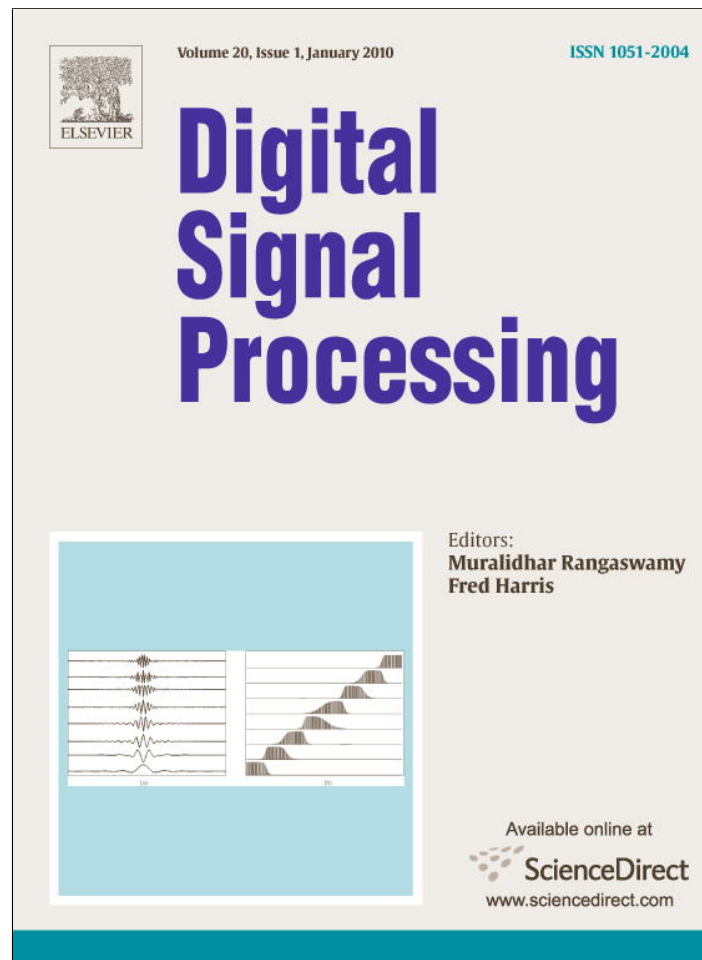


Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

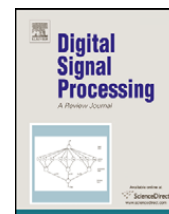
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

A diffusion framework for detection of moving vehicles

A. Schclar*, A. Averbuch, N. Rabin, V. Zheludev, K. Hochman

School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel

ARTICLE INFO

Article history:

Available online 21 February 2009

Keywords:

Acoustic detection
Dimensionality reduction
Diffusion maps
Geometric harmonics
Classification

ABSTRACT

We introduce a novel real-time algorithm for automatic acoustic-based vehicle detection. Commonly, surveillance systems for this task use a microphone that is placed in a target area. The recorded sounds are processed in order to detect vehicles as they pass by. The proposed algorithm uses the wavelet-packet transform in order to extract spatio-temporal characteristic features from the recordings. These features constitute a unique acoustic signature for each of the recordings. A more compact signature is derived by the application of the *Diffusion Maps (DM)* dimensionality reduction algorithm. A new recording is classified according to its compact acoustic signature in the DM reduced-dimension space. The signature is efficiently obtained via the *Geometric Harmonics (GH)* algorithm. The introduced algorithm is generic and can be applied to various signal types for solving different detection and classification problems.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Sounds emitted by a device can be utilized to identify it via characteristic acoustic features that are inherent in these sounds (assuming that different features are associated with different devices). We refer to these characteristic features as *acoustic signatures*. Motorized vehicles emit sounds whose characteristics are determined by many factors. These factors can be divided into two: factors that depend on the vehicle e.g. engine type, gear system, tires, body and traveling speed and factors that are due to the environment conditions where the vehicles are traveling e.g. road type (dirt, paved, etc.) and condition. A characteristic acoustic signature can be extracted from the sounds of every vehicle. We assume that similar vehicles have similar acoustic signatures while vehicles of different types do not. For example, cars, trucks, airplanes and motorized boats each have a unique acoustic signature. Furthermore, the environment where the vehicles travel influences the sounds they emit. However, we do not try to characterize this influence and regard it as background noise. This noise makes the identification of vehicles in various environmental conditions a challenging task. Moreover, background noise that is present in the recordings that are used for the training phase of a detection system poses yet another challenge. Another difficulty is the variability of vehicles' speed and the distance of the vehicles from the microphone. In this paper we wish to detect land vehicles e.g. cars, trucks by separating them from non-land vehicles e.g. planes, helicopters via their acoustic signatures.

A common approach for the extraction of acoustic signatures utilizes signal processing techniques e.g. Fourier analysis. An effective tool for this purpose is the wavelet packet transform which we use in the proposed algorithm. We analyze every recording by inspecting the short-term dynamics of the acoustic signal which we represent as vectors.

The proposed algorithm is composed of two steps: (a) an offline training step; and (b) an online detection step. The input to the training step is a set of vehicle and non-vehicle recordings. The acoustic signature of every recording in the set

* Corresponding author.

E-mail address: alon.schclar@gmail.com (A. Schclar).

is extracted by applying the wavelet packet transform to it followed by reduction of its dimensionality using the Diffusion Maps algorithm [6]. The signatures are stored for use during the detection step. In the classification step, the features of a new recording are extracted and they are embedded in the dimension-reduced space via the Geometric Harmonics out-of-sample extension algorithm [7]. Although we apply the proposed algorithm for the detection of land vehicles, it is generic and can be used for a multitude of detection and identification tasks. The contribution of this paper is two-fold: first, the utilization of the Diffusion Maps and the Geometric Harmonics algorithms for classification and detection; second, the proposed method fortifies (in addition to [14]) the effectiveness of incorporating dimensionality reduction in classification algorithms.

The rest of this paper is organized as follows: in Section 2 we survey previous work on vehicle detection. We describe the proposed algorithm in Section 3. Experimental results are given in Section 4. We conclude in Section 5.

2. Related work

In the following, we describe papers that propose different solutions to the problem at hand—many of them are designated for military use.

In [11], speech recognition techniques were utilized for the classification of vehicle types. Several speech recognition techniques were compared after the application of the short-time Fourier transform to the recordings.

Choe et al. [4] utilized the discrete wavelet transform in order to extract the acoustic features. The classification to vehicle types was achieved by a comparison of the feature vectors to a reference database via statistical pattern matching.

In [8], a time-varying autoregressive modeling approach was used for the analysis of the signals following the application of the discrete cosine transform. Averbuch et al. [2] discriminate between different types of vehicles using features that are derived from the wavelet packet coefficients. The classification of new signals was based on Classification and Regression Trees (CARTs). A similar approach was proposed in [1], however, different techniques were used for the feature extraction and classification. Specifically, the features were extracted using a multiscale local cosine transform that was applied to the frequency domain of the acoustic signal. The classifier was based on the “Parallel Coordinates” [9] methodology.

Another paper which makes use of the wavelet packet transform to extract features is [3]. The algorithm distinguishes between vehicles and background. The algorithm incorporates a procedure which searches for a near-optimal footprint.

