

# Applications of Algebraic Topology to Video Compression

JASMINE BAYROOTI & GUNNAR CARLSSON  
STANFORD UNIVERSITY

In this study, we applied Algebraic Topology techniques to extract information about the shape of data and applied these insights to the research problem of video compression. Specifically, we applied a computational tool known as Persistent Homology to point-cloud data sets and extracted insights on the data from the induced barcodes. We generalized results from a study of the local behavior of spaces of natural images by Carlsson et al [1] to the study of videos. To this end, we considered an ambient space of 81-dimensional points containing arrangements of  $3 \times 3$  patches of pixels extracted from the frames within a video. We developed a computational model for the high-contrast dense sub-manifolds of this point cloud and found that these sub-manifolds have the topological properties of a connected bouquet of spheres. The reduction of dimension to a bouquet of spheres could have potential applications for video compression.

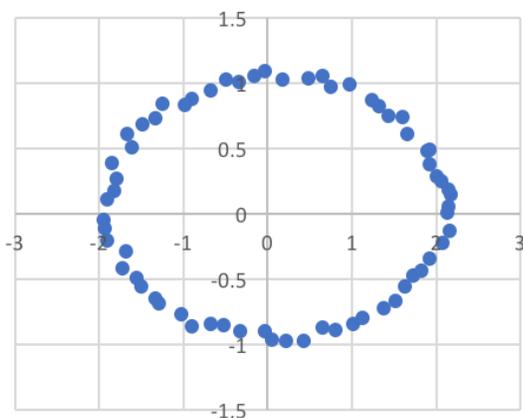
## Introduction

Topology is the study of properties of spaces and homology is a tool that mathematicians use to characterise the shape of a space. Algebraic Topology is a branch of mathematics that requires background knowledge in subjects including Groups, Rings, and Fields as well as Metric Spaces, Topological Spaces, and Analysis. It works by assigning algebraic invariants, such as a group, to topological spaces. There are a number of ways of manipulating point cloud data to transform the data into a representative topological space. Recently, Carlsson et al [1] at Stanford University have developed a computational software called Plex [2] that enables efficient computations on point cloud data, including calculations of Persistent Homology and barcodes.

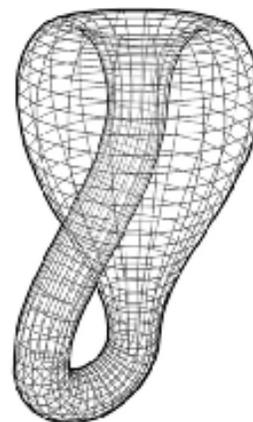
The algebraic invariant in Persistent Homology, called Betti numbers, can be thought of as identifying the shape of the data.

For instance, if a point cloud is in the shape of a circle, Persistent Homology would identify that as a 1-dimensional hole in  $\mathbb{R}^2$  as shown in Figure 1. In higher dimensions, there may be  $n$ -dimensional holes or spheres embedded in the data. In addition to holes, Persistent Homology is well known for a large class of objects and these characterizations may be used to better understand the shape of the data. For instance, objects like the Klein Bottle shown in Figure 2 are amongst a vast database of objects for which the homology is well understood.

In this section, we give a brief introduction to Persistent Homology and barcodes largely following the Math Review from a JavaPlex tutorial [2] and online paper [3]. Readers interested in Algebraic Topology should refer to Armstrong [4] and to Zomorodian and Carlsson [5] for more details on Persistent



**Figure 1.** Given a high-dimensional point cloud, it is often difficult to know if linear statistics can be applied. For instance, in this shape the linear trendline would be a misleading representation of the dataset. On the other hand, Persistent Homology would characterize this data with one 1-dimensional hole.



**Figure 2.** The Klein Bottle is 2-dimensional surface which can be embedded in 4-dimensional space or higher. The shape may be captured in  $\mathbb{R}^3$  as well but then the self-intersection is necessary and adding another dimension creates the Klein Bottle proper without self-intersection.

Homology.

### Simplicial Complexes

In a topological analysis, we first replace a set of data points with a family of simplicial complexes to convert the point cloud into a topological space. An abstract simplicial complex is given by the following data.

