Unsupervised Dense Shape Correspondence using Heat Kernels

Mehmet Aygün   Zorah Lähner   Daniel Cremers

Technical University of Munich

{mehmet.ayguen,zorah.laehner,cremers}@tum.de

Abstract

In this work, we propose an unsupervised method for learning dense correspondences between shapes using a recent deep functional map framework. Instead of depending on ground-truth correspondences or the computationally expensive geodesic distances, we use heat kernels. These can be computed quickly during training as the supervisor signal. Moreover, we propose a curriculum learning strategy using different heat diffusion times which provide different levels of difficulty during optimization without any sampling mechanism or hard example mining. We present the results of our method on different benchmarks which have various challenges like partiality, topological noise and different connectivity.

1. Introduction

With the growing market of 3D scanners and capturing systems, applications around 3D scanning are becoming more popular. One of the main building blocks of these applications in the domain of Augmented Reality/Virtual Reality is aligning scans and calculating correspondences between them. Correspondences are necessary to apply detected user motion to an avatar or transfer properties of objects in the virtual world. Finding correspondences between shapes is a difficult problem in computer vision and graphics. Especially challenging cases are non-rigid motion, scanning noise (e.g. partiality, topological noise), different resolutions or connectivity between source and target shapes.

In recent years, various approaches were suggested for the non-rigid shape correspondence problem. Similar to the development in image processing, the first approaches chose to solve a descriptor matching problem [4, 10, 46, 48, 49]. However, designing descriptors which are invariant to various kinds of deformations and noise, as well as sufficiently discriminative is not easy. Nowadays better results can be achieved with learned descriptors instead of hand-crafted ones [7, 8, 31, 16, 14].

Recent methods have adapted to include other parts of correspondence pipelines into the learning process instead of just descriptors. With the functional map framework [37], shape matching can be formulated as a learning problem in function space with the aim of minimizing distance distortion [28]. This approach uses ground truth correspondences for training the network in order to determine the distortion. Later, [21, 45] introduced networks which can be trained in an unsupervised way. This is a big advantage, because ground-truth correspondences only exist for a very limited amount of shape collections. While Halimi *et al.* [21] used the geodesic distance matrix as the supervisor signal, Roufosse *et al.* [45] used constraints on functional maps to optimize their deep networks, but both are based on the network architecture introduced in [28]. Not relying on ground-truth correspondences eliminates the need for labelled data, and Halimi *et al.* [21] used geodesic distance matrices for this purpose. However, geodesic distances are computationally expensive, not stable on degenerated meshes, and sensitive to topological noise on the surface. As a solution, we suggest to use heat kernels instead of geodesic distance matrices for unsupervised learning.

In this paper, we argue that using heat kernels is almost always beneficial to geodesic distances. Heat kernel are fast to compute, more stable on noisy meshes, have advantageous properties during optimization [50], and their time parameter offers control on the amount of localization. A heat kernel with a large time parameter includes information on the entire shape whereas a small time parameter is limited to only the close vicinity around one point. We use this property to design a curriculum learning strategy for the training process and show that it benefits finding a good optimum. This can be interpreted as a coarse-to-fine optimization strategy similar to approaches introduced in [50, 33]. However, in our proposed curriculum learning the coarse-to-fine optimization is only needed during learning and not at inference time. This makes the final operations even more efficient.

Our main contribution can be summarized as **i)** showing that heat kernels, which are extremely cheap to compute compared to geodesic distances, can be used as a supervision signal for learning shape correspondence using the

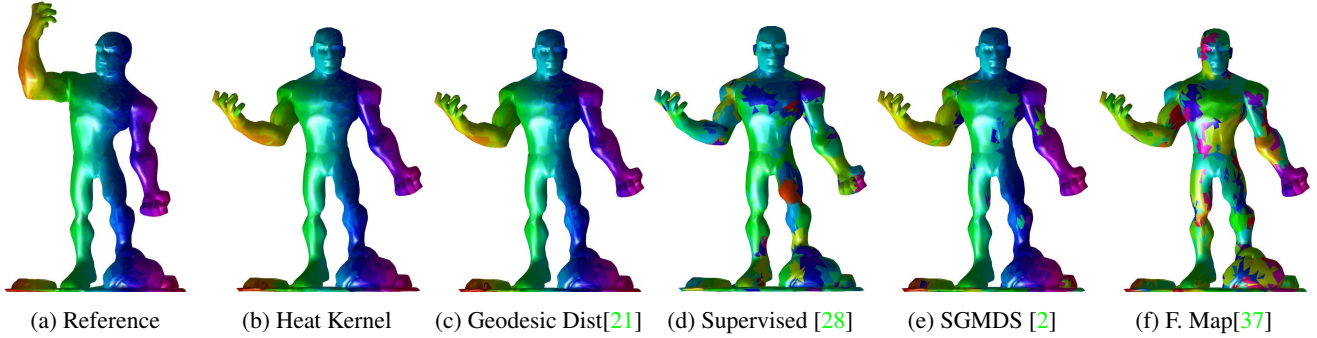| (a) Reference | (b) Heat Kernel | (c) Geodesic Dist[21] | (d) Supervised [28] | (e) SGMDS [2] | (f) F. Map[37] |

Figure 1: Dense correspondence between two shapes optimized only on a single pair of shapes. Since there is no ground-truth available, our method (1b) and 1c are trained in an unsupervised manner. Moreover, while [21] took 15 minutes to optimize, our method took just 1 min. We did not employ post-processing method to any of the results.

deep functional framework and produce even better results, **ii)** using different time parameters for heat diffusion allows us to design a curriculum learning approach which boost the correspondence accuracy on different benchmarks.

The rest of the paper is organized as follows: In the Sec. 2 we discuss previous approaches and the latest deep learning based shape correspondence methods. In Sec. 3 we give theoretical background about the shape representation that we used, heat kernels and (deep) functional maps, which are fundamental blocks of our method. Sec. 4 explains how we used heat kernels as a supervisor signal for training and introduces our curriculum learning approach. In Sec. 5 we present our results on a variety of benchmarks.