The “eigenfaces method” [13], which was originally used for recognition of human faces, was utilized by Wu et al. [14] in order to classify the acoustic signatures of vehicles. A sliding window was used in order to decompose the signal into a series of short-time blocks. Following the decomposition, the blocks were transformed into the frequency domain. Finally, the dimensionality of the windows in the frequency domain was reduced via Principle Component Analysis (PCA). The classification procedure projected the new signal onto the principle components (after applying a similar decomposition and transform into the frequency domain) that were found during the training.

3. The proposed algorithm

The proposed algorithm can be categorized as a supervised learning algorithm i.e. it consists of two steps: training and classification. In the training, feature extraction is applied to the training set whose classification is known. This step obtains the acoustic signatures of the recordings and its novelty lies in its feature extraction procedure—specifically, the utilization of the DM algorithm. The second step classifies new recordings by embedding them into the feature space that was constructed during the training step in order to determine whether or not they belong to a vehicle according to the trained acoustic signatures.

The training step analyzes a training set of T recordings whose classifications are known a priori. The recordings do not necessarily have the same size. Every recording is decomposed into overlapping segments $S_i = \{s_j^i\}$. A segment size is denoted by $l = 2^z$, where $l, z \in \mathbb{N}$. The segments are grouped into a single set $\Sigma = \bigcup_{i=1}^T S_i$. For notational convenience, a single index is used in s_j^i and the output is denoted by $\Sigma = \{s_j\}_{j=1}^{n_s}$ where the total number of segments resulting from the decomposition of all the signals is $n_s \triangleq |\Sigma|$.

Following the decomposition, feature extraction is applied to the segments via the wavelet packet transform. In order to classify a new segment, only a short segment at a time is required (not the entire signal as a batch). Adding the high efficiency of the algorithm makes the proposed scheme suitable for real-time applications.

The wavelet packet transform is an efficient tool that is commonly used for signal processing applications in which both the frequency and the time of the signal are of interest. The proposed algorithm extracts features by applying the 6th order spline wavelet packet, which is described in Appendix A and in more detail in [2,3]. Specifically, the wavelet coefficients are divided to frequency equi-width bands. The features are set to the total magnitude (l_1 norm) of the coefficients in each band—one feature for each band. The total number of features is equal to the number of bands. The feature vectors are then embedded into a dimension-reduced space using the *Diffusion Maps* algorithm ([6] and Appendix B). New recordings undergo the same feature extraction procedure and their corresponding feature vectors (which belong to recordings we wish to classify) are embedded into the *same* space via the Geometric Harmonics [7] algorithm and are classified by using a nearest-neighbor scheme.

Diffusion Maps and *diffusion distances* provide a method for finding meaningful geometric structures in datasets. In most cases, the dataset holds data points in a high-dimensional space \mathbb{R}^n . The Diffusion Maps algorithm constructs coordinates

Algorithm 1. Training step

1. Application of the wavelet packet transform to every segment. The transform uses spline wavelet of order 6.
 2. Feature extraction: Construct a feature vector in which every feature is the total magnitude of the wavelet coefficients in every frequency band.
 3. Every β time-consecutive feature vectors are averaged in order to reduce noise.
 4. Dimensionality reduction: The Diffusion Maps algorithm is applied to further reduce the dimensionality of every averaged feature vector.
-

that parameterize the dataset according to the *diffusion distance* [6]—a metric that preserves the geometry of local neighborhoods (within a ball of 2ε radius) within the data. The global geometric structure is revealed by combining the local overlapping neighborhoods. The Diffusion Maps and Geometric Harmonics algorithms are described in Appendix B. For further details on the Diffusion Maps and Geometric Harmonics algorithms, see [6] and [7].

3.1. The training step

The first part in the training step extracts features from every recording. Each recording is divided into segments. Features are extracted from each segment and are grouped into feature vectors—one vector per segment. The second part in the training step reduces the dimensionality of the feature vectors via the application of the Diffusion Maps algorithm. This training step is summarized in Algorithm 1.

Below is a detailed description of every step in Algorithm 1:

- Step 1:** The sixth order spline wavelet packet is applied up to a scale $D \in \mathbb{N}$ to every segment $s_j \in \Sigma$. Typically, if $l = 2^{10} = 1024$, then $D = 6$ and if $l = 2^9 = 512$ then $D = 5$. The coefficients are taken from the last scale D . This scale contains $l = 2^2$ coefficients that are arranged into 2^D blocks of length 2^{2-D} . Each block is associated with a certain frequency band whose width is $F_N/2^{2-D}$ where F_N is the Nyquist frequency. Thus, these bands form a near uniform partition (equi-width frequency bands) of the Nyquist frequency into 2^D parts. We denote the outcome of this step by $U \triangleq \{u_j\}_{j=1}^{n_s}$, where $u_j \in \mathbb{R}^l$. At the end of this step, each segment $s_j \in \Sigma$ is substituted for the set of its spline wavelet packet coefficients.
- Step 2:** The acoustic signature of the vehicle is constructed according to the wavelet coefficients of the segments. The total magnitude of the wavelet coefficients in every band from the previous step is calculated producing a feature vector with 2^D elements for every segment. The result of this step is denoted by $E = \{e_j\}_{j=1}^{n_s}$ where $e_j \in \mathbb{R}^{2^D}$. This operation reduces the dimension by a factor of 2^{2-D} .
- Step 3:** This step is applied to reduce perturbations and noise. Every β time-consecutive feature vectors that belong to the same recording are averaged. Every feature in an averaged feature vector is the average of its corresponding (β) features in the β time-consecutive feature vectors. The output is denoted by $\tilde{E} = \{\tilde{e}_j\}_{j=1}^{n_a}$ where $n_a = n_s - T \cdot (\beta - 1)$. The number of segments β that are averaged is given as a parameter to the algorithm.
- Step 4:** The Diffusion Maps algorithm ([6] and Appendix B) is applied to the set \tilde{E} . This reduces further the dimensionality of the (averaged) feature vectors. First, a Gaussian kernel $k(\tilde{e}_i, \tilde{e}_j) = \exp(-\frac{\|\tilde{e}_i - \tilde{e}_j\|^2}{2\varepsilon})$ of size $n_a \times n_a$ is constructed. The kernel is normalized by $P(\tilde{e}_i, \tilde{e}_j) = \frac{k(\tilde{e}_i, \tilde{e}_j)}{d(\tilde{e}_i)}$ where $d(\tilde{e}_i) = \sum_{\tilde{e}_l \in \tilde{E}} k(\tilde{e}_i, \tilde{e}_l)$. Thus, $P = p(\tilde{e}_i, \tilde{e}_j)$ is a Markov transition matrix that corresponds to a random walk on the data with the following eigen-decomposition $p(\tilde{e}_i, \tilde{e}_j) = \sum_{\rho \geq 0} \lambda_\rho \psi_\rho(\tilde{e}_i) \phi_\rho(\tilde{e}_j)$. The family of Diffusion maps $\Psi(\tilde{e}_j) = (\lambda_1 \psi_1(\tilde{e}_j), \lambda_2 \psi_2(\tilde{e}_j), \lambda_3 \psi_3(\tilde{e}_j), \dots)$, embeds the dataset $\tilde{E} = \{\tilde{e}_j\}_{j=1}^{n_a}$ into an Euclidean space. In these new coordinates, the Euclidean distance between two points is equal to the diffusion distance between the two corresponding high-dimensional points. The diffusion distance measures the connectivity between the two points within the dataset (see Appendix B for the formal definition of the diffusion distance).