- A set  $Z$  of vertices or 0-simplices
- For each  $k \geq 1$ , a set of  $k$ -simplices  $\sigma = [z_{i_0}, z_{i_1}, \dots, z_{i_k}]$ , where  $z_i \in Z$ .
- Each  $k$ -simplex has  $k+1$  faces obtained by deleting one of the vertices. The following membership property must be satisfied: if  $\sigma$  is in the simplicial complex, then all faces of  $\sigma$  must be in the simplicial complex.

We think of 0-simplices as vertices, 1-simplices as edges, 2-simplices as triangle faces, and 3-simplices as tetrahedrals.

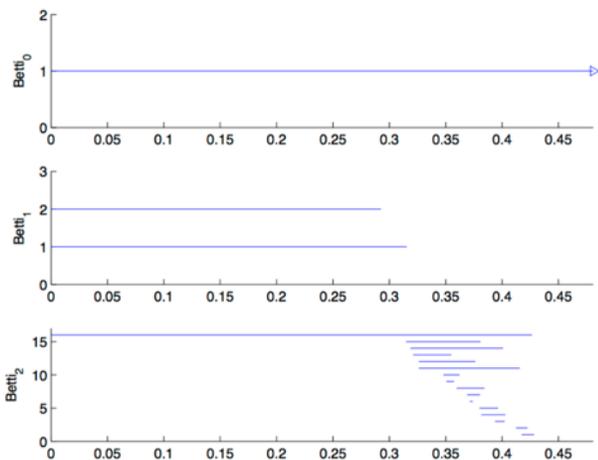
### Homology

Betti numbers describe the homology, which one can think of as holes, of a simplicial complex  $X$ . The value  $Betti_k$ , where  $k \in \mathbb{N}$ , is equal to the rank of the  $k^{th}$  homology group of  $X$ .  $Betti_k$  can be thought of as giving the number of  $k$ -dimensional holes and  $Betti_0$  is the number of connected components. For instance, the circle has a shape described by  $Betti_0 = 1$  and  $Betti_1 = 1$ . In  $\mathbb{Z}/2\mathbb{Z}$ , the Klein Bottle has a shape described by  $Betti_0 = 1$  and  $Betti_1 = 2$ ,  $Betti_2 = 1$ .

When one works with a point cloud of data, there is often noise embedded in the dataset. For instance, large amounts of financial data, readings from sensors, and pixels from images, all contain some amount of noise due to a variety of factors. Standard homology of a simplicial complex does not offer a way to deal with that noise. Persistence and barcodes, whose definitions rely on filtered simplicial complexes, are rigorous responses to this problem.

### Filtered Simplicial Complexes

A filtration on a simplicial complex  $X$  is a collection of sub-complexes  $\{X(t) | t \in \mathbb{R}\}$  of  $X$  such that  $X(t) \subset X(t')$  whenever  $t \geq t'$ . The filtration



**Figure 3.** This is an example of a barcode sequence for a point cloud corresponding to a Klein Bottle from figure 2, which confirms that the Betti numbers are  $Betti_0 = 1$  and  $Betti_1 = 2$ ,  $Betti_2 = 1$ . The short barcodes are noise.

value of a simplex  $\sigma \in X$  is the smallest  $t$  such that  $\sigma \in X(t)$ . There are many ways one can generate a filtered simplicial complex, for instance by introducing a metric. We start with a vertex  $z_0$  for  $X(0)$  and, define  $X(t)$  as all vertices and edges in the larger simplicial complex that are a distance at most  $t$  away from  $z_0$ .

### Persistent Homology and Barcodes

Given a filtered simplicial complex, those topological features which persist over a significant parameter range of  $t$  are considered as signal while short-lived features are noise. Barcodes show holes on an interval, with short intervals corresponding to potential noise and longer intervals corresponding to topological features that persist. Betti intervals help describe how the homology  $X(t)$  changes with  $t$ . A  $k$ -dimensional Betti interval, with endpoints  $[t_{start}, t_{end}]$ , corresponds roughly to a  $k$ -dimensional hole that appears at filtration value  $t_{start} < t < t_{end}$  and closes at the value  $t_{end}$ . This can be viewed in the form of a barcode graph, as shown in Figure 3.

