Eurographics Symposium on Geometry Processing 2015
Mirela Ben-Chen and Ligang Liu
(Guest Editors)

Volume 34 (2015), Number 5

Designing class-specific shape descriptors by vertex-frequency analysis

submission id 1052

Abstract

In this paper we explore the use of localized frequency analysis (a generalization of the windowed Fourier transform to manifolds) for the design of intrinsic shape descriptors. Applying the windowed Fourier transform to some dense intrinsic descriptor adds a local "context" capturing richer geometric structures. The resulting local frequency representations are then passed through a bank of filters whose coefficient are determined by a learning procedure minimizing a task-specific cost. Conceptually, our approach is reminiscent of convolutional neural networks (CNN), and also generalizes several previous methods such as spectral CNN and GPS embeddings. Our experimental results show that the proposed approach allows learning class-specific shape descriptors significantly outperforming recent state-of-the-art methods on standard benchmarks.

Categories and Subject Descriptors (according to ACM CCS):

1. Introduction

Point based matching plays an important role in shape analysis and geometry processing. The key approach consists in defining an effective point *signature* (or descriptor) able to capture the most notable local shape characteristic or *feature* in the neighborhood of the point. In general a feature descriptor is encoded in a vector lying in some multi-dimensional feature space. Given this representation the matching between points can be easily computed by comparing their respective descriptors in such space (usually the standard L2-norm is used for this purpose). The ideal descriptor should be *discriminative*, highlighting distinctive attributes, *robust*, invariant with respect to noise and deformations, *compact*, to reduce the dimension of the descriptor space, and *computationally-efficient*. In general is very difficult to design a descriptor able to simultaneously satisfy all of these properties. This

To this aim, it is desirable for the descriptor to adapt for a particular application so that the trade-off between the aformentioned requirements can be ameliorated. This can be achieved estimating the parameters of the descriptor on some exemplar points (learning) w.r.t. some problem dependent cost function.

Previous work Several work has been proposed for shape description (see [LGB*et al.*13] for a recent survey). Early methods like spin images [JH99], shape distributions [OFCD02], integral volume descriptors [M*06], and multi-scale features [PKG03] are capable of encoding only the extrinsic structure of the shape and therefore they are invariat under Euclidean variations. A drastic improvement has been shown when descriptors are able to encode the shape intrinsic properties to deal effectively with non rigid transformations [EK03]. A popular class of these methods is based on spectral shape analysis [Lév06] like the shapeDNA [RWP06] or the Global Point Signature (SGP) [Rus07] descriptors. In particular, a further refinement of spectral methods has been proposed by exploiting diffusion geometry [BBG94, CL06] on shapes. Important descriptors in this class are the heat kernel signatures (HKS) [SOG09,GBAL09], their scale-invariant version [BK10] (SI-HKS), and the wave kernel signatures (WKS) [ASC11]. A relevant characteristic of shape descriptors is related to their capability to encode *local* or *global* shape information [LGB*et al.*13]. Being inspired by the success of local image signatures [Low04,BMP00], several methods have been proposed on point-based feature description for the shape domain [SB11, KBLB12]. Differently from local methods like HKS or WKS where the local information is encoded implicitly from the diffusion process, these approaches require an explicit evaluation of the point neighbourhood that make their estimate computational demanding. Indeed, the collection of local context around a point is not trivial for non-Euclidean spaces like shapes or graphs. A possible solution has been introduced recently in [SRV13], where a method to perform a localized frequency analysis is generalized to graphs, relying on the analogy between the classical Fourier transform and the Laplace-Beltrami eigenbasis.

[below needs a rephrasing] An open issue of spectral based descriptors is due to the strong assumption that isometric shapes share their spectral decomposition. It is well known that in the real world case, when we work on a discretized domain, this hypothesis is violated by the possible switching and sign flips of the eigenfunctions of the Laplace-Beltrami operator. This is caused by the loss of perfect isometry and by numerical inaccuracies. Some works tried to cope with this ambiguity [MHK*08, PBB*13]. In [MHK*08] the effect of switching is attenuated by introducing a clustering procedure among eigenfunctions. In [PBB*13] a matching between eigenfunctions is estimated by exploiting the functional mapping [O*12] procedure on a sparse coding framework.

Following a recent trend from the image domain that prefers learning based to hand-crafted signatures, a few works are proposed on shape descriptors defined by learning by example paradigm [LB14, R*14, LBBC14]. For instance in [LB14] an optimal data dependent trasfer function is estimated by showing that HKS and WKS can be considered as particular parametric families of this approach. In [R*14] a deep learning framework based on random forest is proposed. In [LBBC14] a learning version of the bag-of-word descriptor is introduced by supervised sparse coding.

Contribution The paper proposes a new method for point-to-point matching between heterogeneous shapes which extends the windowed Fourier transform to the manifold analysis domain. The main idea is to formulate the classical time-frequency analysis to signals lying on a 3D surface, and use the output of this transform to define a new shape descriptor. This gives a local frequency representation which effectively and efficiently encodes the local "context" of a given point. There are different important factors that must be defined to design the descriptor: i) the initial signal, ii) the window function, iii) the range of frequences. In order to optimize the choice of these factors for a specific class of shapes we propose to combine the analytic construction of the descriptor with a learning by example procedure. Moreover, we show that by choosing a proper setting of such factors our method can reproduce previous descriptors like spectral CNN [BZSL14] and GPS embeddings [Rus07].

More precisely, this work has the following main contributions: i) it defines a generalized convolution, translation and modulation operators for signals on shapes, thus obtaining a new definition that further extends the WFT for graph proposed in [SRV13] to the 3D surfaces, ii) it introduces a learning procedure which addresses the sign ambiguity problem and ordering of eigenfunctions in the spectrum of near-isometric shapes, iii) we show that this descriptors overcome state-of-the-art performance on several synthetic and real datasets such as SCAPE [A*05] and FAUST [BRLB14].

2. Background

Manifold We model a 3D shape as a connected smooth compact two-dimensional surface (manifold) X, possibly with a boundary ∂X. Locally around each point x the manifold is homeomorphic to the *tangent plane* T_xX. The *exponential map* $\exp_x : T_xX \to X$ maps tangent vectors onto the surface. A *Riemannian metric* is an inner product $\langle \cdot, \cdot \rangle_{T_xX} : T_xX \times T_xX \to \mathbb{R}$ on the tangent space depending smoothly on x.

