

Semi-Supervised Multiresolution Classification Using Adaptive Graph Filtering with Application to Indirect Bridge Structural Health Monitoring

Siheng Chen, *Student Member, IEEE*, Fernando Cerda, Piervincenzo Rizzo, Jacobo Bielak, James H. Garrett and Jelena Kovačević, *Fellow, IEEE*

Abstract—We present a multiresolution classification framework with semi-supervised learning on graphs with application to the indirect bridge structural health monitoring. Classification in real-world applications faces two main challenges: reliable features can be hard to extract and few labeled signals are available for training. We propose a novel classification framework to address these problems: we use a multiresolution framework to deal with nonstationarities in the signals and extract features in each localized time-frequency region and semi-supervised learning to train on both labeled and unlabeled signals. We further propose an adaptive graph filter for semi-supervised classification that allows for classifying unlabeled as well as unseen signals and for correcting mislabeled signals. We validate the proposed framework on indirect bridge structural health monitoring and show that it performs significantly better than previous approaches.

Index Terms—multiresolution classification, semi-supervised learning, discrete signal processing on graphs, adaptive graph filter, indirect bridge structural health monitoring

I. INTRODUCTION

CLASSIFICATION is a signal processing task whose goal is to design a map that associates each input signal with a predefined class label. It is widely used in a number of real-world applications, such as geophysical waveform classification [1], radar signal classification [2], structural health monitoring [3], computer-aided diagnosis of medical images and classification of biological images [4]–[6], among others. A generic classification system consists of a feature extractor and a classifier: a feature extractor reduces the dimensionality of the problem, while a classifier labels the features. In many real-world problems, however, reliable features can be hard to extract; for example, different lighting conditions can derail robust face recognition [7]. Moreover, few labeled signals could be available for training as it could be either too time consuming or expensive to label signals; for example, a research pathologist can spend hours labeling one histology image only.

S. Chen is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213 USA. Email: si-henge@andrew.cmu.edu. F. Cerda is with Universidad de Concepción, Concepción, Chile. Email: facera@udec.cl. P. Rizzo is with the Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, PA, 15261, USA. Email: pir3@pitt.edu. J. Bielak and J. H. Garrett are with the Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213 USA. Emails: {jbielak, garrett}@cmu.edu}. J. Kovačević is with the Departments of Electrical and Computer Engineering Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA. Email: jelenak@cmu.edu.

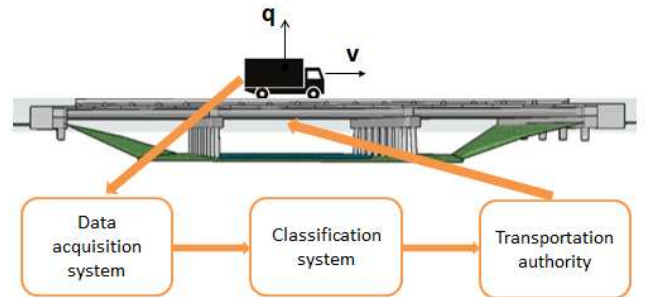


Fig. 1: Indirect bridge structural health monitoring system. Acceleration signals are collected from a moving vehicle and sent to a classification system, which identifies the bridge status and reports it to a transportation authority.

We study one such real-world classification problem, indirect bridge structural health monitoring, in this paper. Assessing and monitoring bridges have been intense areas of interest for some time, especially in the aftermath of several bridge collapses, such as the I-35 bridge over the Mississippi River on Aug. 1, 2007, Shershan Bridge, Pakistan, Sep. 1 2007, Harp Road Bridge, USA, Aug 15 2007, Loncomilla Bridge, Chile, Nov 18, 2004. Significant efforts have been made to reduce the number of structurally deficient bridges, however, near 11% of bridges in the United States are actually structurally deficient; there is thus a need for bridge health monitoring. The current standard in bridge structural health monitoring is based on visual inspections, which are time-consuming, subjective and cannot be done frequently on all the bridges.

Sensor-based structural health monitoring systems have been proposed to automate and improve on the visual inspection process. One approach is to install an array of different sensors, such as strain gauges and accelerometers, directly on the bridge. The drawback is that such sensors still require a sophisticated and expensive electronic infrastructure with installation, maintenance and power support. Recently, indirect approaches have been proposed [8]–[10] based on using moving vehicles to collect acceleration signals from accelerometers inside the vehicles, a more efficient solution that is expected to be economically feasible (see Figure 1).