### Using Algebraic Topology to Compress Optical Images

#### Summary of Optical Image Study

In the paper *On the Local Behavior of Natural Images* [1], Dr. Carlsson et al showed that a large subset of 9-dimensional data from natural images lies on the surface of a Klein Bottle, as observed through the “three circle” model (see p. 8-9 of [1] for details). In the paper *The Ring of Algebraic Functions on Persistent Barcodes* [6], this result was extended with the goal of compressing storage of optical photos. In this study, we verify the findings in [1] on a different set of images and generalize the approach to videos.

#### Procedure for Image Analysis

Carlsson et al [1] performed their analysis on photos from a database of images constructed by H. van Hateren [7]. In our study, we randomly select different optical photos from a private photo library to determine whether or not the results continue to hold. Given the different set of photos, we follow the steps taken in [1]:

**Step 1:** Randomly select approximately 50,000 size  $3 \times 3$  patches of pixels from a set of unconnected grayscale images in the database. Each  $3 \times 3$  patch corresponds to a 9-dimensional vector  $x$  where  $x = (x_1, x_2, \dots, x_9) = [I_{11}, I_{12}, I_{13}, I_{21}, I_{22}, \dots, I_{33}] \in \mathbb{R}^9$  where  $I_{ij}$  is the intensity of the  $i_j$ th pixel.

**Step 2:** Take the natural logarithm of each coordinate to obtain  $x = (\log x_1, \log x_2, \dots, \log x_9) \in \mathbb{R}^9$ . According to Weber’s Law, there is an inverse relationship between ambient illumination and human sensitivity to light. Therefore, the ratio of  $\frac{dL}{L}$  is constant for a wide range of luminances (see [8] for more details).

**Step 3:** Compute the D-norm of each vector to obtain a measure of contrast of a patch. The D-norm is calculated by summing the differences between the vertically and horizontally adjacent neighbors in a  $3 \times 3$  patch and then taking the square root. The D-norm is defined in this way because it is commonly believed that regions of photos with high contrast convey the most significant content of a scene (p. 3 [8]). If the 9-dimensional patch is  $x$ , the D-norm is given by  $\sqrt{x^T D x}$  where  $D$  is a positive definite symmetric matrix shown below (see [8] for more details on derivation). For instance, to obtain the 5th row of the matrix in Figure 4, we take  $(x_5 - x_2) + (x_5 - x_4) + (x_5 - x_6) + (x_5 - x_8) = 4x_5 - x_2 - x_4 - x_6 - x_8$ .

$$D = \begin{bmatrix} 2 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 3 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 3 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 3 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 2 \end{bmatrix}$$

**Step 4:** Select the patches with a D-norm in the top T percent of the entire sample. This is done to maintain the high contrast patches since they follow a different distribution than low-contrast patches and contain the most important information about an image (p. 3 [8]). The resulting point cloud is denoted  $X_T$ .

**Step 5:** Subtract from each vector the average of its coordinates, to get  $x = (x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_9 - \bar{x})$  where  $\bar{x} = \Sigma x_i$ . This reduces the dimension from nine to eight because the sum of the new points will always be zero, meaning if we know eight of the values, the ninth one is determined. Next, divide by the D-norm to normalize the selected vectors. This makes each vector have a unit length so that they all lie on a unit sphere in  $R^8$ . Note that the unit sphere is given by the equation  $x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2 + x_6^2 + x_7^2 + x_8^2 = 1$ . Therefore, knowing seven of the coordinates determines the eighth. Hence, the point cloud lies on a 7-dimensional sphere in  $R^8$ .

**Step 6:** Filter out the outlier points in the remaining point cloud using the kth nearest neighbor density function  $p_k(x)$ . Select p percent of the points whose Euclidean distance to their kth nearest

neighbor are smallest. This gives a dense subset of the point cloud denoted by  $X_T(k,p)$ . A small choice of k results in a local density estimate while a larger k value provides a more global estimate.

**Step 7:** Use the resulting dense point cloud  $X_T(k,p)$  as input for topological analysis. This involves running JavaPLEX to create the filtered simplicial complexes and find the corresponding persistent homology and barcodes described in Section 1 for the point cloud. Examining the barcodes will show patterns or submanifolds in the dense point cloud  $X_T(k,p)$ .

