Multiview Kernels for Low-Dimensional Modeling of Seismic Events

Ofir Lindenbaum, Yuri Bregman, Neta Rabin, and Amir Averbuch

***Abstract*— The problem of learning from seismic recordings has been studied for years. There is a growing interest of developing automatic mechanisms for identifying the properties of a seismic event. One main motivation is the ability to have a reliable identification of man-made explosions. The availability of multiple high-dimensional observations has increased the use of machine learning techniques in a variety of fields. In this paper, we propose to use a kernel-fusion-based dimensionality reduction framework for generating meaningful seismic representations from raw data. The proposed method is tested on 2023 events that were recorded in Israel and Jordan. The method achieves promising results in the classification of event type as well as the estimation of the event location. The proposed fusion and dimensionality reduction tools may be applied to other types of geophysical data.**

***Index Terms*— Diffusion maps (DMs), dimensionality reduction, multiview, seismic discrimination.**

I. Introduction

MACHINE learning techniques play a central role in data analysis, data fusion, and visualization. As geophysical acquisition tools become more sophisticated and gather more information, data analysts relay more on machine learning techniques for generating meaningful representations of the data. A coherent representation of complex data often includes a feature extraction step followed by a dimensionality reduction step, which results in a compact and visual model. Analysis tasks, such as clustering, classification, anomaly detection, or regression, may be carried out in the constructed low-dimensional space. Common dimensionality reduction methods, such as principal component analysis (PCA) [1] and linear discriminant analysis (LDA) [2], project the feature space into a low-dimensional space by constructing meaningful coordinates that are linear combinations of the original feature vectors. PCA is widely used for low-dimensional modeling of geoscience data sets. Jones and Christopher [3] applied PCA to infer aerosol specification for research of oceans or more complex land surfaces. Griparis *et al.* [4] applied a linear dimensionality reduction tool LDA for a projection of earth observations into a low-dimensional space. Their low-dimensional representation resulted in a cluster organization of the image data by land types. PCA and self-organizing map (SOM) [5] were applied for pattern recognition in volcano seismic spectra by Unglert *et al.* [6] and for geologic pattern recognition by Roden *et al.* [7].

Another key issue in processing large amounts of data is the ability to fuse data from different sensors. Typical seismometers record data using three channels. These three channels capture the motion in the horizontal and perpendicular directions to the earth. Each channel may be processed separately, and the results can be combined. Alternatively, a fused representation may be formed for common analysis. Recent advances in machine learning and, in particular, the use of nonlinear kernel-based algorithm enable to construct data-driven fusions and to compute geometry-preserving low-dimensional embeddings. Such kernel-based embedding techniques are known as manifold learning methods; they include local linear embedding [8], Lapacian eigenmaps [9], and diffusion maps (DMs) [10]. Manifold learning methods overcome the limitations of linear dimensionality reduction tools, such as PCA and LDA [11]. When the relationship between the original high-dimensional points is complex and nonlinear, linear projections may fail to organize the data in a way that is loyal to the intrinsic physical parameters that drives the observed phenomena.

This paper focuses on extending manifold learning techniques for low-dimensional modeling and kernel-based data-driven fusion of seismic data. Identifying the characteristic of seismic events is a challenging and important task. This includes the discrimination between earthquakes and explosions, which is not only an essential component of nuclear test monitoring but it is also important for maintaining the quality of earthquake catalogs. For example, wrong classification of explosions as earthquakes may cause the erroneous estimation of seismicity hazard. The discrimination task is typically performed based on some extracted seismic parameters. These parameters include the focal depth, the ratio between surface wave magnitude and body wave magnitude, and the spectral ratio between different seismic phases [12], [13]. Discrimination methods based on seismic parameters give only a partial solution to the problem. For instance, a larger half of seismic events reported by the Comprehensive Nuclear-Test-Ban Treaty Organization are not screened out as natural

Manuscript received May 26, 2017; revised October 5, 2017 and December 4, 2017; accepted January 2, 2018. Date of publication February 14, 2018; date of current version May 21, 2018. This work was supported by the Research Grant of Pazy Foundation 2015–2017. *(Corresponding author: Ofir Lindenbaum.)*

O. Lindenbaum is with the Applied Mathematics Program, Yale University, New Haven, CT 06520 USA (e-mail: ofirlin@gmail.com).

Y. Bregman is with the Israel National Data Center, Soreq Nuclear Research Center, Yavne 70600, Israel.

N. Rabin is with the Unit of Mathematics, Afeka-Tel-Aviv Academic College of Engineering, Tel Aviv-Yafo 6910717, Israel.

A. Averbuch is with the School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2018.2797537

events or even are not considered for the discrimination at all although most of those events are typically earthquakes [14].

Recently, this problem and other geophysical challenges have been approached using machine learning frameworks. The hidden Markov model was proposed in [15]–[17] and modeled the data in an unsupervised manner. Artificial neural networks [18]–[20] or support vector machines [21], [22] were also used to construct a classifier in a supervised manner. The study in [23] utilizes SOM to distinguish microearthquakes from quarry blasts in the vicinity of Istanbul, Turkey. Manifold learning is used in [24] for seismic phase classification. In [25], a graph is used to detect sea mines in side-scan sonar images. The DM method is used in [26] for visualization of meteorological data. A nonlinear dimensionality reduction is proposed in [27] to discriminate between earthquakes and explosions.

In this paper, the manifold learning approach that was presented in [27] is extended by using a kernel-based fusion method for identification of seismic events. The method is model free, and it is based on signal processing for feature extraction followed by manifold learning techniques for embedding the data. Furthermore, the method reveals the underlying intrinsic physical properties of the data, which results in a natural organization of the events by type. Since seismic data are recorded at multiple channels, we suggest fusing the information to extract a more reliable representation for the seismic recordings. The fusion framework is based on a recent work by Lindenbaum *et al.* [28], [29]. This study extends the DM framework, which has been successfully applied for phase classification [24], estimation of arrival times [30], and event discrimination [31]. Other constructions for fusing kernels were proposed in [32]–[34].

The proposed framework begins with a preprocessing stage in which a time–frequency representation is extracted from each seismic event. The training phase includes the construction of a normalized graph that holds the local connections between the seismic events. A low-dimensional map is then obtained by the eigendecomposition of the graph. The constructed embedding is distance preserving. Thus, the geometry of the data set is kept in the new embedding coordinates. By utilizing the low-dimensional embedding, we demonstrate capabilities of classification, location estimation, and anomaly detection of seismic events.

This paper is organized as follows. Sections II and III present the machine learning frameworks for manifold learning and data fusion. In Section IV, the data set is described. The mathematical methods required for the analysis of seismic data are provided in Section V. The proposed framework and experimental results are presented in Section VI. We conclude this paper in Section VII.

II. Manifold Learning

This section reviews the manifold learning method that is applied in this paper for nonlinear dimensionality reduction, DMs. The main ingredient of the method is a kernel function. Here, radial basis kernel functions are used; their construction is described in detail.

A. Radial Basis Kernel Function

Kernel functions are vastly utilized in machine learning. Classification, clustering, and manifold learning use some affinity measure to learn the relations among data points. A kernel is a predefined similarity function designed to capture the fundamental structure of a high-dimensional data set. Given a high-dimensional data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_M\}, \boldsymbol{x}_i \in \mathbb{R}^D$, a kernel $\mathcal{K} : \boldsymbol{X} \times \boldsymbol{X} \longrightarrow \mathbb{R}$ is an affinity function over all pairs of points in $\boldsymbol{X}$. The discrete kernel is represented by a matrix $\boldsymbol{K}$ with the following properties.

1) Symmetry $K_{i,j} = \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \mathcal{K}(\boldsymbol{x}_j, \boldsymbol{x}_i)$.
2) Positive semidefiniteness: $\boldsymbol{v}_i^T \boldsymbol{K} \boldsymbol{v}_i \geq 0$ for all $\boldsymbol{v}_i \in \mathbb{R}^M$ and $\mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_j) \geq 0$.

These properties guarantee that the matrix $\boldsymbol{K}$ has real eigenvectors and nonnegative real eigenvalues. In this paper, radial basis functions (RBFs) are used for constructing the kernel. The RBF kernel function is defined by

$$K_{i,j} = \exp\left\{-\frac{\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2}{2\sigma^2}\right\}. \tag{1}$$

Applying the Euclidean distance to high-dimensional pairs of distant vectors could somewhat be misleading, as data are typically sparse in the high-dimensional space. For this reason, the decaying property of the Gaussian kernel is beneficial. The Gaussian tends to 0 for distant points, whereas its value is close to 1 for adjacent points.

