

# Persistence Codebooks for Topological Data Analysis

Bartosz Zieliński<sup>1</sup> Mateusz Juda<sup>1</sup> Matthias Zeppelzauer<sup>2</sup>

## Abstract

Topological data analysis, such as persistent homology has shown beneficial properties for machine learning in many tasks. Topological representations, such as the persistence diagram (PD), however, have a complex structure (multiset of intervals) which makes it difficult to combine with typical machine learning workflows. We present novel compact fixed-size vectorial representations of PDs based on clustering and bag of words encodings that cope well with the inherent sparsity of PDs. Our novel representations outperform state-of-the-art approaches from topological data analysis and are computationally more efficient.

## 1. Introduction

Topological data analysis (TDA) provides a powerful framework for the structural analysis of high-dimensional data which currently gains increasing importance in natural data analysis (Ferri, 2017). An important tool in TDA is Persistence Homology, PH (Edelsbrunner et al., 2002) which provides methods to study qualitative properties of objects based on the theory of size functions introduced by Frosini et al. (Frosini, 1992; Verri et al., 1993). PH has gained increasing importance in data science, since efficient algorithms exist for its computation (Edelsbrunner & Harer, 2010; De Silva et al., 2011; Chen & Kerber, 2011; Bauer et al., 2017; Dey et al., 2016).

PH has recently been successfully applied to computer vision problems, such as shape and texture analysis (Li et al., 2014; Reininghaus et al., 2015), 3D surface analysis (Adams et al., 2017; Zeppelzauer et al., 2017), 3D shape matching (Carrière et al., 2015), mesh segmentation (Skraba et al., 2010), and motion analysis (Vejdemo-Johansson et al., 2015). Further application areas include time series analysis (Seversky et al., 2016), music tagging (Liu et al., 2016)

<sup>1</sup>Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland <sup>2</sup>Media Computing Group, Institute of Creative Media Technologies, St. Pölten University of Applied Sciences, St. Pölten, Austria. Correspondence to: Bartosz Zieliński <bartosz.zielinski@uj.edu.pl>.

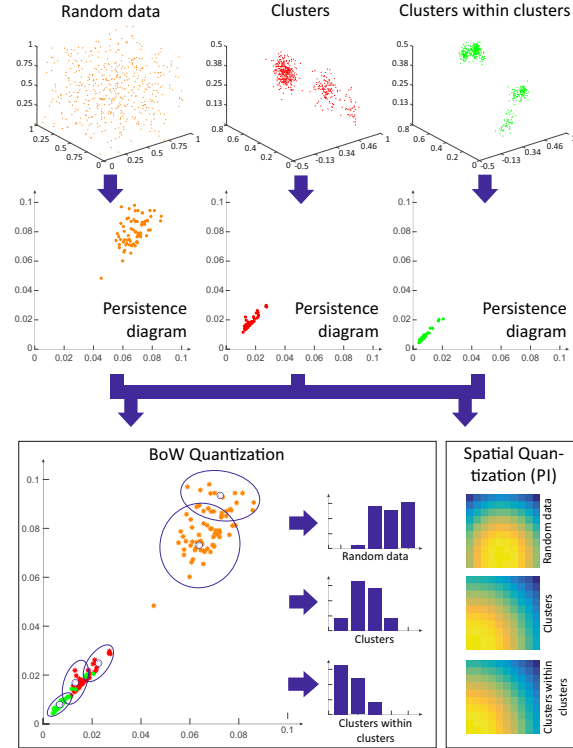


Figure 1. Point clouds of data distributions with different topological structure from a synthetic dataset together with their persistence diagrams. Spatial quantization (persistence image, PI) generates very similar representations especially for “clusters” and “clusters within clusters”. The persistence bag of words (PBoW) quantization proposed in this paper yields a more selective sampling of the space and thus more discriminative representations.

and social-network analysis (Hofer et al., 2017) as well as applications from the bio-medical domain, e.g. biomolecular analysis (Cang & Wei, 2017), brain network analysis (Lee et al., 2012), protein investigation (Gameiro et al., 2015) and material science (Nakamura et al., 2015).

The idea behind PH is to gradually build a topological space upon the input data by growing shapes from the input data. During this process, it builds a sequence of topological spaces called a filtration and tracks when new holes are born and when existing holes die. The length of the associated birth-death intervals (persistence intervals) indicates if the holes are structurally relevant or more likely to be noise.

The standard representation in persistence homology is the persistence diagram (PD) (Edelsbrunner et al., 2002) as well as later introduced persistence barcodes (Carlsson et al., 2005) which are equivalent to PDs. The PD is a two-dimensional representation that visualizes the lifetime of all identified holes during the filtration. Each point in the PD represents a topological structure and its location in the diagram specifies its life span (birth time goes along the horizontal axis and death time along the vertical axis). Examples of PDs for three different input point clouds is shown in Figure 1. The PD compactly represents the geometric properties of multi-dimensional data in a simple two-dimensional representation.

The PD is a complex data structure which is not easy to integrate in common data analysis and machine learning workflows because it has variable length and multiset structure.

**Related work.** A number of approaches have been introduced for PDs to make them compatible with statistical analysis and machine learning methods. There are basically two different types of approaches: kernel-based approaches and vectorized representations.