2. Related Work

Finding correspondences between two shapes is often formulated as a minimization problem using similarity measures between points. This can be done via matching of point-wise descriptors. These can be hand-crafted [4, 10, 46, 48, 49] or learned [7, 8, 31, 16, 14]. A similar direction is to match pair-wise descriptors like distances to find the optimal matching [13, 34, 50].

Functional maps [37] had a big impact by modelling the correspondence problem as a mapping between functions on the surface instead of the points of the surface directly. The main advantages is that using the Laplace-Beltrami eigenbasis for representing these functions reduces the dimensionality of the problem drastically. The idea of functional maps was refined for various applications [39, 41, 23, 26, 2, 17, 30, 29, 36] and recently adapted in deep learning frameworks [28, 21, 45]. Different from descriptor based learning approaches, the optimization focuses on functions represented in a limited basis which is low dimensional compared to the number of vertices. After obtaining functional map between shapes, correspondences between vertices can be extracted in various ways [37, 43].

2.1. Deep Learning Based Shape Correspondence

Monti *et al*. [35] proposed an unified framework for extending convolutional neural networks on non-euclidean data such as manifolds and graphs. The proposed network learns task specific local features. They modelled dense correspondence as a classification task.

Deep Functional Maps [28] introduced a framework for learning dense shape correspondence between 3D shapes using deep learning and functional maps. Instead of depending on hand-crafted local descriptors such as SHOT [47], the optimal descriptor is learned directly during training. Moreover, the learned descriptors are optimal as functions for finding functional correspondences between shapes which can be transformed into point correspondences. Since our method is based on this framework, we give a more detailed introduction in Sec. 4. The first version of this network, as it was introduced in [28], is trained using ground truth correspondences between shapes.

Groueix *et al*. [19] proposed to learn correspondences and encoding of 3D shapes together. Their *Shape Deformation Networks* take the input shape and try to align a template with it. For finding correspondences between pairs of shapes, they deform the template twice, and extract the correspondence through the template. This requires a suitable template and ground-truth correspondences for training. However, obtaining ground truth dense correspondences between shapes is expensive, since it requires either manual labeling via human annotators or special capturing setups, like [6]. To overcome the need for labeled data, Halimi *et al*. [21] proposed to use geodesic distance matrices as supervisor signals during training. They used FM-NET [28] to obtain a functional map between shapes, and tried to minimize the geodesic distance distortion between the resulting correspondences instead of relying on ground-truth error. Our method follows a similar methodology, but we advocate for heat kernels instead of geodesic distances as the supervisor signal.

In parallel to [21], Roufosse *et al*. [45] proposed another

unsupervised learning method for learning dense correspondences using FM-NET. They used the functional maps itself as the supervisor signal. The idea is based on putting constraints on functional maps, such as bijectivity, orthogonality and Laplacian commutativity which are assumed to be properties of the optimal map. They predict the functional map between shapes in both directions (source-to-target and target-to-source) while preserving the aforementioned properties in the best possible way.

2.2. Heat Kernels

Heat kernels are a popular tool in shape analysis and have been utilized for many applications in 3D correspondence [13, 38, 42]. They have also been used as an approximation of adjacency matrices in [22]. Moreover, they are the basis for the famous Heat Kernel Signature [48] and can provide a good replacement for Gaussian kernels [50]. It was also shown in [50] that using heat kernels gives the bistochistic relaxation of the Quadratic Assignment Matching beneficial properties for optimization. In this paper, we also use heat kernels as a pair-wise descriptor. We explain how we calculate heat kernels on shape and how we use them for training correspondence networks in Sec. 3.2 and Sec. 4.

3. Background

This section will introduce the necessary background to understand the rest of the paper. We model shapes as Riemannian 2-manifolds $\mathcal{X}$ with a distance function $d_{\mathcal{X}} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ and define $\pi : \mathcal{X} \to \mathcal{Y}$ as an *isometry* map which satisfies

$$d_{\mathcal{X}}(x_1, x_2) = d_{\mathcal{Y}}(\pi(x_1), \pi(x_2)). \quad (1)$$

for any pair $x_1, x_2 \in \mathcal{X}$. The correspondence problem for isometric pairs tries to find a map π which fulfills the distance criterion defined in Eq. (1). The optimization aims to minimize the distortion error:

$$L(\pi) = \sum_{x_1, x_2 \in \mathcal{X}} (d_{\mathcal{X}}(x_1, x_2) - d_{\mathcal{Y}}(\pi(x_1), \pi(x_2)))^2. \quad (2)$$

Generally, the geodesic distance describes the length of the shortest path between two vertices on the surface.

3.1. Functional Maps

The idea of functional maps is to replace finding an point-to-point correspondence π between $\mathcal{X} \to \mathcal{Y}$ with seeking a functional map $\mathcal{T}$, which represents a correspondence between functions on the surfaces, namely $\mathcal{F}(\mathcal{X}) \to \mathcal{F}(\mathcal{Y})$ [37]. This can be shown to be a linear mapping and given the right basis for the function spaces $\mathcal{F}(\mathcal{X}), \mathcal{F}(\mathcal{Y})$ formulated as a very low dimensional problem in comparison the point-to-point correspondence. The basis proposed by [37], and chosen by basically every subsequent work, are

the eigenfunctions of the Laplace-Beltrami operator (LBO), because they are invariant under isometries and frequency ordered.

[28] combines deep learning methods and functional maps to solve the dense shape correspondence problem. Instead of depending on hand-crafted descriptors as a functions to guide the functional map optimization, SHOT descriptors [49] are processed with a multi-layer perceptron network, and a new feature representation is learned during training. In forward time SHOT descriptors are processed with fully connected layers wherein their dimension is preserved. After, they are projected onto the eigenbasis, and point correspondences are extracted using functional maps. Finally, the loss is calculated with the aim of optimizing distortion error (see Eq. 2), and the error is propagated while training the network.