3.2. Classification step

Denote by $r_t = (r(t - \zeta + 1), r(t - \zeta + 2), \dots, r(t))$ the sequence of ζ signal values that were received up to time t and have to be classified. ζ is chosen such that r_t is decomposed into exactly β overlapping segments of size l with the same overlapping quantity that was used in the training step. First, features are extracted from the elements of every segment using a feature extraction procedure that is similar to the one used in the training step. Then, the embedding, which was created in the training step (Section 3.1), is extended to include the new averaged feature vector via the *geometric harmonics* out-of-sample extension algorithm. Finally, the sequence r_t is classified according to the most frequent class of its nearest neighbors. These steps are described in Algorithm 2.

- Step 1:** The 6th order spline wavelet packet is applied to every segment in r_t up to level D .

Algorithm 2. Classification step

1. Application of the wavelet packet transform (spline wavelet of order 6) to the overlapping segments.
2. Construct a feature vector in which each feature is the total magnitude of the wavelet coefficients in every frequency band.
3. Averaging of the β consecutive feature vectors.
4. Extension of the embedding obtained by the Diffusion Maps algorithm in the training step via the Geometric Harmonics algorithm in order to include the result of Step 3.
5. Classification of the new sample according to its nearest neighbors in the embedding space.

Step 2: The wavelet coefficients are divided to equi-width frequency bands and the total magnitude of the coefficients in each band is calculated (in a similar manner to the second step of the training step). The output is denoted by $\{v_j\}_{j=1}^{\beta}$ where $v_j \in \mathbb{R}^{2^D}$.

Step 3: The β feature vectors from the previous step are averaged to a single feature vector.

Step 4: The diffusion embedding of the training set \tilde{E} is extended to include the result of Step 3 by applying the Geometric Harmonics (Appendix B.2 and [7]) out-of-sample extension algorithm. The classification of r_t is determined using a nearest neighbors scheme. Specifically, r_t is classified to the highest occurring class among the $\mu \geq 1$ nearest neighbors of the embedded point in the embedding space. We found empirically that $\mu = 11$ produces satisfactory results.

4. Experimental results

The training set, (*TS*), that was used to train the classifier was composed of 42 recordings: 21 of them were recordings of vehicles such as cars, trucks and vans. The remaining part of the training set consisted of non-vehicle recordings such as airplanes, helicopters, silence, speech, wind, etc. The recordings were sampled at 1000 Hz.

The following parameters were used for the training and classification steps of the algorithm: $l = 512$, the overlap between segments was set to 50%, $\beta = 5$, $D = 6$ and the number of diffusion coordinates (dimension of the embedding space) was set to 9. The value of ε in the kernel was determined according to procedure that is described in Appendix B.1.1. The parameter values for the GH part of the classification step were set to $\delta = 0.1$ and $err = 10^{-5}$. These parameters were determined empirically after several experiments were conducted.

The classification step was tested on recordings that were obtained in various road and wind conditions. For every recording that was used in the classification step we show: (a) the plot of the original signal and (b) the classification probability from the classifier. The classification probability is defined as the number of nearest neighbors classified as vehicles divided by μ . A probability value that is above 0.5 signifies a detection of a vehicle.

Since we need to detect vehicles, an effective algorithm should detect them as soon as they appear in the recording. Furthermore, the number of false positives should be minimal. Once detected, it is of less importance that the probabilities will be above 0.5 throughout the duration the vehicle is heard in the recording. Figs. 1–5 below show that the proposed algorithm exhibits all the required properties.

Figure 1 contains the results of a recording in which a car and a truck pass at times $t = 30$ sec and $t = 50$ sec, respectively. The sections between seconds 37 and 43 and between seconds 50 and 60 were part of the *TS* in the training step.

Figure 2 shows the results of a recording that contains sounds emitted by a car (around $t = 11$ sec) and by another car (around $t = 27$ sec). A helicopter is heard starting from $t = 28$ sec until the end of the recording and a third car appears at $t = 70$ sec while the helicopter is still heard. The car at $t = 70$ sec is detected in spite of the helicopter sound in the background.

Figure 3 shows results for a recording in which two helicopters are heard from $t = 15$ sec until $t = 100$ sec and two trucks are driven at high speed around $t = 110$ sec and around $t = 160$ sec.

Figure 4 shows the results for a recording in which a van passes far away from the recording device starting at $t = 15$ sec.

Figure 5 contains the results for a recording in which a heavy vehicle is heard twice: first, between $t = 20$ sec and $t = 50$ sec and then, between $t = 90$ sec and $t = 110$ sec. During the last 30 seconds of the recording a plane is heard.

The recording in Fig. 1 was part of the training set while the recordings in Figs. 2–5 were not. The classifier exhibits very good results, successfully filtering out background noise and non-vehicle sounds such as wind.

5. Conclusion and future work

In this paper, we presented a novel supervised learning algorithm for the detection of vehicles according to their acoustic recordings. The acoustic signature is derived using the wavelet packet transform and the Diffusion Maps dimensionality reduction algorithm. The classification is obtained by a nearest-neighbor scheme that is employed in the dimension-reduced space. In order to classify a new recording, features are extracted and the result is embedded in the low-dimensional space