Acceleration signals collected from existing structures are often nonstationary and have relevant information that is

Generic classification		
$\mathcal{X} = \{x^{(i)}\}$	input dataset	$i = 1, \dots, N$
$\mathcal{L} = \{x^{(i)}\}$	labeled dataset	$i = 1, \dots, L$
$\mathcal{U} = \{x^{(i)}\}$	unlabeled dataset	$i = L + 1, \dots, N$
$\mathcal{Y} = \{y^{(i)}\}$	ground-truth labels for \mathcal{L}	$i = 1, \dots, L$
$\hat{\mathcal{Y}} = \{\hat{y}^{(i)}\}$	estimated labels for \mathcal{U}	$i = L + 1, \dots, N$
F	feature extraction function	
$f^{(i)}$	feature vector	$i = 1, \dots, N$
$q^{(i)}$	ground-truth vector	$i = 1, \dots, L$
Q	ground-truth matrix	$L \times C$
$\hat{q}^{(i)}$	confidence vector	$i = L + 1, \dots, N$

TABLE I: Parameters used in a generic classification system.

present in localized time-frequency regions; long-term behavior such as the natural frequency and the harmonic frequencies can be observed in the frequency domain, while the short-term behavior such as bumps and local roughness can be observed in the time domain. It is well known that *multiresolution techniques*, such as wavelets, are suited to the analysis of such signals, allowing for signal-adapted decompositions. In addition, on operational bridges, signals are gathered frequently but visual inspections typically occur only every two years, which naturally leads to a small number of labeled signals and a large number of unlabeled ones; this constraint calls for *semi-supervised learning*.

We thus propose a novel classification framework that takes advantage of multiresolution classification [4], [11], which extracts hidden features in localized time-frequency regions (subbands), and semi-supervised learning [12], which uses both labeled and unlabeled signals for classification; we make this possible by developing a *semi-supervised* weighting algorithm. In the new framework, (1) each localized subband contributes to the classification by its discriminative power; and (2) both labeled and unlabeled signals provide information.

Among semi-supervised classifiers, graph-based ones are often used, because they are able to represent data with complex structure. We thus propose a novel semi-supervised classifier, *adaptive graph filter*; it extends the applications of discrete signal processing on graphs [13] to classification [14]. A graph structure is built by defining each node to be a signal in a given dataset and each edge to be the similarity between each pair of signals. The adaptive graph filter classifies signals by filtering on the graph structure and propagating labels from labeled to unlabeled signals. It thus allows for classifying unlabeled, but also unseen signals (in which case we also add regression) and for correcting mislabeled signals. Furthermore, we establish the connection to the theory of diffusion maps [15] as well as that of diffusion wavelets [16], which allows us to view the adaptive graph filter as a multiresolution classifier on graphs; the multiresolution nature of the framework is thus felt twofold.

We validate our proposed framework and algorithm on a lab-scale bridge-vehicle dynamic system, and show excellent performance. To show generality of the system without changing the focus of the paper, we include results on additional datasets on the reproducible research page for the paper [17].

Previous Work. Multiresolution classification was orig-

Algorithm 1 Generic classification

Input	\mathcal{X}	input dataset
	\mathcal{Y}	ground-truth labels for \mathcal{L}
Output	$\hat{\mathcal{Y}}$	estimated labels for \mathcal{U}
Function	$C(\mathcal{X})$	
	$f^{(i)} = F(x^{(i)})$	feature extraction
	$\hat{q}^{(i)} = C(f^{(i)}, Q)$	classification
	$\hat{y}^{(i)} = \arg \max_c \hat{q}_c^{(i)}$	
	return $\hat{\mathcal{Y}}$	

inally proposed for bioimaging applications with excellent performance on classifying images of protein subcellular locations [4], developmental stages of Drosophila embryos [5], germ layer components in teratomas [6], and even fingerprint recognition [18]. Previous work on semi-supervised learning includes generative mixture models with expectation maximization, co-training, transductive support vector machine and graph-based approaches [12], each of which makes specific assumptions on how to use unlabeled signals to help classification. Signal processing on graphs has been proposed as a framework to build tools to analyze structured signals and is a rather recent development [13], [19]. Indirect bridge structural health monitoring determines the state of the bridge by using advanced signal processing techniques to analyze vibrational signals collected from the dynamic responses of vehicles traversing a bridge [9], [10], [20]–[22].