3.2. Heat Kernels

On a manifold $\mathcal{X}$ the *heat diffusion equation* is

$$\frac{\partial u(t, x)}{\partial t} = \Delta_{\mathcal{X}} u(t, x), \quad (3)$$

with the initial condition $u(0, x) = u_0(x)$ and additional boundary conditions if applicable. The equation describes how a given initial heat distribution u_0 on $\mathcal{X}$ diffuses over time t. $u : [0, \infty) \times \mathcal{X} \to \mathbb{R}$ represents the amount of heat at point x at time t. In more informal terms, we can imagine putting a heat source onto a point and letting the heat diffuse over the surface of the shape without anything escaping into the surrounding space. If a point is far away from the heat source in terms of geodesic distance, the amount of heat arriving there in a certain time is normally less than to a close point. See Fig. 3 for a visualization.

The solution for the heat diffusion equation can be calculated by using the equation below which is linear in the initial distribution:

$$u(t, x) = \int_{\mathcal{X}} k(t, x, x') u_0(x') dx', \quad (4)$$

where $k : \mathbb{R}^+ \times \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is the *heat kernel*. The values of k represent the the amount of heat transported from x' to x in time t. With the help of Laplace-Beltrami eigendecomposition heat kernels can be calculated in closed form using the eigenfunctions and -values:

$$k(t, x, x') = \sum_{k=0}^{\infty} e^{-\lambda_k \cdot t} \phi_k(x) \phi_k(x') \quad (5)$$

λ_k, ϕ_k denote the k-th LBO eigenvalue and -function. If the sum is restricted to a fixed k, this equation can be written as a matrix multiplication.

One caveat of heat kernels is that you need to choose the time parameter t carefully, if it is too big the heat diffuses
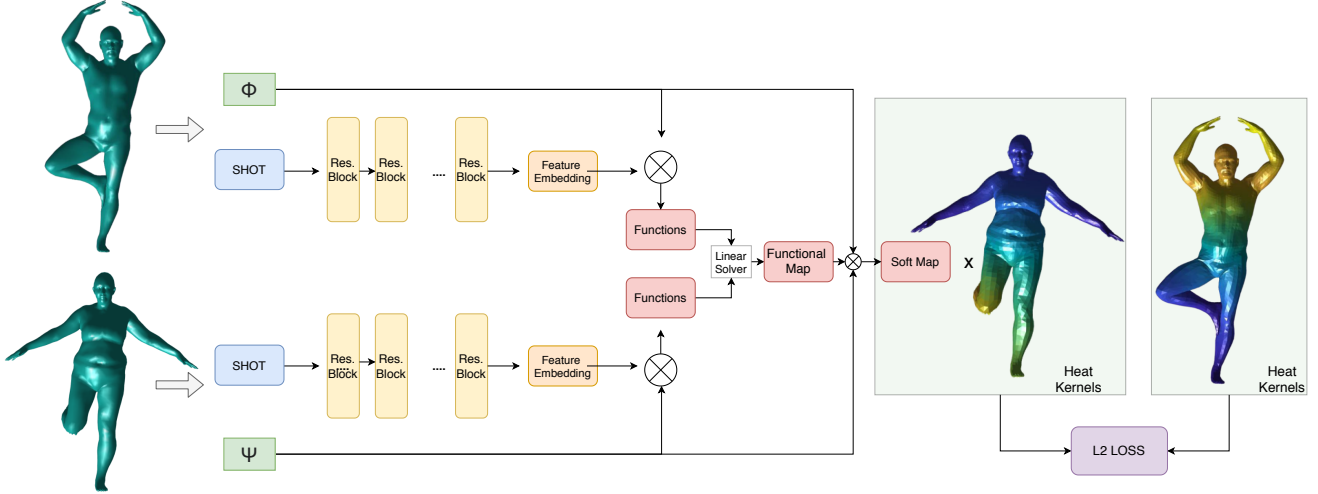
Figure 2: Overview of our proposed network. The network extracts feature embeddings from their SHOT descriptors. Afterwards, the features are projected onto the eigenfunctions and, using a linear solver, the functional map and the soft correspondence map are calculated. The network is trained using the distortion between heat kernels as the supervisor signal.

Figure 3: Visualisation of solutions of heat diffusion equation with different diffusion times (0.1, 0.03, 0.01, 0.003 from left to right). Yellow are high values, blue is zero.

equally to every point or if its too small numerical issues arise. One possible option is choosing t by visual inspection. Another option is looking at the variance of heat values of some vertices or all vertices to adjust that parameter automatically without visual inspection.

4. Method

In this section, we will explain how we construct our loss formulation in an unsupervised way and used heat kernels for curriculum learning.

4.1. Network loss

We use FM-NET which is based on deep functional framework introduced in [28]. We use the same architecture with fully-connected residual layers and exponential linear units (ELU) [12]. See Fig. 2 for an overview. The input layer takes SHOT descriptors $S \in \mathcal{R}^{N_s \times D}$, where N_s

is the number of vertices and D is the dimension of SHOT descriptor(352 in our all experiments). All fully-connected layers preserve the dimension which means the output of the network is again $N_s \times D$. Next, the learned point descriptors are projected onto the eigenvectors $\Phi \in \mathcal{R}^{E \times N_s}$ (E is number of used eigenvalues) to obtain the final function coefficient representation of the shape $F \in \mathcal{R}^{E \times D}$. The same process is applied to target shape T. T can have a different number of vertices but its function representation $G \in \mathcal{R}^{E \times D}$ has the same dimensions as F due to being projected onto the same number of eigenfunctions. Note that both shapes are processed with shared layers as in siamese networks. Afterwards, the functional map C is obtained by solving the linear equation

$$\mathbf{G} = \mathbf{CF}. \tag{6}$$

By using this functional map formulation instead of finding point-to-point correspondence we find a lower dimensional correspondence between functions of shapes. Instead of solving a linear equation with dimensions $\mathcal{R}^{N \times D}$, the result is in $\mathcal{R}^{E \times D}$, where $N >> E$, which is more efficient since linear solvers generally depend on the Cholesky decomposition with time complexity $O(n^3)$.