Laplace-Beltrami operator (LBO) Let us denote by $L^2(X)$ the space of square-integrable real functions on X and by $\langle f, g \rangle_{L^2(X)} = \int_X f(x)g(x)dx$ the standard inner product on $L^2(X)$, where dx is the infinitesimal area element induced by the Riemannian metric. Given a smooth function $f \in L^2(X)$, the composition $f \circ \exp_x : T_xX \to \mathbb{R}$ is a function on the tangent plane. We define the *Laplace-Beltrami operator* (LBO) as a positive semidefinite operator $\Delta_X : L^2(X) \to L^2(X)$ given by

$$\Delta_X f(x) \;=\; \Delta(f \circ \exp_x)(0), \tag{1}$$

where Δ is the Euclidean Laplacian operator on the tangent plane. The LBO is *intrinsic*, i.e., expressible entirely in terms of the Riemannian metric. As a result, it is invariant to isometric (metric-preserving) deformations of the surface.

Spectral analysis on manifolds The LBO of a compact manifold admits an eigendecomposition $\Delta_X \phi_k = \lambda_k \phi_k$ with a countable set of real eigenvalues $\lambda_1 \leq \lambda_2 \leq \dots$ and the corresponding eigenfunctions $\phi_1, \phi_2, \dots$ form an orthonormal basis on $L^2(X)$. Consequently, a function $f \in L^2(X)$ can be represented as the *Fourier series*

$$f(x) \;=\; \sum_{k \geq 1} \langle f, \phi_k \rangle_{L^2(X)} \phi_k(x), \tag{2}$$

where the analysis $\hat{f}_k = \langle f, \phi_k \rangle_{L^2(X)}$ can be regarded as the forward Fourier transform and the synthesis $\sum_{k \geq 1} \hat{f}_k \phi_k(x)$ is the inverse one; the eigenvalues $\{\lambda_k\}_{k \geq 1}$ play the role of frequencies. The first eigenvalue $\lambda_1 = 0$ corresponds to a constant eigenvector ('DC component'). The Laplacian eigenbasis is a generalization of the classical Fourier basis to non-Euclidean domains, and one can easy verify that $e^{i\omega x}$ are eigenfunctions of the Euclidean Laplacian operator $-\frac{d^2}{dx^2} e^{i\omega x} = \omega^2 e^{i\omega x}$.

The *generalized convolution* of f and g on the manifold can be defined by analogy to the classical case as the inverse transform of the product of forward transforms,

$$\begin{aligned}(f \star g)(x) \;&=\; \sum_{k \geq 1} \langle f, \phi_k \rangle_{L^2(X)} \langle g, \phi_k \rangle_{L^2(X)} \phi_k(x) \\ &=\; \sum_{k \geq 1} \hat{f}_k \hat{g}_k \phi_k(x), \end{aligned} \tag{3}$$

and is in general *non-shift-invariant*.

Heat diffusion on manifolds is governed by the *diffusion equation*,

$$(\Delta_X + \partial_t)\, u(x,t) = 0;\ u(x,0) = u_0(x), \tag{4}$$

where $u(x,t)$ denotes the amount of heat at point x at time t, $u_0(x)$ is the initial heat distribution (if the surface has a boundary, appropriate boundary conditions must be added). The solution of (4) is obtained by applying the *heat operator* $H^t = e^{-t\Delta_X}$ to the initial condition,

$$u(x,t) \ = \ H^t u_0(x) = \int_X u_0(x') h_t(x,x') dx', \tag{5}$$

where $h_t(x,x')$ is called the *heat kernel*. Since H^t has the same eigenfunctions as Δ_X with the eigenvalues $\{e^{-t\lambda_k}\}_{k\geq 1}$, the solution of (4) can be expressed a generalized convolution (3),

$$\begin{aligned} u(x,t) \ &= \ H^t u_0(x) = \sum_{k\geq 1} \langle u_0, \phi_k \rangle_{L^2(X)} e^{-t\lambda_k} \phi_k(x) \qquad (6) \\ &= \ \int_X u_0(x') \underbrace{\sum_{k\geq 1} e^{-t\lambda_k} \phi_k(x)\phi_k(x')}_{h_t(x,x')} dx', \end{aligned}$$

where the coefficients $e^{-t\lambda}$ play the role of a transfer function corresponding to a low-pass filter sampled at frequencies $\{\lambda_k\}_{k\geq 1}$.

3. Spectral descriptors

In this section, we provide a concise overview of several spectral shape descriptors that will be important for our further discussion. This overview is by no means exhaustive, and we refer the reader to the cited related works and references therein for a more complete picture.

Global Point Signature (GPS) Rustamov [Rus07] proposed the *global point signature (GPS)* embedding, a dense shape descriptor constructed using scaled LBO eigenfunctions,

$$\mathbf{f}(x) \ = \ (\lambda_1^{-1/2}\phi_1(x), \ldots, \lambda_Q^{-1/2}\phi_Q(x))^\top, \tag{7}$$

thus associating each point x with a Q-dimensional descriptor (see [BBG94,CL06] for earlier constructions in the theoretical math community).

Due to an inherent ambiguity in the definition of the LBO eigenbasis, GPS descriptors cannot be matched in a simple-minded manner. First, an eigenfunction is defined up to sign, $\Delta_X(\pm\phi_i) = \lambda_i(\pm\phi_i)$. Second, if an eigenvalue with non-trivial multiplicity is present in the spectrum of Δ_X, any rotation in the corresponding subspace produces valid eigenfunctions. Third, noise and non-isometric deformations may alter the eigenvalues and eigenfunctions of the LBO. Trying to cope with these ambiguities, several techniques have been proposed trying to match GPS descriptors (see, e.g. [MHK*08]).

Heat/Wave Kernel Signature (HKS/WKS) Several popular spectral shape descriptor take a generic form of the diagonal of a parametric kernel diagonalized by the LBO eigenbasis. Notable examples include the *heat kernel signature (HKS)* [SOG09, GBAL09] and the *wave kernel signature (WKS)* [ASC11]. More specifically, such methods construct at each point a descriptor

$$\mathbf{f}(x) \ = \ \sum_{k\geq 1} \tau(\lambda_k)\phi_k^2(x) \tag{8}$$

expressed by a bank of transfer functions $\tau(\lambda) = (\tau_1(\lambda), \ldots, \tau_Q(\lambda))^\top$. Such descriptors have several appealing properties making their use popular in numerous applications. First, they are intrinsic and hence invariant to isometric deformations of the manifold by construction. Second, they are dense. Third, (8) can be efficiently computed using the first few eigenvectors and eigenvalues of the Laplace-Beltrami operator.

