denote the set of all videos by the sequence $\mathbf{X}(t) \in \mathbb{R}^{(3 \cdot W \cdot H)}$ for integer-valued $t \in [1, N]$. For steps (a-c), we do not rely on the *extrinsic* geometry of the set of points $\mathbf{X}(t)$, only its set of *pairwise distances*. Thus, we can project all of the points into a lower-dimensional space and apply the set of techniques to the projection, assuming that pairwise distances are preserved. Since $N \ll (3 \cdot W \cdot H)$, a distance-preserving projection into an $N - 1$-dimensional space is always possible. Denoting this projection by $\mathbf{P} \in \mathbb{R}^{(N-1) \times (3 \cdot W \cdot H)}$, our method for the steps in (a) operates on $\mathbf{Z}(t) = \mathbf{P}\mathbf{X}(t)$. We also use the third layer of a Gaussian pyramid on each frame in place of the raw frames before applying $\mathbf{P}$ to mitigate drift.
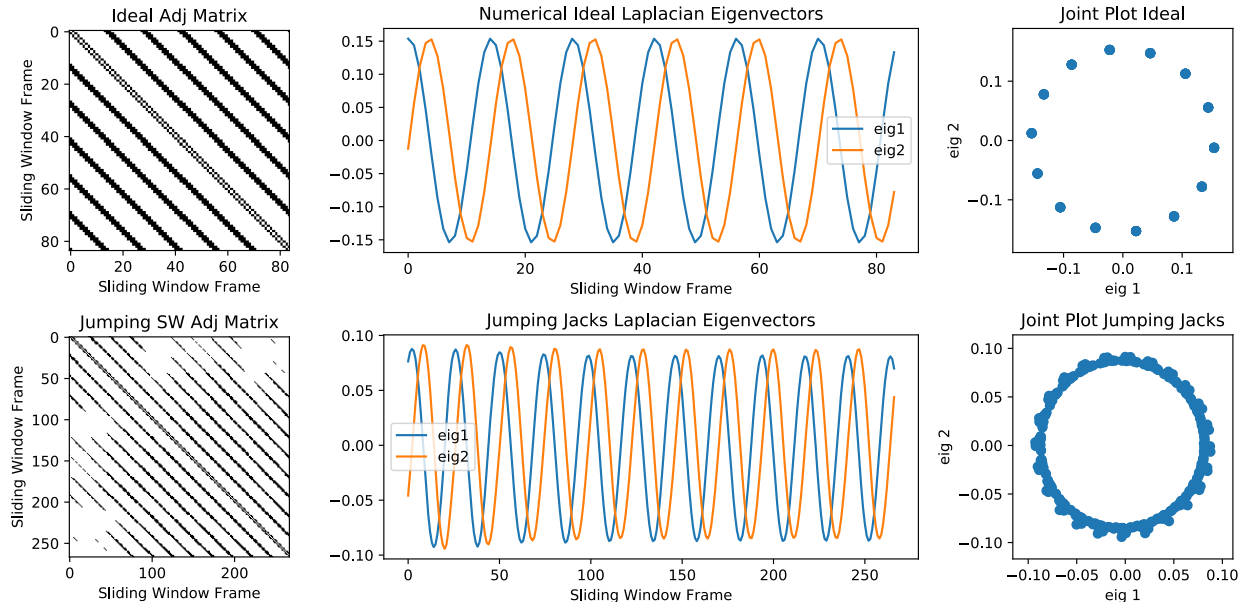
Figure 2: An example of the unweighted graph Laplacian on an ideal adjacency matrix corresponding to 6 periods of length 14 (top row), and the adjacency matrix from a sliding window embedding of two men doing jumping jacks (bottom row).

2.1 Sliding Window Embeddings

The sliding window video embedding [3, 29, 31] for $\mathbf{Z}$ is defined as

$$SW_d[\mathbf{Z}(t)] = \begin{bmatrix} \mathbf{Z}(t) \\ \mathbf{Z}(t+1) \\ \vdots \\ \mathbf{Z}(t+(d-1)) \end{bmatrix} \in \mathbb{R}^{(N-1)\cdot(d)}, \tag{1}$$

where d is the dimension of the embedding. As shown by Takens [27], a sliding window of dimension $2m + 1$ of even a single generic observation function of a dynamical system of intrinsic dimension m is sufficient to reconstruct a topological embedding of the underlying trajectory in the original state space. For periodic signals, this state space is a torus, and d should be twice the number of harmonics present to injectively reconstruct loops on that torus [24]. The same is true of videos in general [31]. Furthermore, the sliding window length $(d-1)$ maximizes the topological persistence (Section 2.3) of the embeddings when $(d-1) = kT$ for some integer k, where T is the fundamental period [24, 31]. To estimate T, and hence d, we perform 1D ISOMAP [28] on $\mathbf{Z}$ to generate a 1D surrogate signal, and perform an autocorrelation-based fundamental frequency estimation [22].