### Results for Image Analysis

We found that the barcode representation of persistent homology in Figure 8 as well as the 2-dimensional cross sections (Figures 9-11) corroborate the results from *On the Local Behavior of Natural Images* ([1], p. 5-8) when run on a random assortment of optical images including Figures 5-7 and more.

### Discussion for Image Analysis

These results and their corresponding homology aligns well with the study by Carlsson et al [1]. In particular, we also observed the three circles in the (e1, e2) plane, (e1, e3) plane, and (e1, e5) plane of the Three Circle Model of the Klein Bottle. This indicates that their conclusion that the high contrasting and dense sets of the original patches lie largely on the surface of a Klein Bottle may apply to other datasets beyond the Van Hateren database. This conclusion makes the compression of images more efficient since the dense subset lies on a 4-dimensional shape rather than a point cloud in 9 dimensions [6], potentially allowing up to a 50% reduction of storage on dense subsets.

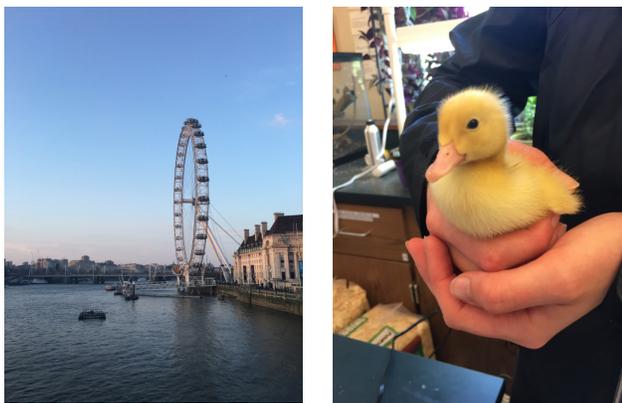
### Using Algebraic Topology to Compress Optical Videos

#### Procedure for Video Analysis

For the video project, a similar method was followed with a few key differences. The main difference is that a video is a collection of frames or pictures that are taken in linear time and are consequently similar to neighboring frames. To address this difference, we assign consecutive frames to groups and look for areas of large variance or contrast in each group. The steps for the method are detailed below:

**Step 1:** Process the video in grayscale and divide all of the scenes in the video into groups of nine consecutive frames. For videos with a frame rate of 30 frames per second, as used in this study, nine frames corresponds to approximately 330 milliseconds of video footage.

**Step 2:** Construct a point cloud by randomly picking 900 3 x



Figures 5, 6. Sample optical images from the database.