B. Setting the Kernel's Bandwidth

The kernel's bandwidth σ controls the number of points taken into consideration by the kernel. A simple choice for σ is based on the standard deviation of the data. This approach is good when the data are sampled from a uniform distribution. In this paper, we use a max–min measure. The method was proposed in [35] and aims to find a small scale to maintain local connectivities. The scale is set to

$$\sigma^2_{\text{MaxMin}} = \mathcal{C} \cdot \max_j \left[\min_{i, i \neq j} (\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2)\right] \tag{2}$$

where $\mathcal{C} \in [2, 3]$. Alternative methods, such as [36] and [37], have demonstrated similar results in our experiments.

C. Nonlinear Dimensionality Reduction

Most dimensionality reduction methods are unsupervised frameworks that seek for a low-dimensional representation of complex, high-dimensional data sets. Each method preserves a certain criteria while reducing the dimension of the data. PCA [38] reduces the dimension of the data while preserving most of the variance. Nonlinear methods, such as local linear embedding [8], Laplacian eigenmaps [39], and DMs [10], preserve the local structure of the high-dimensional data. In particular, in DM [10], a metric that describes the intrinsic connectivity between the data points is defined. This metric is preserved in the low-dimensional space, resulting in a distance-preserving embedding. The metric is referred to as diffusion distance; it is defined later in this section.

The DM framework enforces a fictitious random walk on the graph of a high-dimensional data set $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\}$,

$\boldsymbol{x}_i \in \mathbb{R}^D$. This results in a Markovian process that travels in the high-dimensional space only in areas where the sampled data exist. The method has been demonstrated useful when applied to audio signals [37], image editing [40], medical data analysis [41], and other types of data sets.

Reducing the dimension of a data set by construction of DM coordinates is performed using the following steps.

1) Given a data set $\boldsymbol{X}$ to compute an RBF kernel $\boldsymbol{K}$ based on (1).
2) Normalize the kernel using $\boldsymbol{D}$ where $D_{i,i} = \sum_j K_{ij}$. Construct the row stochastic matrix $\boldsymbol{P}$ by
$$P_{i,j} \triangleq \mathcal{P}(\boldsymbol{x}_i, \boldsymbol{x}_j) \triangleq [\boldsymbol{D}^{-1}\boldsymbol{K}]_{i,j}. \tag{3}$$
3) Compute the spectral decomposition of the matrix $\boldsymbol{P}$ to obtain an ordered sequence of eigenvalues $\{\lambda_m\}$ and normalized right eigenvectors $\{\boldsymbol{\psi}_m\}$ that satisfy $\boldsymbol{P}\boldsymbol{\psi}_m = \lambda_m \boldsymbol{\psi}_m, m = 0, \ldots, M-1$.
4) Define the d-dimensional ($d \ll D$) DM representation as
$$\boldsymbol{\Psi}(\boldsymbol{x}_i) : \boldsymbol{x}_i \longmapsto [\lambda_1 \psi_1(i), \ldots, \lambda_d \psi_d(i)]^T \in \mathbb{R}^d \tag{4}$$
where $\psi_m(i)$ denotes the ith element of $\boldsymbol{\psi}_m$.

The power of the DM framework stems from the diffusion distance [see (5)]. It was shown in [10] that the Euclidean distance in the embedded space $\boldsymbol{\Psi}(\boldsymbol{x}_i)$ is equal to a weighted distance between the rows of the probability matrix $\boldsymbol{P}$. This distance is defined as the diffusion distance
$$\mathcal{D}_t^2(\boldsymbol{x}_i, \boldsymbol{x}_j) = \|\boldsymbol{\Psi}_t(\boldsymbol{x}_i) - \boldsymbol{\Psi}_t(\boldsymbol{x}_j)\|^2 = \|\boldsymbol{P}_{i,:} - \boldsymbol{P}_{j,:}\|^2_{\boldsymbol{W}^{-1}} \tag{5}$$
where $\boldsymbol{W}$ is a diagonal matrix with elements $W_{i,i} = ((D_{i,i})/(\sum_{i=1}^M D_{i,i}))$. Thus, the DM embedding is distance preserving, meaning that neighboring points in the high-dimensional space are embedded close to each other in the diffusion coordinates.

III. Data Fusion

Many physical phenomena are sampled using multiple types of sensing devices. Each sensor provides a noisy measurement of a latent parameter of interest. Data fusion is the process of incorporating multiple observation of the same data points to find a more coherent and accurate representation.

Problem Formulation: Given L sets of data points $\boldsymbol{X}^l, l = 1, \ldots, L$. Each view is a high-dimensional data set $\boldsymbol{X}^l = \{\boldsymbol{x}_1^l, \boldsymbol{x}_2^l, \boldsymbol{x}_3^l, \ldots, \boldsymbol{x}_M^l\}, \boldsymbol{x}_i^l \in \mathbb{R}^D$. Find a reliable low-dimensional representation $\boldsymbol{\Psi}(\boldsymbol{X}^1, \ldots, \boldsymbol{X}^L) \in \mathbb{R}^d$.

A. Multiview Diffusion Maps (Multiview DM)

An approach for fusion kernel matrices in the spirit of DM framework was presented in [28]. The idea is to enforce a random walk model based on the kernels that model each view by restraining the random walker to "hop" between views in each time step.

The construction requires to compute a Gaussian kernel for each view
$$K_{i,j}^l = \exp\left\{-\frac{\|\boldsymbol{x}_i^l - \boldsymbol{x}_j^l\|^2}{2\sigma_l^2}\right\}, \quad l = 1, \ldots, L. \tag{6}$$
Then, the multiview kernel is formed by the following matrix:
$$\widehat{\boldsymbol{K}} = \begin{bmatrix} \boldsymbol{0}_{M\times M} & \boldsymbol{K}^1\boldsymbol{K}^2 & \boldsymbol{K}^1\boldsymbol{K}^3 & \ldots & \boldsymbol{K}^1\boldsymbol{K}^L \\ \boldsymbol{K}^2\boldsymbol{K}^1 & \boldsymbol{0}_{M\times M} & \boldsymbol{K}^2\boldsymbol{K}^3 & \ldots & \boldsymbol{K}^2\boldsymbol{K}^L \\ \boldsymbol{K}^3\boldsymbol{K}^1 & \boldsymbol{K}^3\boldsymbol{K}^2 & \boldsymbol{0}_{M\times M} & \ldots & \boldsymbol{K}^3\boldsymbol{K}^L \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ \boldsymbol{K}^L\boldsymbol{K}^1 & \boldsymbol{K}^L\boldsymbol{K}^2 & \boldsymbol{K}^L\boldsymbol{K}^3 & \ldots & \boldsymbol{0}_{M\times M} \end{bmatrix}. \tag{7}$$
Next, renormalizing using the diagonal matrix $\widehat{\boldsymbol{D}}$, where $\widehat{D}_{i,i} = \sum_j \widehat{K}_{i,j}$, the normalized row-stochastic matrix is defined as
$$\widehat{\boldsymbol{P}} = \widehat{\boldsymbol{D}}^{-1}\widehat{\boldsymbol{K}}, \quad \widehat{P}_{i,j} = \frac{\widehat{K}_{i,j}}{\widehat{D}_{i,i}} \tag{8}$$
where the m, l block is a square $M \times M$ matrix located at $[1 + (m-1)M, 1 + (l-1)M]$, $l = 1, \ldots, L$. This block describes the probability of transition between view $\boldsymbol{X}^m$ and $\boldsymbol{X}^l$. The multiview DM representation for $\boldsymbol{X}^l$ is computed by
$$\widehat{\boldsymbol{\Psi}}_t(\boldsymbol{x}_i^l) : \boldsymbol{x}_i^l \longmapsto \left[\lambda_1^t \psi_1(i + \bar{l}), \ldots, \lambda_d^t \psi_d(i + \bar{l})\right]^T \in \mathbb{R}^d \tag{9}$$
where $\bar{l} = (l-1)\cdot M$. The final low-dimensional representation is defined by a concatenation of all low-dimensional multiview mappings
$$\vec{\widehat{\boldsymbol{\Psi}}}(\vec{\boldsymbol{X}}) = [\widehat{\boldsymbol{\Psi}}(\boldsymbol{X}^1), \widehat{\boldsymbol{\Psi}}(\boldsymbol{X}^2), \ldots, \widehat{\boldsymbol{\Psi}}(\boldsymbol{X}^L)]. \tag{10}$$
This multiview framework enforces a cross-domain probabilistic model. The transition probabilities depend on the connectivities in all views. This probability model in (8) takes into consideration all the various connectivities of nodes $\boldsymbol{x}_i^l$–$\boldsymbol{x}_s^l$ and the connectivities of the corresponding node $\boldsymbol{x}_s^m$ to the destination node $\boldsymbol{x}_j^m$. The probability of transitioning from $\boldsymbol{x}_i^l$ to $\boldsymbol{x}_j^m$ could be larger than 0 even if $K_{i,j}^l = 0$ and $K_{i,j}^m = 0$. Assume that there is a subset $\mathcal{S} = \{s_1, \ldots, s_F\}$ such that $K_{i,s_f}^l > 0$ and $K_{s_f,j}^m > 0$, $f = 1, \ldots, F$, by definition of the multiview probability, we get that $\widehat{p}_1(\boldsymbol{x}_i^l, \boldsymbol{x}_j^m) > 0$. This property enables to overcome view specific noise and extract only the common parameters, whereas using a concatenation could enhance the noise from all views. A similar property was demonstrated in [33].