2.2 Laplacian Eigenmaps

Since sliding window embeddings of periodic videos lie on a topological loop, this means that, at an appropriate scale, a graph built on the sliding window point cloud will be approximately circular. Based on this, we use tools from spectral graph theory [5], in particular the *graph Laplacian*, to estimate the *phase* $\phi \in [0, 2\pi]$ of each sliding window. We first construct an adjacency matrix $\mathbf{A}$ in which the entry at $\mathbf{A}_{ij}$ represents the similarity between windows at times i and j. Given a scale σ, we define the unweighted adjacency as $\mathbf{A}_{ij} = 1$ if $||SW_d[\mathbf{Z}(i)] - SW_d[\mathbf{Z}(j)]||_2 \leq \sigma$ and 0 otherwise. We also define a weighted version in which $\mathbf{A}_{ij} = exp(-||SW_d[\mathbf{Z}(i)] - SW_d[\mathbf{Z}(j)]||_2/2\sigma^2)$. In both cases, we define the graph Laplacian as:

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \tag{2}$$

3

where $\mathbf{D}_{ii} = \sum_{j=1}^{N} \mathbf{A}_{ij}, \mathbf{D}_{i \neq j} = 0$ is the degree matrix representing the sum of all outgoing weights on the diagonal.

We now examine the unweighted Laplacian of an ideal model to motivate our approach. Suppose we have extracted a sliding window embedding SW_d for a video $\mathbf{Z}$ which repeats itself exactly every T frames for k times, for a total of $N = kT$ windows. Let us further suppose it is possible to choose a σ so that the unweighted adjacency matrix contains an entry of 1 for edges corresponding to windows that are adjacent in time, as well as for the corresponding windows at, before, and after repetitions of the window at intervals of T; that is, the Laplacian is a symmetric, *circulant* matrix defined as

$$L_{ij} = \begin{cases} 3k-1 & i = j \\ -1 & |i-j| = lT, l \in \mathbb{Z}^+ \\ -1 & |i-j| = lT \pm 1, l \in \mathbb{Z}^+ \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Circulant matrices are diagonalized by the Discrete Fourier Transform [12], and their nonzero eigenvalues come in pairs with multiplicity two, with corresponding eigenvectors $v_1[n] = 1$, $v_{2m}[n] = \cos(2\pi mn/N)$, $v_{2m+1}[n] = \sin(2\pi mn/N)$, $m \geq 1$. In the case of Equation 3, it can be shown using the DFT that the eigenvalues are $\lambda_1 = 0$ and

$$\begin{matrix} \lambda_{2m}, \\ \lambda_{2m+1} \end{matrix} = \begin{cases} 3k - k\left(1 + 2\cos\left(\frac{2\pi}{kT}m\right)\right) & m = lk, l \in \mathbb{Z}^+ \\ 3k & \text{otherwise} \end{cases} \tag{4}$$

The smallest two nonzero eigenvalues occur when $m = k$, corresponding to the eigenvectors [1]

$$v_{2k}[n] = \cos(2\pi n/T), v_{2k+1}[n] = \sin(2\pi n/T) \tag{5}$$

Therefore, the smallest two numerically nonzero eigenvalues each correspond to sinusoids with period T which are mutually orthogonal. When plotted against each other, they form a circle with an arbitrary phase offset. Therefore, we compute the circular phase numerically $\phi = \tan^{-1}(\hat{v}_1[n]/\hat{v}_2[n])$, where $\hat{v}_i$ is the eigenvector corresponding to the ith smallest numerically nonzero eigenvalue.

In practice, graphs of the sliding window embedding may deviate from the ideal Laplacian model in Equation 3, though they do so gracefully. Figure 2 shows an example of $\hat{v}_1[n]$ and $\hat{v}_2[n]$ for the jumping jacks video. To improve robustness, we default to the weighted Laplacian so that small changes in the threshold lead to small changes in $\mathbf{A}$. In this case, harmonics of the actual frequency of interest occasionally have smaller eigenvalues when the corresponding harmonics are strongly present in the video. To mitigate this, we search through the 10 eigenvectors corresponding to the smallest 10 eigenvalues, sorted by eigenvalue, and we use the pair of adjacent eigenvectors with the smallest number of zero crossings whose zero crossing counts are within a factor of 20% of each other.