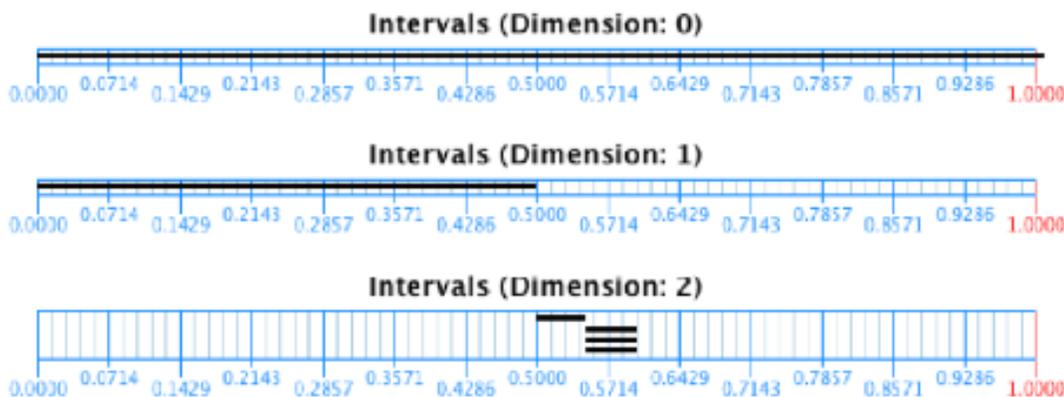
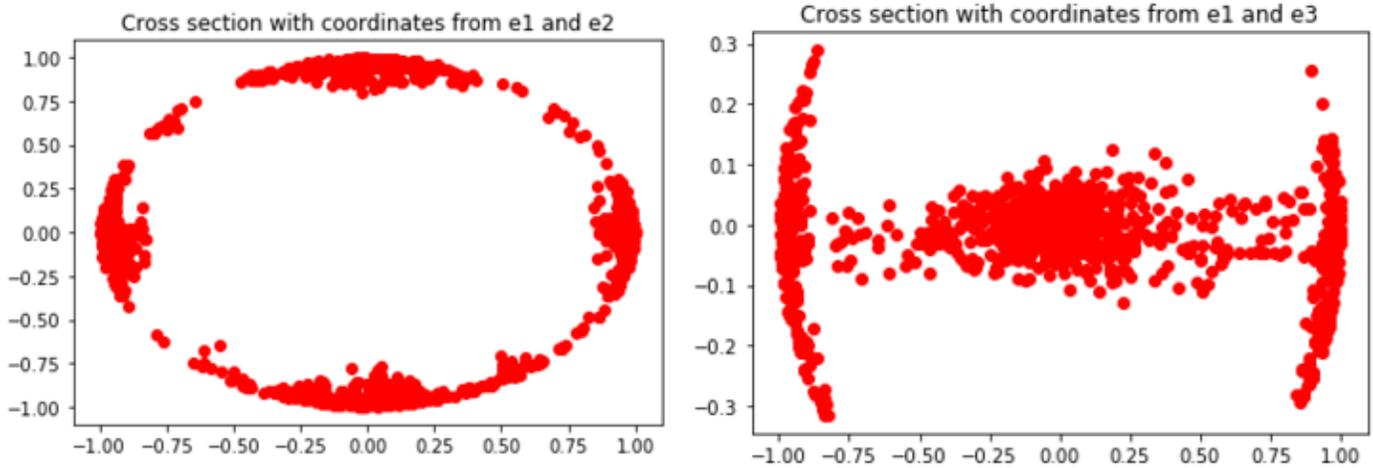
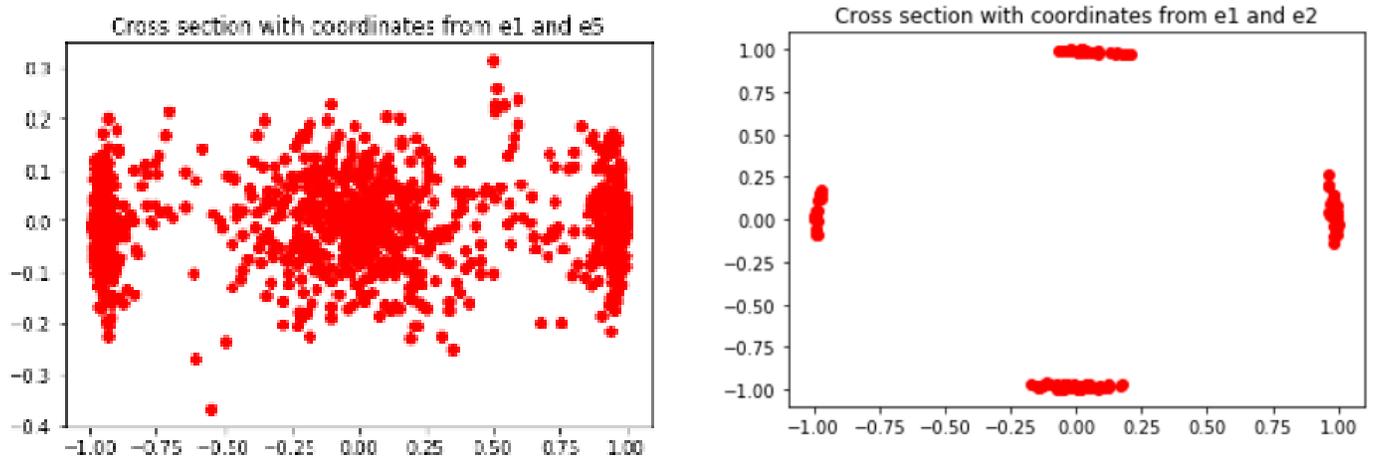


Figure 7. Barcode results from optical image analysis.



**Figure 8.** Representations of the shape of data from two different cross sections for optical images with  $k=15$ ,  $p=20\%$  and  $T=10\%$ ,  $X_{0,1}(15, 0.2)$ . These two views are part of the Three Circle Model.