However, quantifying the error induced by the functional map without knowing the ground-truth map is hard. We convert C into a soft-correspondence map $P \in \mathcal{R}^{N_s \times N_t}$ by transferring indicator functions from S to T using C,

$$\mathbf{P} = |\mathbf{\Psi C \Phi^T A}|_{\|\cdot\|} \tag{7}$$

where A is the mass matrix of the source shape. Afterwards, we convert the mapped indicator functions into probability distributions by taking $Q = P \circ P$, where $\circ$ is the Hadamard

product. The ith row of j_{th} column of Q represents the probability of the ith vertex of the source corresponding to the j_{th} vertex of the target. We used same unsupervised loss function as Halimi *et al.* [21] proposed. It is defined as:

$$\ell_{\text{uns}}(\mathcal{X}, \mathcal{Y}) = \frac{1}{|\mathcal{X}|^2} \left\| \mathbf{D}_{\mathcal{X}} - \mathbf{Q}^T \mathbf{D}_{\mathcal{Y}} \mathbf{Q} \right\|_{\text{F}}^2 \qquad (8)$$

where $\mathbf{D}_{\mathcal{X}}$ and $\mathbf{D}_{\mathcal{Y}}$ represent the geodesic distance matrices on source and target shape.

Heat Kernel. Instead of using geodesics distance matrices, we used heat kernels as the pairwise supervisor signal. Preserving the geodesic distances, as it is done in [21], implies that an isometry is found. We saw in Eq. (5) that heat kernels can be calculated using the eigenfunctions and values of the Laplace-Beltrami operator only which implies that heat kernel are also invariant under isometries. However, geodesic distances assume large values in points far away from each other whereas heat kernel achieve their maximal values on the point itself and its immediate neighborhood. In the case of approximate isometries, due to noise or different classes, a change on one vertex will affect not only its neighborhood but geodesic distances on the entire shape. Due to the Frobenius norm in Eq. (8) a single large error will have a lot more influence on the loss than a collection of small errors. This will lead to many local distortions when the isometry assumption is violated, see Fig. 9. The influence of heat kernels can be controlled via the diffusion time and for small times the values for further points are nearly zero which means large outliers do not dominate the loss. Additionally, heat kernels are computationally more efficient compared to geodesic distances since they can be calculated in a closed form, as explained in Sec. 3.2. The drawback is that the global information of geodesic distances guides the optimization to the correct local optimum whereas more local information is prone to get stuck in unfortunate local minima. To counteract this behavior we introduce a curriculum learning approach in the next section, and show in Sec. 5 that our results are equal or better to [21].

4.2. Curriculum Learning

Curriculum learning is a teaching strategy where in every step the task to solve gets more challenging with the aim of solving a complex task in the end. This learning strategy is used for training deep neural networks and showed effective performance on various problems [20, 5, 14]. As discussed in Sec. 3.2 we can control the locality of the heat diffusion via the diffusion time parameter which determines how far the heat will spread. For instance, as can be seen in Fig.3, higher diffusion values lead to a more spread out and global solution. When the time parameter is small, the information is very localized and only solutions in which
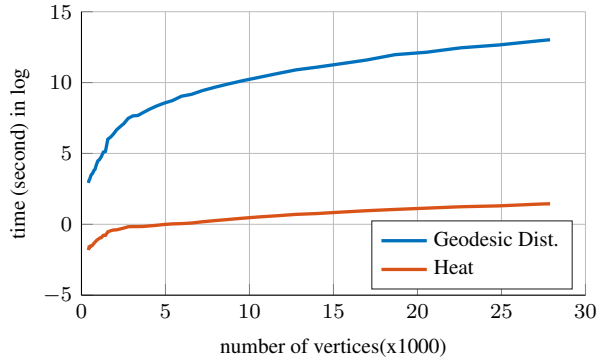

Figure 4: Comparison of times for calculating heat kernels and geodesic distance matrices with varying number of vertices. Please note that the y-axis is in log scale.

neighborhoods are well preserved give a good loss. Unfortunately, this makes the loss function harder to optimize due to stronger local optima. We use this property and start training with higher diffusion times to get good gradient information. Then, we gradually decrease the time when we are already in the vicinity of the optimal solution to get a more precise placement.

The main idea in curriculum learning settings is that easy examples are optimized first then harder examples are added. Hard examples are normally obtained via hard mining which is a time consuming process. However, our heat kernel formulation becomes naturally more challenging, and this allows us to control difficulty without any mining process. To create such curriculum learning procedure, we start training with heat kernels with a high time parameter and decrease the time after some iterations. In our experiments we called this way of training heat decay. Moreover, since it is fast to calculate heat kernels, this extra computation step does not lead to longer training times. In contrast, geodesic distance matrices need to be pre-calculated before training because of their complexity. We measured calculation times of heat kernel and geodesic distance matrices for shapes that contain different number of vertices. As can be seen in Fig.4, calculating heat kernels is extremely fast compared to geodesic distances.

Our curriculum learning follows the idea of coarse-to-fine optimization in spectral space that has also been explored in [50] and [33]. However, both are non-learning methods which means they need to run multiple iterations during inference time. Our method only applies this strategy during training, and the run time of inference does not suffer from additional calculations.

4.3. Implementation Details

We implemented our framework in Tensorflow [1] based on the original Deep Functional Maps implementation [1].

---

[1] https://github.com/orlitany/DeepFunctionalMaps

For all shapes, we extracted 150 eigenvectors and 352-dimensional SHOT descriptors. We calculate SHOT descriptors using 10 bins and a radius equal to 5% of total area of the shape. We pre-calculated all heat kernels using eigenvectors and eigenvalues before training. If the shapes are too big for training ($\sim 100k$) we down-sampled them via edge contraction [18] to $\sim 7k$ vertices. However, during testing we used the original resolution. We did not use any post processing steps for refining the correspondences. We use the Adam [25] optimizer with an initial learning rate of 0.001 for all of our training. In our experiments, using fancy learning rate schedules did not improve our results significantly.

