High Dimensional Geometry of Sliding Window Embeddings of Periodic Videos

Christopher J. Tralie[1]

1 Department of Electrical and Computer Engineering, Duke University
Durham, NC USA `chris.tralie@gmail.com`

Abstract

We explore the high dimensional geometry of sliding windows of periodic videos. Under a reasonable model for periodic videos, we show that the sliding window is necessary to disambiguate all states within a period, and we show that a video embedding with a sliding window of an appropriate dimension lies on a topological loop along a hypertorus. This hypertorus has an independent ellipse for each harmonic of the motion. Natural motions with sharp transitions from foreground to background have many harmonics and are hence in higher dimensions, so linear subspace projections such as PCA do not accurately summarize the geometry of these videos. Noting this, we invoke tools from topological data analysis and cohomology to parameterize motions in high dimensions with circular coordinates after the embeddings. We show applications to videos in which there is obvious periodic motion and to videos in which the motion is hidden.

1 Video Dynamics And Subspace Geometry

1.1 Video Delay Embeddings

We use a sliding window through time, also known as a "delay embedding" in the dynamical systems literature [3], to capture the dynamics of a periodic video. More precisely,

▶ **Definition 1.** Given a discrete video $X[n] \in \mathbb{R}^{W \times H}$, where $W \times H$ are the dimensions of each frame in pixels and $n \in \mathbb{Z}^+$ is a discrete time index, the video is *periodic* if there exists a $T \in \mathbb{Z}^+$ so that $X[n] = X[n + T]$ for all n.

▶ **Definition 2.** Given a video $X[n] \in \mathbb{R}^{W \times H}$ and a window size M, *delay embedding* $Y[n]$ is formed as

$$Y[n] = \begin{bmatrix} X[n] \\ X[n+1] \\ \vdots \\ X[n+(M-1)] \end{bmatrix} \in \mathbb{R}^{W \times H \times M} \tag{1}$$

As n varies, $Y[n]$ traces out a samples of a 1-manifold in $\mathbb{R}^{W \times H \times M}$, though for a video of F frames, it lies on a $F - 1$ dimensional subspace, which we exploit to speed up processing. Figure 1 shows a pictorial depiction of this scheme.
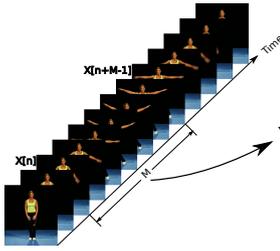
Figure 1 A depiction of a discrete delay embedding of a video of a woman doing jumping jacks with a sliding window embedding of length M
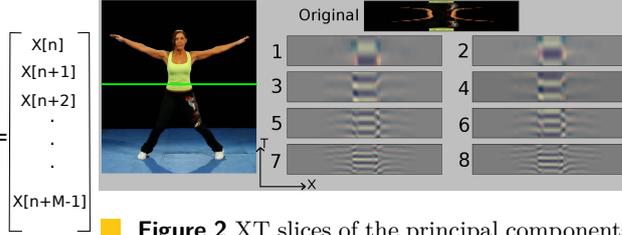
Figure 2 XT slices of the principal components of a sliding window of length 34 on the jumping jacks videos. The green line on the left image shows the X slice that is represented in the plots. Each row corresponds to the two axes of an independent ellipse in the delay embedding

1.2   Hypertorus Video Model

We now characterize the high dimensional geometry of sliding window embeddings of periodic videos, following a similar analysis to recent on delay embeddings of 1D time series [4] (which we also summarize in our narrated video). We start by assuming very general model for periodic videos. For a period T and for constants A and ϕ, and for an arbitrary function g_i at each pixel X_i, define the grayscale level at pixel X_i as

$$X_i[n] = g_i\left(A\cos\left(\frac{2\pi}{T}n + \phi\right)\right) \tag{2}$$

That is, each pixel is an arbitrary function composed with a scale of the same cosine. Though the function at each pixel may differ, the functions across all pixels are globally *in phase*. This means that the model has mirror symmetry built in. In particular $X_i[n] = X_i\left[T - \left(n + \phi\frac{T}{\pi}\right)\right]$. That is, each pixel repeats itself during the second half of its period, but in reverse, making it impossible to disambiguate "going there" from "coming back." On the other hand, a sliding window size of appropriate length can turn this path into a topological loop by taking a different trajectory in the embedding space during the second half of the period. A similar observation was made in early work on video textures [5]. To see this mathematically, express each pixel as a discrete cosine transform with T terms, which is sufficient to summarize it over its period. Storing all T terms for all N pixels in a period in the $N \times T$ matrix D, all pixels can be combined into a column vector of the following form:

$$X[n] = \sum_{k=0}^{T-1} \cos\left(\frac{2\pi}{T}kn\right) D^k \tag{3}$$

where D^k is the kth column of the matrix of DCT coefficients. A sliding window of length M then takes the following form:

$$Y[n] = \sum_{k=0}^{T-1} \begin{bmatrix} D^k \cos\left(\frac{2\pi}{T}kn\right) \\ D^k \cos\left(\frac{2\pi}{T}k(n+1)\right) \\ \vdots \\ D^k \cos\left(\frac{2\pi}{T}k(n+M-1)\right) \end{bmatrix} \tag{4}$$
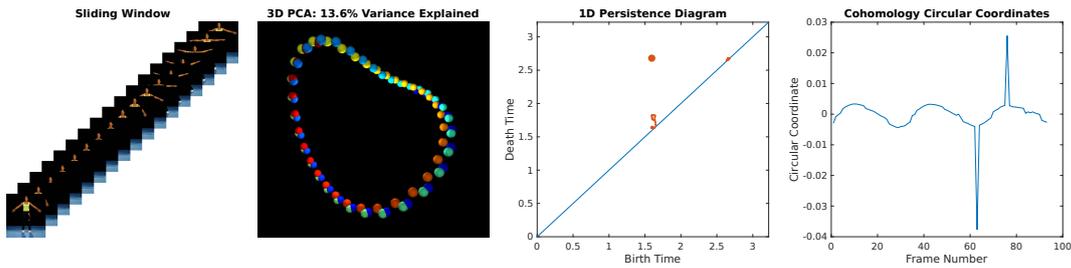
Figure 3 Sliding window embedding of a woman doing jumping jacks
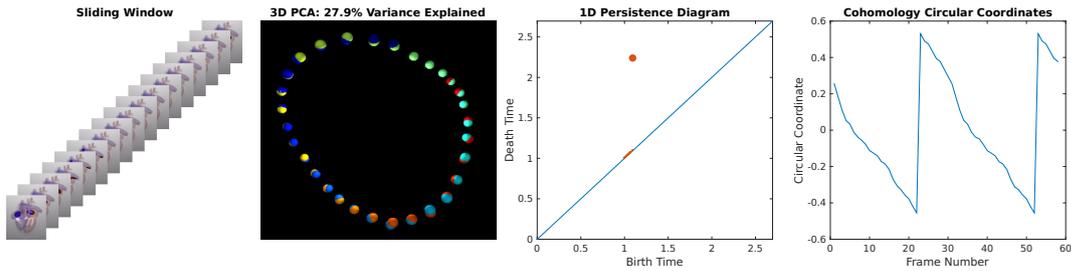
Figure 4 Sliding window embedding of a heartbeat animation

using the cosine sum identity, this can be rewritten as

$$
Y[n] = \sum_{k=0}^{T-1} \cos\left(\frac{2\pi}{T}kn\right) \begin{bmatrix} D^k \\ D^k \cos\left(\frac{2\pi k}{T}\right) \\ \vdots \\ D^k \cos\left((M-1)\frac{2\pi k}{T}\right) \end{bmatrix} + \sin\left(\frac{2\pi}{T}kn\right) \begin{bmatrix} 0^N \\ D^k \sin\left(\frac{2\pi k}{T}\right) \\ \vdots \\ D^k \sin\left((M-1)\frac{2\pi k}{T}\right) \end{bmatrix} \tag{5}
$$

Hence, the path that is traced out by varying n is the sum of $d \leq T/2$ independent ellipses, each spanned by a plane. Such paths live on a topological d-torus, corresponding to $2d$ nonzero columns in D, with 2 dimensions for each independent ellipse.