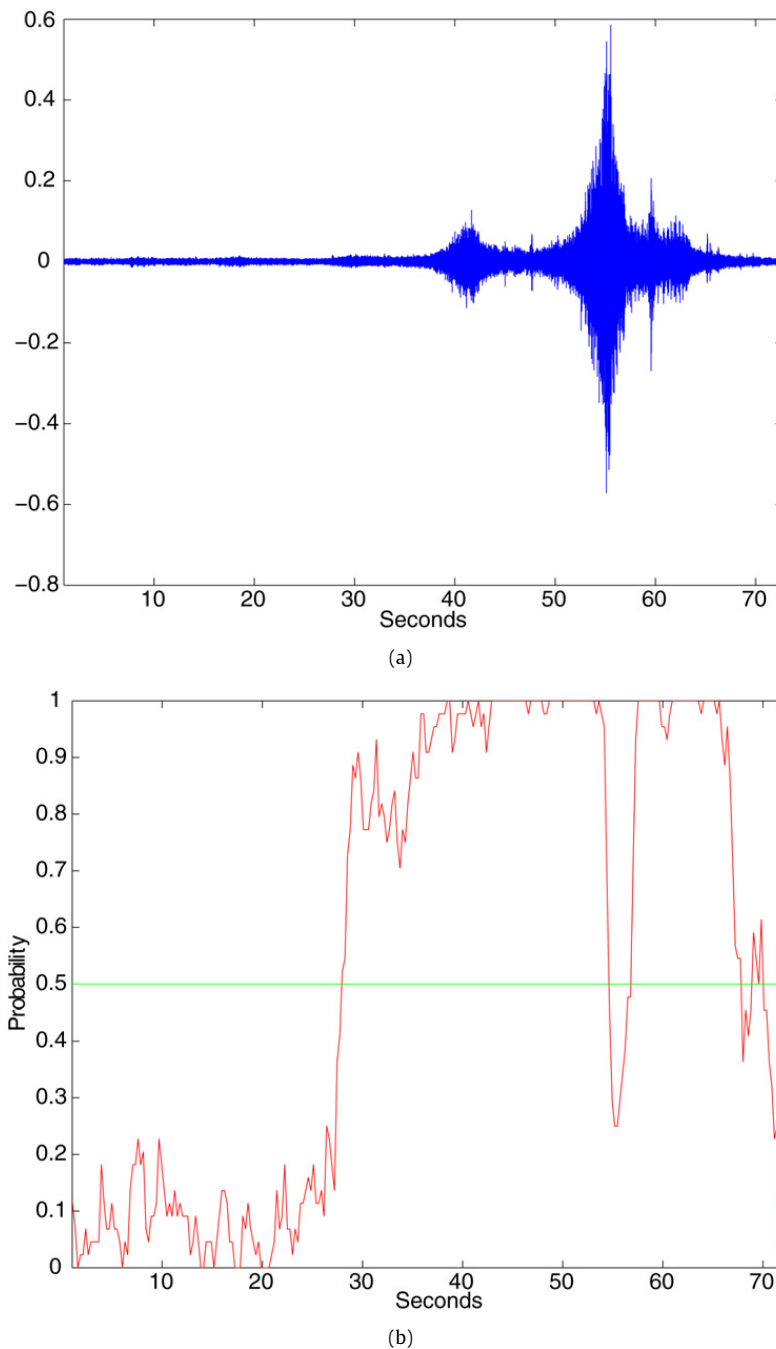


Fig. 1. (a) A recording that contain sounds emitted by a car at $t = 30$ sec and by a truck at $t = 50$ sec. The car and truck sections of this recording were part of the training set. (b) The vehicle detection probability.

via the Geometric Harmonics out-of-sample extension algorithm. The proposed algorithm is robust to background noise and achieves high detection rates in noisy environments.

Utilizing the Diffusion Maps and Geometric Harmonics algorithms in a classification problem is new and the promising results motivate the application of these methods to other detection and classification problems. Furthermore, other dimensionality reduction schemes along with out-of-sample extension algorithms should be investigated as tools for detection and classification problems. In this paper, we used the nearest neighbor classification method. Other classification methods should be examined as well.

Appendix A. Wavelet packet transforms

The wavelet packet transform is carried out by an iterated application of quadrature mirror filters (*QMFs*) and followed by downsampling. For each such pair of *QMFs*, the high pass filter $G(t)$ and the low pass filter $H(t)$ are applied successively to a finite sampled signal, in order to build a binary tree of coefficients. For example, the application of the first iteration of the wavelet packet transform on a signal f of length 2^n produces two blocks, w_0^1 and w_1^1 , of size $\frac{n}{2}$. The w_0^1 block

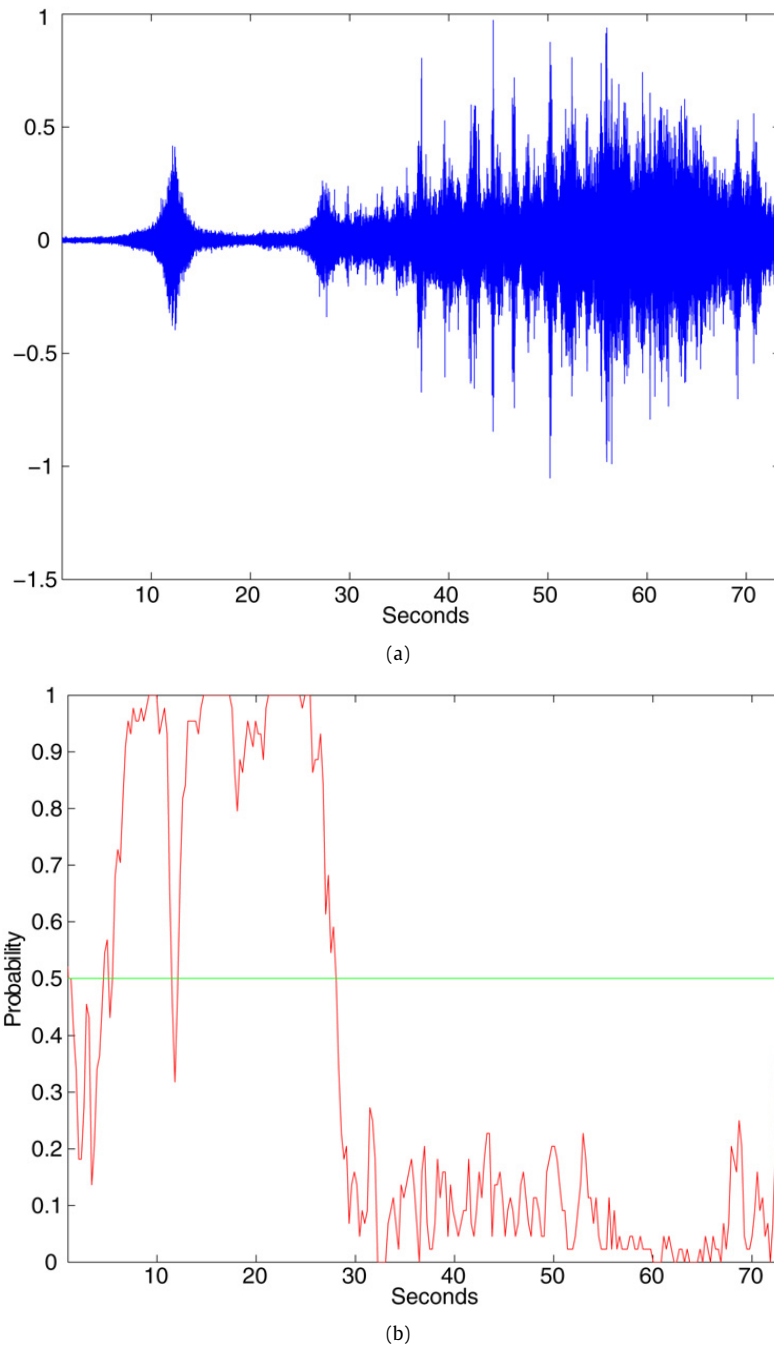
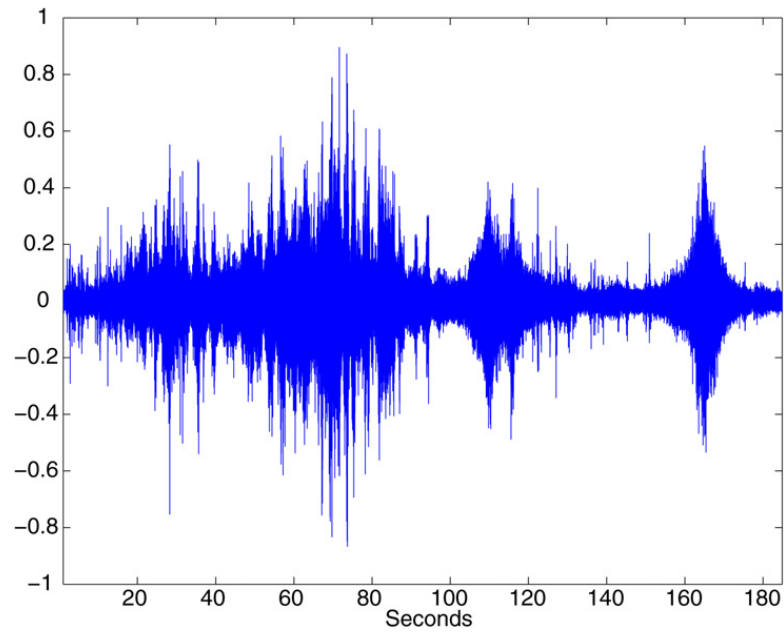


