SIAM J. IMAGING SCIENCES
Vol. 2, No. 1, pp. 110–117

On the Nonlinear Statistics of Range Image Patches*

Henry Adams† and Gunnar Carlsson‡

Abstract. In [A. B. Lee, K. S. Pedersen, and D. Mumford, *Int. J. Comput. Vis.*, 54 (2003), pp. 83–103], the authors study the distributions of 3×3 patches from optical images and from range images. In [G. Carlsson, T. Ishkanov, V. de Silva, and A. Zomorodian, *Int. J. Comput. Vis.*, 76 (2008), pp. 1–12], the authors apply computational topological tools to the data set of optical patches studied by Lee, Pedersen, and Mumford and find geometric structures for high density subsets. One high density subset is called the *primary circle* and essentially consists of patches with a line separating a light and a dark region. In this paper, we apply the techniques of Carlsson et al. to range patches. By enlarging to 5×5 and 7×7 patches, we find core subsets that have the topology of the primary circle, suggesting a stronger connection between optical patches and range patches than was found by Lee, Pedersen, and Mumford.

Key words. range images, topology, persistent homology

AMS subject classifications. 62H35, 65D18

DOI. 10.1137/070711669

1. Introduction.

An optical image has a grayscale value at each pixel, whereas a range image pixel contains a distance: the distance between the laser scanner and the nearest object in the correct direction. In either case, we can think of an $n \times m$ pixel patch as a vector in $\mathbb{R}^{n \times m}$ and of a set of patches as a set of points in $\mathbb{R}^{n \times m}$. Figure 1 shows two range images from the Brown database by Lee and Huang which contains a variety of indoor and outdoor shots. Darker regions are closer than lighter regions, except for out-of-range data such as the sky that is colored black.

In [7], Lee, Pedersen, and Mumford describe the distributions of optical and range patches. They begin their analysis by constructing data sets of *high-contrast* patches, where high contrast is defined by thresholding a natural measure of contrast given in section 2. They find that high-contrast 3×3 range patches are densely clustered around the binary patches. A pixel in a binary range patch is one of two values: foreground or background. For optical patches, Lee, Pedersen, and Mumford find a strikingly different distribution: the majority of the high-contrast patches lie near a two-dimensional annulus. Each patch on this manifold is a linear step edge, a few of which are shown in Figure 2. The annulus is parameterized by the angle of the edge and by the distance of the edge from the center of the patch.

In [2], Carlsson et al. expand upon the optical image findings of [7]. They use persistent homology, which we introduce in section 3, to identify the topologies of high density subsets of

*Received by the editors December 26, 2007; accepted for publication (in revised form) October 7, 2008; published electronically January 30, 2009. This research was supported in part by DARPA HR 0011-05-1-0007 and NSF DMS 0354543.
http://www.siam.org/journals/siims/2-1/71166.html
†Department of Mathematics, Caltech, Pasadena, CA 91106 (henryhughadams@gmail.com).
‡Department of Mathematics, Stanford University, Stanford, CA 94305 (gunnar@math.stanford.edu).

Figure 1. *Sample range images.*

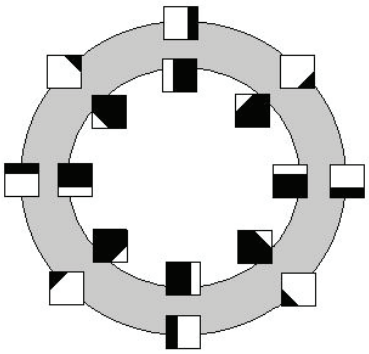

Figure 2. *Step-edge annulus.*

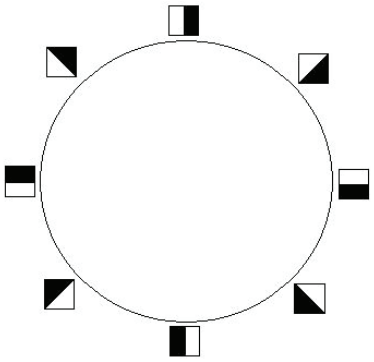

Figure 3. *Primary circle.*

the optical patch space. When using one choice of density estimator, the collection of densest points has the topology of a circle, called the primary circle (see Figure 3). The primary circle is homotopy equivalent to the step-edge annulus. When using a different (finer) density estimator, the collection of densest points has the topology of a Klein bottle that contains the primary circle. In this paper we apply the topological methods of [2] to study range patches.