Contributions. Our contributions are as follows:

- A novel classification framework that combines multiresolution classification with semi-supervised learning.
- A novel semi-supervised classifier, adaptive graph filter, which allows for classifying unlabeled as well as unseen signals and for correcting mislabeled signals.
- A promising solution to indirect bridge health monitoring validated on a lab-scale bridge-vehicle dynamic system.

Outline of the Paper. Section II states the problem and briefly reviews multiresolution classification, semi-supervised learning and signal processing on graphs; Section III describes our proposed framework for semi-supervised multiresolution classification, while Section IV describes our proposed adaptive graph filter for semi-supervised classification. The algorithms are validated in Section V on acceleration signals collected from a lab-scale bridge-vehicle dynamic system. Section VI concludes with discussion and pointers to future directions.

II. BACKGROUND AND PROBLEM FORMULATION

In this section, we cover the background material necessary for the rest of the paper. We start with the classification problem and then a supervised classification framework, multiresolution classification system. Next, we introduce signal processing on graphs, which lays a foundation for our proposed semi-supervised classifier. Finally, we overview semi-supervised learning, which we will use in Section III.

A. Classification

The goal of classification is to label signals as belonging to one of a number of given classes [23]. Let $\mathcal{X} = \{x^{(i)} \in$

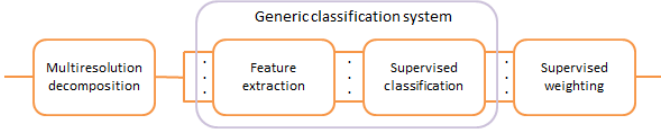


Fig. 2: Supervised multiresolution classification decomposes a signal into localized space-frequency subbands using a given multiresolution transform, followed by feature extraction and supervised classification in each subband, yielding a local classification decision. A supervised weighting algorithm combines all local decisions into a global decision.

$\mathbb{R}^D\}_{i=1}^N$ be the given dataset with $N = L+U$ signals, the first L belonging to the *labeled dataset* $\mathcal{L} = \{x^{(i)} \in \mathcal{X}\}_{i=1}^L$, and the last U belonging to the *unlabeled dataset* $\mathcal{U} = \{x^{(i)} \in \mathcal{X}\}_{i=L+1}^N$. Thus, the inputs to the classifier are the dataset \mathcal{X} and the *ground-truth labels* $\mathcal{Y} = \{y^{(i)} \in \{1, 2, \dots, C\}\}_{i=1}^L$ for the labeled dataset \mathcal{L} , while the outputs are the *estimated labels* $\hat{\mathcal{Y}} = \{\hat{y}^{(i)} \in \{1, 2, \dots, C\}\}_{i=L+1}^N$ for the unlabeled dataset \mathcal{U} (see Table I and Algorithm 1). Note that we categorize signals as labeled, unlabeled and unseen. Both labeled and unlabeled signals are observed in the dataset; while unseen signals are not in the current dataset.

We formulate the problem as designing a map that associates an input signal to a class label with a certain confidence. That is, we view the label as a *confidence vector* $\hat{q} \in \mathbb{R}^C$, where the c th component, \hat{q}_c , is the confidence that a signal belongs to the c th class. The confidence vector for a labeled signal $i = 1, 2, \dots, L$, is the *ground-truth vector*, $q \in \mathbb{R}^C$, with 0s everywhere except 1 in position c indicating membership in class c . The *ground-truth matrix* $Q = [q^{(1)} \ q^{(2)} \ \dots \ q^{(L)}]^T$ of size $L \times C$ collects all L ground-truth vectors as its rows.

Typically, a generic classification system will have an intermediate block between the input and the output, a *feature extractor* F , aimed at reducing the dimensionality of the problem; this is followed by a *classifier* C . If the classifier is supervised, we denote it by SC (see Figure 2 and Algorithm 1).

B. Multiresolution Classification

Multiresolution classification is a supervised classification framework (see Figure 2). It decomposes a signal into S localized space-frequency subbands using multiresolution transforms, both bases and frames [24]–[29]. In each subband, multiresolution classification extracts features, classifies them, and produces a local classification decision. A supervised weighting algorithm combines all local decisions into a global decision. This process implicitly mimics the use of wavelet packets, a data-adaptive multiresolution technique [30], and is summarized in Algorithm 2.