The goal of **kernel-based approaches** is to define computationally efficient metrics (more precisely kernel functions) to compare PDs and thereby make them compatible with kernel-based machine learning methods, such as kernel support vector machines (SVMs) and kernel PCA. Li et al. combine the traditional bag-of-features approach with PDs (Li et al., 2014). The authors use various distance functions between 0-dimensional persistence diagrams (bottleneck distance, Wasserstein distances, distance function for persistence landscape) to generate kernels and thereby integrate them into the classification process. On different datasets (SHREC 2010, TOSCA, hand gestures, Outex) they show that topological information is complementary to the information captured by BoF. Reininghaus et al. propose an SVM kernel for persistence diagrams to make them compatible with statistical machine learning methods (Reininghaus et al., 2015). The authors smooth the PDs using Gaussian kernels. They apply topological descriptors together with the novel kernel to shape retrieval and texture classification. Carrière et al. proposes another kernel based on Sliced Wasserstein approximation of the Wasserstein distance (Carrière et al., 2017). The authors show that the kernel is not only stable, but also mimics bottleneck distances between persistence diagrams. They further develop an approximation technique to reduce the kernel computation time. They apply it to 3D shape segmentation, texture classification, and orbit recognition in dynamical systems. A widely used approach for the representation of PDs is Persistence Landscape, PL (Bubenik, 2015). PL is a functional representation of a PD where the PD is transformed into a

set of real-valued, piecewise linear, Lipschitz functions. The representation lies in a Banach space and is stable. To compare two landscapes, the authors have introduced a distance measure in (Bubenik, 2015) which can be used, similarly as the approaches above, to compute a kernel matrix. Note that PL can also be transformed into a fixed-length vectorized representation by sampling the values of the landscape function. Previous results, however, have shown that the use of the distance measure is favorable (Bubenik & Dłotko, 2017).

Approaches for **vectorized representations** derive fixed-sized variants of the PD that can easily be used in current machine learning methods. One approach to leverage topological information for machine learning tasks is the persistent image (PI) proposed by Adams et al. (Adams et al., 2017). It is a vector-based representation of PD which is built upon earlier work on size functions (Ferri et al., 1998; Donatini et al., 1998). The PI can be employed directly as a feature vector in conventional machine learning techniques. PI lives in Euclidean space, which makes it easy to combine with other (non-topological) descriptors. Anirudh et al. propose an approach based on Riemannian manifold (Anirudh et al., 2016). Their method first transforms a PD into a Gaussian kernel which is a Riemannian manifold with Fisher-Rao metric. Next, they transform the manifold to a vector space and use PCA to reduce its dimension to a fixed size which allows to use SVM more effectively.

Recently, a third type of approach has been introduced, which tries to learn which points in the PD are of particular importance for the given task in a supervised manner (Hofer et al., 2017). This approach has promising properties, however, can only be used in cases where supervisory information (labels) is available during the construction of the representation. In our work we focus on representations that can be build independently from class labels.

Proposed approaches exhibit different advantages and disadvantages. Kernel-based approaches are well-defined in theory, in practice they become highly inefficient when the number of training vectors becomes large, as the entire kernel matrix must be computed explicitly (in case of SVM leading to roughly cubic complexity). Furthermore, they are limited to those data analysis approaches that allow for the use of kernels. Vectorized representations are compatible with a much wider range of methods and applicable even to large-scale data. They, however, may lack in representational power due to the required spatial quantization of the PD. The reason for this is that they do not cope well with the sparseness of the PDs.

**Contribution.** We present a novel more adaptive and thus more precise representation for PDs which aims at combining the advantages of kernel-based approaches and vector-

ized representations. We extend popular bag of word (BoW) encodings to PH to cope with the inherent sparsity of PDs (see Figure 1 for an illustration). The proposed approach generates powerful discriminative representations and results in a universally applicable fixed-sized feature vector of low-dimension. Furthermore, it can be computed efficiently.

We evaluate our representations on synthetic as well as on real-world data and compare them to other state-of-the-art methods (kernels and vectorized representations). Results show that we obtain comparable or even better performance in significantly less time.

The remaining paper is structured as follows. We provide the necessary background on persistence homology in Section 2. The novel representations are introduced in Section 3. Finally, we describe the experimental setup and the results in Section 4 and conclude our work in Section 5.

## 2. Background on Persistent Homology

In this section we present basic background on persistent homology. We refer the interested reader to (Edelsbrunner et al., 2002; Zomorodian & Carlsson, 2005; Edelsbrunner & Harer, 2010) for more information.

Topological spaces appearing in data analysis are typically constructed from small pieces called cells. Natural tools in the study of multi-dimensional point cloud data (see first row of Fig. 1) with topological methods are simplices. They are building blocks for structures called simplicial complexes. A *simplex*  $\sigma$  is a non-empty set of points and a *simplicial complex*  $\mathcal{X}$  is a set of simplices, such that if  $\sigma \in \mathcal{X}$  and a simplex  $\tau \subset \sigma$ , then  $\tau \in \mathcal{X}$ . For a point cloud data  $X$  with a metric  $d$  and a parameter  $r \in \mathbb{R}$  one can define a *Vietoris-Rips complex*  $\mathcal{X}_r$  as a simplicial complex whose each simplex  $\sigma = \{v_0, v_1, \dots, v_k\}$  satisfies  $d(v_i, v_j) \leq r$ . Simplicial complexes, in particular Vietoris-Rips complexes, give topology a combinatorial flavour and make it a natural tool in the study of multi-dimensional data sets.

In general an arbitrary function  $f : X \rightarrow \mathbb{R}$ , called the Morse function or the measurement function, may be used to control the order in which the complex is built, starting from low values of  $f$  and increasing subsequently. This way we obtain a sequence of topological spaces, called a filtration,

$$\emptyset = \mathcal{X}_{r_0} \subset \mathcal{X}_{r_1} \subset \mathcal{X}_{r_2} \subset \dots \subset \mathcal{X}_{r_n} = \mathcal{X},$$

where  $\mathcal{X}_r := f^{-1}((-\infty, r])$  is a simplicial complex and  $r_i$  is a growing sequence of values of  $f$  at which the complex changes. As the space is gradually constructed, holes are born, persist for some time and eventually may die. The length of the associated birth-death  $[b_i, d_i]$  intervals (persistence intervals) indicates if the holes are relevant or merely noise. The lifetime of holes is usually visualized by the

persistence diagram (PD).