One conclusion of the work in [7] was that range and optical 3×3 patches were essentially different. The range patches simply broke up into clusters without an obvious simple geometry, while the optical patches were organized in a clearly geometric way. Our hypothesis concerning this conclusion is that it is due to the clustering of the range patches around binary patches: after normalizing the contrast, essentially only two values are taken in range patches, whereas up to nine values are taken in optical patches. The 3×3 binary patches are simply too coarse grained to encode an analogue of the annulus or circle found in [2] and [7], and so one cannot expect to find any reasonable geometry. This suggests that on larger scales, which have a greater capacity to encode, the range patches might show this primary circle behavior. This is indeed what we find, for when we study 5×5 and 7×7 range patches, an analogue to the primary circle present in optical patches occurs.

2. Preparing the spaces of range patches. We analyze 3×3, 5×5, and 7×7 pixel patches from the Brown database by Lee and Huang, a set of about 200 range images. The operational range for the Brown scanner is typically 2–200 meters, and the distance values for

each pixel are stored in units of 0.008 meters. More details about the Brown database can be found at http://www.dam.brown.edu/ptg/brid/index.html. We obtain the spaces of patches through the following steps, which are nearly identical to the procedures used in [2] and [7].

Step 1. We randomly select nearly $4 \cdot 10^5$ size $m \times m$ patches from the images in the database, where m is 3, 5, or 7.

Step 2. Regarding each $m \times m$ patch as an m^2-dimensional vector $\vec{x}$, we take the logarithm of each coordinate. This step helps provide shape invariance (see [6]).

Step 3. We compute the D-norm, $\|\vec{x}\|_D$, of each vector. This is a measure of the contrast of a patch. Two coordinates of $\vec{x}$ are neighbors, denoted $i \sim j$, if the corresponding pixels in the $m \times m$ patch are adjacent. We calculate the D-norm for a vector by summing the squared difference between all neighboring coordinates and then taking the square root: $\|\vec{x}\|_D = \sqrt{\sum_{i \sim j}(x_i - x_j)^2}$.

Step 4. We select the patches that have a D-norm in the top T percent of the entire sample. The rationale is that high-contrast patches are believed to contain the most important information of an image but to follow a different distribution than low-contrast patches. We use $T = 20\%$, as done in [2] and [7].

Step 5. For computational feasibility, we randomly select 50,000 of the above patches in the top T percent.

Step 6. So that images of distant and close objects are comparable, we subtract from each vector the average of its coordinates. This maps our vectors to a hyperplane of dimension $m^2 - 1$.

Step 7. We map to a sphere of dimension $m^2 - 2$ by dividing each vector by its Euclidean norm, which is nonzero because the patches are high contrast. We do not change to the discrete cosine transform (DCT) basis, as done in [7] for convenience.

Step 8. From the first seven steps, we have a set of 50,000 high-contrast, normalized, range image patches. However, hoping to approximate the topology of such a space is still a daunting task: the outlier points may significantly alter the computed topology. A more modest goal is to describe the topology of core subsets of the space, in the hope that the core subsets will reflect important patterns of the entire space. We estimate the density at a point $\vec{x}$ using the function ρ_k, where $\rho_k(\vec{x})$ is defined to be the distance from $\vec{x}$ to the kth nearest neighbor of $\vec{x}$. A small choice of the parameter k produces a local density estimate, whereas a larger k-value gives a more global estimate. We select the points whose densities are in the top p percent.

The core subset, which depends on the patch size $m \times m$, the density parameter k, and the density cut p, shall be denoted $X^m(k, p)$. Here we consider core subsets with density parameter $k = 300$ and cut percentage $p = 30\%$.

3. Persistent homology.

Our core subsets are finite samplings from an unknown underlying space. Persistent homology is a method that uses only a finite sampling to estimate the underlying space's topology.

We present a simplified example. Suppose we are given 10 points which (unknown to us) are sampled from a circle. Can we identify the underlying space? Recall that the kth Betti number of a space is a topological invariant, equal to the rank of the kth homology group and roughly equal to the number of k-dimensional holes. For a circle, $Betti_0 = Betti_1 = 1$. For more information, see [1] and [5].

We build a family of Rips simplicial complexes, nested by the parameter R, as follows. The vertices are our 10 given points. The included higher simplexes are spanned by any set of vertices within distance R of each other. Figure 4 shows three such nested complexes, with R increasing from left to right. For more information on Rips complexes, see [3] and [8].