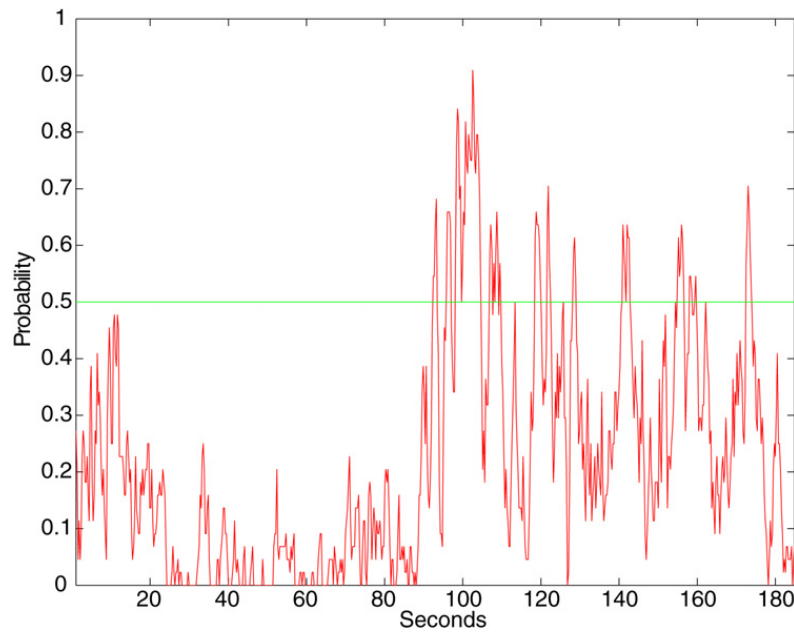
Fig. 2. (a) A recording that contain sounds emitted by a car (around $t = 11$ sec) and by another car (around $t = 27$ sec). A helicopter is heard starting from $t = 28$ sec until the end of the recording and a third car appears at $t = 70$ sec while the helicopter is still heard. The car at $t = 70$ sec is detected in spite of the helicopter sound in the background. (b) The vehicle detection probability.

contains the coefficients for the reconstruction of the low frequency component of the signal. Similarly, the high frequency component can be reconstructed from the w_0^1 block. A second application of the G and H filters on w_0^1 and w_1^1 produces the second level blocks $w_0^2, w_1^2, w_2^2, w_3^2$. All of these blocks are stored in the second level and transformed into eight blocks in the third level after the third iteration, etc. The involved waveforms are well localized in the time and frequency domains. Their spectra form a refined partition of the frequency domain (into 2^j parts in scale j) where each block of the wavelet packet coefficients describes a certain frequency band. The flow of the wavelet packet transform is illustrated in Fig. 6. The partition of the frequency domain corresponds approximately to the location of the blocks in the diagram.

There are many wavelet packet libraries. They differ by their generating filters H and G , the shape of the basic waveforms and their frequency content. In the proposed algorithm, we use the 6th order spline wavelet packets. Fig. 7 shows the frequency and time waveform of the third scale wavelet packets generated by wavelet splines of the 6th order. While the splines do not have a compact support in the time domain, they are localized fairly well in the frequency domain. They produce perfect splitting of the frequency domain.



(a)



(b)

Fig. 3. (a) A recording in which two helicopters are heard from $t = 15$ sec until $t = 100$ sec and two trucks pass by at high speed around $t = 110$ sec and $t = 160$ sec. (b) The vehicle detection probability.

Appendix B. Diffusion framework

B.1. Diffusion maps

Let Γ be a set of points in \mathbb{R}^n . A graph (Γ, K) is constructed with a weight function that is defined by the kernel $k(x, y)$. The weights measure the pairwise similarity between the points. The weight function is:

symmetric: $k(x, y) = k(y, x)$;

non-negative: $k(x, y) \geq 0$ for all x and y in Γ ; and

positive semi-definite: for every real-valued bounded function f defined on Γ $\sum_{x \in \Gamma} \sum_{y \in \Gamma} k(x, y) f(x) f(y) \geq 0$.

The following normalization transforms the kernel into a Markov transition matrix P : $P \triangleq p(x, y) = \frac{k(x, y)}{d(x)}$ where $d(x) = \sum_{y \in \Gamma} k(x, y)$ is the degree of the node x . This normalization is known as the weighted Graph Laplacian normalization [5].

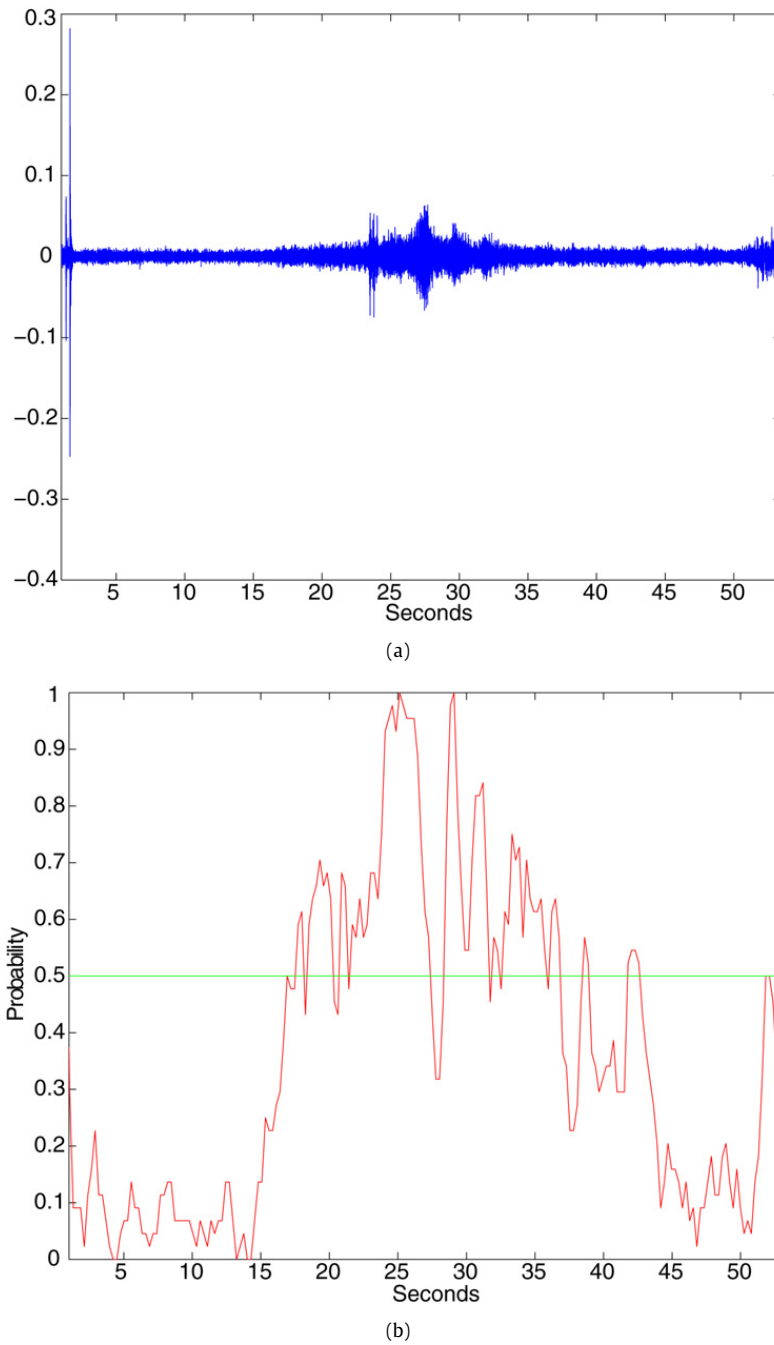


Fig. 4. (a) A recording in which a van passes far from the recording device starting at $t = 15$ sec. (b) The vehicle detection probability.