2.3   Persistent Homology

It remains to find the spatial scale σ for the Laplacian. To adapt to the data, we leverage 1D persistent homology from TDA [8] to find the scale at which the primary topological feature – the single cycle – exists. Specifically, we compute the 1D Vietoris Rips Filtration on our sliding window data $\mathbf{Z}$, which tracks equivalence classes of loops, known as *homology classes* [14]. The algorithm returns a so-called persistence diagram, a multiset of points on a 2D birth/death grid, with each point corresponding to a loop class. The birth value b_i indicates the scale at which the ith loop class forms, and the death value d_i indicates the scale at which that class no longer exists. The difference $d_i - b_i$ is known as the *persistence* of the ith class. For our scenario, one point in the diagram should have a much larger persistence than the others (e.g. Figure 1) [24, 31], and this ideally reflects the single cycle of motion that we seek. We take our scale to be $\alpha b_i + (1-\alpha)d_i$ for the largest $d_i - b_i$, where $\alpha \in [0,1]$ is a parameter which will be explored experimentally in Section 3.

Finally, homology computation requires the specification of coefficients that belong to a user-given field [14, 21]. Coefficients in $\mathbb{Z}_2$ are commonly chosen, however in our scenario this is problematic. For instance, the motion of a jumping jack contains a second harmonic, since an individual jumps twice per cycle, and

---

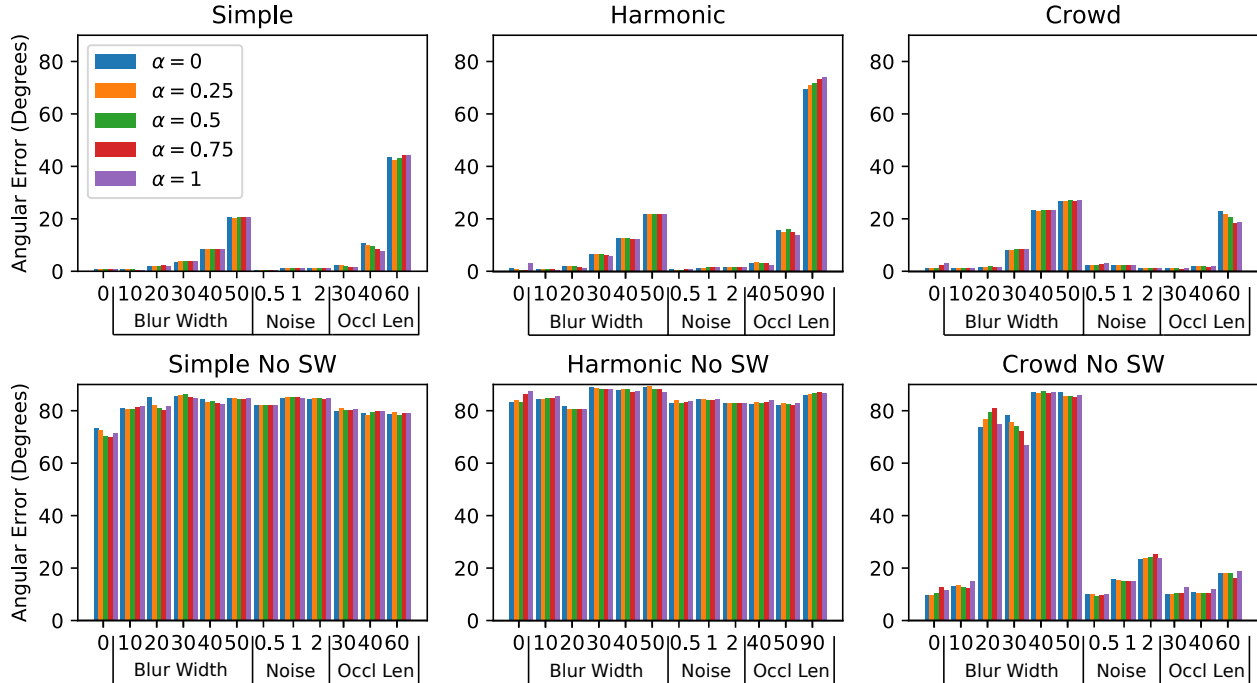[1] This generalizes the circle graph used in [1], in which $T = N, k = 1$

Figure 3: Average angular errors of circular coordinates under different sources of corruption on 3 different ground truth videos, with and without a sliding window.

using coefficients in $\mathbb{Z}_2$ would lead to so-called Möbius splitting [24, 32, 30]. Thus, we use coefficients in the field $\mathbb{Z}_{49}$ in order to capture these types of complex motions.