B. Alternative Methods

Here, we provide a brief description of several methods for fusing the views before the application of a spectral decomposition.

1) Kernel Canonical Correlation Analysis (KCCA): This method detailed in [42] and [43] extends the well-known canonical correlation analysis. Two kernels $\boldsymbol{K}^1$and $\boldsymbol{K}^2$ are constructed in each view as in (6), and the canonical vectors $\boldsymbol{v}_1$and $\boldsymbol{v}_2$ are computed by solving the following generalized eigenvalue problem:
$$\begin{bmatrix} \boldsymbol{0}_{M\times M} & \boldsymbol{K}^1 \cdot \boldsymbol{K}^2 \\ \boldsymbol{K}^2 \cdot \boldsymbol{K}^1 & \boldsymbol{0}_{M\times M} \end{bmatrix} \begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix} = \rho \cdot \begin{bmatrix} (\boldsymbol{K}^1 + \gamma \boldsymbol{I})^2 & \boldsymbol{0}_{M\times M} \\ \boldsymbol{0}_{M\times M} & (\boldsymbol{K}^2 + \gamma \boldsymbol{I})^2 \end{bmatrix} \begin{pmatrix} \boldsymbol{v}_1 \\ \boldsymbol{v}_2 \end{pmatrix} \tag{11}$$
where $\gamma \boldsymbol{I}$ are regularization terms which guarantee that the matrices $(\boldsymbol{K}^1 + \gamma \boldsymbol{I})^2$ and $(\boldsymbol{K}^2 + \gamma \boldsymbol{I})^2$ are invertible.

2) Kernel Product: Multiplying the kernel matrices element wise $\boldsymbol{K}^{\circ} \triangleq \boldsymbol{K}^1 \circ \boldsymbol{K}^2 \circ \cdots \circ \boldsymbol{K}^L$, $K^{\circ}_{ij} \triangleq K^1_{ij} \cdot K^2_{ij} \cdot \ldots \cdot K^L_{ij}$, then normalizing by the sum of rows. The resulting row stochastic matrix is denoted by $\boldsymbol{P}^{\circ}$. This kernel corresponds to the approach in [10].

In the case of a Gaussian kernel with $\sigma_1 = \sigma_2 = \cdots = \sigma_L$ in (6), the resulting matrix $\boldsymbol{K}^{\circ}$ is equal to the matrix $\boldsymbol{K}^w$ constructed using the concatenated vector $\boldsymbol{w}_i = [(\boldsymbol{x}_i^1)^T, \ldots, (\boldsymbol{x}_i^L)^T]^T$, such that $K^w_{i,j} = \exp\{-((\|\boldsymbol{w}_i - \boldsymbol{w}_j\|^2)/(2\sigma_w^2))\}$. The scale is set to

$$\sigma_w = \sqrt{\sum_{i=1}^{L} \sigma_i^2} = \sqrt{L \cdot \sigma_1^2}. \tag{12}$$

3) Kernel Sum (KS): Defining the sum kernel $\boldsymbol{K}^{+} \triangleq \sum_{l=1}^{L} \boldsymbol{K}^l$. Normalizing the sum kernel by the sum of rows to compute $\boldsymbol{P}^{+}$. This random walk sums the step probabilities from each view. This approach is proposed in [44].

IV. Seismic Data Set

The data set that is used for demonstrating the proposed kernel-based approaches includes 2023 explosions and 105 earthquakes. The 1654 of the explosions occurred at the Shidiya phosphate quarry in the Southern Jordan between 2005 and 2015 (see a map of the region in Fig. 1). These events were reported by the Israel National Data Center at the Soreq Nuclear Research Center with magnitudes $2 \leq \mathrm{ML} \leq 3$ seismic. The rest of the events were taken from the seismic catalog of the Geophysical Institute of Israel between 2004 and 2014. All events were reported in Israel between latitudes 29–32.5°N and long 34.2–35.7°E with duration magnitudes $\mathrm{Md} \geq 2.5$.

Most of the earthquakes in the data set occurred in the Dead Sea transform [45]. The data set includes the February 11, 2004, earthquake with the duration magnitude of $\mathrm{Md} = 5.1$. This has been the strongest event in this area since 1927 [46]. Twelve aftershocks that are included in the data set are associated with this main shock. The majority of the explosions in the data set are ripple-fire query blasts. Moreover, the data set consists of several one-shot explosions, for instance, two experimental underwater explosions in the Dead Sea [46], and surface and near-surface experimental explosions at the Oron quarry [47] and the Sayarim Military Range [48] in the Negev desert.

The data set consists of seismogram recordings from the HRFI (Harif) station in Israel. The station is part of the Israel National Seismic Network [46]. It is equipped with a three-component broadband STS-2 seismometer and a Quanterra data logger. The seismograms are sampled at a frequency of 40 Hz. Waveform segments of 2.5 min (6000 samples) have been selected for every event. In each waveform, the first P phase onsets reside 30 s after the beginning of each waveform. Fig. 1 displays the events on the regional map.

V. Seismic Preprocessing and Feature Extraction Methods

This section provides background on typical methods that are used for seismic signal processing as well as the description of the feature extraction method that was applied here.

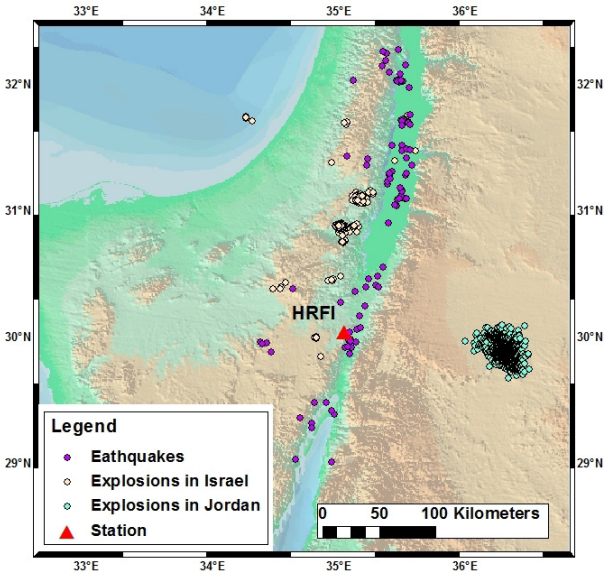


Fig. 1. Seismic events in the data set and the HRFI station.

First, the short-time average/long-time average (STA/LTA) detector is reviewed. Next, we describe how the alignment between the different waveforms was implemented. Last, the feature extraction step, which results in a time–frequency representation of the seismic signal, is described.

A. Short- and Long-Time Averages

Detection of seismic signal embedded in the background noise is a classical problem in the signal processing theory. In the context of statistical decision theory, it may be formulated as a choice between two alternatives: a waveform contains solely the noise or it contains a signal of interest superimposed on the noise. The STA/LTA trigger is a most widely accepted detection algorithm in seismology [49]. It relies on the assumption that a signal is characterized by a concentration of higher energy level compared with the energy level of the noise. This is done by comparing short-time energy average to a long-time energy average using an STA/LTA detector. Usually, a bandpass filter is applied before the STA/LTA test.

Given a time signal $\boldsymbol{y}(n)$, the ratio $R(i)$ at each time instance i is computed as follows:

$$R(i) = \frac{L \cdot \left[\sum_{j=i-S}^{i} y^2(j)\right]}{S \cdot \left(\sum_{j=i-L}^{i} y^2(j)\right)}, \quad i = L+1, \ldots, M \tag{13}$$

where $L \gg S$ are the number of samples used for the long and short averages correspondingly. The ratio $R(i)$ is compared to a threshold δ to identify time windows suspected as seismic events.

B. Seismic Event Alignment

All waveform segments in the data set were extracted according to the first P phase onset time. Those onset times were manually picked by the analysts. However, our selective waveform inspection showed that the P onsets often have

Algorithm 1 Seismic Trigger Alignment

Input: Input time signals $\boldsymbol{y}[n]$.

Output: Estimated time sample $\hat{n}_P$ for P onset of seismic event.