Since P consists of non-negative real numbers, where each row is summed to 1, the matrix P can be viewed as the Markov transition matrix of a random walk through the points in Γ . The probability to move from x to y in *one* time step is given in $p(x, y)$. These probabilities measure the connectivity of the points within the graph.

The transition matrix P is conjugate to a symmetric matrix A given by $a(x, y) = \sqrt{d(x)}p(x, y)\frac{1}{\sqrt{d(y)}}$. Using a matrix notation, $A \triangleq D^{\frac{1}{2}}PD^{-\frac{1}{2}}$, where D is the diagonal matrix with the values $\sum_y k(x, y)$ on its diagonal. The symmetric matrix A has n real eigenvalues $\{\lambda_l\}_{l=0}^{n-1}$ and a set of orthonormal eigenvectors $\{v_l\}$ in \mathbb{R}^n , thus, it has the following spectral decomposition:

$$a(x, y) = \sum_{k \geq 0} \lambda_k v_k(x)v_k(y). \tag{B.1}$$

Since P is conjugate to A , the eigenvalues of both matrices are identical. In addition, if $\{\phi_l\}$ and $\{\psi_l\}$ are the corresponding left and right eigenvectors of P , then we have the following equalities:

$$\phi_l = D^{\frac{1}{2}} v_l, \quad \psi_l = D^{-\frac{1}{2}} v_l. \tag{B.2}$$

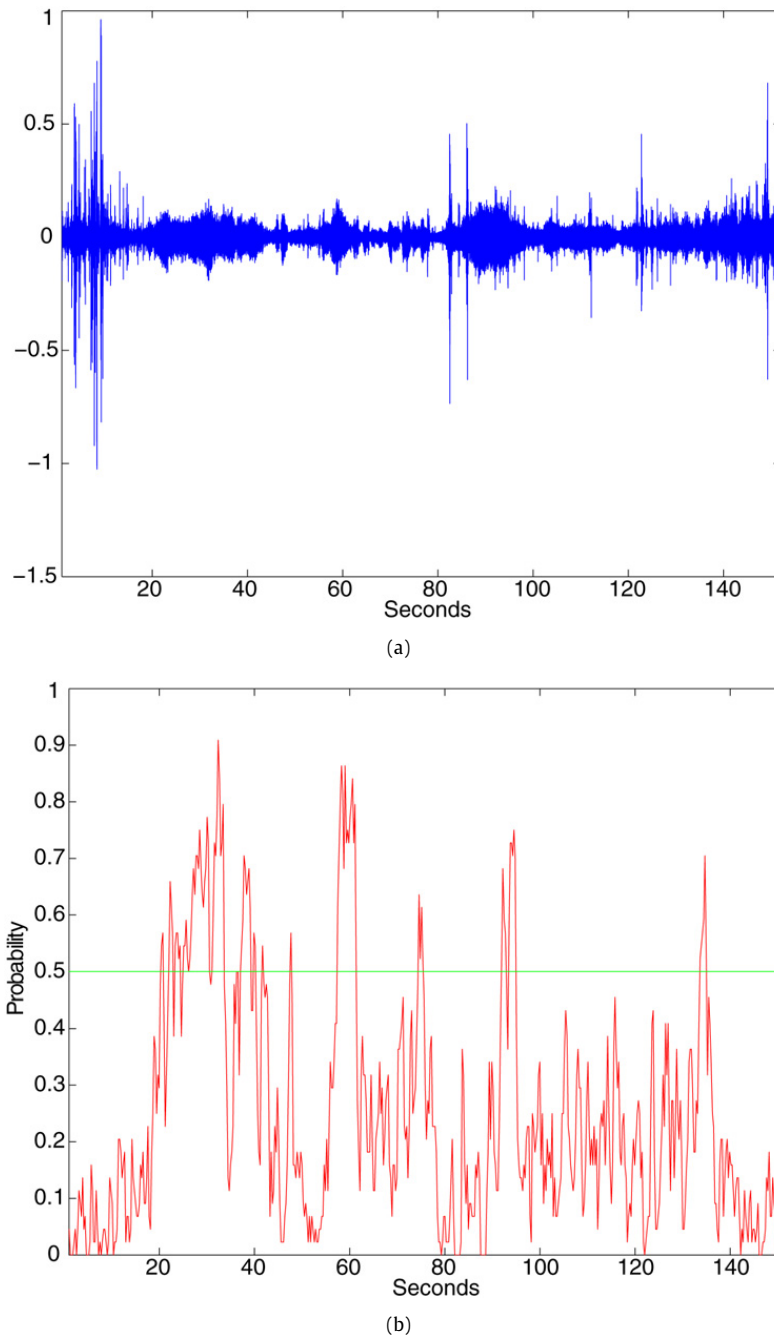


Fig. 5. (a) A recording in which a heavy vehicle is heard twice: first between $t = 20$ sec and $t = 50$ sec and then between $t = 90$ sec and $t = 110$ sec. During the last 30 seconds of the recording a plane is heard. (b) The vehicle detection probability.

From the orthonormality of $\{v_i\}$ and Eq. (B.2) it follows that $\{\phi_l\}$ and $\{\psi_l\}$ are biorthonormal i.e. $\langle \phi_m, \psi_l \rangle = \delta_{ml}$. Combing Eqs. (B.1) and (B.2) together with the biorthogonality of $\{\phi_l\}$ and $\{\psi_l\}$ leads to the following eigen-decomposition of the transition matrix P

$$p(x, y) = \sum_{l \geq 0} \lambda_l \psi_l(x) \phi_l(y). \tag{B.3}$$

When the spectrum decays rapidly (provided ε is appropriately chosen—see Appendix B.1.1), only a few terms are required to achieve sufficient accuracy in the sum.

The family of Diffusion Maps $\{\Psi(x)\}$ defined by $\Psi(x) = (\lambda_1 \psi_1(x), \lambda_2 \psi_2(x), \lambda_3 \psi_3(x), \dots)$ embeds the dataset into a Euclidean space. This embedding constitutes a new parametrization of the data in a low-dimensional space. We recall the diffusion distance between two data points x and y as it was defined in [6]:

$$D^2(x, y) = \sum_{z \in \Gamma} \frac{(p(x, z) - p(z, y))^2}{\phi_0(z)}. \tag{B.4}$$

WAVELET PACKETS

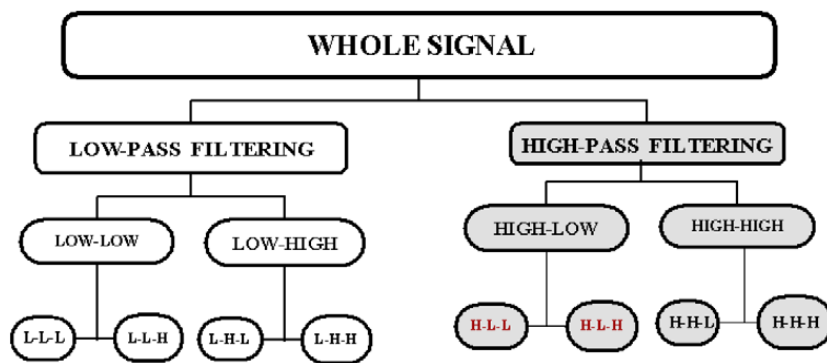


Fig. 6. The multiscale decomposition resulting from the application of the wavelet packet transform.