2.4 Cycle Reordering And Median Voting

Given the phase estimates ϕ_i, we can now synthesize the final slow motion template by lining up the N sliding windows $SW_d[\mathbf{X}(i)]$ by ϕ. For a template with M frames, we choose a set of M equally spaced angles $\theta_1, \theta_2, ..., \theta_M$ around the circle at which to sample the template. Let Φ_i be the unwrapped phase of ϕ_i. Based on this, we estimate the number of periods $\mathbf{X}$ goes through as $k = (N \times \max_i(\Phi_i)/(2\pi(N-d+1)))$. Assuming k is correct and that ϕ_i has a constant slope over i, there is a phase gap of $2\pi k/N$ between adjacent stacked frames in each sliding window. Thus, $SW_d[\mathbf{X}(i)]$ spans the interval $[\phi_i, \phi_i + 2\pi kd/N(\mathrm{mod}2\pi)]$. A given θ_j will potentially be contained in the intervals for many different windows, though it is unlikely to coincide exactly with any of the frames in each window, so we use linear interpolation to fill in a frame corresponding to θ_j in each window overlapping it. The frame at angle ϕ_j is taken as the median of all interpolated frames. Furthermore, we only use interpolated frames that are a user-prescribed amount F away in time for each given reordered frame, to avoid ghosting artifacts [23].

3 Experiments

3.1 Quantitative Tests of Circular Coordinates

We first experimentally quantify the accuracy of our circular coordinate inference, since we cannot hope to get a good slow motion template without accurate circular coordinates. We generate 3 different 600 frame synthetic periodic videos for which we know the ground truth circular coordinates, using software from [15][2]. Each video is roughly 300×300 pixels. We vary the number of cycles that the videos undergo between $3, 5, 10, 15, 20, 25, 30, 40, 50$. We then vary the "shake" of the video (width of a motion blur kernel) from

---

[2]Please see supplementary material for these videos and simulated errors.
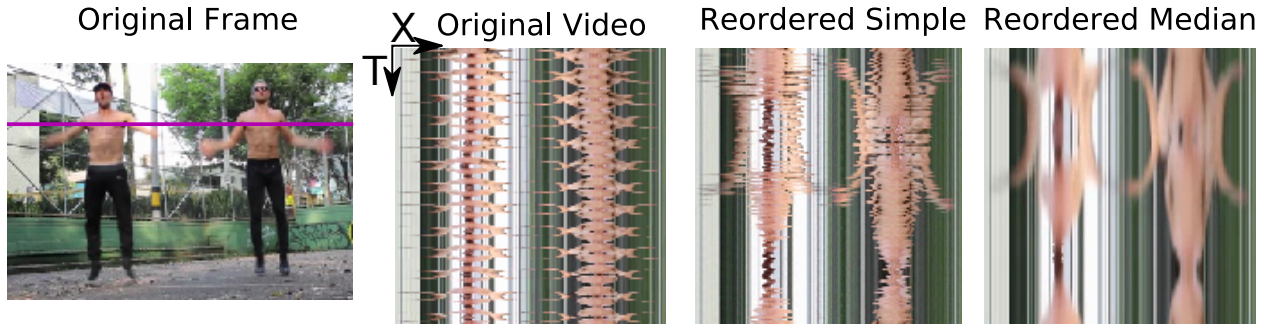
Figure 4: An XT slice of a line of pixels (magenta line, upper left) over time for an input video of men doing jumping jacks and for reordered videos with and without median consensus.
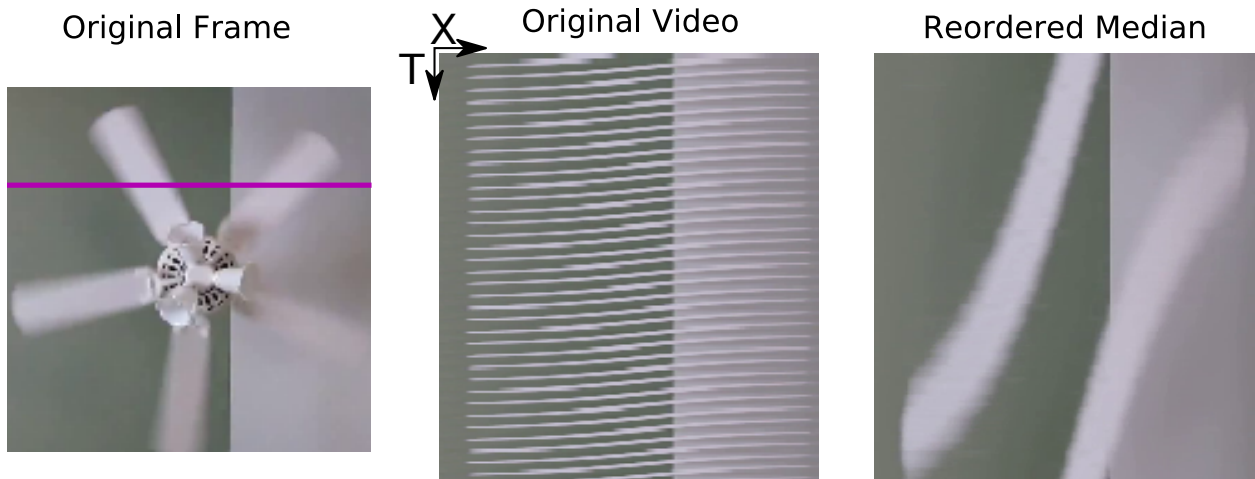