Further, for choosing the decay values and intervals, we used a validation set. For calculating validation loss, we calculated distortion error using heat kernels with initial time diffusion value. However, on most benchmarks we observed no additional improvements after decaying the time value twice.

5. Experiments

In this section we show that heat kernels can obtain a similar or better performance than geodesic distances while they provide vastly different energy landscape than geodesic distances. First, we show that our method is working even under extreme conditions like having only one training example without ground-truth information. Then, we try our proposed method on the FAUST Synthetic data set [6] which is a commonly used data set for evaluating shape correspondence performance. We also evaluate on the SHREC 16' [15, 27] and SHREC 19' data sets which include challenges such as partiality, topological noise and different connectivity respectively.

We use geodesic error curves as the evaluation measure [24]. This curve contains threshold values on the x-axis and the percentage of correct correspondences on the y-axis. For every vertex on the source shape the difference between the match and the ground-truth match is considered. The curve plots the percentage of differences which are lower than the threshold value on the x-axis. The perfect solution would show a constant curve at 1.

5.1. Single Shape

One of the most extreme case in learning for shape correspondences is not having any training examples or other priors. We repeat the experiment from Halimi *et al.* [21] and calculate a correspondence for a single shape pair. The result and a comparison can be seen in Fig. 1. We used 150 LBO eigenfuctions, the SHOT descriptor and optimize our network using heat kernels. Our method achieved comparable, in some parts even better results, than the geodesic distance version and other methods. Since no ground truth is available, there is no quantitative evaluation.

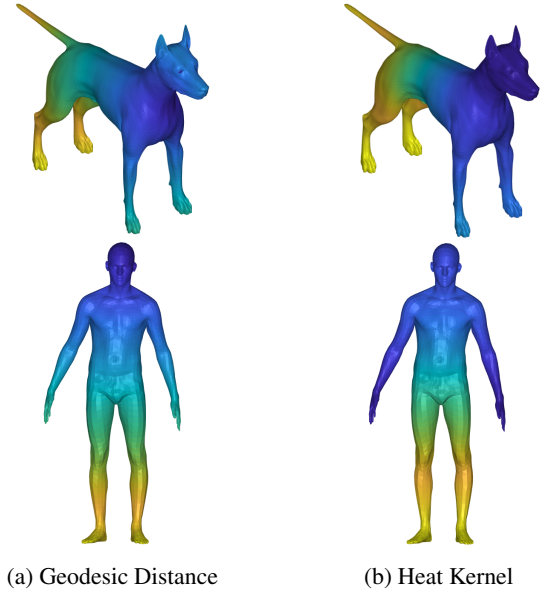(a) Geodesic Distance   (b) Heat Kernel

Figure 5: Visualisation of heat kernels and geodesic distances on FAUST [6] and SHREC 16' [15] data sets. While blue represent close areas, yellow represent far away areas. As can be seen on both shapes, both heat kernel and geodesic distances provide similar information about closeness.

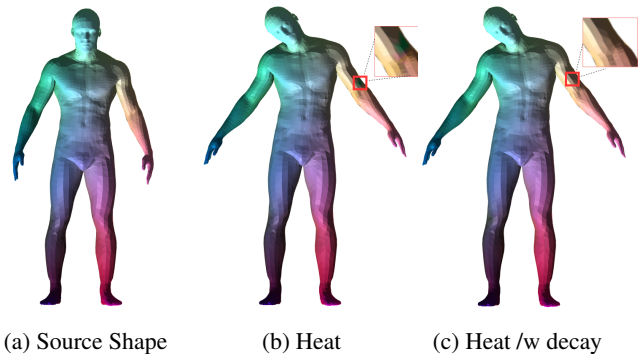(a) Source Shape   (b) Heat   (c) Heat /w decay

Figure 6: Visualisation of correspondences on a FAUST [6] shape. Same colors represent the correspondences. While the model trained with heat kernel produce very good result, on the left arm there is a small error. However, the model with decayed training did not generate a similar error, since it trained with refined heat kernels also. Please note that, the person in the shape was not included in the training.

5.2. FAUST - Synthetic data set

The FAUST data set [6] contains 100 shapes which show 10 different people in 10 different poses. The original scans are registered to a template with 6890 vertices, and we use these registrations. For training, we used the first eight persons and all of their poses. We used the remaining the two persons for testing our method. We used 120 LBO eigen-
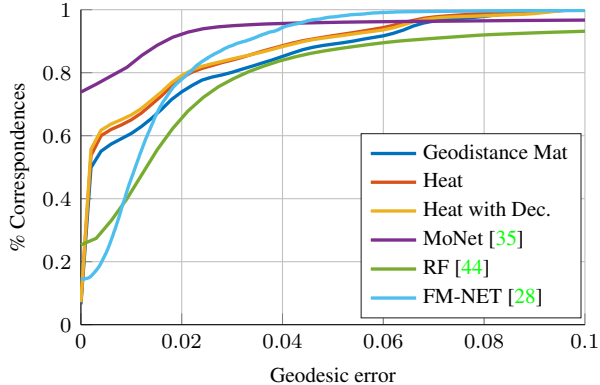
Figure 7: Comparison of heat kernel and geodesic distance matrices on FAUST Synthetic shapes. Training with Heat Kernels and its decayed version obtains better results than with the geodesic distance matrix.

functions, and 352-dimensional SHOT descriptor as the input to our model. We evaluate three different unsupervised signals: the geodesic distance matrix as proposed in Halimi *et al*. [21], heat kernel with a fixed temperature of $t = 0.1$ and decayed heat training with initially $t = 0.1$ and later $t = 0.01$. When we use the geodesic distance matrix, we train our method with $10k$ iterations similar to [21]. For heat kernels, we also trained our model with $10k$ iterations and decayed the heat kernel values after $5k$ iterations.