**Figure 9.** A representation of the shape of data from a third cross section for optical images with  $k=15$ ,  $p=20\%$  and  $T=10\%$ ,  $X_{0,1}(15, 0.2)$ . This is the third part of the Three Circle Model.

**Figure 10.** A representation of the shape of the data for optical images with  $k=300$ ,  $p=3\%$ , and  $T=10\%$ ,  $X_{0,1}(300, 0.03)$ . This shows the four dense corners of the circle.

3 patches of pixels in the same location across all nine frames in the group. As in the picture study, take the natural logarithm of each coordinate of the  $3 \times 3$  patch to obtain a 9-dimensional vector  $x$ , where  $x = (x_1, x_2, \dots, x_9) = [I_{11}, I_{12}, I_{13}, I_{21}, I_{22}, \dots, I_{33}] \in R^9$ . By considering the  $3 \times 3$  patches from all nine frames in a group, obtain points in  $R^{81}$ .

**Step 3:** Calculate the D-norm for each point by averaging the D-norms of the nine 9-dimensional vectors from each frame in the group. If the standard deviation of the D-norms from the group is small, then the contrast in that region of the group is small and should be eliminated. Therefore, the point is kept provided that the standard deviation among the nine D-norms is larger than a specified threshold value to ensure that there is significant change within the group itself.

**Step 4:** In addition to including patches where there is high variation among the nine frames from step 3, we select the patches with a D-norm in the top  $T$  percent of the entire sample to produce the point cloud  $X_r$ . This ensures that we consider areas of high contrast within the individual frame.

**Step 5:** Subtract from each vector the average of its coordinates

to reduce the dimension from 81 to 80. Next, divide by the D-norm to normalize the selected vectors. This makes each vector have a unit length so that they all lie on a unit sphere in  $R^{80}$ . Note that the unit sphere is given by the equation  $x_1^2 + x_2^2 + \dots + x_{80}^2 = 1$ . Hence, the point cloud lies on a 79-dimensional sphere in  $R^{80}$ .

**Step 6:** Filter out the outlier points in the remaining point cloud using the  $k$ th nearest neighbor density function  $pk(x)$ . Select  $p$  percent of the points whose Euclidean distance to their  $k$ th nearest neighbor are smallest. This gives us a dense subset of the point cloud denoted by  $XT(k, p)$ . A small choice of  $k$  results in a local density estimate while a larger  $k$  value provides a more global estimate. Because we have many points in high dimensions, finding the  $k$ th nearest neighbor is computationally expensive and is an active area of research. The brute force approach runs in  $O(NDk)$  where  $N$  is the size of the training set,  $D$  is the dimension of each point, ie 79, and  $k$  is the number of neighbors to take. We use the KDTree algorithm from the SKLearn package in SciPy. This is a C package that is available within Python and is very efficient as it runs in  $O(\log N)$  for searching.

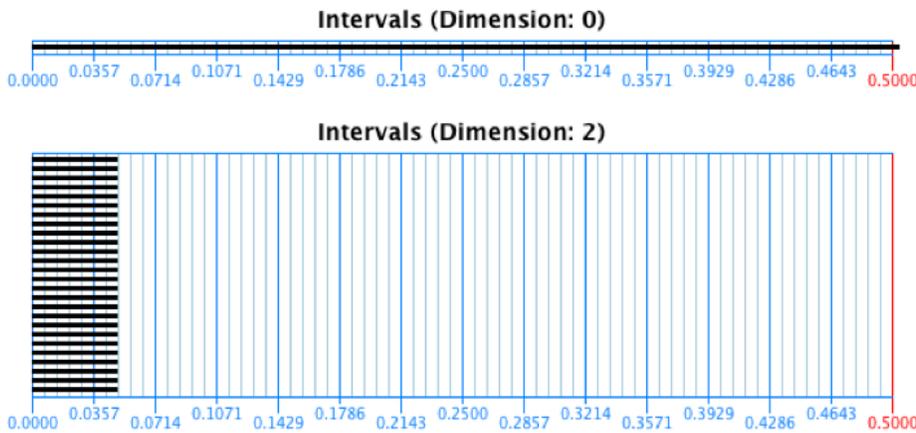
**Step 7:** Run a topological analysis on the resulting dense point cloud  $X_T(k,p)$  to look for patterns or submanifolds through an examination of the barcodes. Visualize the shapes using 2-dimensional cross sections with Python's matplotlib.