1: Apply a finite impulse response band pass filter to $\boldsymbol{y}[n]$. The filter $\boldsymbol{h}_1$ is designed to pass the signal between $f_L^{(1)} = 2[Hz]$ and $f_H^{(1)} = 4[Hz]$. The filtered signal is denoted as $\tilde{\boldsymbol{y}}^{(1)}[n]$

2: Compute the STA/LTA ratio based on Eq. (13).

3: Set $n^{(1)} \triangleq \min(n)$, s.t. $R(n) > \delta$. The threshold δ is computed based on the following formula $\delta = \min(4, 0.3 \cdot \max(R(n)))$.

4: Repeat steps 1-3 using $f_L^{(2)} = 4[Hz]$, $f_H^{(2)} = 8[Hz]$ and $f_L^{(3)} = 8[Hz]$, $f_H^{(3)} = 12[Hz]$. Denote the trigger indexes as $n^{(2)}$ and $n^{(3)}$.

5: Set the estimated trigger as $\hat{n} \triangleq \min(n^{(1)}, n^{(2)}, n^{(3)})$.

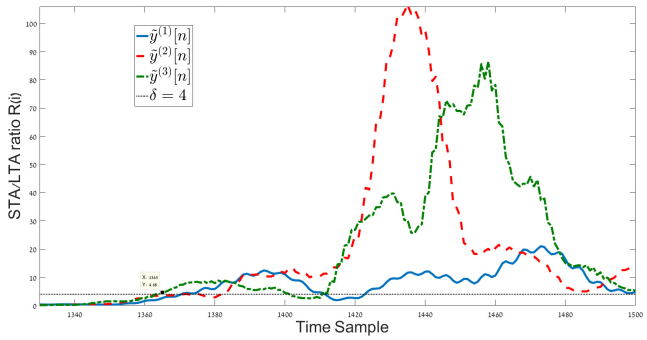


Fig. 2. STA/LTA ratios [see (13)] computed for an earthquake. Each ratio $R(i)$ is computed using one of the three filtered signals $\tilde{\boldsymbol{y}}^{(1)}(n)$, $\tilde{\boldsymbol{y}}^{(2)}(n)$, and $\tilde{\boldsymbol{y}}^{(3)}(n)$. The filters are designed as explained in step 2 of Algorithm 1. The constant black line is an example of a threshold $\delta = 4$. The onset $\hat{n}_P$ is defined as the first cross point of the threshold δ.

actual offsets of several seconds, sometimes even of ten seconds. In order to increase the accuracy of the alignment, Algorithm 1 is proposed to detect the first P onsets.

Algorithm 1 aligns the seismic events based on the STA/LTA ratios which are computed using three filtered versions of the input signal. We assume that most of the energy of the seismic signature is between 2 and 12 Hz. Fig. 2 presents a visual example for the application of Algorithm 1.

C. Feature Extraction

In this paper, a time–frequency representation, named sonograms [50], is used, with some modification. The sonogram is a normalized short-time Fourier transform (STFT) rearranged to be equal tempered on a logarithmic scale. Each raw single-trace seismic waveform input is denoted by $\boldsymbol{y}(n) \in \mathbb{R}^{\bar{N}}$. The length of the signals in this paper is $N = 6000$ with a sampling rate of $F_s = 40$ Hz. An example of seismic signals recorded using three channels is presented in Fig. 3.

The sonogram is extracted from $\boldsymbol{y}(n)$ based on the following steps.

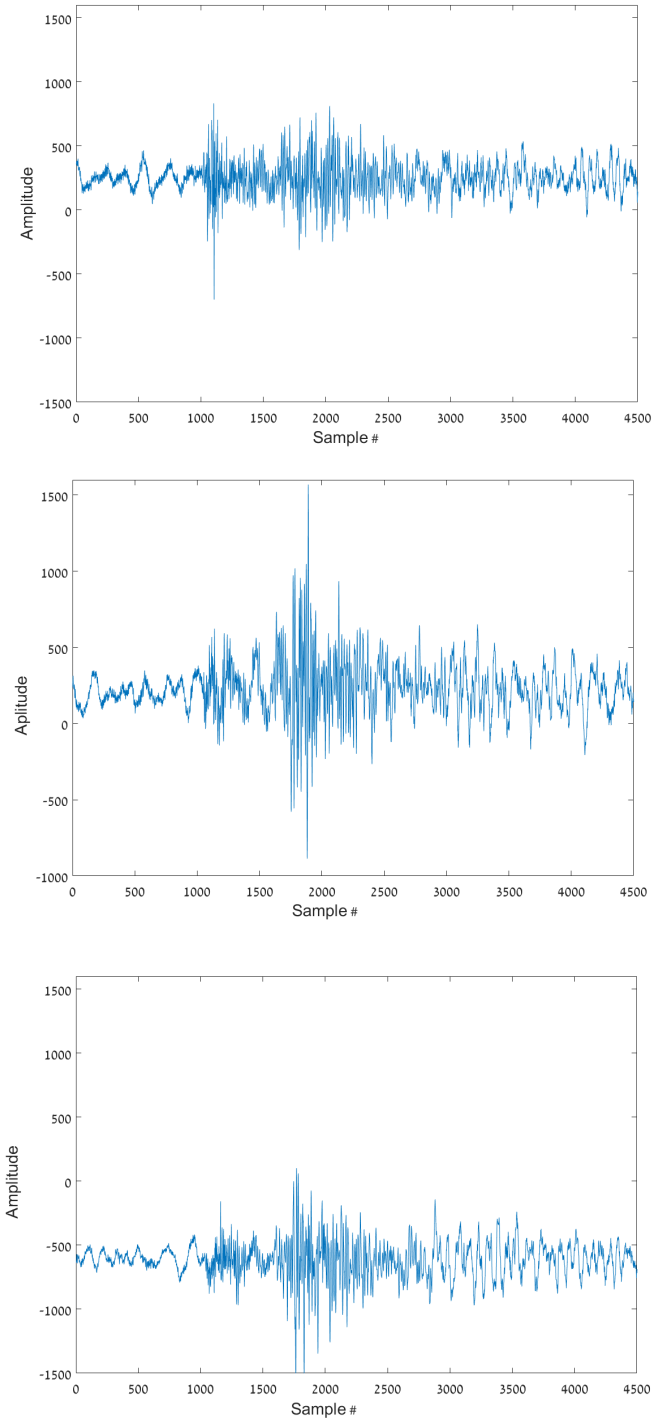


Fig. 3. Four thousand five hundred samples from a recording of an explosion. (Top) Z-channel. (Middle) E-channel. (Bottom) N-channel.

1) Given a recorded signal $\boldsymbol{y}(n) \in \mathbb{R}^N$, the STFT is computed by

$$\boldsymbol{STFT}(f,t) = \sum_{n=1}^{N} w(n - \ell(t)) \cdot y(n) \cdot e^{-j2\pi f n} \quad (14)$$

where $w(n-t)$ is the Hann window function with a length of $N_0 = 256$ and $s = 0.8$ overlap. The time indexes are $\ell(t) = (1-s) \cdot N_0 \cdot t, t = 1, \ldots, T$. The number of time bins is computed using the following

TABLE I
LIST OF FREQUENCY BANDS USED FOR THE SONOGRAM COMPUTATION

Band Number	f-start	f-end
#1	0 [Hz]	0 [Hz]
#2	0.157 [Hz]	0.315 [Hz]
#3	0.315 [Hz]	0.630 [Hz]
#4	0.630 [Hz]	1.102 [Hz]
#5	1.102 [Hz]	1.889 [Hz]
#6	1.889 [Hz]	2.992 [Hz]
#7	2.992 [Hz]	4.567 [Hz]
#8	4.567 [Hz]	6.772 [Hz]
#9	6.772 [Hz]	9.921 [Hz]
#10	9.921 [Hz]	14.331 [Hz]
#11	14.331 [Hz]	20 [Hz]

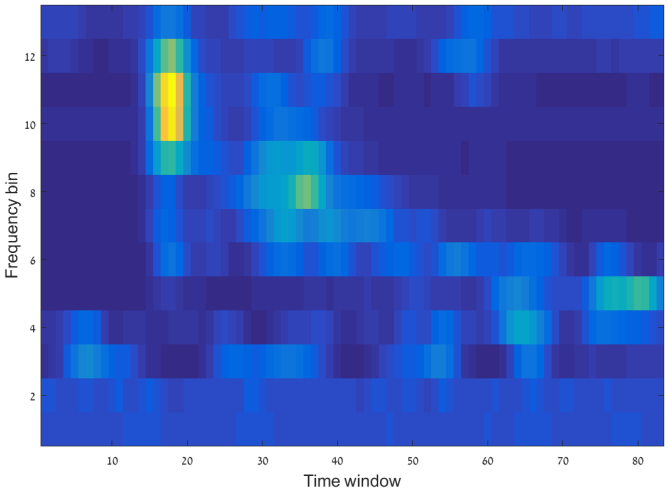


Fig. 4. Sonogram extracted from the E-channel of an explosion seismogram.

equation:

$$T = \left\lceil \frac{N - N_0}{(1-s)\cdot N_0} \right\rceil + 1. \tag{15}$$

2) The spectrogram is the normalized energy of $\boldsymbol{STFT}(f,t)$

$$\boldsymbol{R}(f,t) = \frac{|\boldsymbol{STFT}(f,t)|^2}{N_0}. \tag{16}$$

The spectrogram $\boldsymbol{R}(f,t)$ contains T time bins and $F = N_0$ frequency bins.