During training, we shuffled vertices of all shapes because the ground-truth is one-to-one, and the network could just learn to return the identity. As can be seen in Fig.7 heat kernels obtained better results compared to geodesic distances. Our interpretation is that decayed heat kernels provide different details on each scale which helps the network to refine the solution. Moreover, heat kernels provide similar distance preservation criteria as geodesic distances. We visualised heat kernel and geodesic distance on FAUST shape to validate this claim, which can be seen on Fig.5.

5.3. SHREC 16' - Partiality

The SHREC 16' partiality data set [15] contains shapes from different classes like dogs and humans. The main challenge of this data set is that the shapes are partial, i.e contain holes or cuts. In our experiment, we choose the dog class with holes similar to [21]. The data set also has a null shape for each category and correspondences are calculated to that shape for evaluation. In the dog category there are 10 training and 26 test shapes. For each shape we extracted 120 LBO eigenbases and SHOT descriptors. We trained our models using i) fixed heat kernel value ($t = 1.0$), ii) dynamic heat kernel value ($t = 1.0$ then $t = 0.1$), and iii) geodesic distance matrix. We used FM-NET with 5 layers since this data set is small compared to other data sets and trained the model with only $1k$ iterations.
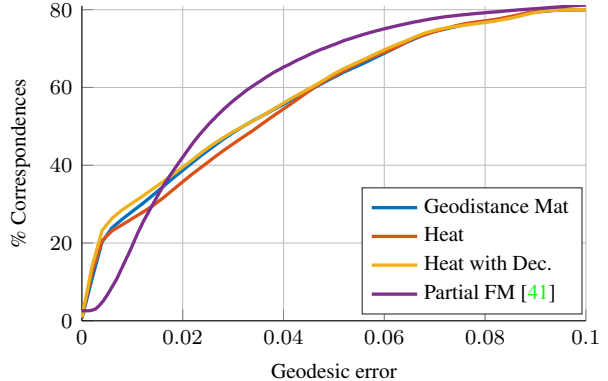

Figure 8: Comparison of heat kernel and geodesic distance matrices on SHREC'16 benchmark for the dog class. Shapes contains holes and this lead to partial matching of deformable shapes. Heat kernel with decayed training obtained comparable results with the geodesic distances.

As illustrated in Fig.8, while fixed heat kernels obtain lower scores than geodesic distance matrix, heat kernels with decayed training obtain better scores. In addition to the correspondence error curve we visualised correspondences on two challenging poses in Fig.9.

5.4. SHREC 16' - Topological Noise

The SHREC 16' - Topological Noise data set [27] contains shapes from KIDS data set [44] with topological merging. The data set contains 16 training and 9 test shapes. For training we used 150 LBO eigenbases and the original resolution of shapes. During evaluation we looked all pairs of test shapes and calculate similar plot as in previous sections. We compare our method with geodesic distances.

The results of our experiment can be seen in Fig. 10. While decayed heat kernels achieved better results than geodesic distance with a big error margin, the decayed version obtained better result on small error margin. This show that the model which was trained with decayed heat kernels was able to find correspondences on finer details while fixed heat kernels obtained more rough matches.

5.5. SHREC 19' - Different Connectiviy

The Shrec'19 Matching Humans with Different Connectivity challenge [32] is designed to evaluate correspondence methods under different density (5k to 50k) and different meshing distribution (uniform and nonuniform). The challenge contains different human shapes from variety of data sets like FAUST [6], TOSCA [9], SCAPE [3].

In our experiments, we remeshed all shapes to have about $7k$ vertices due to memory constraints. Afterwards, we extracted 150 LBO eigenfunctions and 352-dimensional SHOT descriptors from these resized shapes. We used FM-NET[28] with 7 residual layers. We trained our models with $10k$ iterations, and decayed after $5k$ iterations.

(a) Heat   (b) Heat /w decay   (c) Geodesic[21]

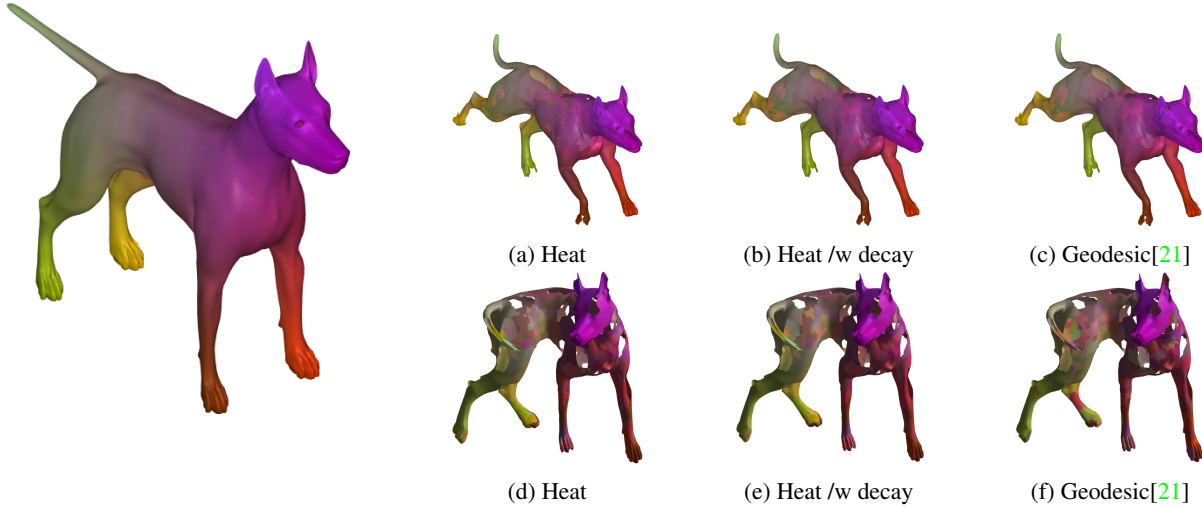(d) Heat   (e) Heat /w decay   (f) Geodesic[21]