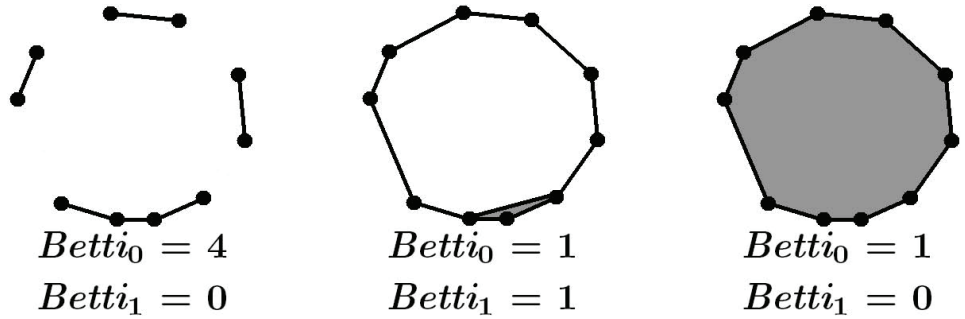

Figure 4. *Nested simplicial complexes built on* 10 *circle points.*

Clearly the choice of R is an important one, which we do not know how to make without knowledge of the underlying space. However, using the ideas of [9] and the computer software package PLEX, created by Carlsson et al., we compute the Betti numbers over a range of R-values and display the result in a Betti barcode (Figure 5).

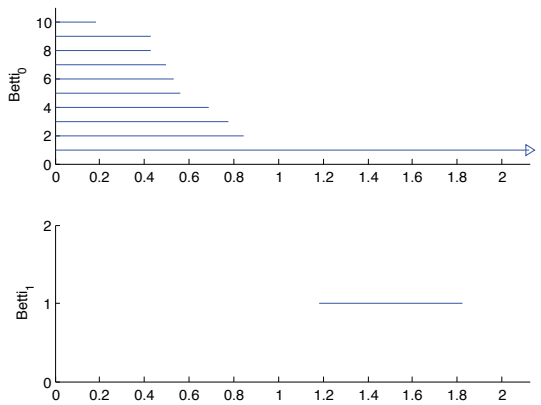

Figure 5. *Betti barcode for the* 10 *circle points.*

The R-values are on the horizontal axis, and $Betti_k$ is the number of intervals in the $Betti_k$ plot that intersect the vertical line through R. The $Betti_0$ plot shows the 10 vertices joining into one connected component as R increases. The single interval in the $Betti_1$ plot begins when the loop forms and ends when the loop fills to a disk. For a large range of R-values, the correct circular profile $Betti_0 = Betti_1 = 1$ is obtained. The idea of persistent homology is that long intervals in the $Betti_k$ plot generally correspond to real topological features of the underlying space, whereas short intervals are considered to be noise.

In practice we do not use the Rips complex: if every point in a large finite sampling is used as a vertex, the computations quickly become infeasible. Instead we use the weak witness complex defined in [3]. We choose 50 vertices (called *landmark points*) using sequential

maxmin, which guarantees that the landmarks are spread throughout the data set. The remaining data points serve as witnesses for inclusions of higher simplexes. This is the same topological machinery used in [2]. The weak witness complex depends on a parameter $v \in \{0, 1, 2\}$. Roughly speaking, a higher v-value means that the witness complex connects more quickly. De Silva and Carlsson in [3] find $v = 0$ to be generally less effective, and $v = 2$ has the disadvantage of connecting every landmark point to at least one other at $R = 0$, which is problematic for our data set when a cluster of patches contains only one landmark. Hence we choose $v = 1$.

4. Results. Many of our persistent homology results for a core subset $X^m(300, 30)$ can be illuminated by the following projection. We change $X^m(300, 30)$ to the DCT basis for $m \times m$ patches, with DCT basis vectors normalized as in Steps 6 and 7 of section 2. Two of the DCT basis vectors are horizontal and vertical linear gradients (see Figure 6), and we project onto these two coordinates (the x and y axes, respectively, in Figures 9, 12, and 13).

Figure 6. *Horizontal and vertical linear gradient DCT basis vectors for the* 5×5 *patch case.*

There are several long intervals in the $Betti_0$ barcode plot for the core subset $X^3(300, 30)$ (see Figure 7), which is evidence of disjoint clusters. With one exception, these clusters are centered on binary approximations of linear step edges, a few of which are shown in the bottom row of Figure 8. The exceptional cluster, at the far right in Figure 9, centers on the horizontal linear gradient. This is consistent with the observation that many range patches are shots of the ground.