3) The frequency scale is then rearrange to be equally tempered on a logarithmic scale, such that the final spectrogram contains 11 frequency bands. The frequency bands are presented in Table I.

4) The bins are normalized such that the sum of energy in every frequency band is equal to 1. The resulted sonogram is denoted by $\boldsymbol{S}(k,t)$, where k is the frequency band number, and t is the time window number. Finally, we transpose the sonogram matrix into a Sonovector $\boldsymbol{x}$ by concatenating the columns such that

$$\boldsymbol{x} = \boldsymbol{S}(:). \tag{17}$$

An example of a sonogram extracted from an explosion is presented in Fig. 4.

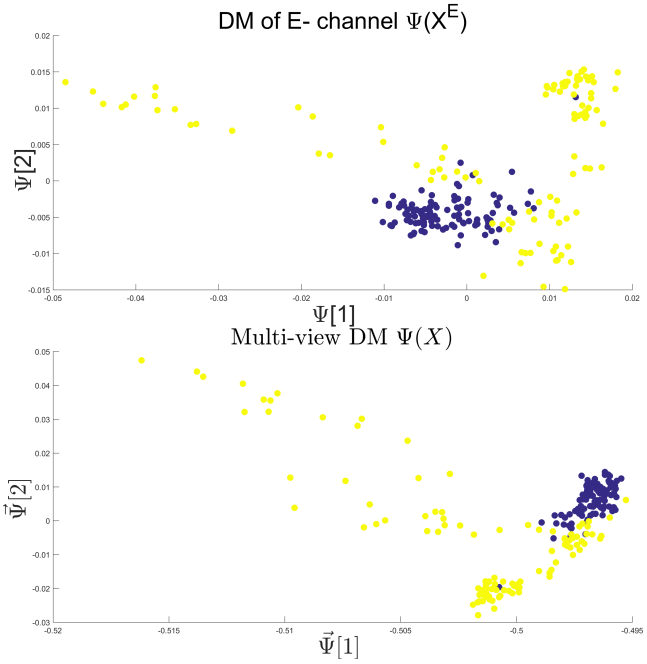


Fig. 5. (Top) Two-dimensional DM mapping extracted from recordings of the E-channel. (Bottom) Two-dimensional multiview DM mapping. Blue points represent man-made explosions from a variety of sources. Yellow points represent recordings of earthquakes, most of which were originated in southern part of Israel.

VI. CASE STUDIES

To evaluate the strength of multiview DM for identifying the properties of seismic events, we perform the following experiments.

A. Discrimination Between Earthquakes and Explosions

We consider the earthquake–explosion discrimination problem as a supervised binary classification task. A homogeneous evaluation data set is constructed by using data from 105 earthquakes and a random sample of 420 explosions. The sampling is repeated 200 times, and the results are the average of all trials. Algorithm 2 is applied to extract a low-dimensional representation of the seismic data. The number of data samples used for each event is 6000, where $N_1 = 1199$ (samples before onset) and $N_2 = 3800$ (samples after onset). Two examples of a 2-D DM mapping are presented in Fig. 5, classes appear more concentrated when using the multiview representation (bottom). In this example, the explosions seem geometrically concentrated, while the earthquakes are spread out. This spread out structure may be associated with the diversity of the time-spectral information describing earthquakes, as oppose to the explosions that were mostly generated in specific quarries. The separation is clearly visible in this example. An evaluation of the separation is performed using a leave-one-out procedure. Test points are classified by using a simple K-NN classifier in a $d = 4$ dimensional representation. The optimal dimension ($d = 4$) for classification was found empirically based on our data set. The average accuracy of classification for various values of K is presented in Fig. 6. Thus, the multiview approach shows better performance with 95% of correct discrimination.

Algorithm 2 Mapping of Seismic Data

Input: Three sets of time signals $\boldsymbol{Y}_E, \boldsymbol{Y}_N, \boldsymbol{Y}_Z$. One for each seismic channel.

Output: A low dimensional multi-view mapping $\vec{\boldsymbol{\Psi}}(\boldsymbol{Y}_E, \boldsymbol{Y}_N, \boldsymbol{Y}_Z)$.

1: Apply Algorithm 1 to each time signal $\boldsymbol{y}_Z^{(i)}$ and estimate the P onset $\hat{n}^{(i)}$.
2: For each channel define the aligned truncated signal as $\bar{\boldsymbol{y}}^{(i)}(n) \triangleq [y^{(i)}(\hat{n}^{(i)} - N_1), \ldots, y^{(i)}(\hat{n}^{(i)} + N_2)]$.
3: Compute the Sonovecs based on Eqs. (14), (16) and (17).
4: Compute the multi-view DM mapping $\vec{\boldsymbol{\Psi}}$ (Eq. (10)).

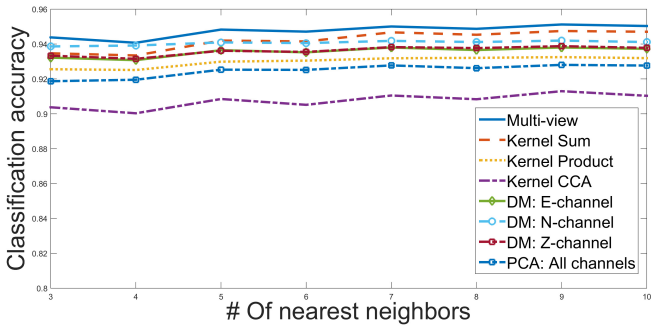


Fig. 6. Classification accuracy for two classes: 105 earthquakes and 420 explosions.

TABLE II
DESCRIPTION OF QUARRY CLUSTERS

Quarry Name	# of events	Center Lat	Center Lon	Dist. to HRFI
Shidiya	250	29.91°	36.32°	125[Km]
Oron	222	30.82°	35.04°	86.7[Km]
Rotem	115	31.09°	35.19°	117.7[Km]
M. Ramon	8	30.46°	34.95°	47.3[Km]
Har Tuv	7	31.68°	35.05°	128.2[Km]

B. Quarry Classification

Identification and separation of quarries by attributing the explosions to the known sources are a challenging task in observational seismology [51], [52]. Here, we demonstrate how the DM representation can be utilized to identify the origin of an explosion. For this paper, 602 seismograms of explosions are used. The explosions occurred in five quarry clusters (see Table II and Fig. 7) and the label data were taken from seismic catalogs. It should be noted that the quarry clusters may include several neighboring quarries and the quarry area may be of several kilometers (like Rotem) or more than ten kilometers (like Shidiya). Moreover, the precise (ground truth) location for most of the explosions inside a quarry is not known. We estimate that the hypocenter accuracy in the used seismic catalogs is about a few kilometers for the explosions in Israel and it is more than 10 km for the explosions in Jordan, which are located outside the Israeli seismic network. The mean latitude and longitude are computed for the explosions belonging to each cluster and referred them to the nearby quarry (see Table II).

The application of Algorithm 2 yields a low-dimensional representation of the seismic recordings. Examples of 3-D single-view and multiview mappings are presented in Fig. 8.

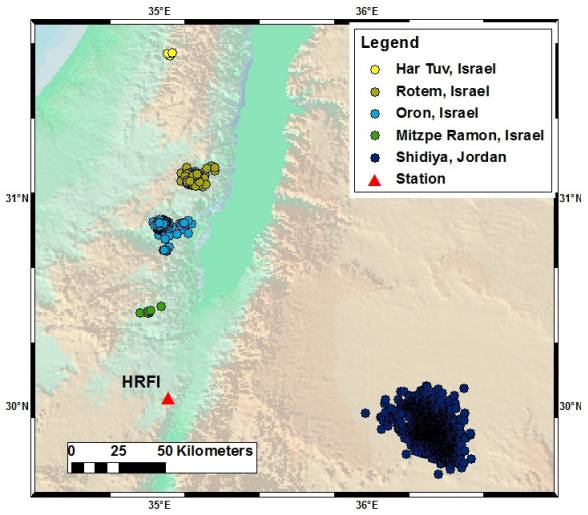


Fig. 7. Map of quarry clusters.