HKS uses low-pass transfer functions $\tau_t(\lambda) = e^{-t\lambda}$ for various values of the parameter $t \in \{t_1, \ldots, t_Q\}$, giving rise to the *autodiffusivity function* $h_t(x,x)$, whose physical interpretation is the amount of heat remaining at point x after time t. A notable drawback of HKS is poor spatial localization, which is a consequence of the uncertainty principle: good localization in the Fourier domain (large value of t) results in a bad localization in the spatial domain.

WKS uses band-bass transfer functions $\tau_\nu(\lambda) = \exp\left(\frac{\log\nu - \log\lambda}{2\sigma^2}\right)$ for various values of the parameter $\nu \in \{\nu_1, \ldots, \nu_Q\}$. The physical interpretation of WKS is the probability to find a quantum particle at point x, given that it has an initial log-normal energy distribution with mean value ν and variance σ. Typically, WKS exhibits oscillatory behavior and has a better localization compared to HKS.

Optimal spectral descriptors (OSD) Litman and Bronstein [LB14] used parametric transfer functions expressed as

$$\tau_q(\lambda) \ = \ \sum_{m=1}^{M} a_{qm}\beta_m(\lambda) \tag{9}$$

in some fixed (e.g. B-spline) basis $\beta_1(\lambda), \ldots, \beta_M(\lambda)$, where a_{qm} $(q = 1, \ldots, Q, m = 1, \ldots, M)$ are the parametrization coefficients. Plugging (9) into (8) one can express the qth component of the spectral descriptor as

$$f_q(x) \ = \ \sum_{k\geq 1} \tau_q(\lambda_k)\phi_k^2(x) = \sum_{m=1}^{M} a_{qm} \underbrace{\sum_{k\geq 1} \beta_m(\lambda_k)\phi_k^2(x)}_{g_m(x)}, \tag{10}$$

where $\mathbf{g}(x) = (g_1(x), \ldots, g_M(x))^\top$ is a vector-valued function referred to as *geometry vector*, dependent only on the intrinsic geometry of the shape. Thus, (8) is parametrized by the $Q \times M$ matrix $\mathbf{A} = (a_{lm})$ and can be written in matrix form as $\mathbf{f}(x) = \mathbf{A}\mathbf{g}(x)$.

The main idea of [LB14] is to *learn* the optimal parameters

$\mathbf{A}$ by minimizing a task-specific loss. Given a training set consisting of a pair of geometry vectors $\mathbf{g},\mathbf{g}^+$ representing knowingly similar points (*positives*), and $\mathbf{g},\mathbf{g}^-$ representing knowingly dissimilar points (*negatives*), one tries to find $\mathbf{A}$ such that $\|\mathbf{f}-\mathbf{f}^+\| = \|\mathbf{A}(\mathbf{g}-\mathbf{g}^+)\|$ is as small as possible and $\|\mathbf{f}-\mathbf{f}^-\| = \|\mathbf{A}(\mathbf{g}-\mathbf{g}^-)\|$ is as large as possible. The authors show that the problem boils down to a simple Mahalanobis-type metric learning.

4. Windowed Fourier transform

A central piece to our construction of shape descriptors is the notion of *vertex-frequency analysis* or *windowed Fourier transform (WFT)*, generalizing these constructions from classical signal processing to non-Euclidean domains. WFT has previously been extended to graphs in [SRV13].

Classical WFT The main idea of classical WFT is to analyze the frequency content of a signal that is localized by means of multiplication by a window. Given a function $f \in L^2(\mathbb{R})$ and some 'mother window' g localized at zero, one computes the WFT as

$$(Sf)_{x,\omega} = \int_{\mathbb{R}} f(x')g(x'-x)e^{-ix'\omega}dx'. \qquad (11)$$

Note that the WFT has two indices: spatial location x' of the window and frequency ω of the signal in that window. Alternatively, it can be presented as an inner product with a translated an modulated window, $(Sf)_{x,\omega} = \langle f, M_\omega T_x g\rangle_{L^2(\mathbb{R})}$, where T_x and M_ω denote the translation and modulation operators, respectively.

Translation operator in the Euclidean setting is simply $(T_{x'}f)(x) = f(x-x')$. In order to generalize it to manifolds, translation to point x' can be replaced by convolution with a delta-function centered at x', yielding

$$\begin{aligned}(T_{x'}f)(x) &= (f\star\delta_{x'})(x)\\ &= \sum_{k\geq 1}\langle f,\phi_k\rangle_{L^2(X)}\langle\delta_{x'},\phi_k\rangle_{L^2(X)}\phi_k(x)\\ &= \sum_{k\geq 1}\hat{f}_k\phi_k(x')\phi_k(x), \qquad (12)\end{aligned}$$

where convolution is understood in the generalized sense 3. Note that such a translation is not shift-invariant in general, i.e., the window would change when moved around the manifold (see Figure 1).

Modulation operator simply amounts to multiplication by a basis function,

$$(M_k f)(x) = f(x)\phi_k(x). \qquad (13)$$

Manifold WFT Combining the two operators together, we have the modulated and translated window (transform 'atom'; see examples in Figure 1) expressed as

$$g_{x',k}(x) = (M_k T_{x'} g)(x) = \phi_k(x)\sum_{l\geq 1}\hat{g}_l\phi_l(x')\phi_l(x). \qquad (14)$$

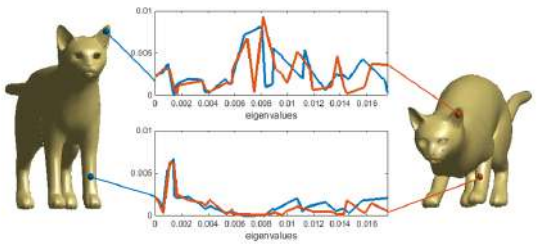

Figure 2: WFT descriptors.