Persistence diagrams constitute the main tool of topological data analysis. They visualize geometric properties of a multi-dimensional object in a simple two-dimensional diagram.

The PD is an expressive representation of the qualitative features in a data set. Unfortunately, the structure of the PD is complex and cannot easily be combined with traditional data analysis and machine learning methods which usually require fixed-size input vectors. This hinders the application of computational topology to machine learning tasks such as computer vision.

In this paper we introduce a novel fixed-size representation of PDs that leverages the strong representative power of bag of words quantization as well as first- and second-order bag of words encodings.

## 3. Bag of Words Quantizations for Persistent Homology

Motivated by the great success of bag of word approaches (Sivic & Zisserman, 2003) in computer vision, we present three novel representations of PDs based on bag of words quantization and different types of word encodings: traditional bag of words (BoW) encoding, vector of locally aggregated descriptors (VLAD), and Fisher vector.

Input to all three approaches is a set of PDs extracted for all (or a subset of) instances of a given training data set. In a first step we rotate the PDs by 45 degrees. This transforms the points in the PD into pairs of birth (horizontal) and persistence (vertical):  $(b_i, d_i - b_i)$ . We will refer this representation as  $\widetilde{PD}$  in the following. The rotated PD has favorable properties when it comes to the clustering (code-word quantization, see Section 3.3). In the following, we cluster the points in  $\widetilde{PD}$  to build persistence codebooks and compute different word encodings from it as described in the following sections. Based on the encodings different codeword histograms can be computed which provide an adaptive fixed-size representation for PDs.

### 3.1. Persistence Bag of Words

Let  $\widetilde{PD} = \{x_t = (b_t, d_t - b_t) \in \mathbb{R}^2, t = 1..T\}$  be the multiset of birth-lifetime pairs. Similarly to traditional bag of words (Sivic & Zisserman, 2003), the *persistence bag of words* (PBoW) assigns points of  $\widetilde{PD}$  to the precomputed codebook  $\{\mu_i \in \mathbb{R}^2, i = 1..k\}$ . The codebook is obtained by  $k$ -means clustering on all the points from all PDs from the training set selected for codebook construction.

Let  $NN(x_t)$  be the nearest word (cluster center) in the codebook, then the PBoW representation is obtained as the histogram of the assignments  $\mathbf{v} = (v_i = |x_t : NN(x_t) =$

$i|)_{i=1..k}$ . This  $k$ -dimensional vector can be normalized in various ways. In our experiments, we use square-rooting normalization, which has been found effective for BoW representations (Perronnin et al., 2010).

### 3.1.1. WEIGHTING FUNCTION

The straight application of BoW to  $\widetilde{PD}$  would neglect an important property of persistence diagrams. A typical interpretation is that points in a PD with high persistence are more important than points with low persistence because points with low persistence (short lifetime) are likely to correspond to unimportant structures originating from noise. This important property should be considered during codebook generation. We integrate this assumption, by performing  $k$ -means clustering on a subset of points obtained by weighted sampling. We define a piecewise linear weighting function which depends only on the second coordinate (persistence) in the  $\widetilde{PD}$ . Given  $a, b > 0$ , we define the weighting function  $w_{a,b} : \mathbb{R} \rightarrow \mathbb{R}$  as follows:

$$w_{a,b}(t) = \begin{cases} 0 & \text{if } t < a \\ (t - a)/(b - a) & \text{if } a \leq t < b \\ 1 & \text{if } t \geq b \end{cases}$$

In our experiments we set  $a$  and  $b$  to the persistence values corresponding to 0.05 quantile of birth time and 0.95 quantile of death time, respectively. Example  $\widetilde{PD}$  with standard (unweighted) and weighted codebooks are shown in Fig. 2. The unweighted clustering produces larger and less adaptive clusters to the strongest topological structures. The weighted clustering yields a more adaptive codebook with a more uniform sampling of the space. Points with strong persistence have more impact during clustering which leads to separate codewords generated for such points or point clusters.

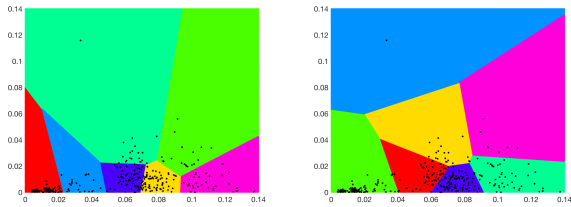


Figure 2. Clustering results for a set of  $\widetilde{PD}$ s with  $k = 7$  codewords (left: unweighted clustering, right: weighted clustering). Colors encode the partition of the space by the individual clusters (codewords). Weighting gives more importance to the points exhibiting strong topological structures and yields more balanced cluster sizes. Note that a cluster can also be created for only one point, as is the case here for the point with the longest lifetime and thus the most persistent topological structure.

### 3.2. Persistence VLAD

*Persistence VLAD* (PVLAD) is based on vector of locally aggregated descriptors (VLAD) by (Jégou et al., 2010). The first computation step is similar to PBoW: a codebook  $\{\mu_i \in \mathbb{R}^2, i = 1..k\}$  is learned using  $k$ -means clustering and each point  $x_t$  is associated to its nearest codeword  $NN(x_t)$ . In the second step, the accumulated distance from points  $x_t$  to the nearest codeword is computed:  $v_i = \sum_{x_t: NN(x_t)=i} x_t - \mu_i$ , and then concatenated, resulting in a  $2k$ -dimensional vector  $\mathbf{v}$ . Intuitively, this vector should capture more information than PBoW, because it encodes the first order moments of the points assigned to a codeword instead of simply counting those points. This idea is illustrated in Fig. 3 for  $\widetilde{PD}$ s of two classes of shapes (“random data” and “clusters”). The same codebook as in Fig. 2 is used for easier comparison. See also Fig. 1 for example point cloud distributions and PDs of both shape classes.