1) *Multiresolution Decomposition*: Multiresolution classification starts with the decomposition of signals using a given multiresolution transform into several localized space-frequency subbands. For the i th signal, the multiresolution

coefficients in the s th subband are

$$a_s^{(i)} = D_s(x^{(i)}),$$

where D_s is the multiresolution transform function in that subband. It is implemented using a signal processing device called a *filter bank*; we have a choice of a number of multiresolution transforms D available, both nonredundant (bases) as well as redundant ones (frames) [24]–[29].

2) *Feature Extraction*: In each subband, features are extracted depending on the application at hand. These features can be generic features, such as texture, Gabor, etc [31]–[38], or can be designed using expert knowledge [39]–[42]. For the i th signal, the feature vector in subband s is

$$f_s^{(i)} = F_s(a_s^{(i)}),$$

where F_s is the feature extraction function in that subband; different subbands can use different feature extraction functions.

3) *Supervised Classification*: In each subband, the features extracted in the previous stage are fed into a supervised classifier. The classifier can be any state-of-the-art classifier, such as logistic regression or support vector machine [23]. For the i th signal, the confidence vector in the s th subband is

$$\hat{q}_s^{(i)} = SC_s(f_s^{(i)}, Q),$$

where SC_s is the supervised classification function in that subband; different subbands can use different classification functions.

4) *Supervised Weighting*: To combine the subbands' classification decisions, we collect subbands' individual confidence vectors $\hat{q}_s^{(i)}$ into a $C \times S$ confidence matrix $\hat{Q}^{(i)}$, and define the weighting function SW as that taking weighted subbands' individual confidence vectors and producing a single confidence vector,

$$\hat{q}^{(i)} = SW(\hat{Q}^{(i)}, Q) = \hat{Q}^{(i)}w.$$

Here, the $S \times 1$ weight vector w assigns a weight to each subband according to its discriminative power; w_s thus tells us how reliable subband s is. The weight vector is found by optimizing a supervised weighting objective function

$$w = \arg \min_{\omega} \left\{ \sum_{i=1}^L \|\hat{q}^{(i)} - \hat{Q}^{(i)}\omega\| \right\}, \quad (1)$$

with the constraint $\|\omega\|_1 = 1$. In other words, the optimal weight vector is the one found to be the most reliable over labeled signals only. After weighting, we compute the global decision as

$$\hat{y}^{(i)} = \arg \max_c \hat{q}_c^{(i)}. \quad (2)$$

Note that in this section we overloaded the symbol \hat{q} , hopefully without confusion; \hat{q}_c is the c th component of the confidence vector \hat{q} , while \hat{q}_s is the confidence vector of subband s .

Algorithm 2 Multiresolution classification

Input	\mathcal{X}	input dataset
	\mathcal{Y}	ground-truth labels for \mathcal{L}
Output	$\hat{\mathcal{Y}}$	estimated labels for \mathcal{U}
Parameters	per subband s	$s = 1, \dots, S$
	D_s	multiresolution function
	$a_s^{(i)}$	multiresolution coefficients
	F_s	feature extraction function
	$f_s^{(i)}$	feature vector
	SC_s	supervised classification function
	$\hat{q}_s^{(i)}$	confidence vector
	for all subbands	
	SW	supervised weighting function
	$\hat{Q}^{(i)}$	confidence matrix
	w	weight vector, $\ w\ _1 = 1$
	$\hat{q}^{(i)}$	final confidence vector
Function	MRC(\mathcal{X})	
	$a_s^{(i)} = D_s(x^{(i)})$	multiresolution decomposition
	$f_s^{(i)} = F_s(a_s^{(i)})$	feature extraction
	$\hat{q}_s^{(i)} = SC_s(f_s^{(i)}, Q)$	supervised classification
	$\hat{q}^{(i)} = SW(\hat{Q}^{(i)}, Q)$	supervised weighting
	$\hat{y}^{(i)} = \arg \max_c \hat{q}_c^{(i)}$	
	return $\hat{\mathcal{Y}}$	

C. Signal Processing on Graphs

With the development of social, biological, and physical networks, signals with complex structure are arising. Traditional discrete signal processing is mainly suited to processing regularly sampled low-dimensional signals, such as discrete time and space signals. To mitigate the problem, signal processing on graphs is emerging as a tool to analyze high-dimensional signals with irregular structure [19], [43], [44], defined on a more general domain.