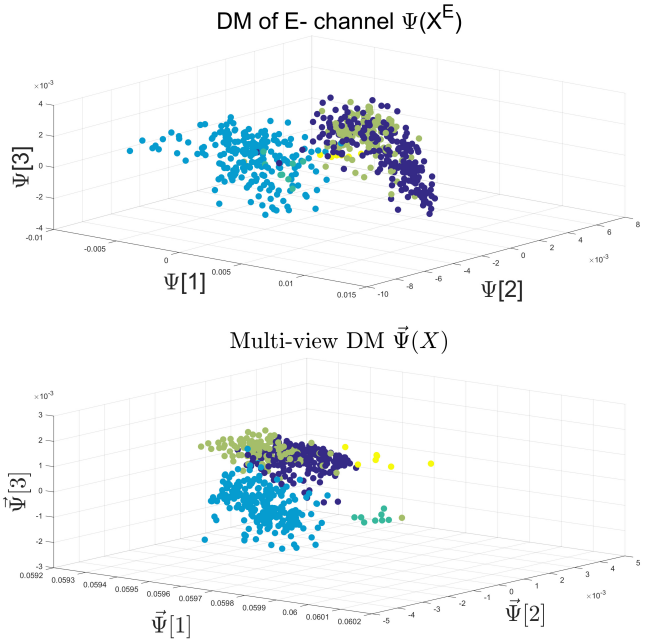


Fig. 8. (Top) 3-D diffusion mapping of 602 explosions based on the E-channel. (Bottom) 3-D multiview diffusion mapping of 602 explosions.

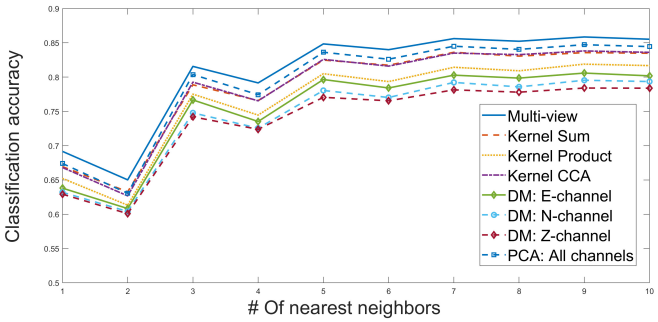


Fig. 9. Classification accuracy for five source locations.

The mapping is followed by a classification step that is performed based on a leave-one-out scheme using K-NN with $K = 3$. The accuracy of the classification is presented in Fig. 9. The multiview approach shows a peak performance of 85% of correct classification rate.

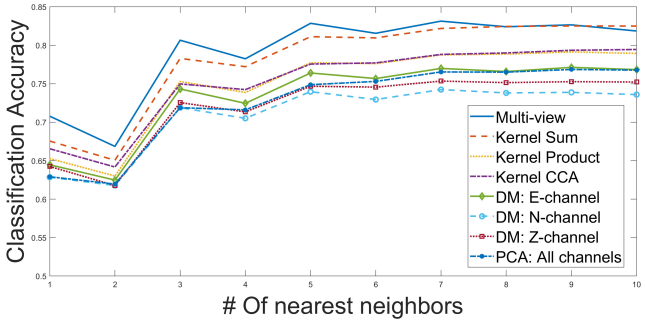


Fig. 10. Classification accuracy for five source locations and earthquakes.

C. Quarry and Earthquakes Classification

In the final classification experiment, we slightly modify the example from Section VI-B by adding the 105 earthquakes. We use Algorithm 1 followed by Algorithm 2. Classification is performed using K-NN in the embedding space. Classification results for different types of embeddings are presented in Fig. 10. We find that once the earthquakes are added, PCA does not perform as well as the kernel-based methods. This could be because of a large variability in the spectral structure of the earthquakes.

D. Location Estimation

The following case study demonstrates how the diffusion coordinates extract underlying physical properties of the sampled signal. In particular, we show that the low-dimensional representation that is generated by DMs organizes the events with respect to their source location, even though this was not an input parameter of the algorithm. The original high-dimensional space holds the sonogram of each event. Nearly, colocated events with the similar source mechanisms and magnitudes should have a similar time–frequency content and, consequently, have similar sonograms. Therefore, we expect them to lie close to each other in the high-dimensional space. The diffusion distance, which is the metric that is preserved in DM, embeds the data while keeping its geometrical structure. Thus, physical properties (such as the source location) that characterize the sonogram and, therefore, define the geometric structure of the points in the high-dimensional space are preserved in low-dimensional DM embedding. Note that such a geometry preserving metric does not exist in linear dimensionality reduction methods like PCA.

The data set for this paper includes 352 explosions that occurred in 4 quarry clustering Israel out 5 clusters above. The explosions in Jordan were removed since they are located at a large distance from the HRFI station. We show that the location of seismic events can be evaluated from the DM embedding coordinates. A similar evaluation based on a linear projection that was calculated with PCA yields a less accurate correlation to the events' true location.

Fig. 11 (top image) displays the longitude and latitude coordinates of catalog locations of the events. These are the source locations of the seismic events. The points are colored by distance in kilometers from HRFI station. The middle and bottom images of Fig. 11 present the 2-D PCA and

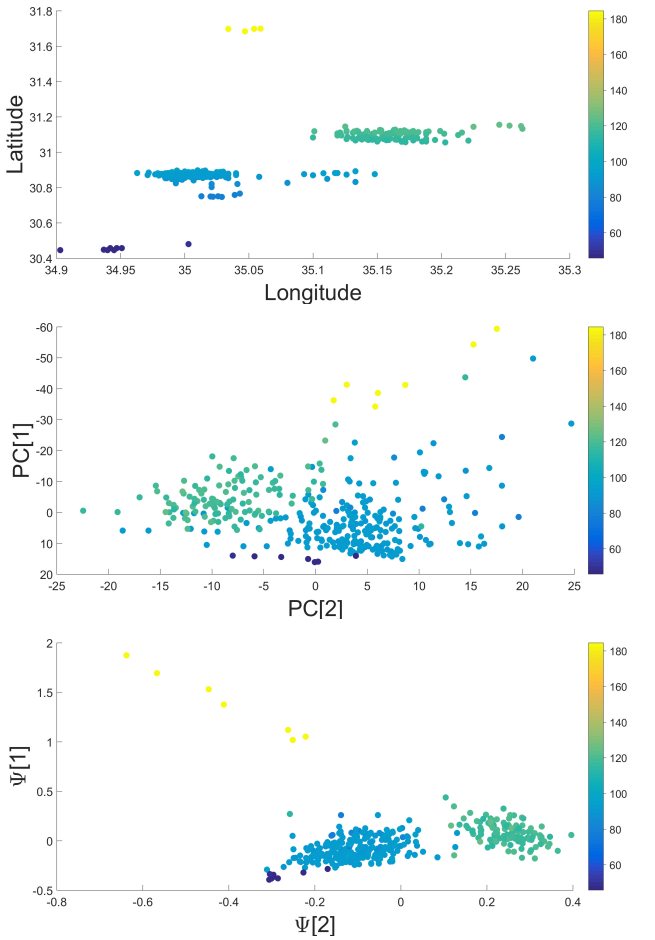


Fig. 11. (Top) Manually estimated location of events. (Middle) First two principle components of the N-channel. (Bottom) First two diffusion coordinates of the N-channel. Color represents the distance from HRFI station.

DM embeddings of the data set, respectively. It is clearly evident that the DM (bottom image in Fig. 11) representation has captured the location variability, while in the PCA representation, this intrinsic factor is less obvious (middle image in Fig. 11). In the DM embedding, the clusters are well separated with respect to the event's location. In PCA, the separation is not as clear, meaning that the low-dimensional PCA representation does not reveal this property. The Pearson correlation coefficients between first two diffusion coordinates and relative latitude and longitude are 0.82 and 0.77 for both dimensions, respectively. The Pearson correlation coefficients between first two principle components and relative latitude and longitude are 0.56 and 0.39, respectively.

E. Detecting Anomalous Events

This case study demonstrates the diffusion representation's ability to detect anomalous events among the set of events at specific site. When two events are nearly colocated and have close magnitudes but with different source mechanisms, then their sonograms should be quite different as well.

Ripple-fire explosions are part of routine mining production cycles at the Oron phosphate quarry in Israel. In July 2006, three experimental one-shot explosions were conducted by

Algorithm 3 K-NN-Based Anomaly Detection

Input: Low dimensional mapping $\boldsymbol{\Psi}$.

Output: A set of indexes $\mathcal{I}$ of suspected anomalies.