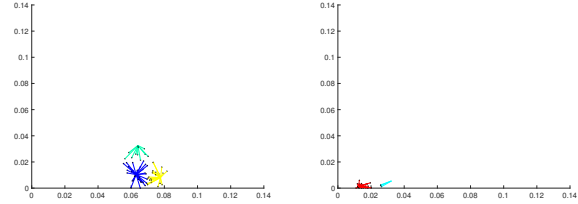


Figure 3. Quantization the plot of  $\widetilde{PD}$  from our synthetic dataset by persistence VLAD (left: “random data”, right: “clusters”). PVLAD is constructed by summing up vectors pointing from cluster center to points assigned to this cluster.

Similarly like in case of PBoW, PVLAD can be modified in order to give more impact to points with stronger persistence. For this purpose, the codebook can be obtained by running  $k$ -means clustering for a subset of points obtained by weighted sampling.

### 3.3. Persistence Fisher Vector

The idea of *persistence Fisher vector* (PFV), similarly like in (Perronnin & Dance, 2007), is to characterize  $\widetilde{PD}$  with a gradient vector derived from a generative probability model (obtained for all  $\widetilde{PD}$ s selected for codebook generation). This is similar to PVLAD, however in case of PFV one computes not only the first order, but also the second order moments of the points assigned to a codeword (i.e. not only the gradient for  $\mu_i$  but also for  $\Sigma_i$ ).

Let  $\lambda$  be the set of parameters of Gaussian mixture models (GMMs) obtained for all input  $\widetilde{PD}$ s:

$$\lambda = \{w_i, \mu_i, \Sigma_i, i = 1..k\},$$

where  $w_i$ ,  $\mu_i$  and  $\Sigma_i$  denote the weight, Gaussian center and

covariance matrix of Gaussian  $i$ , respectively. Let  $p_i(x_t|\lambda)$  be the likelihood that point  $x_t$  was generated by Gaussian  $i$ :

$$p_i(x_t|\lambda) = \frac{\exp\{-\frac{1}{2}(x_t - \mu_i)' \Sigma_i^{-1} (x_t - \mu_i)\}}{(2\pi)^{\frac{D}{2}} |\Sigma_i|^{\frac{1}{2}}},$$

where  $D = 2$  in case of  $\widetilde{PD}$ . Let  $\mathcal{L}(\widetilde{PD}|\lambda) = \log p(\widetilde{PD}|\lambda)$ , where under the independence assumption:

$$\mathcal{L}(\widetilde{PD}|\lambda) = \sum_{x_t} \log p(x_t|\lambda),$$

and the likelihood that point  $x_t$  was generated by the GMM is:

$$p(x_t|\lambda) = \sum_{i=1}^N w_i p_i(x_t|\lambda).$$

Assuming that the covariance matrices are diagonal, the derivations  $\frac{\partial \mathcal{L}(\widetilde{PD}|\lambda)}{\partial \mu_i^d}$  and  $\frac{\partial \mathcal{L}(\widetilde{PD}|\lambda)}{\partial \sigma_i^d}$  (where  $\sigma_i^d = \text{diag}(\Sigma_i)$  and superscript  $d$  denotes the  $d$ -th dimension of a vector) can be effectively computed (Perronnin & Dance, 2007). In this case the gradient vector is just a concatenation of the partial derivatives with respect to all the parameters. Note that, rotating the input PDs by 45 degrees facilitates the use of diagonal covariance matrices. Fig. 4 shows an example of a GMM generated on the  $\widetilde{PD}$  point distribution from Fig. 2.

To normalize the dynamic range of the different dimensions of the gradient vectors, the diagonal of the Fisher information matrix  $F$  is computed. The normalized partial derivatives equal  $f_{\mu_i^d}^{-1/2} \frac{\partial \mathcal{L}(X|\lambda)}{\partial \mu_i^d}$  and  $f_{\sigma_i^d}^{-1/2} \frac{\partial \mathcal{L}(X|\lambda)}{\partial \sigma_i^d}$ , where  $f_{\mu_i^d}$  and  $f_{\sigma_i^d}$  are the corresponding terms on the diagonal of  $F$ . Vector  $\mathbf{v}$  is the concatenation of  $k$  components containing  $D$  values for every Gaussian component (gradient for  $\mu_i$  and gradient for  $\Sigma_i$ ), therefore it is  $2Dk$ -dimensional. While in case of PD ( $D = 2$ ), the dimension of vector  $\mathbf{v}$  equals  $4k$ .

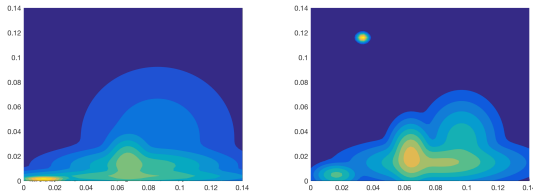


Figure 4. Gaussian mixture model for  $k = 5$  (left: unweighted, right: weighted) generated based on the  $\widetilde{PD}$ s from Fig. 2. As the result of weighting, a Gaussian can be created for only one point, as in the case here for the point with the longest lifetime.

Similarly like in case of PBoW, PFV can be modified in order to give more impact to points with stronger persistence.

For this purpose, the GMM can be obtained by running Expectation Maximization algorithm (Dempster et al., 1977) on the subset of points obtained by weighted sampling.

## 4. Experiments and Results

We evaluate our representation by comparing them with state-of-the-art approaches from TDA for two different tasks. Firstly, the discrimination of synthetically generated shape classes by Adams et al. (Adams et al., 2017) and secondly, the classification of different activities in 3D motion capture data published in (Ali et al., 2007). We compare our method with different kernel-based approaches: 2-Wasserstein distance<sup>1</sup> (Kerber et al., 2017), the multi-scale kernel<sup>2</sup> (Reininghaus et al., 2015), and sliced Wasserstein kernel<sup>3</sup> (Carrière et al., 2017). Furthermore, we employ the persistence landscape<sup>4,5</sup> representation and generate a kernel matrix by the distance metric defined in (Bubenik, 2015). Additionally, we employ two vectorized representations: persistence image<sup>6</sup> (Adams et al., 2017) and the Riemannian manifold approach<sup>7</sup> (Anirudh et al., 2016).