Figure 9: Visualisation of some correspondences on SHREC 16 test set. The shape on the left is the source shape. (a) & (d) is obtained using the model which trained using a fix heat kernel, (b) & (e) with decayed heat kernel. While test shapes are not complete and in different poses, our method is able to find reasonable correspondences which is similar or better than [21].
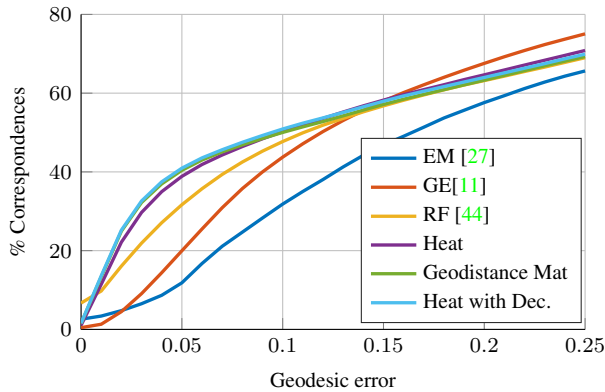

Figure 10: Comparison of Heat Kernels and Geodesic distance matrix on SHREC'16 topological noise benchmark. Shapes contains topological artifacts around their body.

During inference we used the original shapes which have high variety on the number of vertices. In our experiments we explored three different unsupervised signals: i) fixed heat kernel value ($t = 0.1$), ii) dynamic heat kernel value ($t = 0.1$ then $t = 0.01$), and iii) geodesic distance matrix. An overview of the results can be seen in Fig.11. As depicted in the figure heat kernel with decay obtain better performance than geodesic distances.

6. Conclusion

We proposed to use heat kernels instead of geodesic distances for training correspondence networks without any ground-truth matches in the FM-Net architecture. We show that heat kernel are an equal but more efficient replacement for geodesic distances, even though they provide a less optimal energy landscape for optimization. Networks trained
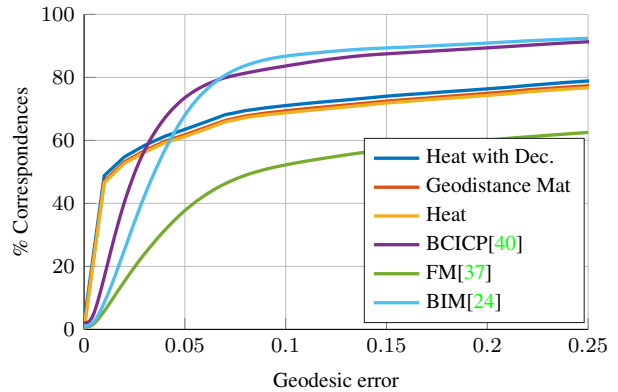

Figure 11: Comparison of Heat Kernels and Geodesic distance matrix on SHREC'19 different connectivity benchmark. Shapes vary in number of vertices and triangles. In this benchmark fixed heat kernel did not outperform geodesic distance matrix, but the decay version did.

with heat kernels obtain equal or better results on various benchmarks while improving run time significantly. Moreover, heat kernel can be recalculated during training, and we proposed a curriculum learning approach with different diffusion times that improved correspondence accuracy on every benchmark.

Acknowledgement

References

[1] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283, 2016.

[2] Y. Aflalo, A. Dubrovina, and R. Kimmel. Spectral generalized multi-dimensional scaling. *International Journal of Computer Vision*, 118(3):380–392, 2016.

[3] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: shape completion and animation of people. In *ACM SIGGRAPH 2005 Papers*, pages 408–416. 2005.

[4] M. Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature: A quantum mechanical approach to shape analysis. In *2011 IEEE international conference on computer vision workshops (ICCV workshops)*, pages 1626–1633. IEEE, 2011.

[5] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, 2009.

[6] F. Bogo, J. Romero, M. Loper, and M. J. Black. Faust: Dataset and evaluation for 3d mesh registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3794–3801, 2014.

[7] D. Boscaini, J. Masci, S. Melzi, M. M. Bronstein, U. Castellani, and P. Vandergheynst. Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks. In *Computer Graphics Forum*, volume 34, pages 13–23. Wiley Online Library, 2015.

[8] D. Boscaini, J. Masci, E. Rodolà, M. M. Bronstein, and D. Cremers. Anisotropic diffusion descriptors. In *Computer Graphics Forum*, volume 35, pages 431–441. Wiley Online Library, 2016.

[9] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. *Numerical geometry of non-rigid shapes*. Springer Science & Business Media, 2008.

[10] M. M. Bronstein and I. Kokkinos. Scale-invariant heat kernel signatures for non-rigid shape recognition. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1704–1711. IEEE, 2010.

[11] O. Burghard, A. Dieckmann, and R. Klein. Embedding shapes with green's functions for global shape matching. *Computers & Graphics*, 68:1–10, 2017.

[12] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.

[13] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. *PNAS*, 102(21):7426–7431, 2005.

[14] É. Corman, M. Ovsjanikov, and A. Chambolle. Supervised descriptor learning for non-rigid shape matching. In *European Conference on Computer Vision*, pages 283–298. Springer, 2014.

[15] L. Cosmo, E. Rodolà, M. M. Bronstein, A. Torsello, D. Cremers, and Y. Sahillioglu. Shrec'16: Partial matching of deformable shapes. *Proc. 3DOR*, 2(9):12, 2016.

[16] L. Cosmo, E. Rodola, J. Masci, A. Torsello, and M. M. Bronstein. Matching deformable objects in clutter. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 1–10. IEEE, 2016.

[17] D. Eynard, E. Rodola, K. Glashoff, and M. M. Bronstein. Coupled functional maps. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 399–407. IEEE, 2016.

[18] M. Garland and P. S. Heckbert. Surface simplification using quadric error metrics. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 209–216, 1997.

[19] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry. 3d-coded: 3d correspondences by deep deformation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 230–246, 2018.