Figure 5: An XT slice of a line of pixels (magenta line, upper left) over time for an input video of a fan with only 6 frames per period and the corresponding reordered template.

10×10 to 50×50 pixels to assess the effect of drift. We then add Gaussian noise width standard deviations of $0.5, 1, 2$ (original RGB ranges are in $[0, 1]$. Finally, we add color drifting, occluding squares of varying lengths taking a random walk to assess the effect of occlusions / background motion. We also compare the α which weights the birth time and death time from TDA. Figure 3 shows the average angular error in degrees for our pipeline under these variations, over 50 trials per condition. Overall, performance is stable to the choice of α, and the errors are low for severe noise and for moderate shake and occlusions with a sliding window. Without a sliding window, the only video that performs reasonably is the "crowd" video, though the errors increase more rapidly with shake/noise/occlusion than with the sliding window, validating the "time regularization" aspect of sliding windows.

3.2 Qualitative Video Template Results

We now qualitatively examine the results of our slow motion templates on some examples. Figure 4 shows the difference between a simple reordering and a median consensus reordering. Due to natural variation from cycle to cycle, the simple reordering has many temporal discontinuities when interleaving these cycles. By contrast, the median voting is clean, and it has the added benefit of removing nonperiodic background components. Figure 5 shows an extreme example in which an original spinning fan video has only 6 frames per period at framerate. Please refer to our supplementary materials for these videos, as well as an exercise video [18] and videos of amplified blood flow in the neck [36] and face [37].

4   Conclusions

We have presented an approach that combines topological data analysis with spectral geometric analysis to reorder a video consisting of repetitive periodic motion into a single, slow motion template cycle. Our quantitative results demonstrate robustness to noise, drift, and background outliers, and our qualitative results reveal motion that is challenging to visually perceive from the raw input video. For future work, we plan on exploring applications of our technique to detecting motion irregularities, visualizing subtle motions from repetitive motion that is temporally aliased, and reconstructing templates for videos with large amounts of missing data.

References

[1] Hadar Averbuch-Elor and Daniel Cohen-Or. Ringit: Ring-ordering casual photos of a temporal event. *ACM Trans. Graph.*, 34(3):33, 2015.

[2] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[3] Liangyue Cao, Alistair Mees, and Kevin Judd. Dynamics from multivariate time series. *Physica D: Nonlinear Phenomena*, 121(1-2):75–88, 1998.

[4] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009.

[5] Fan RK Chung. *Spectral graph theory*. Number 92. American Mathematical Soc., 1997.

[6] Carmeline J. Dsilva, Bomyi Lim, Hang Lu, Amit Singer, Ioannis G Kevrekidis, and Stanislav Y. Shvartsman. Temporal ordering and registration of images in studies of developmental dynamics. *Development*, 142(9):1717–1724, MAY 1 2015 2015.

[7] Herbert Edelsbrunner and John Harer. Persistent homology-a survey. *Contemporary mathematics*, 453:257–282, 2008.