The code of all experiments is written in Matlab. For external approaches we use the publicly available implementations of the original authors. For clustering and bag of words encoding we employ the VLFeat library (Vedaldi & Fulkerson, 2008). Our code is available at <https://github.com/bziiuj/pcodebooks>.

### 4.1. Synthetic Shape Dataset

For the initial comparison of the different representations and to verify the expressive power of our approach we use the synthetic dataset introduced by Adams et al. (2017). It consists of six shape classes: a random data from unit cube, a circle of diameter one, a sphere of diameter one, three clusters with centers randomly chosen in the unit cube, three clusters within three clusters (where the centers of the minor clusters are chosen as small perturbations from the major cluster centers), and a torus (see Fig. 5 for example shapes). The dataset contains 50 point clouds with 500 points for each of the six classes with the level of  $\sigma$  in Gaussian noise equals 0.1. This gives 300 point clouds in total.

We then compute the  $H_1$  PDs for a Rips filtrations obtained from the point clouds. We use an approximating method proposed by Dey et al. (Dey et al., 2016) and SimBa imple-

<sup>1</sup>[https://bitbucket.org/grey\\_narn/hera](https://bitbucket.org/grey_narn/hera)

<sup>2</sup><https://github.com/rkwitt/persistence-learning>

<sup>3</sup>code obtained from Mathieu Carrière

<sup>4</sup><https://www.math.upenn.edu/~dlotko/persistenceLandscape.html>

<sup>5</sup><https://github.com/queenBNE/Persistent-Landscape-Wrapper>

<sup>6</sup><https://github.com/CSU-TDA/PersistenceImages>

<sup>7</sup><https://github.com/rushilanirudh/pdsphere>



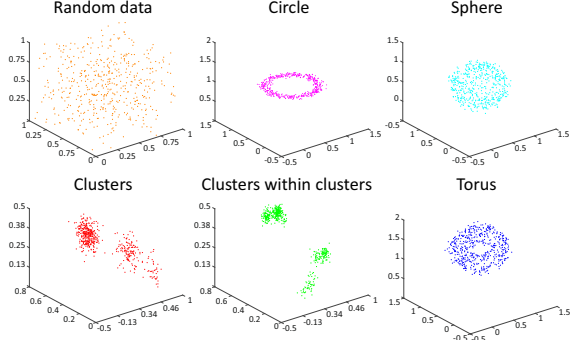


Figure 5. Example shapes from the 6 shape classes of the syntactic dataset.

mentation based on the paper developed by Dayu Shi<sup>8</sup>. We focus on the  $H_1$  PDs (cycles) because they best express the internal structure in the data and yield the most promising results in related works (Adams et al., 2017).

Similarly like in (Adams et al., 2017), our goal is to compare different PD representations with this dataset. For each representation, we generate distance matrices of size  $300 \times 300$  and use them as input for  $k$ -medoids clustering with  $k = 6$  (the number of shape classes). We use partitioning around medoids (PAM), which is the classical algorithm for solving the  $k$ -medoids problem (Kaufman & Rousseeuw, 1990). The score of such a clustering is the sum of the distances from each point to its closest medoid (the desired clustering is the one with the minimal clustering score). A typical approach to search for the global minimum is to cluster many times (in our case 100) with different random initial selections and return the identified final clustering as the one with the lowest score. The final accuracy is computed as the percentage of point clouds identified with a medoid of the same class. We repeat this procedure 50 times with different random seeds. In order to obtain optimal parameters for tested approaches, we used grid search methodology with adequate ranges presented in Table 1. For all approaches where weighting is an option, we evaluated with and without weighting.

The clustering scores for all evaluated PD representations and kernels are presented in Table 1. Persistence fisher vector significantly outperforms all vectorized representations while being only slightly worse than the best kernel-based approach. Fig. 6 shows how the score of the codebook representations depends on the number of clusters. PBoW improves with increasing number of  $k$  and yields an optimum around  $k = 20$ , representing a 20-dimensional vector. Larger codebook sizes do not further improve performance. We assume that mostly redundant information is added to the vector with larger  $k$ . PFV yields the best overall re-

sults and has its peak at  $k = 10$ . This corresponds to a 40-dimensional vector (PFV dimension is  $4k$ ). For larger  $k$  performance slightly decreases but stays at a high level. PVLAD is not able to compete with the other representations. We assume that the reason for the low performance of PVLAD is the hard assignment of the first-order information (distances to clusters). Soft assignment strategies as in PFV may mitigate this problem. For PVLAD, however, weighting has the most positive influence. For the other representations the effect of weighting decreases with increasing  $k$ . A higher  $k$  seems to compensate the weighting due to the finer spatial partition of the space.

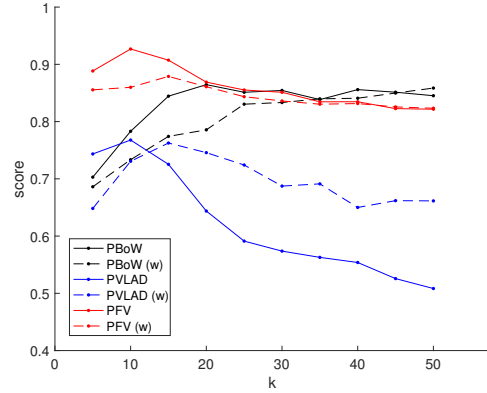


Figure 6. Scores for the three proposed representations (PBoW, PVLAD, and PFV with weighting (dashed lines) and without weighting (solid lines) for  $k = 5..50$ . PFV and PBoW clearly outperform PVLAD. Optimal results are obtained around  $k = 10$  for PFV and for  $k = 20$  for PBoW. Weighting brings most benefit for PVLAD.