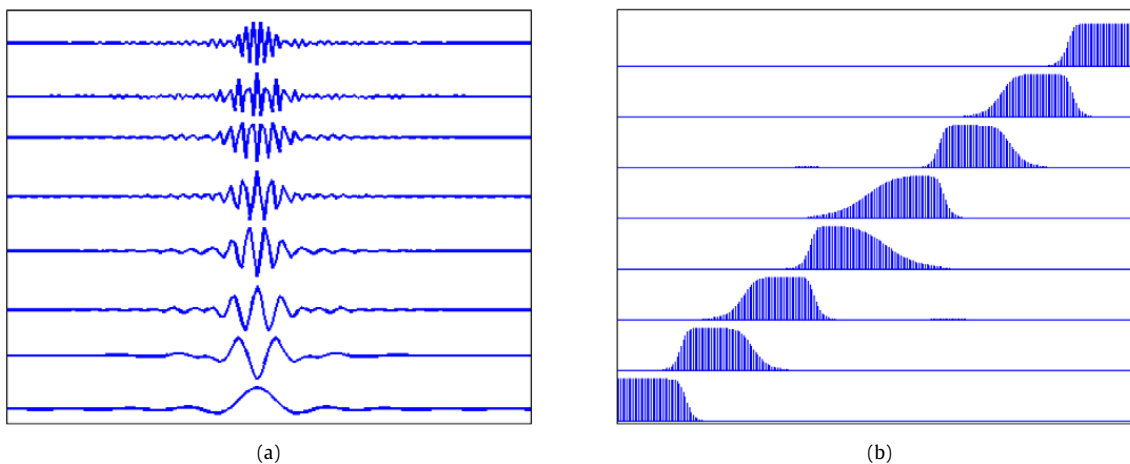


Fig. 7. Frequency (left) and time (right) waveforms of spline-6 wavelet packets of the third scale and their spectra.

This distance reflects the geometry of the dataset and is inverse-proportional to the number of paths connecting x and y . Substituting Eq. (B.3) in Eq. (B.4) together with the biorthogonality property allows to express the diffusion distance using the right eigenvectors of the transition matrix P :

$$D^2(x, y) = \sum_{l \geq 1} \lambda_l^2 (\psi_l(x) - \psi_l(y))^2. \tag{B.5}$$

In these new coordinates, the Euclidean distance between two points in the embedding space is equal to the diffusion distance between the two high dimensional points as defined by the random walk. Moreover, this facilitates the embedding of the original points in a low-dimensional Euclidean space \mathbb{R}^η by:

$$\mathcal{E}_t : x_i \rightarrow (\lambda_2^t \psi_2(x_i), \lambda_3^t \psi_3(x_i), \dots, \lambda_\eta^t \psi_\eta(x_i)), \tag{B.6}$$

which also provides coordinates on the set Γ . Usually, $\eta \ll n$ due to the fast decay of the eigenvalues of P . Furthermore, η depends only on the dimensionality of the data as captured by the random walk and not on the original dimensionality of the data.

B.1.1. Choosing ε

The choice of ε is critical to achieve the optimal performance by the DM algorithm since it defines the size of the local neighborhood of each point. On one hand, a large ε produces a coarse analysis of the data as the neighborhood of each point will contain a large number of points. In this case, the diffusion distance will be close to one for most pairs of points. On the other hand, a small ε might produce many neighborhoods that contain only one point. In this case, the diffusion distance is zero for most pairs of points.

The best choice lies somewhere between these two extremes. A heuristic that produces satisfactory results is the median of all pairwise Euclidean distances of the points in Γ : $\varepsilon = \text{median}\{\|x - y\|\}_{x, y \in \Gamma}$.

B.2. Geometric harmonics

Geometric Harmonics (GH) is a method that facilitates the extension of any function $f : \Gamma \rightarrow \mathbb{R}$ to a set of new points that are added to Γ . In our setting, every coordinate of the low-dimensional embedding constitutes such a function. Let Γ

be a set of points in \mathbb{R}^n and Ψ be its diffusion embedding. Let $\bar{\Gamma}$ be a set in \mathbb{R}^n such that $\Gamma \subseteq \bar{\Gamma}$. The GH scheme extends Ψ_Γ into the dataset $\bar{\Gamma}$ and generalizes the Nyström extension [12] method which we describe next.

B.2.1. The Nyström extension

The eigenvectors and eigenvalues of a Gaussian kernel on the training set Γ with width ε are computed by

$$\lambda_l \varphi_l(x) = \sum_{y \in \Gamma} e^{-\frac{\|x-y\|^2}{2\varepsilon}} \varphi_l(y), \quad x \in \Gamma. \quad (\text{B.7})$$

If $\lambda_l \neq 0$, the eigenvectors in Eq. (B.7) can be extended to any $x \in \mathbb{R}^n$ by

$$\bar{\varphi}_l(x) = \frac{1}{\lambda_l} \sum_{y \in \Gamma} e^{-\frac{\|x-y\|^2}{2\varepsilon}} \varphi_l(y), \quad x \in \mathbb{R}^n. \quad (\text{B.8})$$

This is known as the Nyström extension—a common method for extension of functions to out-of-the-training-set points. Such extensions are required in online processes in which new samples arrive and we need to extrapolate a function f on Γ to the new points.

Let f be a function on the training set Γ . In the DM setting, we are interested in extending each of the coordinates of the embedding function $\Psi(x) = (\lambda_1 \psi_1(x), \lambda_2 \psi_2(x), \lambda_3 \psi_3(x), \dots)$. The eigenfunctions $\{\varphi_l\}$ are the outcome of the spectral decomposition of a symmetric positive matrix, thus, they form an orthonormal basis in $\mathbb{R}^{|\Gamma|}$ where $|\Gamma|$ is the number of points in Γ . Consequently, any function f can be written as a linear combination of this basis $f(x) = \sum_l \langle \varphi_l, f \rangle \varphi_l(x)$, $x \in \Gamma$. Using the Nyström extension, as given in Eq. (B.8), f can be defined for any point in \mathbb{R}^n by

$$\bar{f}(x) = \sum_l \langle \varphi_l, f \rangle \bar{\varphi}_l(x), \quad x \in \mathbb{R}^n. \quad (\text{B.9})$$

The above extension allows to decompose every diffusion coordinate ψ_i in the same way using $\psi_i(x) = \sum_l \langle \varphi_l, \psi_i \rangle \varphi_l(x)$, $x \in \Gamma$. In addition, the embedding of a new point $\bar{x} \in \bar{\Gamma} \setminus \Gamma$ can be evaluated in the embedding coordinate system by $\bar{\psi}_i(\bar{x}) = \sum_l \langle \varphi_l, \psi_i \rangle \bar{\varphi}_l(\bar{x})$. The Nyström extension, which is given in Eq. (B.8), has two major drawbacks:

1. The extension distance is proportional to the value of ε used in the kernel. This extension numerically vanishes beyond this distance.
2. The scheme is ill conditioned since $\lambda_l \rightarrow 0$ as $l \rightarrow \infty$.