Let us now empirically examine the embedding space of these motions. Let A be a matrix containing all sliding windows, with the jth column of A containing $Y[j]$, and take the k eigenvectors of the matrix AA^T with the largest eigenvalues, sorted in descending order by eigenvalue. These are known as the first k *principal components* of the sliding window point cloud, and they capture the maximum variance in the data over all possible k dimensional subspaces in the embedding space. Note that vector in the embedding space in sliding window videos is itself a video with M frames. Figure 2 shows the first 8 principal component vectors in a real video, pulling out a line of pixels along the x-axis in each principal component and plotting its evolution over the M frames. Consistently with the vector part of Equation 5, lower frequency ellipse axes correspond to smooth sinusoidal motions, while higher axes correspond to higher harmonics.

2 Persistent Homology And Circular Coordinates

Since sliding window embeddings of videos lie on a highly curved topological loop on a hypertorus, high dimensional data analysis tools are necessary. We invoke 1D persistent homology to measure the geometric prominence of the loops [2]. Briefly, homology is an algebraic framework for describing equivalence classes of loops in a fixed space, where equivalence is defined up to a boundary in that space (loops in the same class can be stretched into each other along the space without gluing or tearing). *1D Persistent homology* is an adaptation of the theory to point cloud data in which a sequence of simplicial complexes are constructed on top of the point cloud, and classes of loops are tracked as edges and triangle are added to the complex. In a *Rips filtration*, edges are added between
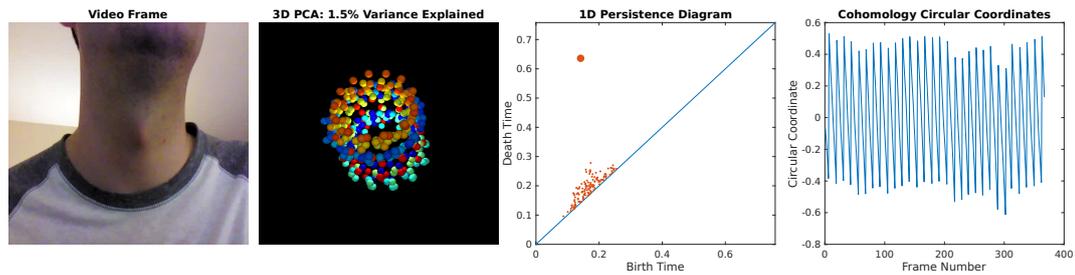
■ **Figure 5** Sliding window embedding of a video of a person sitting still, which has hidden periodic motion due to the person's heartbeat

points in ascending order of distance, and a triangle between three points is added the moment all three edges have been added. The "birth time" of a class is the distance in the Rips filtration at which the loop class first appears, and the "death time" is the distance at which it "fills in" (i.e. is expressible as a boundary of triangles). In this way, persistent homology is a mix of topological and metric information about the loop, as loops which are born earlier are more finely sampled, and loops which die later are rounder and/or geometrically larger. Birth/death pairs are plotted for all classes in a "persistence diagram," as shown in Figures 3, 4, 5. In all of these examples, there is one clear loop which has a much higher "persistence" (death - birth) than all other loops, and this loop corresponds to the periodic motion encoded geometrically with the sliding windows. We also use a related theory of 1D persistent cohomology to find maps of a point cloud to the circle for a loop class, thereby parameterizing the motion of the video embeddings with circular coordinates [1].

Figure 3 shows an example of applying 1D Rips and persistent cohomology to extract circular coordinates for the jumping jack example. The loop is visible after projecting the data onto its first three principal components, but little of the variance is explained since this video has sharp transitions from foreground to background, which need to be represented with many harmonics, as shown in our video. Circular coordinates, on the other hand, are able to parameterize the motion in high dimensions. A similar pattern is visible for a synthetic beating heart video in Figure 4. We can also apply our techniques to videos with very subtle motions. Figure 5 shows such an example with a person sitting still in front of a camera. Hardly any motion is visible, but the persistence diagram and circular coordinates indicate the presence of a cycle. In fact, this cycle corresponds to twice the person's heartbeat, which exists as a low magnitude vibration in the video. To show that there indeed is a periodic process going on, we use phase-based video motion amplification ([6]) to amplify all motions within the frequency band consistent with the parameterization found.

References

1   Vin De Silva, Dmitriy Morozov, and Mikael Vejdemo-Johansson. Persistent cohomology and circular coordinates. *Discrete & Computational Geometry*, 45(4):737–759, 2011.

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

3   Holger Kantz and Thomas Schreiber. *Nonlinear time series analysis*, volume 7. Cambridge university press, 2004.

4   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, 2013.

5   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.

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