Note that the 'mother window' is defined here in the frequency domain by the coefficients $\hat{g}_l$. We thus readily have the WFT of a signal $f \in L^2(X)$

$$(Sf)_{x,k} = \langle f, g_{x,k}\rangle_{L^2(X)} = \sum_{l\geq 1}\hat{g}_l\phi_l(x)\langle f,\phi_l\phi_k\rangle_{L^2(X)} \qquad (15)$$

which can be regarded as a meta-descriptor: given some dense descriptor f (e.g. one of the components of HKS, WKS, or a geometry vector (10)), we construct $\mathbf{f}(x) = ((Sf)_{x,1},\dots,(Sf)_{x,K})^\top$ taking the first K frequencies of the WFT. The WFT allows to capture the local context of a signal on the manifold, making it roughly analogous to taking the values of the signal in a small "patch" (the analogy is not complete since the patch is represented in the frequency domain rather than in the spatial domain).

Special cases We would like to point to the following two special cases of the WFT.

Case I: when $\hat{g}_k = \delta_{k1}$, we simply have $g_{x',k}(x) = \phi_k(x)\phi_1(x')\phi_1(x)$. Since the first LBO eigenvector is constant, the atom (up to scaling) is $g_{x',k}(x) \propto \phi_k(x)$, i.e., the standard LBO eigenbasis element independent on the location x'. The WFT thus simply reduces to a simple Fourier transform (2). This result is an intuitive consequence of the uncertainty principle: when the window is perfectly localized in the frequency domain, it is perfectly delocalized in the spatial domain (and without localization, WFT is equivalent to standard FT).

Case II: when $f \equiv 1$, the WFT contains information only about the geometric structure of the manifold. In this setting,

$$(S1)_{x,k} = \sum_{l\geq 1}\hat{g}_l\phi_l(x)\underbrace{\langle\phi_k,\phi_l\rangle_{L^2(X)}}_{\delta_{kl}} = \hat{g}_k\phi_k(x), \qquad (16)$$

and for a particular choice of $\hat{g}_k = \lambda_k^{-1/2}$ we get Rustamov's GPS descriptor (7).

Case III: when $f = \delta_x$, the DC frequency of the WFT has the form of (8),

$$(S\delta_x)_{x,0} = \sum_{l\geq 1}\hat{g}_l\phi_l^2(x), \qquad (17)$$

and in particular for $\hat{g}_l = e^{-t\lambda_l}$ we obtain the HKS and for $\hat{g}_l = \exp\left(\frac{\log v - \log\lambda_l}{2\sigma^2}\right)$ the WKS at point x, respectively.

Figure 1: Examples of different WFT atoms $g_{x,k}$ shown for two different windows (top and bottom rows).

We can therefore conclude that the WFT is a rich construction, containing as particular settings some of the well-known descriptors.

5. Learning

In general, the WFT is not completely invariant, i.e., the descriptors at corresponding points of isometric shapes may differ. In this section we describe a learning framework that allows to account for this variability.

The architecture of our descriptor bears similarity with convolutional neural networks (CNN) that are now very popular in computer vision applications, and has a layered structure, wherein the output on one layer is used as the input of the next one. For the sake of simplicity, the neural network architecture considered in the following consists of only two layers. The first layer is a *fully connected layer*, producing outputs as weighted sums of the inputs, followed by a nonlinear function. The second layer applies the WFT, acting as a "patch operator" that extracts local structure of the input around each point and represents it in the frequency domain. Since each input dimension might contain features of different scale, we employ a different window for each input dimension. The WFTs are then passed through a bank of filters applied in the frequency domain, producing the outputs used as the descriptor dimensions. As the input to the first layer, any intrinsic descriptor can be used (specifically, we use geometry vectors (10)). All the parameters of the layers (weights, windows coefficients, and filters) are variables that are found by means of supervised learning.

Fully connected layer Let us be given a P-dimensional input $\mathbf{f}^{\mathrm{in}}(x) = (f_1^{\mathrm{in}}(x), \dots, f_P^{\mathrm{in}}(x))$. The fully connected layer produces a Q-dimensional output defined as

$$f_q^{\mathrm{out}}(x) \;=\; \xi\left(\sum_{p=1}^{P}\sum_{k=1}^{K} w_{qp} f_p^{\mathrm{in}}(x)\right), \quad q = 1,\dots,Q, \tag{18}$$

where $\xi(t) = \max(0,t)$ is the *ReLU activation function*. Note that without ReLU, if the inputs are geometry vectors, learning the weights of the fully connected layer is equivalent to the OSD [LB14]. Fixing weights corresponding to low- or band-pass filters, the fully connected layer implements the HKS and WKS, respectively.

Convolutional layer Next, the output of the fully connected layer acts as the input into the convolutional layer; we denote the input again by $\mathbf{f}^{\mathrm{in}}(x)$ and its dimension by P. For each input dimension, we use a different window. The family of P windows is parametrized in some fixed interpolation basis in the frequency domain as in (9),

$$\gamma_p(\lambda) \;=\; \sum_{m=1}^{M} b_{pm}\beta_m(\lambda), \qquad p = 1,\dots,P, \tag{19}$$

where the $P \times M$ matrix (b_{pm}) of weights defines the windows. The WFT of the pth input dimension uses the respective pth window,

$$(Sf_p^{\mathrm{in}})_{x,k} \;=\; \sum_{l\geq 1} \gamma_p(\lambda_l)\phi_l(x)\langle f_p^{\mathrm{in}}, \phi_l\phi_k\rangle_{L^2(X)}, \tag{20}$$

producing at each point a K-dimensional vector for each of the P input dimensions. Our goal is to produce a Q-dimensional output, and for this purpose, the WFTs are passed through a bank of filters, roughly resembling the structure of a convolutional layer used in CNNs. The qth dimension of the output is given by

$$f_q^{\mathrm{out}}(x) \;=\; \sum_{p=1}^{P}\sum_{k=1}^{K} a_{qpk}|(Sf_p^{\mathrm{in}})_{x,k}|, \quad q = 1,\dots,Q \tag{21}$$

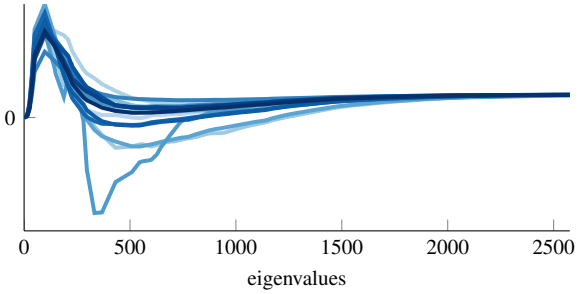

Figure 3: Example of a family of windows $\hat{g}_1,\dots,\hat{g}_Q$ learned on the FAUST dataset.