The second issue can be solved by cutting off the sum in Eq. (B.9) keeping only the eigenvalues (and the corresponding eigenfunctions) satisfying $\lambda_l \geq \delta \lambda_0$ (where $0 < \delta \leq 1$ and the eigenvalues are in descending order of magnitude)

$$\bar{f}(x) = \sum_{\lambda_l \geq \delta \lambda_0} \langle \varphi_l, f \rangle \bar{\varphi}_l(x), \quad x \in \mathbb{R}^n. \quad (\text{B.10})$$

The result is an extension scheme with a condition number δ . In this new scheme, f and \bar{f} do not coincide on Γ , but they are relatively close. The value of ε controls this error. An iterative method for modifying the value of ε with respect to the function to be extended is introduced in [10]. The outline of the algorithm is as follows:

1. Determine an acceptable error for a relatively big value of ϵ . Denote them by err and ϵ_0 , respectively.
2. Build a Gaussian kernel using ϵ_0 such that $k_{\epsilon_0}(x, y) = \exp(-\|x - y\|^2 / 2\epsilon_0)$.
3. Compute the set of eigenvectors for this kernel. Denote this set by $\varphi_l(x)$ and write f as a linear combination of this basis as $f(x) = \sum_l \langle \varphi_l, f \rangle \varphi_l(x)$.
4. Compute the sum

$$r_{err} = \sqrt{\sum_{\lambda_l \leq \delta \lambda_0} |\langle \varphi_l, f \rangle|^2}.$$

5. If $r_{err} < err$ then expand f with this basis as explained in Eq. (B.10). Otherwise, reduce the value of ϵ_0 and repeat the process.

The sum, which is computed in Step 4 of the algorithm, constitutes the reconstruction error of f on the training set. The error will always be at least δ . By decreasing epsilon, the spectrum decays more slowly, and so longer expansions can be used. Consequently, the reconstruction will become more accurate and the error will decrease.

References

- [1] A. Averbuch, E. Hulata, V. Zheludev, Identification of acoustic signatures for vehicles via reduction of dimensionality, Int. J. Wavelets Multiresolut. Inf. Process. 2 (1) (2004) 1–22.

- [2] A. Averbuch, E. Hulata, V. Zheludev, I. Kozlov, A wavelet packet algorithm for classification and detection of moving vehicles, *Multidimens. Syst. Signal Process.* 12 (1) (2001) 9–31.
- [3] A. Averbuch, V.A. Zheludev, N. Rabin, A. Schclar, Wavelet-based acoustic detection of moving vehicles, *Multidimens. Syst. Signal Process.* 20 (1) (2009) 55.
- [4] H.C. Choe, R.E. Karlsten, T. Meitzler, G.R. Gerhart, D. Gorsich, Wavelet-based ground vehicle recognition using acoustic signals, in: *Proc. SPIE*, vol. 2762, 1996, pp. 434–445.
- [5] F.R.K. Chung, *Spectral Graph Theory*, AMS Reg. Conf. Ser. Math., vol. 92, 1997.
- [6] R.R. Coifman, S. Lafon, Diffusion maps, in: *Special Issue on Diffusion Maps and Wavelets*, *Appl. Comput. Harmon. Anal.* 21 (July 2006) 5–30.
- [7] R.R. Coifman, S. Lafon, Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions, in: *Special Issue on Diffusion Maps and Wavelets*, *Appl. Comput. Harmon. Anal.* 21 (July 2006) 31–52.
- [8] Kie B. Eom, Analysis of acoustic signatures from moving vehicles using time-varying autoregressive models, *Multidimens. Syst. Signal Process.* 10 (1999) 357–378.
- [9] A. Inselberg, The plane with parallel coordinates, *Visual Comput.* 1 (1985) 69–91.
- [10] S. Lafon, Y. Keller, R.R. Coifman, Data fusion and multi-cue data matching by diffusion maps, *IEEE Trans. Pattern Anal. Mach. Intell.* 28 (11) (2006) 1784–1797.
- [11] M.E. Munich, Bayesian subspace method for acoustic signature recognition of vehicles, in: *Proceedings of the 12th European Signal Processing Conference EUSIPCO, 2004*.
- [12] E.J. Nyström, Über die praktische Auflösung von linearen Integralgleichungen mit Anwendungen auf Randwertaufgaben der Potentialtheorie, *Comment. Phys.-Math.* 4 (15) (1928) 1–52.
- [13] L. Sirovich, M. Kirby, Low-dimensional procedure for the characterization of human faces, *J. Opt. Soc. Am. A* 4 (1) (March 1987).
- [14] H. Wu, M. Siegel, P. Khosla, Vehicle sound signature recognition by frequency vector principal component analysis, *IEEE Trans. Instrum. Meas.* 48 (5) (October 1999) 1005–1008.

Alon Schclar holds a Ph.D. degree in computer science from Tel-Aviv University, Israel. He received his M.Sc. degree in computer science (summa cum laude) and his B.Sc. degree in mathematics and computer science from Tel-Aviv University, Tel Aviv, Israel. In 2005, he was a fellow at the applied mathematics department at Yale University supported by the Fulbright grant for Israeli doctoral dissertation students and he was also a fellow at the Institute of Pure and Applied Mathematics at UCLA (2004). His areas of interest include machine learning, unsupervised learning, supervised learning, dimensionality reduction, diffusion processes, signal and image processing, computer vision, computer graphics and hyper-spectral image analysis.

Amir Averbuch is a professor of computer science, Tel Aviv University, Tel Aviv, Israel. He received the B.Sc. and M.Sc. degrees in Mathematics from the Hebrew University in Jerusalem, Israel in 1971 and 1975, respectively. He holds a Ph.D. (1983) in Computer Science from Columbia University, New York. In 1977–1986 he was a Research Staff Member in the department of Computer Science, IBM Corp., T.J. Watson Research Center, Yorktown Heights, New York. His research interests include applied harmonic analysis, signal/image processing, wavelets and efficient numerical computation, fast algorithms for computational problems.

Neta Rabin received her B.Sc. degree in mathematics and computer science and her M.Sc. degree in computer science from Tel Aviv University, Tel Aviv, Israel, in 2002 and 2004, respectively. She is currently pursuing her Ph.D. degree at the School of Computer Science, Tel Aviv University.

Valery A. Zheludev received his M.S. degree in Math. Physics in 1963 from St. Petersburg University, Russia. In 1968 he received Ph.D. degree in Math. Physics from Steklov Math. Inst. of Acad. Sci. USSR and the prize "For the best Ph.D. thesis" at St. Petersburg University. In 1991 he received Dr. Sci. degree in Comput. Math. from Siberia Branch of Acad. Sci. USSR. 1963–1965, lecturer at Pedagogical Univ., St. Petersburg, 1965–1968, Ph.D. student, St. Petersburg University, 1968–1970, Associate Professor, Kaliningrad Univ., 1970–1975, Senior Researcher, Research Inst. for Electric Measuring Devices, St.-Petersburg, 1975–1995, Associate then full Professor, St.-Petersburg Military Institute for Construction Engineering, 1995–present, Researcher then Senior Researcher, currently an Associate Professor, School of Computer Science, Tel Aviv University, Israel. Cooperated with the companies Paradigm Geophysical Ltd, Waves Audio Ltd, Bar-Kal Systems Engineering Ltd and ELTA Systems Ltd. Fields of research: wavelet analysis, approximation theory, signal and image processing, geophysics, pattern recognition.

Keren Hochman received the B.Sc. degree in bioinformatics from Ben Gurion University, Beer Sheva, Israel, and M.Sc. degree in applied mathematics from Tel Aviv University, Tel Aviv, Israel, in 2004 and 2007, respectively. She is currently working in the field of computer science.