Why do range images cluster around binary patches while optical images are more continuously distributed? Perhaps objects in an indoor or outdoor scene are more likely to be a constant distance from the camera than to be monochromatic, and thus an edge between two objects is more likely to produce a binary range patch than a binary optical patch. A second possible explanation lies in the subresolution properties of range scanners and optical cameras described by Lee, Pedersen, and Mumford in [7]. Range scanners record subpixel detail in a single pixel value by selecting the minimum; digital cameras take an average. A patch with subpixel binary values will remain binary after pixel values are chosen via minimums, but may not if pixels are chosen via averages.

The homology cycle producing the longest interval in the $Betti_1$ barcode of Figure 7 is the primary circle of Figure 3. We see pieces of this primary circle in Figure 9. However, the shorter $Betti_1$ intervals cloud the picture and vary with a different selection of 50 landmark points. The topology of the core subset $X^3(300, 30)$ is not clear.

Range images tend to cluster near binary patches, and there are relatively few 3×3 binary patches. Therefore we consider 5×5 and 7×7 patches in hopes of finding a manifold model. In Figures 10 and 11 are sample Betti barcode plots for the core subsets $X^5(300, 30)$ and $X^7(300, 30)$.

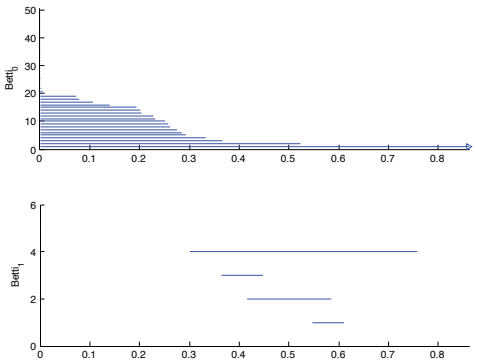

Figure 7. *Barcodes for* $X^3(300, 30)$.

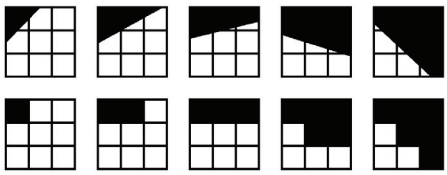

Figure 8. *Linear step edges in the top row and their binary approximations beneath.*

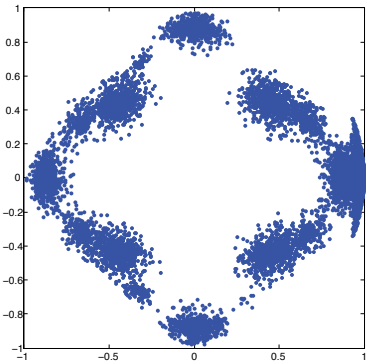

Figure 9. *Projection of* $X^3(300, 30)$ *onto linear gradients.*

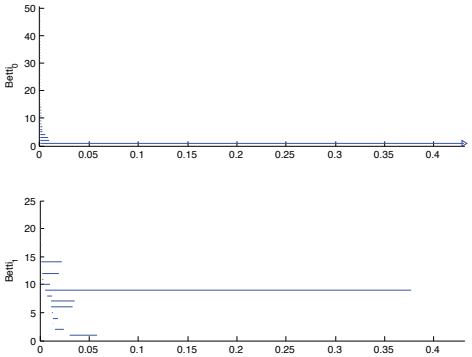

Figure 10. *Barcodes for* $X^5(300, 30)$.

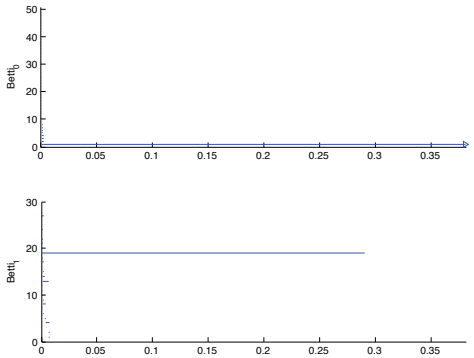

Figure 11. *Barcodes for* $X^7(300, 30)$.

Both the $X^5(300, 30)$ plot and the $X^7(300, 30)$ plot contain a single long $Betti_0$ interval and a single long $Betti_1$ interval, evidence of circular topology. The homology cycle producing the $Betti_1$ interval is the primary circle, visible in Figures 12 and 13. We ran 25 trials each on