Learning Combining the fully-connected and convolutional layer, we get a parametric hierarchical system $\mathbf{f}_\Theta(x)$ producing a Q-dimensional descriptor at each point x (here $\Theta=\{(w_{qp}),(b_{pm}),(a_{qpk})\}$ denotes the set of learnable parameters).

Given a dataset of knowingly similar and dissimilar pairs of points on pairs of shapes, respectively positives $\mathcal{T}^+=\{(x,x^+)\}$ and negatives $\mathcal{T}^-=\{(x,x^-)\}$, we aim at estimating the optimal task-specific parameters of the descriptor minimizing the aggregate loss

$$\ell(\Theta) \;=\; (1-\mu)\ell_+(\Theta)+\mu\,\ell_-(\Theta) \tag{22}$$

where

$$\ell_+(\Theta) \;=\; \frac{1}{|\mathcal{T}^+|}\sum_{(x,x^+)\in\mathcal{T}^+}\|\mathbf{f}_\Theta(x)-\mathbf{f}_\Theta(x^+)\|_2^2, \tag{23}$$

$$\ell_-(\Theta) \;=\; \frac{1}{|\mathcal{T}^-|}\sum_{(x,x^-)\in\mathcal{T}^-}\max\{0,M-\|\mathbf{f}_\Theta(x)-\mathbf{f}_\Theta(x^-)\|^2\}.$$

are the positive and negative losses, respectively, μ is a parameter governing their trade-off, and M is a margin mapping the negatives apart.

We stress that since HKS, WKS, and OSD descriptors are obtained by a particular choice of the parameters Θ, if the training set is designed well and performed correctly, our descriptor can perform only better than the above.

6. Implementation

Discretization In the discrete setting, the surface X is sampled at N points $x_1,\dots,x_N$. On these points, we construct a triangular mesh (V,E,F) with vertices $V=\{1,\dots,N\}$, in which each interior edge $ij\in E$ is shared by exactly two triangular faces ikj and $jhi\in F$, and boundary edges belong to exactly one triangular face.

A function on the surface is represented by an N-dimensional vector $\mathbf{f}=(f(x_1),\dots,f(x_N))^\top$. The inner product is discretized as $\langle\mathbf{f},\mathbf{g}\rangle=\mathbf{f}^\top\mathbf{A}\mathbf{g}$, where $\mathbf{A}=\mathrm{diag}(a_1,\dots,a_N)$ and $a_i=\frac{1}{3}\sum_{jk:ijk\in F}A_{ijk}$ denotes the local area element at vertex i and A_{ijk} denoting the area of triangle ijk. To discretize the LBO as an $N\times N$ matrix $\mathbf{L}=\mathbf{A}^{-1}\mathbf{W}$, we use the classical cotangent formula [PP93, MDSB03], according to which

$$w_{ij} \;=\; \begin{cases} (\cot\alpha_{ij}+\cot\beta_{ij})/2 & ij\in E;\\ -\sum_{k\neq i}w_{ik} & i=j;\\ 0 & \text{else;}\end{cases} \tag{24}$$

where α_{ij},β_{ij} denote the angles $\angle ikj,\angle jhi$ of the triangles sharing the edge ij.

The first $K\le N$ eigenfunctions and eigenvalues of the LBO operator are computed by performing the generalized eigendecomposition $\mathbf{W}\mathbf{\Phi}=\mathbf{A}\mathbf{\Phi}\mathbf{\Lambda}$, where $\mathbf{\Phi}=(\boldsymbol{\phi}_1,\dots,\boldsymbol{\phi}_K)$ is an $N\times K$ matrix containing as columns the discretized eigenfunctions and $\mathbf{\Lambda}=\mathrm{diag}(\lambda_1,\dots,\lambda_K)$ is the diagonal matrix of the corresponding eigenvalues. Note that the eigenvectors are $\mathbf{A}$-orthonormal, i.e. $\mathbf{\Phi}^\top\mathbf{A}\mathbf{\Phi}=\mathbf{I}$.

Therefore, including the discretized inner product to equations 2 and 3 the translation operator becomes:

$$T(f)=\mathbf{\Phi}\hat{f}\odot\mathbf{\Phi}', \tag{25}$$

where $\odot$ is the element-wise product. Also the modulation operator is modified accordingly as:

$$M(f)=f\odot\mathbf{\Phi} \tag{26}$$

Finally the WFT in the discrete settings is defined as:

$$(Sf)=f'\,\boldsymbol{A}\,M(T(g)) \tag{27}$$

Settings In all our experiments, we used $K=300$ LBO eigenfunctions and eigenvalues computed using MATLAB *eigs* function. For OSD, we used $M=150$-dimensional geometry vectors computed according to (9)–(10) using B-spline bases [LB14].

Training Our approach was implemented in Theano [B*10, B*12] and source code for training and reproducibility of the experiments will be released upon publication.

7. Results

Datasets We used two public-domain datasets of scanned human shapes in different poses: SCAPE [A*05] and FAUST [BRLB14], the latter being the most recent and particularly challenging, given a high variability of non-isometric deformations as well as significant variability between different human subjects. The meshes in SCAPE were resampled to 12.5K vertices, whereas for FAUST we used the registration meshes without further pre-processing. In addition we scaled all shapes to have unit geodesic diameter. In both datasets, groundtruth point-wise correspondence between the shapes was known for all points.

[DAVIDE? I think we do NOT ignore intrinsic symmetry right?] Since we consider intrinsic descriptors, in all our evaluations we ignore intrinsic symmetry.

Methods and Settings We compared the performance of the proposed approach to HKS [SOG09], WKS [ASC11], and OSD [LB14] using the code and settings provided by the respective authors. To make the comparison fair, all the descriptors were $Q = 16$-dimensional as in [LB14].

We used three different configurations:

1. Convolutional layer with fixed $\hat{g} = 1e-5$ applied to the constant input function with 16 output maps. In this configuration the set of parameters which are learned is $\Theta = \{w_{qp}\}$;
2. Geometry vectors are passed through a FC layer to reduce the number of dimensions to 16 followed by a convolutional layer with fixed $\hat{g} = 1e-5$ and 16 output maps. In this configuration $\Theta = \{(w_{qp}), (a_{qpk})\}$;
3. Same function as in 2 but with a fully learnable $\hat{g}$ via the coefficients β. This is the more complex descriptor which has the largest flexibility, in fact we have $\Theta = \{(w_{qp}), (b_{pm}), (a_{qpk})\}$. In fact, while for some applications it is reasonable to fix the window by hand, because of some prior on the task for example, but generally speaking this is usually suboptimal and learning the windows allows to better adapt to the various application domains.