1: For each point $\boldsymbol{y}_i, i = 1, \ldots, M$ find the K nearest neighbors from $\boldsymbol{\Psi}(\boldsymbol{y}_i)$. Denote the set of K neighbors for point i as $\mathcal{J}_i = \boldsymbol{\Psi}^i_1, \ldots, \boldsymbol{\Psi}^i_K$.

2: Define the K-NN average distance as $\hat{\boldsymbol{D}}_i \triangleq \sum_{l=2}^{K} \frac{\|\boldsymbol{\Psi}(\boldsymbol{y}_i) - \boldsymbol{\Psi}^i_l\|^2}{K}$.

3: Find all points with average distance $\hat{\boldsymbol{D}}_i$ larger then a threshold δ.

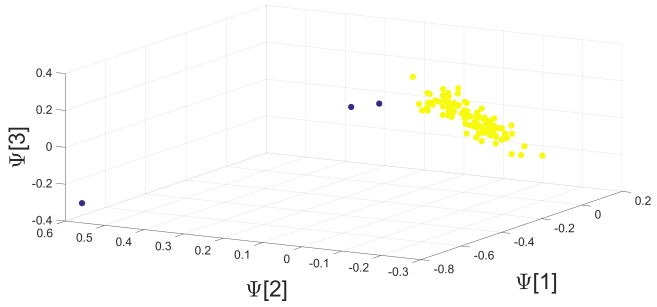


Fig. 12. Diffusion representation of 98 explosions recorded using the Z-channel. The suspected anomalies are colored in blue.

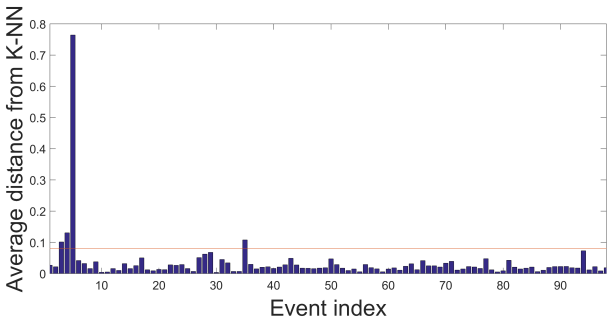


Fig. 13. Average K-NN distance for each explosion. The distance is computed using $d = 3$ coordinates and $K = 5$ nearest neighbors.

The Geophysical Institute of Israel, Lod, Israel, at the Oron quarry [47]. Our goal is to distinguish between the one-shot explosions and the ripple-fire quarry blasts. This is not a trivial task, as all the events were conducted at very close distances.

To remove the variability created by the location of the events, 98 blasts from a small region surrounding the ground truth location of the experimental explosions as reported in [47] are used. Algorithm 2 is applied and a mapping extracted from the Z-channel is used to find the suspected anomalies. The DM embedding is presented in Fig. 12. The three anomaly points are colored in blue; they are clearly separated from the main cluster. The anomalies are automatically identified using Algorithm 3 with $K = 4$ and a threshold set as four times the median of all distances $\hat{D}_i, i = 1, \ldots, M$. The average K-NN distance for the 98 blasts is presented in Fig. 13. The four events that were suspected as anomalies include the three experimental explosions (which are described in [47]). The one false positive appears as events #36 and is anomalous due to the fact that there is a secondary blast within the analyzed time window.

VII. Conclusion

In this paper, we have adapted a multiview manifold learning framework for fusion of seismic recordings and low-dimensional modeling. The abilities of kernel fusion methods for extracting meaningful seismic parameters were demonstrated on various case studies. Various algorithms for the classification of seismic events type, location estimation, and anomaly detection were presented. These algorithms can be used as decision support tools for analysts who need to determine the source, location, and type of recorded seismic events. Correct classification of events results in improved and more accurate seismic bulletins.

The proposed method is model free; thus, it does not require knowledge of physical parameters. The underlying physical parameters are revealed by the DMs and multiview constructions. This type of kernel-based sensor fusion is new in seismic signal processing, and it overcomes some of the limitation of the traditional model-based fusion methods.

Acknowledgment

The authors would like to thank Y. B. Horin for his advice and suggestions. They would also like to thank D. Zakosky and B. Reich for providing the seismic catalog of The Geophysical Institute of Israel, Lod, Israel.

References

[1] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.

[2] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York, NY, USA: Academic, 1990.

[3] T. A. Jones and S. A. Christopher, "Multispectral analysis of aerosols over oceans using principal components," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 9, pp. 2659–2665, Sep. 2008.

[4] A. Griparis, D. Faur, and M. Datcu, "Dimensionality reduction for visual data mining of Earth observation archives," *IEEE Trans. Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1701–1705, Nov. 2016.

[5] T. Kohonen, "The self-organizing map," *Neurocomputing*, vol. 21, nos. 1–3, pp. 1–6, 1998.

[6] K. Unglert, D. Radić, and A. M. Jellinek, "Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra," *J. Volcanol. Geothermal Res.*, vol. 320, pp. 58–74, Jun. 2016.

[7] R. Roden, T. Smith, and D. Sacrey, "Geologic pattern recognition from seismic attributes: Principal component analysis and self-organizing maps," *Interpretation*, vol. 3, no. 4, pp. SAE59–SAE83, 2015.

[8] S. T. Roweis and L. K. Sau, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[9] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[10] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, 2006.

[11] Y. Ma and Y. Fu, Eds., *Manifold Learning Theory and Applications*. London, U.K.: CRC Press, 2012.

[12] R. R. Blandford, "Seismic event discrimination," *Bull. Seismol. Soc. Amer.*, vol. 72, no. 6B, pp. S69–S87, 1982.

[13] A. J. Rodgers, T. Lay, W. R. Walter, and K. M. Mayeda, "A comparison of regional-phase amplitude ratio measurement techniques," *Bull. Seismol. Soc. Amer.*, vol. 87, no. 6, pp. 1613–1621, 1997.

[14] Y. Ben-Horin and Y. Bregman, "Evaluation of the comprehensive nuclear-test-ban treaty seismic monitoring system performance in the Middle East region," Soreq Nucl. Res. Center, Yavne, Israel, Tech. Rep. N4583, 2015.

[15] M. Ohrnberger, "Continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia," Ph.D. dissertation, Inst. Earth Environ. Sci., Univ. Potsdam, Potsdam, Germany, 2001.