### Results for Video Analysis

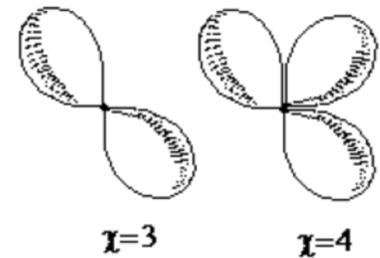
We obtained the following results on a grayscale video. See Figures 11-15.

### Discussion for Video Analysis

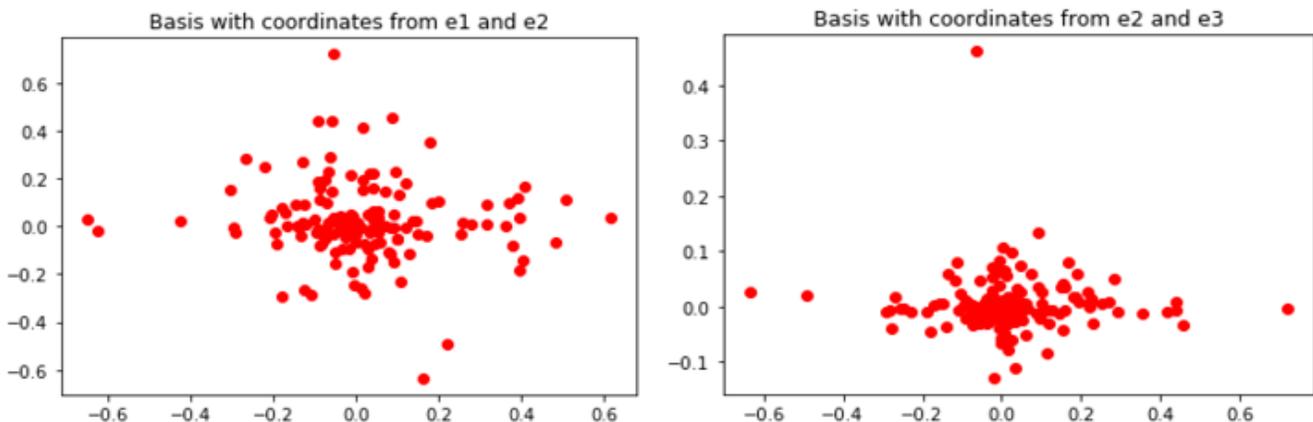
The persistent barcodes corresponding to dense subsets  $X_T(k,p)$  of a point cloud obtained from video footage appears to lie on a surface that is topologically homeomorphic to a bouquet of spheres.



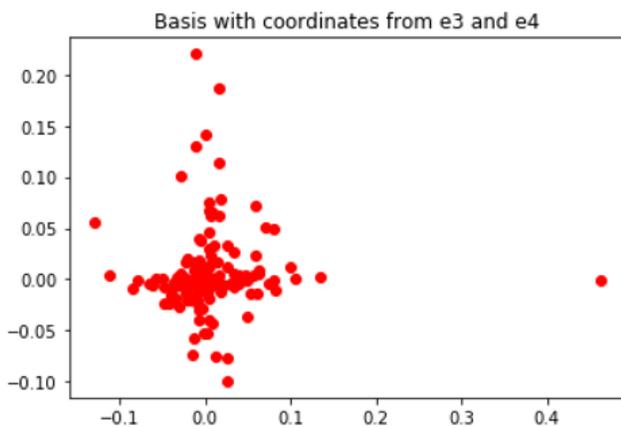
**Figure 11.** These are the barcode results of the analysis. Since there is only one interval in dimension 0, this means that the surface is connected. Note that there are no holes in dimension 1 and there are a number of distinct holes in dimension 2. 2-dimensional holes are topologically equivalent to spheres. Therefore, the barcodes suggest a shape called a "bouquet of spheres" with the number of intervals  $n$  in the second dimension being the number of spheres attached to a common point.