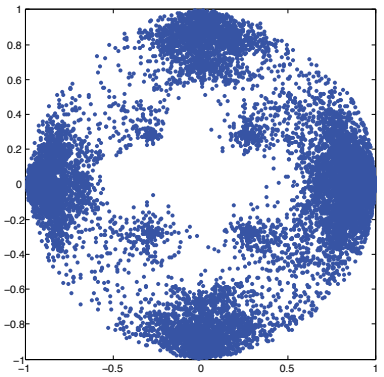

Figure 12. *Projection of* $X^5(300, 30)$ *onto linear gradients.*

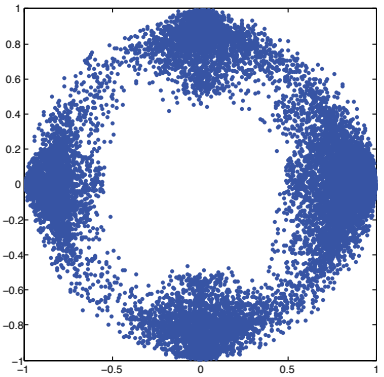

Figure 13. *Projection of* $X^7(300, 30)$ *onto linear gradients.*

$X^5(300, 30)$ and $X^7(300, 30)$, selecting different landmark points. In each trial, the circular profile $Betti_0 = Betti_1 = 1$ is obtained for a range of R-values of length greater than 0.25, and no other Betti plot interval has length greater than 0.05.

5. Application to compression. The understanding of the statistical distribution of range image patches can be used to devise methods of compression of range patch data. We first demonstrate how one can use the intuition that a given patch is well approximated by one containing only two values, foreground or background. Rather than storing all patch values, one should recoordinatize as follows.

- Compute the mean value μ of the range patch.
- Compute the standard deviation

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}},$$

  where there are n pixels in the patch and the x_i's are the pixel values. The values $\mu \pm \sigma$ are now a useful approximation to the two values taken by the range patch.
- Each patch will be stored as the quadruple $(\mu, \sigma, \{\varepsilon_i\}, \{\beta_i\})$. The mean μ is a real number, the standard deviation σ is a nonnegative real number, the residue vector $\{\varepsilon_i\}$ has one entry for each pixel, and the direction vector $\{\beta_i\}$ has coordinates in $\{0, \pm 1\}$. Given a patch with values x_i, compute values ε_i and β_i as follows. For each i, determine if x_i is greater than or less than μ. If $x_i < (>)\ \mu$, then we set $\beta_i = -1$ $(+1)$. If $x_i = \mu$, we set $\beta_i = 0$. Finally, we set $\varepsilon_i = x_i - (\mu + \beta_i \sigma)$.

It is clear how to decompress the patch given the quadruple $(\mu, \sigma, \{\varepsilon_i\}, \{\beta_i\})$. The information about the values of a range patch means that the values ε_i will tend to be small in absolute value.

If we are dealing with 5×5 patches or larger, since the density analysis is consistent with the optical result that the most frequently occurring high-contrast patches are linear patches, then one should be able to perform a further compression along the lines of the wedgelet method of Donoho [4].

6. Conclusions. The primary circle is a good model for core subsets of 5×5 and 7×7 range image patches, providing evidence that the behaviors of optical patches and larger scale range patches may really be quite similar. This suggests that were one to attempt to develop sophisticated compression schemes for range images, methods such as wedgelets [4] could indeed be extended to the range image case. However, the binary bias of range patches suggests that wavelet-based schemes for compression might use different "mother wavelets" for the encoding. Further investigation may determine if range patches also tend to lie along other portions of the Klein bottle model for optical patches of Carlsson et al. [2].

REFERENCES

[1] M. A. Armstrong, *Basic Topology*, Springer, New York, Berlin, 1983.
[2] G. Carlsson, T. Ishkhanov, V. de Silva, and A. Zomorodian, *On the local behavior of spaces of natural images*, Int. J. Comput. Vis., 76 (2008), pp. 1–12.
[3] V. de Silva and G. Carlsson, *Topological estimation using witness complexes*, in Proceedings of the Symposium on Point-Based Graphics, ETH, Zürich, Switzerland, 2004, pp. 157–166.
[4] D. Donoho, *Wedgelets: Nearly minimax estimation of edges*, Ann. Statist., 27 (1999), pp. 859–897.
[5] A. Hatcher, *Algebraic Topology*, Cambridge University Press, Cambridge, UK, 2002.
[6] J. Huang, A. B. Lee, and D. Mumford, *Statistics of range images*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2000, pp. 324–332.
[7] A. B. Lee, K. S. Pedersen, and D. Mumford, *The nonlinear statistics of high-contrast patches in natural images*, Int. J. Comput. Vis., 54 (2003), pp. 83–103.
[8] L. Vietoris, *Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen*, Math. Ann., 97 (1927), pp. 454–472.
[9] A. Zomorodian and G. Carlsson, *Computing persistent homology*, Discrete Comput. Geom., 33 (2005), pp. 247–274.