[16] M. Beyreuther, C. Hammer, M. Wassermann, M. Ohrnberger, and M. Megies, "Constructing a Hidden Markov Model based earthquake detector: Application to induced seismicity," *Geophys. J. Int.*, vol. 189, no. 1, pp. 602–610, 2012.
[17] C. Hammer, M. Ohrnberger, and D. Fäh, "Classifying seismic waveforms from scratch: A case study in the alpine environment," *Geophys. J. Int.*, vol. 192, no. 1, pp. 425–439, 2013.
[18] T. Tiira, "Discrimination of nuclear explosions and earthquakes from teleseismic distances with a local network of short period seismic stations using artificial neural networks," *Phys. Earth Planetary Interiors*, vol. 97, nos. 1–4, pp. 247–268, 1996.
[19] E. Del Pezzo, A. Esposito, F. Giudicepietro, M. Marinaro, M. Martini, and S. Scarpetta, "Discrimination of earthquakes and underwater explosions using neural networks," *Bull. Seismol. Soc. Amer.*, vol. 93, no. 1, pp. 215–223, 2003.
[20] A. M. Esposito, F. Giudicepietro, S. Scarpetta, L. D'Auria, M. Marinaro, and M. Martini, "Automatic discrimination among landslide, explosion-quake, and microtremor seismic signals at Stromboli volcano using neural networks," *Bull. Seismol. Soc. Amer.*, vol. 96, no. 4A, pp. 1230–1240, 2006.
[21] J. Kortström, M. Uski, and T. Tiira, "Automatic classification of seismic events within a regional seismograph network," *Comput. Geosci.*, vol. 87, pp. 22–30, Feb. 2016.
[22] A. E. Ruano, G. Madureira, O. Barros, H. R. Khosravani, M. G. Ruano, and P. M. Ferreira, "Seismic detection using support vector machines," *Neurocomputing*, vol. 135, pp. 273–283, Jul. 2014.
[23] H. S. Kuyuk, E. Yildirim, E. Dogan, and G. Horasan, "An unsupervised learning algorithm: Application to the discrimination of seismic events and quarry blasts in the vicinity of Istanbul," *Natural Hazards Earth Syst. Sci.*, vol. 11, no. 1, pp. 93–100, 2011.
[24] J. Ramirez, Jr., and F. G. Meyer, "Machine learning for seismic signal processing: Phase classification on a manifold," in *Proc. IEEE 10th Int. Conf. Mach. Learn. Appl. Workshops (ICMLA)*, vol. 1. Dec. 2011, pp. 382–388.
[25] G. Mishne, R. Talmon, and I. Cohen, "Graph-based supervised automatic target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2738–2754, May 2015.
[26] Á. Fernández, A. M. González, J. Díaz, and J. R. Dorronsoro, "Diffusion maps for dimensionality reduction and visualization of meteorological data," *Neurocomputing*, vol. 163, pp. 25–37, Sep. 2015.
[27] N. Rabin, Y. Bregman, O. Lindenbaum, Y. Ben-Horin, and A. Averbuch, "Earthquake-explosion discrimination using diffusion maps," *Geophys. J. Int.*, vol. 207, no. 3, pp. 1484–1492, 2016.
[28] O. Lindenbaum, A. Yeredor, M. Salhov, and A. Averbuch. "Multi-view diffusion maps," unpublished paper, 2015. [Online]. Available: https://arxiv.org/abs/1508.05550
[29] O. Lindenbaum, A. Yeredor, and M. Salhov, "Learning coupled embedding using multiview diffusion maps," in *Proc. Int. Conf. Latent Variable Anal. Signal Separat.*, 2015, pp. 127–134.
[30] K. M. Taylor, M. J. Procopio, C. J. Young, and F. G. Meyer, "Estimation of arrival times from seismic waves: A manifold-based approach," *Geophys. J. Int.*, vol. 185, no. 1, pp. 435–452, 2011.
[31] Y. Bregman, N. Rabin, Y. Ben-Horin, and I. Lev, "Seismic event discrimination using diffusion maps," in *Proc. Sci. Technol. Conf. (CTBT)*, 2015. [Online]. Available https://www.ctbto.org/fileadmin/user_upload/SnT2015/SnT2015_Posters/T3.3-P33.pdf
[32] M. Salhov, O. Lindenbaum, A. Silberschatz, Y. Shkolnisky, and A. Averbuch. "Multi-view kernel consensus for data analysis and signal processing," unpublished paper, 2016. [Online]. Available: https://arxiv.org/abs/1606.08819
[33] R. R. Lederman and R. Talmon, "Learning the geometry of common latent variables using alternating-diffusion," *Appl. Comput. Harmon. Anal.*, 2015.
[34] T. Michaeli, W. Wang, and K. Livescu, "Nonparametric canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2016, pp. 1967–1976.
[35] S. Lafon, Y. Keller, and R. R. Coifman, "Data fusion and multicue data matching by diffusion maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1784–1797, Nov. 2006.
[36] A. Singer, R. Erban, I. G. Kevrekidis, and R. R. Coifman, "Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 38, pp. 16090–16095, 2009.
[37] O. Lindenbaum, A. Yeredor, and I. Cohen, "Musical key extraction using diffusion maps," *Signal Process.*, vol. 117, pp. 198–207, Dec. 2015.
[38] I. T. Jolliffe, "Principal component analysis and factor analysis," in *Principal Component Analysis*. New York, NY, USA: Springer, 1986, pp. 115–128.
[39] W. Luo, "Face recognition based on Laplacian Eigenmaps," in *Proc. IEEE Int. Conf. Comput. Sci. Service Syst. (CSSS)*, Jun. 2011, pp. 416–419.
[40] Z. Farbman, R. Fattal, and D. Lischinski, "Diffusion maps for edge-aware image editing," *ACM Trans. Graph.*, vol. 29, no. 6, 2010, Art. no. 145.
[41] L. Haghverdi, F. Buettner, and F. J. Theis, "Diffusion maps for high-dimensional single-cell analysis of differentiation data," *Bioinformatics*, vol. 31, no. 18, pp. 2989–2998, 2015.
[42] P. L. Lai and C. Fyfe, "Kernel and nonlinear canonical correlation analysis," *Int. J. Neural Syst.*, vol. 10, no. 5, pp. 365–377, 2000.
[43] S. Akaho. (2006). "A kernel method for canonical correlation analysis." [Online]. Available: https://arxiv.org/abs/cs/0609071
[44] D. Zhou and C. J. C. Burges, "Spectral clustering and transductive learning with multiple views," in *Proc. 24th Int. Conf. Mach. Learn.*, 2007, pp. 1159–1166.
[45] Z. Garfunkel, Z. Ben-Avraham, and E. Kagan, Eds., *Dead Sea Transform Fault System: Reviews* (Modern Approaches in Solid Earth Sciences). New York, NY, USA: Springer, 2014.
[46] R. Hofstetter, Y. Gitterman, V. Pinsky, N. Kraeva, and L. Feldman, "Seismological observations of the northern Dead Sea basin earthquake on 11 February 2004 and its associated activity," *Israel J. Earth Sci.*, vol. 57, no. 2, pp. 101–124, 2008.
[47] Y. Gitterman, "Source phenomenology experiments with borehole explosions of special design in Israel," *Bull. Seismol. Soc. Amer.*, vol. 99, no. 3, pp. 1892–1905, 2009.
[48] D. Fee *et al.*, "Overview of the 2009 and 2011 Sayarim infrasound calibration experiments," *J. Geophys. Res., Atmos.*, vol. 118, no. 12, pp. 6122–6143, 2013.
[49] A. Trnkoczy, "Understanding and parameter setting of STA/LTA trigger algorithm," in *New Manual of Seismological Observatory Practice (NMSOP)*, P. Bormann, Ed. Potsdam, Germany: Deutsches GeoForschungs Zentrum GFZ, 2009, pp. 1–20.
[50] M. Joswig, "Pattern recognition for earthquake detection," *Bull. Seismol. Soc. Amer.*, vol. 80, no. 1, pp. 170–186, 1990.
[51] D. B. Harris, "A waveform correlation method for identifying quarry explosions," *Bull. Seismol. Soc. Amer.*, vol. 81, no. 6, pp. 2395–2418, 1991.
[52] D. B. Harris and T. Kvaerna, "Superresolution with seismic arrays using empirical matched field processing," *Geophys. J. Int.*, vol. 182, no. 3, pp. 1455–1477, 2010.

Ofir Lindenbaum received the B.Sc. degree (*suma cum laude*) in electrical engineering and physics from Technion, Haifa, Israel, in 2010, and the Ph.D. degree from the Department of Electrical Engineering, Tel Aviv University, Tel Aviv, Israel, in 2017.

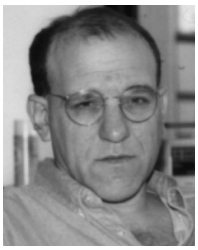
He is currently a Post-Doctoral Associate with the Genetics and the Applied Mathematics Departments, Yale University, New Haven, CT, USA. His research interests include signal processing, music and audio analyses, manifold learning, spectral methods for data mining, and dimensionality reduction.

Yuri Bregman received the M.Sc. degree in applied mathematics from the Latvian State University of Riga, Riga, Latvia, in 1979, and the Ph.D. degree in mathematics from the Peoples' Friendship University of Russia, Moscow, Russia, in 1984.

He held several research and lecturer positions in applied mathematics at the Latvian State University of Riga, from 1984 to 1986, at Riga Technical University, Riga, from 1986 to 1990, and at Haifa University, Haifa, Israel, from 1991 to 1994. Since 1994, he has been a Research Staff Member with the Soreq Nuclear Research Center, Yavne, Israel. His research interests include signal processing and machine learning in geophysics, especially the detection and the characterization of seismic events.

Neta Rabin received the B.Sc. degree in mathematics and computer science and the M.Sc. and Ph.D. degrees in computer science from Tel-Aviv University, Tel Aviv, Israel.

She was a Gibbs Assistant Professor with the Applied Mathematics Department, Yale University, New Haven, CT, USA. She is currently a Senior Lecturer with the Mathematics Department, Afeka Tel-Aviv Academic College of Engineering, Tel Aviv. Her research interests include data analysis, signal processing, and computational harmonic analysis.



Amir Averbuch was born in Tel Aviv, Israel. He received the B.Sc. and M.Sc. degrees in mathematics from Hebrew University, Jerusalem, Israel, in 1971 and 1975, respectively, and the Ph.D. degree in computer science from Columbia University, New York, NY, USA, in 1983.

During 1966–1970 and 1973–1976, he served in the Israeli Defense Forces. From 1976 to 1986, he was a Research Staff Member at the Department of Computer Science, IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. In 1987, he joined the School of Computer Science, Tel Aviv University, Tel Aviv, where he is currently a Professor of computer science. Since 2013, he has been the Co-Founder of ThetaRay Ltd., Hod HaSharon, Israel. His research interests include applied harmonic analysis, big data processing and analysis, wavelets, signal/image processing, numerical computation, and scientific computing (fast algorithms).