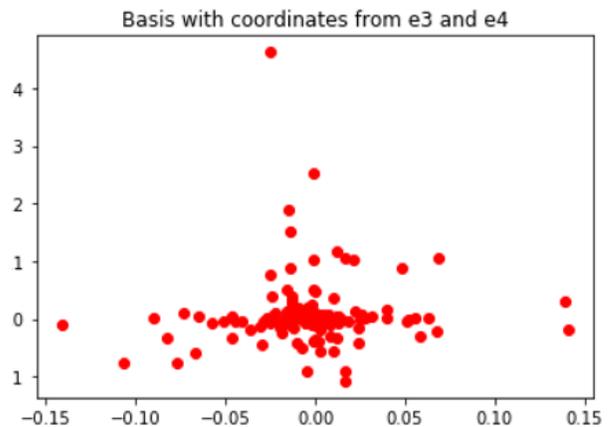
**Figure 12.** Two representations of bouquets of spheres where  $n = 3$  and 4.



**Figure 13.** Representations of the data from an optical video with  $k=15$ ,  $p=10\%$ , and  $T=20\%$ ,  $X_{0.2}(15, 0.1)$ .



**Figure 14.** Representation of the data from an optical video with  $k=15$ ,  $p=10\%$ , and  $T=20\%$ ,  $X_{0.2}(15, 0.1)$ .



**Figure 15.** Representation of the data from an optical video with  $k=300$ ,  $p=10\%$ , and  $T=20\%$ ,  $X_{0.2}(300, 0.1)$ .

The bouquet of spheres is a surface for which the homology is well understood with holes only in dimension 2. Each of the spheres in the bouquet is a 2-dimensional surface. By preprocessing the video frames, one can potentially transform each pixel in the original 81-dimensional space to a 2-dimensional point on an appropriate sphere. For instance, one could map an 81-dimensional object to a 3-dimensional object, the first component of which identifies which sphere from the bouquet it lands on while the other two components specify the location on the sphere. Since this transformation may be performed on a dense subset of the video, there is significant potential for compression of data. The next steps in this study would be to repeat the analysis on additional video footage to confirm the shape of dense subsets and also to define and test a transformation for compressing the data without losing image quality.

## Conclusion

In this study, we applied Algebraic Topology to gather information about the shape of data and used these insights on the research problem of the local behavior of natural videos. We found that the persistent barcodes corresponding to the high-contrast and dense subset of the original 81-dimensional point cloud appeared to lie on a surface that is homeomorphic to a connect bouquet of spheres, which is a topological surface whose homology is well understood. The lower dimensional results could be potentially useful for video compression just as the resulting Klein Bottle shape from the study of local behavior of natural images by Carlsson et al [1] was used for the compression of images. This application of Algebraic Topology shows the great potential in the field for new discoveries and approaches to important research problems.

## Acknowledgements

I would like to thank Dr. Carlsson for all of the meetings, emails, skype sessions, and advice he gave me over the course of this research project. It has been a very rewarding process for me personally and it has motivated me in my further research. I would also like to thank my independent math research instructor, Ms. Keshena Richardson, for her guidance, advice, and many valuable comments and Dr. Andrew Ying for his very helpful feedback and editing. Finally, I would like to thank my parents for their support throughout my education.

## References

- [1] Carlsson G., Ishkhanov T., De Silva V., Zomorodian A., On the local behavior of spaces of natural images, *International Journal of Computer Vision* 76, 1 (2008), pp.1-12.
- [2] Adams H., Tausz A. *JavaPlex Tutorial*. 2017
- [3] Ghrist R., *Barcodes the Persistent Topology of Data*. *Bull. Amer. Math. Soc.* 45 (2008), 2007, pp. 61-75
- [4] Armstrong, M. A. *Basic topology*. Springer, 2011.
- [5] Zomorodian A., Carlsson G. *Computing Persistent Homology*. *Discrete Computational Geometry*, 33:249-274. 2005
- [6] Adcock A., Carlsson E., Carlsson G., *The Ring of Algebraic Functions on Persistent Barcodes*. 2013.
- [7] van Hateren J. and van der Schaff A., *Independent Component Filters of Natural Images*

- Compared with Simple Cells in Primary Visual Cortex, *Proc. R. Soc. London*, B 265 (1998), 359-366.
- [8] Lee A.B., Pedersen K.S., Mumford D. The non-linear statistics of high-contrast patches in natural images. *International Journal of Computer Vision* 54, 1-3(2003), pp. 83-103