Finally, we investigate if the derived vectorial representations (codebook histograms) capture the different structures in the six classes well. Figure 7 shows averaged PBoW histograms for all six classes over all instances. We select PBoW for visualization due to their easier interpretability. We can clearly see different distributions in the histograms for the different classes. Remarkably, the classes “clusters” and “clusters within clusters” have clearly different distributions which shows that the hierarchical cluster structure can be modeled well. For classes “Random data” and “Sphere” the distributions of PDs are similar, what results in similar PBoWs.

## 4.2. Motion Capture Dataset

We further evaluate all approaches on real-world data representing 3-dimensional motion capture sequences of body joints (Ali et al., 2007). The dataset contains a collection of five actions: dancing, jumping, running, sitting and walking with 31, 14, 30, 35 and 48 instances respectively. For each action, a set of 19 3D motion trajectories (each cor-

<sup>8</sup><http://web.cse.ohio-state.edu/~dey.8/SimBa/Simba.html>

Table 1. The average clustering scores for all evaluated PD representations together with the standard deviation, obtained by repeating the procedure 50 times. Upper and lower part of the table contains kernel-based approaches and vectorized representations, respectively. The optimal parameters are marked bold ( $r$  stands for resolution,  $\sigma$  in Gaussian function,  $k$  is the number of words, while  $d$  stands for dimension).

REPRESENTATION	PARAMETERS (IF ANY)	SCORE
2-WASSERSTEIN DISTANCE		$0.79 \pm 0.00$
MULTI-SCALE KERNEL	$\sigma = \{0.5, \mathbf{1}, 1.5\}$	$0.58 \pm 0.00$
SLICED WASSERSTEIN KERNEL	$n = \{\mathbf{50}, 100, 150, 200, 250\}$	<b><math>0.95 \pm 0.00</math></b>
PERSISTENCE LANDSCAPE		$0.67 \pm 0.00$
PERSISTENCE IMAGE	$r = \{\mathbf{10}, 20, 30, 40, 50\}; \sigma = \{\mathbf{0.05}, 0.1, 0.15, 0.2, 0.25\};$ WEIGHTING	$0.77 \pm 0.00$
RIEMANNIAN MANIFOLD	$r = \{\mathbf{20}, 40\}; \sigma = \{\mathbf{0.1}, 0.2, 0.3\}; d = \{\mathbf{25}, 50, 75, 100\}$	$0.83 \pm 0.00$
PBoW	$k = \{10, 20, 30, \mathbf{40}, 50\};$ WEIGHTING	$0.86 \pm 0.04$
PVLAD	$k = \{10, \mathbf{20}, 30, 40, 50\};$ WEIGHTING	$0.76 \pm 0.06$
PFV	$k = \{\mathbf{10}, 20, 30, 40, 50\};$ NO WEIGHTING	<b><math>0.94 \pm 0.04</math></b>

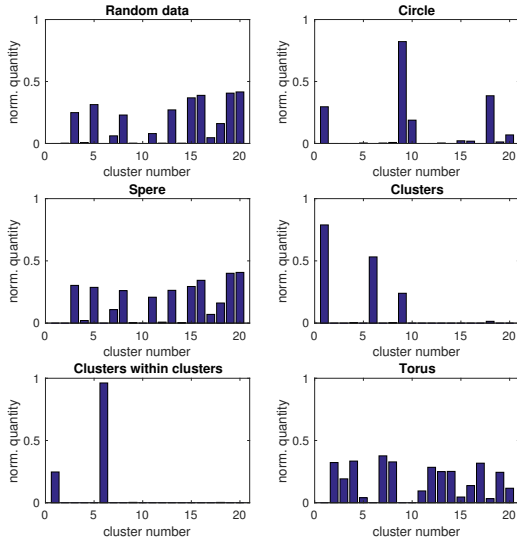


Figure 7. Averaged persistence bag of word histograms for the different shape classes from the synthetic dataset (with a codebook size of  $k = 20$  and without weighting). The different shapes show clearly different codeword distributions. The similar distribution of “Random data” and “Sphere” result from the fact that they produce highly similar PDs

responding to the motion of one joint) are provided. This corresponds to  $57 = 3 \cdot 19$  curves of individual components ( $x$ ,  $y$ , and  $z$  components) for which separate PDs are computed.

For our experiments, with the motion capture dataset we compare with the vectorized representations (PI and the Riemannian manifold). Kernel-based approaches would require the computation of 57 kernel matrices, which is computationally highly demanding. For this reason, we

focus on the vectorized representations which scale much better to such complex data.

For classification we generate 100 random splits where each split has 5 test examples from each action class and use an SVM as classifier. In case of the Riemannian manifold representation (Anirudh et al., 2016), we generate vectorial features by PCA and concatenate them as proposed by the authors. For PI and our persistence bag of words approaches, we use simple concatenation of the 57 vectors obtained for different PDs. We use classification accuracy as performance measure.

Table 2 summarizes the average recognition rates for all evaluated approaches. Overall, PBoW overcomes all fixed-size representations and it is almost five times faster than the Riemann approach and fifty times faster than PI.

Figure 8 shows the performance of the proposed representations for different codebook sizes. The best performance over all codebook sizes is achieved by PBoW (with and without weighting). Similarly, PFV yields constantly good results when weighting is applied. Weighting seems to improve the underlying GMM models. These results further indicates that the codebook size (usually a crucial parameter in bag of word approaches) has only little influence for these representations. PVLAD is clearly outperformed by the other approaches. A closer investigation of the VLAD encoding in the context of  $\widehat{PD}s$  will be a subject of future investigation.