Training Each dataset was split into disjoint training, validation, and test sets. On the FAUST dataset subjects 1–7 were used for training, subject 8 for validation, and subject 9–10 for testing. On SCAPE, we used shapes 20–29 and 50–70 for training, five random shapes from the remaining ones for validation, and the rest for testing. The positive and negative sets of vertex pairs required for training were generated on the fly, to keep the storage requirements for the training algorithm, via uniform stochastic sampling. Each point on the first shape has only a single groundtruth match (given by the one-to-one correspondence) and is assigned to one out of $N-1$ possible negatives: first, sample two shapes, then form the positive set with all corresponding points, and finally, form the negative set with first shape vertices and a random permutation of the ones of the second shape. This strategy differs from [LB14] who considered only points on the same shape. The advantage of our sampling strategy is that it allows learning invariance also across several poses and subjects.

While it is true that in principle some negatives are much easier than others, and thus could inflate the learning performance we found that in practice a simple random selection works well in practice. In case this will be a problem for other application where interesting negative samples are hard to incorporate with this simple procedure, importance sampling (or fracking [**?**]) will be enough to compensate for this bias and steer learning to a better solution space.

Our approach was trained until convergence with the Adadelta [Zei12] optimizer, a first order method which tries to automatically adjust the learning rate (or step size), which delivered a quicker convergence in out experiments.

Timing Typical training times for the more complex descriptor (third configuration) are around two hours on a NVIDIA TITAN Black GPU board and, at test time, the system is able to produce a throughput of approximately 30K vertices per second.

Evaluation times: [what do you mean by this? what other timings do we need?]

Similarity map Figures 4–**??** (compare to Figure 2 in [LB14]) depicts the Euclidean distance in the descriptor space between the descriptor at a selected point and the rest of the points on the same shape as well as its transformations. Our approach shows a good tradeoff between localization (similar to HKS) and accuracy (less spurious minima than WKS and OSD), as well as robustness to different kinds of noise (isometric and non-isometric deformations, geometric and topological noise, different sampling, and missing parts).

Descriptor evaluation Following [LB14], we evaluated the descriptor performance using the *cumulative match characteristic* (CMC) and the *receiver operator characteristic* (ROC). The CMC evaluates the probability of a correct correspondence among the k nearest neighbors in the descriptor space. The ROC measures the percentage of positives and negatives pairs falling below various thresholds of their distance in the descriptor space (*true positive* and *negative rates*, respectively). Figure **??** shows the performance of different descriptors in terms of CMC and ROC characteristics on the FAUST and SCAPE datasets. We observe that our approach significantly outperform other descriptors, and that the more complex model further boosts performance.

To evaluate the correspondence quality possible with our descriptors, we performed a simple experiment assigning points to their nearest neighbor in the descriptor space, thus avoiding bias towards a particular matching method (only the quality of the raw descriptors is considered). Figure **??** evaluates the resulting correspondence using the Princeton protocol [KLF11], plotting the percentage of matches that are at most r-geodesically distant from the groundtruth correspondence. Our approach achieves the best performance among the compared descriptors.

Generalization capability In order to test the generalization capability of the learned descriptors, we applied...

Out of class shapes?

8. Conclusions

In this paper we propose a new shape descriptor for local frequency analysis on the 3D surface domain. We show that the Windowed Fourier Transform can be extended to discretized 3D manifold by leading to a very expressive and

Heat Kernel Signature (HKS)

Wave Kernel Signature (WKS)

Optimal Spectral Descriptor (OSD)

Windowed Fourier Transform (WFT)

Figure 4: Distance map in the descriptor space. A point on the reference shape (leftmost) is compared to all other points on the same and on other shapes which have undergone various transformations of increasing complexity. From left to right we see the behavior for .... Ideally we should see a localized descriptor around the point of interest (colored in blue) and the rest of the shape colored in red with no other minima regions.

well localized intrinsic feature point descriptor. Our approach is very flexible and can be combined with other well known and effective descriptors like HKS and WKS. In particular, the exploration of a learning based framework has demonstrated its strength and versatility in designing the desired behaviour of the proposed descriptor. Broadly speacking we believe that our appraoch will open new perspective of the use of advanced signal processing techniques on non Euclidean domain like the 3D shapes.

Limitations and Extensions This work highlights how important is the effect of LBO eigenbasis ambiguity in the spectral shape decompositions. Our class-specific learnig approach is able to overcome this limit but still some open issues remain, especially for high generalization purposes. In more details, our method works well when the class is composed by a family of similar objects like the case of humans (with possible heterogeneous instances like fat/thin or tall/short shapes, and so on). Conversely, the performance of our method decreases when the class becomes very generic (i.e., quadruped). In this case the only reliable eigenbasis are those associated to the first eigenvalues by leading to a simplified encoding of the shape composed of low frequences (i.e., mean shape). Some method to include a matching of

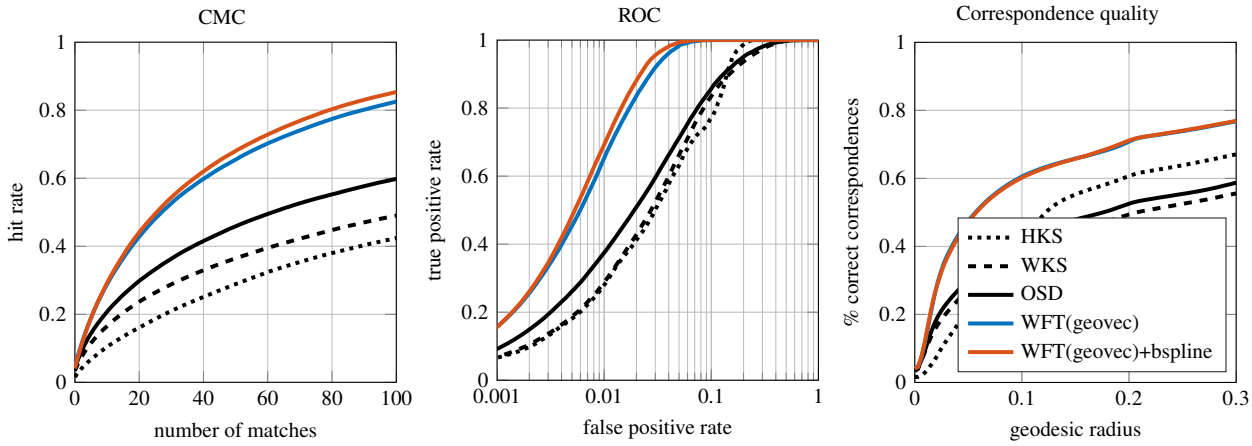

Figure 5: FAUST2FAUST.