[8] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

[9] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. In *Foundations of Computer Science, 2000. Proceedings. 41st Annual Symposium on*, pages 454–463. IEEE, 2000.

[10] Jordan Frank, Shie Mannor, and Doina Precup. Activity and gait recognition with time-delay embeddings. In *AAAI*. Citeseer, 2010.

[11] Robert W Ghrist. *Elementary applied topology*. Createspace, 2014.

[12] Chris Godsil and Gordon F Royle. *Algebraic graph theory*, volume 207. Springer Science & Business Media, 2013.

[13] Runyu L Greene, David P Azari, Yu Hen Hu, and Robert G Radwin. Visualizing stressful aspects of repetitive motion tasks and opportunities for ergonomic improvements using computer vision. *Applied ergonomics*, 65:461–472, 2017.

[14] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 2002.

[15] Alec Jacobson, Ilya Baran, Ladislav Kavan, Jovan Popović, and Olga Sorkine. Fast automatic skinning transformations. *ACM Transactions on Graphics (TOG)*, 31(4):77, 2012.

[16] Firas A Khasawneh and Elizabeth Munch. Chatter detection in turning using persistent homology. *Mechanical Systems and Signal Processing*, 70:527–541, 2016.

[17] Mayank Kumar, Ashok Veeraraghavan, and Ashutosh Sabharwal. Distanceppg: Robust non-contact vital signs monitoring using a camera. *Biomedical optics express*, 6(5):1565–1588, 2015.

[18] Ofir Levy and Lior Wolf. Live repetition counting. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3020–3028, 2015.

[19] Jing Liao, Mark Finch, and Hugues Hoppe. Fast computation of seamless video loops. *ACM Trans. Graph.*, 34(6):197:1–197:10, October 2015.

[20] Zicheng Liao, Neel Joshi, and Hugues Hoppe. Automated video looping with progressive dynamism. *ACM Trans. Graph.*, 32(4):77:1–77:10, July 2013.

[21] Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The gudhi library: Simplicial complexes and persistent homology. In *International Congress on Mathematical Software*, pages 167–174. Springer, 2014.

[22] Philip Mcleod and Geoff Wyvill. A smarter way to find pitch. In *In Proceedings of the International Computer Music Conference (ICMC05*, pages 138–141, 2005.

[23] Simone Meyer, Oliver Wang, Henning Zimmer, Max Grosse, and Alexander Sorkine-Hornung. Phase-based frame interpolation for video. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pages 1410–1418. IEEE, 2015.

[24] Jose A Perea and John Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15(3):799–838, 2015.

[25] Arno Schödl, Richard Szeliski, David H Salesin, and Irfan Essa. Video textures. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 489–498. ACM Press/Addison-Wesley Publishing Co., 2000.

[26] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2, pages 246–252. IEEE, 1999.

[27] Floris Takens et al. Detecting strange attractors in turbulence. *Lecture notes in mathematics*, 898(1):366–381, 1981.

[28] Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

[29] Christopher J Tralie. High dimensional geometry of sliding window embeddings of periodic videos. In *Proceedings of the 32st International Symposium on Computational Geometry (SOCG)*, 2016.

[30] Christopher J Tralie. Moebius beats: The twisted spaces of sliding window audio novelty functions with rhythmic subdivisions. In *18th International Society for Music Information Retrieval (ISMIR), Late Breaking Session*, 2017.

[31] Christopher J. Tralie and Jose A. Perea. (quasi)periodicity quantification in video data, using topology. *SIAM Journal on Imaging Sciences*, 11(2):1049–1077, 2018.

[32] Christopher John Tralie. *Geometric Multimedia Time Series*. PhD thesis, Duke University Department of Electrical And Computer Engineering, 2017.

[33] Mikael Vejdemo-Johansson, Florian T Pokorny, Primoz Skraba, and Danica Kragic. Cohomological learning of periodic motion. *Applicable Algebra in Engineering, Communication and Computing*, 26(1-2):5–26, 2015.

[34] V Venkataraman and P Turaga. Shape descriptions of nonlinear dynamical systems for video-based inference. *IEEE transactions on pattern analysis and machine intelligence*, 2016.

[35] Vinay Venkataraman, Karthikeyan Natesan Ramamurthy, and Pavan Turaga. Persistent homology of attractors for action recognition. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 4150–4154. IEEE, 2016.

[36] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman. Phase-based video motion processing. *ACM Transactions on Graphics (TOG)*, 32(4):80, 2013.

[37] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. 2012.