## 5. Conclusion

We have presented three new representations for persistence diagrams based on bag of word quantizations and different word encodings. We have evaluated the representations on synthetic as well as real-world data and compared their performance with several state-of-the-art topological approaches (both kernel-based approaches and vectorized

Table 2. The average recognition rates for different PD representations. In order to obtain optimal parameters, we used grid search for the parameters with the same ranges as in Table 1. We list the best overall results obtained for each representation (with the respective parameters). Each codebook-based representation is evaluated with and without weighting separately.

REPRESENTATION	PARAMETERS (IF ANY)	ACCURACY	TIME
PERSISTENCE IMAGE	$r = 40; \sigma = 0.1$	$87.40 \pm 5.46$	90.75
RIEMANNIAN MANIFOLD	$r = 20; \sigma = 0.2; d = 25$	$91.76 \pm 5.24$	13.50
PBoW (NO WEIGHTING)	$k = 10$	<b><math>92.88 \pm 5.64</math></b>	<b>2.53</b>
PBoW (WEIGHTING)	$k = 40$	$91.92 \pm 4.72$	2.81
PVLAD (NO WEIGHTING)	$k = 50$	$86.12 \pm 5.55$	2.94
PVLAD (WEIGHTING)	$k = 50$	$87.32 \pm 5.43$	2.97
PFV (NO WEIGHTING)	$k = 50$	$89.52 \pm 5.74$	7.53
PFV (WEIGHTING)	$k = 20$	$91.92 \pm 5.54$	6.10

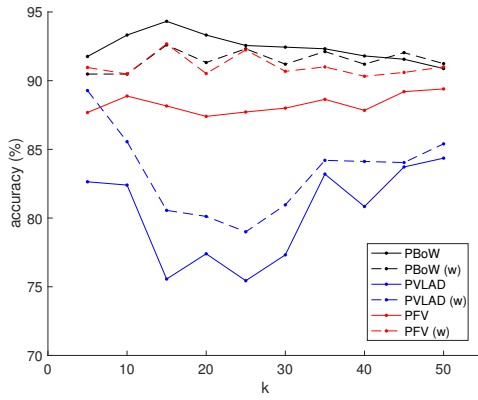


Figure 8. Scores for the three proposed representations (PBoW, PVLAD, and PFV with weighting (dashed lines) and without weighting (solid lines) for  $k = 5..50$  for the motion capture dataset. PBoW clearly outperforms PVLAD and PFV and yields the best result with a codebook dimension of  $k = 15$ . PVLAD is not able to capture the structural differences in the classes. PFV strongly benefits from weighting and yields good results over all codebook sizes.

representations). Our results show that the new representations outperform most alternative approaches and are notably faster. The proposed representations have two major advantages over related ones. Firstly, building a codebook enables our approach to adapt to the structures in the PD and to encode the points in an effective way. This makes the resulting vectorized representation compact and expressive, i.e., the representation does not need to preserve entries in the vector for areas in the PD where no points are available. Secondly, in contrast to related approaches which transform the PDs individually and separately for each input instance, the proposed codebooks can be constructed from a subset or even from the entire training data. Thereby the codebook can model global structures in the data and gets a global support. This facilitates the generation of robust representations. Note that our approaches are still completely unsupervised

and do not need any class labels for the construction of the representation.

## 6. Acknowledgements

The work from this article were supported by the National Science Centre, Poland under grant agreement no 2015/19/D/ST6/01215; 2016-2019.

## References

- Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., and Ziegelmeier, L. Persistence images: a stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18(8):1–35, 2017.
- Ali, S., Basharat, A., and Shah, M. Chaotic invariants for human action recognition. In *ICCV*, pp. 1–8. IEEE Computer Society, 2007. ISBN 978-1-4244-1631-8. URL <http://dblp.uni-trier.de/db/conf/iccv/iccv2007.html#AliBS07>.
- Anirudh, R., Venkataraman, V., Natesan Ramamurthy, K., and Turaga, P. A riemannian framework for statistical analysis of topological persistence diagrams. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 68–76, 2016.
- Bauer, U., Kerber, M., Reininghaus, J., and Wagner, H. Phat–persistent homology algorithms toolbox. *Journal of Symbolic Computation*, 78:76–90, 2017.
- Bubenik, P. Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research*, 16(1):77–102, 2015.
- Bubenik, Peter and Dłotko, Paweł. A persistence landscapes toolbox for topological statistics. *Journal of Symbolic Computation*, 78:91–114, 2017.
- Cang, Z. and Wei, G. Topologynet: Topology based deep convolutional and multi-task neural networks for