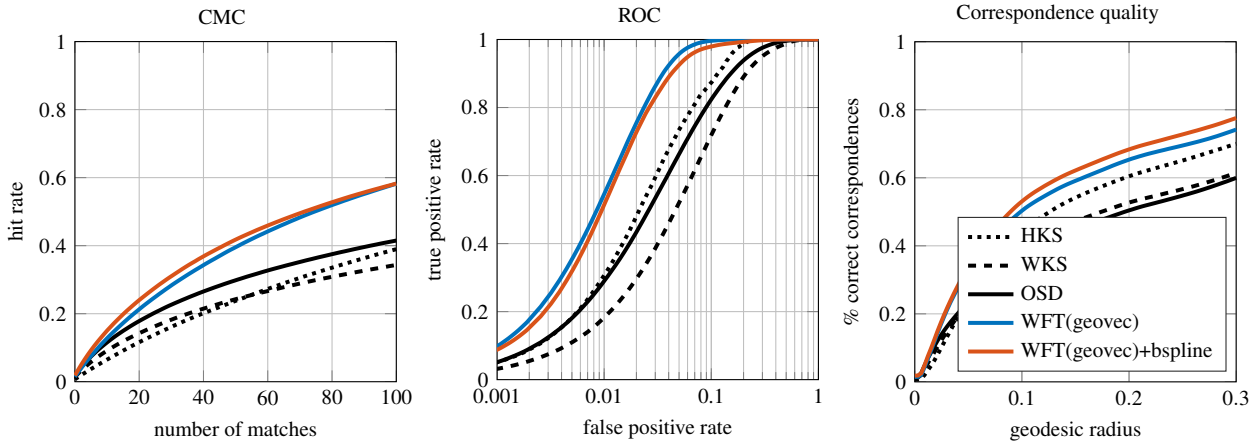

Figure 6: FAUST2SCAPE

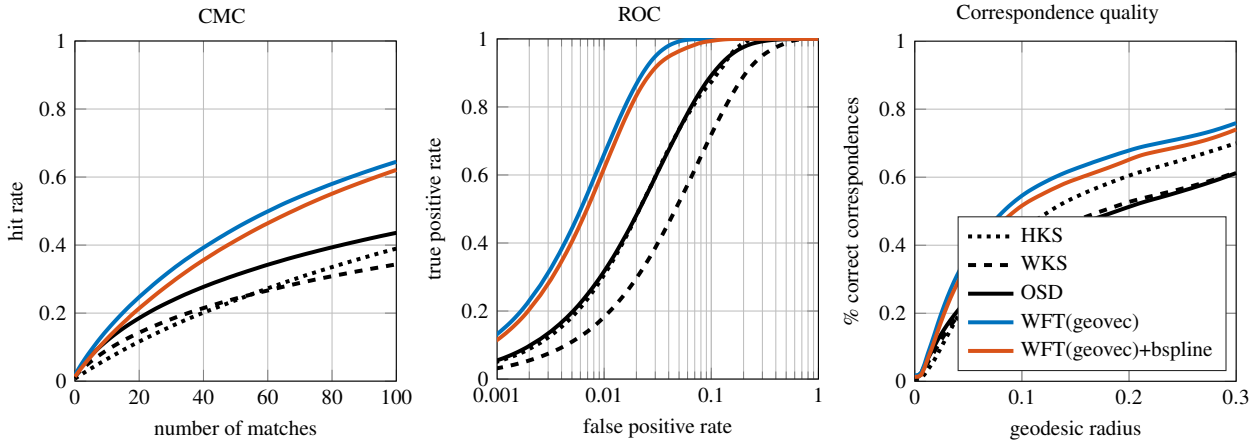

Figure 7: SCAPE2SCAPE

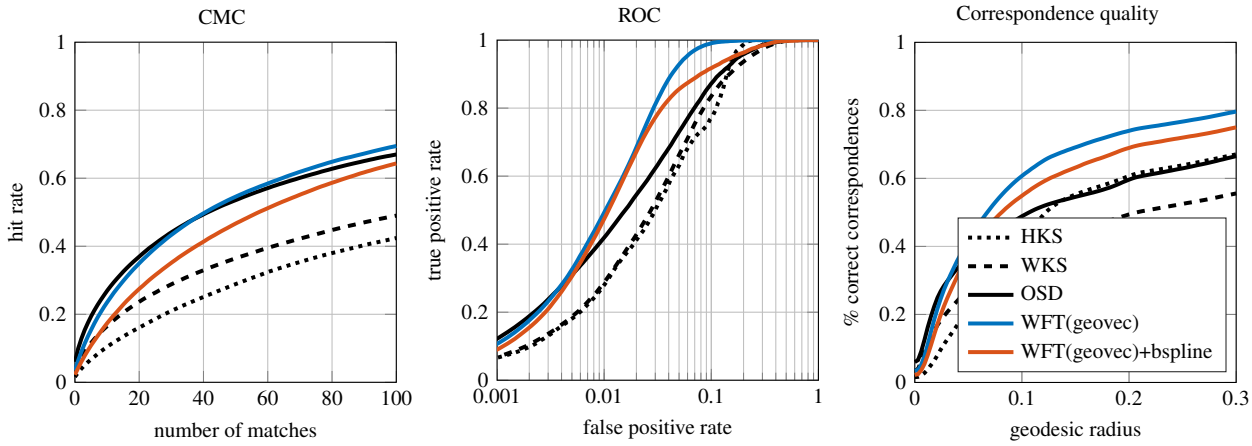

Figure 8: SCAPE2FAUST

eigenbasis in the learning framework will be addressed as future work.