[20] G. Hacohen and D. Weinshall. On the power of curriculum learning in training deep networks. *arXiv preprint arXiv:1904.03626*, 2019.

[21] O. Halimi, O. Litany, E. Rodola, A. M. Bronstein, and R. Kimmel. Unsupervised learning of dense shape correspondence. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4370–4379, 2019.

[22] N. Hu and L. Guibas. Spectral descriptors for graph matching. *arXiv preprint arXiv:1304.1572*, 2, 2013.

[23] Q. Huang, F. Wang, and L. Guibas. Functional map networks for analyzing and exploring large shape collections. *ACM Transactions on Graphics (TOG)*, 33(4):1–11, 2014.

[24] V. G. Kim, Y. Lipman, and T. Funkhouser. Blended intrinsic maps. *ACM transactions on graphics (TOG)*, 30(4):1–12, 2011.

[25] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[26] A. Kovnatsky, M. M. Bronstein, X. Bresson, and P. Vandergheynst. Functional correspondence by matrix completion. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 905–914, 2015.

[27] Z. Lähner, E. Rodolà, M. M. Bronstein, D. Cremers, O. Burghard, L. Cosmo, A. Dieckmann, R. Klein, and Y. Sahillioglu. Shrec'16: Matching of deformable shapes with topological noise. In *Proc. of Eurographics Workshop on 3D Object Retrieval (3DOR)*, 2016.

[28] O. Litany, T. Remez, E. Rodolà, A. Bronstein, and M. Bronstein. Deep functional maps: Structured prediction for dense shape correspondence. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5659–5667, 2017.

[29] O. Litany, E. Rodolà, A. M. Bronstein, and M. M. Bronstein. Fully spectral partial shape matching. In *Computer Graphics Forum*, volume 36, pages 247–258. Wiley Online Library, 2017.

[30] O. Litany, E. Rodolà, A. M. Bronstein, M. M. Bronstein, and D. Cremers. Non-rigid puzzles. In *Computer Graphics Forum*, volume 35, pages 135–143. Wiley Online Library, 2016.

[31] R. Litman and A. M. Bronstein. Learning spectral descriptors for deformable shape correspondence. *IEEE transactions on pattern analysis and machine intelligence*, 36(1):171–180, 2013.

[32] S. Melzi, R. Marin, E. Rodolà, U. Castellani, J. Ren, A. Poulenard, P. Wonka, and M. Ovsjanikov. Matching Humans with Different Connectivity. In *Eurographics Workshop on 3D Object Retrieval*, 2019.

[33] S. Melzi, J. Ren, E. Rodola, A. Sharma, P. Wonka, and M. Ovsjanikov. Zoomout: Spectral upsampling for efficient shape correspondence. *arXiv preprint arXiv:1904.07865*, 2019.

[34] F. Mémoli and G. Sapiro. A theoretical and computational framework for isometry invariant recognition of point cloud data. *Foundations of Computational Mathematics*, 5(3):313–347, 2005.

[35] F. Monti, D. Boscaini, J. Masci, E. Rodola, J. Svoboda, and M. M. Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5115–5124, 2017.

[36] D. Nogneng and M. Ovsjanikov. Informative descriptor preservation via commutativity for shape matching. In *Computer Graphics Forum*, volume 36, pages 259–267. Wiley Online Library, 2017.

[37] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (TOG)*, 31(4):1–11, 2012.

[38] M. Ovsjanikov, Q. Mérigot, F. Mémoli, and L. Guibas. One point isometric matching with the heat kernel. *Computer Graphics Forum*, 29(5), 2010.

[39] J. Pokrass, A. M. Bronstein, M. M. Bronstein, P. Sprechmann, and G. Sapiro. Sparse modeling of intrinsic correspondences. In *Computer Graphics Forum*, volume 32, pages 459–468. Wiley Online Library, 2013.

[40] J. Ren, A. Poulenard, P. Wonka, and M. Ovsjanikov. Continuous and orientation-preserving correspondences via functional maps. *ACM Transactions on Graphics (TOG)*, 37(6):1–16, 2018.

[41] E. Rodolà, L. Cosmo, M. M. Bronstein, A. Torsello, and D. Cremers. Partial functional correspondence. In *Computer Graphics Forum*, volume 36, pages 222–236. Wiley Online Library, 2017.

[42] E. Rodolà, Z. Lähner, A. M. Bronstein, M. M. Bronstein, and J. Solomon. Functional maps representation on product manifolds. *Computer Graphics Forum*, 38, 2019.

[43] E. Rodolà, M. Moeller, and D. Cremers. Point-wise map recovery and refinement from functional correspondence. In *The 20th Intl. Symposium on Vision, Modeling and Visualization*, 2015.

[44] E. Rodolà, S. Rota Bulo, T. Windheuser, M. Vestner, and D. Cremers. Dense non-rigid shape correspondence using random forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4177–4184, 2014.

[45] J.-M. Roufosse, A. Sharma, and M. Ovsjanikov. Unsupervised deep learning for structured shape matching. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1617–1627, 2019.

[46] R. M. Rustamov. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Proceedings of the fifth Eurographics symposium on Geometry processing*, pages 225–233. Eurographics Association, 2007.

[47] S. Salti, F. Tombari, and L. Di Stefano. Shot: Unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding*, 125:251–264, 2014.

[48] J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer graphics forum*, volume 28, pages 1383–1392. Wiley Online Library, 2009.

[49] F. Tombari, S. Salti, and L. Di Stefano. Unique signatures of histograms for local surface description. In *European conference on computer vision*, pages 356–369. Springer, 2010.

[50] M. Vestner, Z. Lähner, A. Boyarski, O. Litany, R. Slossberg, T. Remez, E. Rodola, A. Bronstein, M. Bronstein, R. Kimmel, et al. Efficient deformable shape correspondence via kernel matching. In *2017 International Conference on 3D Vision (3DV)*, pages 517–526. IEEE, 2017.