- biomolecular property predictions. *PLOS Computational Biology*, 13(7):e1005690, 2017.
- Carlsson, G., Zomorodian, A., Collins, A., and Guibas, L. J. Persistence barcodes for shapes. *International Journal of Shape Modeling*, 11(02):149–187, 2005.
- Carrière, M., Oudot, S., and Ovsjanikov, M. Stable topological signatures for points on 3d shapes. In *Computer Graphics Forum*, volume 34, pp. 1–12. Wiley Online Library, 2015.
- Carrière, M., Cuturi, M., and Oudot, S. Sliced wasserstein kernel for persistence diagrams. In *ICML*, 2017.
- Chen, C. and Kerber, M. Persistent homology computation with a twist. In *Proceedings 27th European Workshop on Computational Geometry*, volume 11, 2011.
- De Silva, V., Morozov, D., and Vejdemo-Johansson, M. Dualities in persistent (co) homology. *Inverse Problems*, 27(12):124003, 2011.
- Dempster, A. P., Laird, N., and Rubin, D. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.
- Dey, T., Shi, D., and Wang, Y. Simba: An efficient tool for approximating rips-filtration persistence via simplicial batch-collapse. In Sankowski, P. and Zaroliagis, C. (eds.), *24th Annual European Symposium on Algorithms, ESA 2016, August 22–24, 2016, Aarhus, Denmark*, volume 57 of *LIPIcs*, pp. 35:1–35:16. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016. ISBN 978-3-95977-015-6. doi: 10.4230/LIPIcs.ESA.2016.35. URL <https://doi.org/10.4230/LIPIcs.ESA.2016.35>.
- Donatini, P., Frosini, P., and Lovato, A. Size functions for signature recognition. In *Proc. SPIE*, volume 3454, pp. 178–183, 1998. doi: 10.1117/12.323253. URL <http://dx.doi.org/10.1117/12.323253>.
- Edelsbrunner, H., Letscher, D., and Zomorodian, A. Topological persistence and simplification. *Discrete and Computational Geometry*, 28:511–533, 2002.
- Edelsbrunner, Herbert and Harer, John. *Computational topology: an introduction*. American Mathematical Soc., 2010.
- Ferri, M. Persistent topology for natural data analysis — a survey. In Holzinger, Andreas, Goebel, Randy, and Palade, Vasile (eds.), *Towards Integrative Machine Learning and Knowledge Extraction*, pp. 117–133, Cham, 2017. Springer International Publishing. ISBN 978-3-319-69775-8.
- Ferri, M., Frosini, P., Lovato, A., and Zambelli, C. Point selection: A new comparison scheme for size functions (with an application to monogram recognition). In *Computer Vision — ACCV’98: Third Asian Conference on Computer Vision Hong Kong, China, January 8–10, 1998 Proceedings, Volume I*, pp. 329–337. Springer, 1998.
- Frosini, P. Measuring shapes by size functions. In *Intelligent Robots and Computer Vision X: Algorithms and Techniques*, pp. 122–133. International Society for Optics and Photonics, 1992.
- Gameiro, M., Hiraoka, Y., Izumi, S., Kramar, M., Mischaikow, K., and Nanda, V. A topological measurement of protein compressibility. *Japan Journal of Industrial and Applied Mathematics*, 32(1):1–17, 2015.
- Hofer, C., Kwitt, R., Niethammer, M., and Uhl, A. Deep learning with topological signatures. In *Advances in Neural Information Processing Systems*, pp. 1633–1643, 2017.
- Jégou, H., Douze, M., Schmid, C., and Pérez, P. Aggregating local descriptors into a compact image representation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 3304–3311. IEEE, 2010.
- Kaufman, L. and Rousseeuw, P. Partitioning around medoids (program pam). *Finding groups in data: an introduction to cluster analysis*, pp. 68–125, 1990.
- Kerber, M., Morozov, D., and Nigmetov, A. Geometry helps to compare persistence diagrams. *Journal of Experimental Algorithmics (JEA)*, 22:1–4, 2017.
- Lee, H., Kang, H., Chung, M., Kim, B., and Lee, D. Persistent brain network homology from the perspective of dendrogram. *IEEE transactions on medical imaging*, 31(12):2267–2277, 2012.
- Li, C., Ovsjanikov, M., and Chazal, F. Persistence-based structural recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2003–2010. IEEE, 2014. doi: 10.1109/CVPR.2014.257.
- Liu, J., Jeng, S., and Yang, Y. Applying topological persistence in convolutional neural network for music audio signals. *arXiv preprint arXiv:1608.07373*, 2016.
- Nakamura, T., Hiraoka, Y., Hirata, A., Escobar, E. G., and Nishiura, Y. Persistent homology and many-body atomic structure for medium-range order in the glass. *Nanotechnology*, 26(30):304001, 2015.
- Perronnin, F. and Dance, C. Fisher kernels on visual vocabularies for image categorization. In *Computer Vision and Pattern Recognition, 2007. CVPR’07. IEEE Conference on*, pp. 1–8. IEEE, 2007.

- Perronnin, F., S  nchez, J., and Xerox, Y. Large-scale image categorization with explicit data embedding. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 2297–2304. IEEE, 2010.
- Reininghaus, J., Huber, S., Bauer, U., and Kwitt, R. A stable multi-scale kernel for topological machine learning. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4741–4748, June 2015. doi: 10.1109/CVPR.2015.7299106.
- Seversky, M., Davis, S., and Berger, M. On time-series topological data analysis: New data and opportunities. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2016 IEEE Conference on*, pp. 1014–1022. IEEE, 2016.
- Sivic, J. and Zisserman, A. Video google: A text retrieval approach to object matching in videos. In *Ninth IEEE International Conference on Computer Vision, 2003*, pp. 1470–1477. IEEE, 2003.
- Skraba, P., Ovsjanikov, M., Chazal, F., and Guibas, L. Persistence-based segmentation of deformable shapes. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pp. 45–52, June 2010. doi: 10.1109/CVPRW.2010.5543285.
- Vedaldi, A. and Fulkerson, B. VLFeat: An open and portable library of computer vision algorithms. <http://www.vlfeat.org/>, 2008.
- Vejdemo-Johansson, M., Pokorny, F., Skraba, P., and Kragic, D. Cohomological learning of periodic motion. *Applicable Algebra in Engineering, Communication and Computing*, 26(1):5–26, Mar 2015. ISSN 1432-0622. doi: 10.1007/s00200-015-0251-x. URL <https://doi.org/10.1007/s00200-015-0251-x>.
- Verri, A., Uras, C., Frosini, P., and Ferri, M. On the use of size functions for shape analysis. *Biological Cybernetics*, 70(2):99–107, 1993. ISSN 1432-0770. doi: 10.1007/BF00200823.
- Zeppelzauer, M., Zieliński, B., Juda, M., and Seidl, M. A study on topological descriptors for the analysis of 3d surface texture. *Computer Vision and Image Understanding*, 2017.
- Zomorodian, A. and Carlsson, G. Computing persistent homology. *Discrete & Computational Geometry*, 33(2): 249–274, 2005.