References

[A*05] ANGUELOV D., ET AL.: SCAPE: Shape completion and animation of people. *TOG 24*, 3 (2005), 408–416. 2, 6

[ASC11] AUBRY M., SCHLICKEWEI U., CREMERS D.: The wave kernel signature: A quantum mechanical approach to shape analysis. In *Proc. ICCV* (2011), pp. 1626–1633. 1, 3, 7

[B*10] BERGSTRA J., ET AL.: Theano: a CPU and GPU math expression compiler. In *Proc. SciPy* (June 2010). 6

[B*12] BASTIEN F., ET AL.: Theano: new features and speed improvements. In *Proc. Workshop on Deep Learning and Unsupervised Feature Learning* (2012). 6

[BBG94] BÉRARD P., BESSON G., GALLOT S.: Embedding riemannian manifolds by their heat kernel. *Geometric & Functional Analysis 4*, 4 (1994), 373–398. 1, 3

[BK10] BRONSTEIN M. M., KOKKINOS I.: Scale-invariant heat kernel signatures for non-rigid shape recognition. In *Proc. CVPR* (2010). 1

[BMP00] BELONGIE S., MALIK J., PUZICHA J.: Shape context: A new descriptor for shape matching and object recognition. In *Proc. NIPS* (2000). 1

[BRLB14] BOGO F., ROMERO J., LOPER M., BLACK M. J.: FAUST: Dataset and evaluation for 3D mesh registration. In *Proc. CVPR* (2014). 2, 6

[BZSL14] BRUNA J., ZAREMBA W., SZLAM A., LECUN Y.: Spectral networks and locally connected networks on graphs. In *Proc. ICLR* (2014). 2

[CL06] COIFMAN R. R., LAFON S.: Diffusion maps. *Applied and Computational Harmonic Analysis 21*, 1 (2006), 5–30. 1, 3

[EK03] ELAD A., KIMMEL R.: On bending invariant signatures for surfaces. *PAMI 25*, 10 (2003), 1285–1295. 1

[GBAL09] GEBAL K., BÆRENTZEN J. A., ANÆS H., LARSEN R.: Shape analysis using the auto diffusion function. *CGF 28*, 5 (2009), 1405–1413. 1, 3

[JH99] JOHNSON A. E., HEBERT M.: Using spin images for efficient object recognition in cluttered 3D scenes. *PAMI 21*, 5 (1999), 433–449. 1

[KBLB12] KOKKINOS I., BRONSTEIN M. M., LITMAN R., BRONSTEIN A. M.: Intrinsic shape context descriptors for deformable shapes. In *Proc. CVPR* (2012). 1

[KLF11] KIM V. G., LIPMAN Y., FUNKHOUSER T.: Blended intrinsic maps. *TOG 30*, 4 (2011), 1–12. 7

[LB14] LITMAN R., BRONSTEIN A. M.: Learning spectral descriptors for deformable shape correspondence. *PAMI 36*, 1 (2014), 170–180. 2, 3, 5, 6, 7

[LBBC14] LITMAN R., BRONSTEIN A., BRONSTEIN M., CASTELLANI U.: Supervised learning of bag-of-features shape descriptors using sparse coding. *CGF 33*, 5 (2014), 127–136. 2

[Lév06] LÉVY B.: Laplace-Beltrami eigenfunctions towards an algorithm that "understands" geometry. In *Proc. SMI* (2006). 1

[LGB*et al.*13] LIAN Z., GODIL A., BUSTOS B., *et al.*: A Comparison of Methods for Non-rigid 3D Shape Retrieval. *Pattern Recognition 46*, 1 (2013), 449–461. 1

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *IJCV 60*, 2 (2004), 91–110. 1

[M*06] MANAY S., ET AL.: Integral invariants for shape matching. *PAMI 28*, 10 (2006), 1602–1618. 1

[MDSB03] MEYER M., DESBRUN M., SCHRÖDER P., BARR A. H.: Discrete differential-geometry operators for triangulated 2-manifolds. *Visualization&Mathematics* (2003), 35–57. 6

[MHK*08] MATEUS D., HORAUD R., KNOSSOW D., CUZZOLIN F., BOYER E.: Articulated Shape Matching Using Laplacian Eigenfunctions and Unsupervised Point Registration. In *Proc. CVPR* (2008). 2, 3

[O*12] OVSJANIKOV M., ET AL.: Functional maps: a flexible representation of maps between shapes. *TOG 31*, 4 (2012), 1–11. 2

[OFCD02] OSADA R., FUNKHOUSER T., CHAZELLE B., DOBKIN D.: Shape distributions. *TOG 21*, 4 (2002), 807–832. 1

[PBB*13] POKRASS J., BRONSTEIN A. M., BRONSTEIN M. M., SPRECHMANN P., SAPIRO G.: Sparse modeling of intrinsic correspondences. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 459–468. 2

[PKG03] PAULY M., KEISER R., GROSS M.: Multi-scale feature extraction on point-sampled surfaces. *CGF 22*, 3 (2003), 281–289. 1

[PP93] PINKALL U., POLTHIER K.: Computing discrete minimal surfaces and their conjugates. *Experimental Mathematics 2*, 1 (1993), 15–36. 6

[R*14] RODOLÀ E., ET AL.: Dense non-rigid shape correspondence using random forests. In *Proc. CVPR* (2014). 2

[Rus07] RUSTAMOV R. M.: Laplace-Beltrami eigenfunctions for deformation invariant shape representation. In *Proc. SGP* (2007). 1, 2, 3

[RWP06] REUTER M., WOLTER F.-E., PEINECKE N.: Laplace–beltrami spectra as 'shape-dna'of surfaces and solids. *Computer-Aided Design 38*, 4 (2006), 342–366. 1

[SB11] SIPIRAN I., BUSTOS B.: Harris 3D: a robust extension of the harris operator for interest point detection on 3D meshes. *Visual Computer 27*, 11 (2011), 963–976. 1

[SOG09] SUN J., OVSJANIKOV M., GUIBAS L. J.: A concise and provably informative multi-scale signature based on heat diffusion. *CGF 28*, 5 (2009), 1383–1392. 1, 3, 7

[SRV13] SHUMAN D. I., RICAUD B., VANDERGHEYNST P.: Vertex-frequency analysis on graphs. *arXiv:1307.5708* (2013). 1, 2, 4

[Zei12] ZEILER M. D.: ADADELTA: An adaptive learning rate method. *arXiv:1212.5701* (2012). 7