We focus here on one of the recent developments, discrete signal processing on graphs [13]. The dataset is represented by a *graph* $\mathcal{G} = (\mathcal{V}, P)$, where $\mathcal{V} = \{v_i\}_{i=1}^N$ is the set of *nodes* representing signals and $P \in \mathbb{C}^{N \times N}$ is a *graph shift* describing the relational dependencies among the nodes. The graph shift P is not necessarily an adjacency matrix nor does it necessarily have a probabilistic meaning. A *graph signal* s is then defined as the following map:

$$s : \mathcal{V} \rightarrow \mathbb{C}^{N \times D},$$

where D is the dimension of the graph signal on each node. A linear shift-invariant system, or, a *graph filter*, is defined as

$$H = h(P) = \sum_{k=0}^K h_k P^k, \quad (3)$$

with $h_k \in \mathbb{C}$, $k = 0, 1, \dots, K$.

Then, a graph filter $H \in \mathbb{C}^{N \times N}$ applied to a graph signal $s \in \mathbb{C}^{N \times D}$ produces an output, which is again a graph signal,

$$Hs = h(P)s.$$

Discrete signal processing on graphs then defines a series of standard signal processing concepts including the graph Fourier transform, frequency, spectrum, spectral decomposition, and impulse and frequency responses [13].

D. Semi-Supervised Learning

Traditional classifiers typically fall under supervised learning, with only labeled signals to train. In many real-world applications, however, a large number of labeled signals is not available, which can cause overfitting. Semi-supervised learning is a technique for training classifiers with both labeled and unlabeled signals, which assumes that unlabeled signals can provide distribution information to build a stronger classifier. Some well-known semi-supervised learning algorithms include generative mixture models with expectation maximization, co-training and graph-based approaches [12]. Generative mixture models with expectation maximization assume that classes produce well clustered signals, and that with large number of unlabeled signals, the mixture components can be identified [45]. Co-training assumes that the features are sufficiently discriminative that they can be split into two sets, with each set being able to build a good classifier [46]. Graph-based approaches assume that while the measured signals are defined in a high-dimensional space, they exist in a low-dimensional manifold; a graph is then constructed by measuring the similarity of each pair of signals, and those deemed similar are labeled as belonging to the same class [47].

We focus here on label propagation, one of the graph-based approaches. Label propagation classifies signals by understanding how labels propagate on a graph; two methods are in use, diffusion functions [48] and harmonic functions [49], [50]. Both methods work based on propagating the known labels on the transition matrices. Diffusion functions propagate those labels a finite number of times without any intervention. Harmonic functions, on the other hand, correct the known labels to the initial values after each propagation and propagate an infinite number of times. The advantage of harmonic functions is that the known labels keep pushing the decision boundaries to low-density gaps. The drawback is that if the known labels are not reliable, harmonic functions may keep diffusing wrong information.

III. SEMI-SUPERVISED MULTIREOLUTION CLASSIFICATION

Multiresolution classification analyzes data to uncover hidden information; in its original form, it uses supervised classification, and can thus train only on labeled signals. When the labeled dataset is small or contains improperly labeled signals, the classification boundary and the weights assigned to subbands can be unreliable. Semi-supervised learning, on the other hand, uses the entire dataset to help classification, but works on one resolution level only. We thus propose to merge these two concepts and gain the best of both worlds: a *semi-supervised multiresolution classifier*. We stress here that this is not a simple combination of known techniques, as there is no known way to weigh subband decisions for unlabeled signals; this is one of our contributions.

A. Semi-Supervised Classification

Figure 3 summarizes our proposed framework; the multiresolution decomposition and feature extraction blocks from Figure 2 work as before. The first change is that the supervised

classification block is replaced by a semi-supervised one so we can use both labeled and unlabeled signals to make a labeling decision in each subband. For the i th signal, the confidence vector in the s th subband is now

$$\hat{q}_s^{(i)} = \text{SSC}_s(f_s^{(i)}, Q), \quad (4)$$

where SSC_s is the semi-supervised classification function in subband s . As for supervised classification, SSC can be chosen from a variety of approaches; we propose a new one, adaptive graph filter, described in Section IV.

B. Semi-Supervised Weighting

We now explain how to build a semi-supervised weighting block, that is, how to weigh decisions from all the subbands to get a global decision in a semi-supervised manner.

Labeled signals contribute to weighting directly by fitting their confidence vectors from all the subbands to the ground truth; unlabeled signals cannot do the same as they do not have the ground truth. We thus face the task of finding a way to measure the confidence of labeling an unlabeled signal. The simplest way would be just to assign the label of the largest component in the confidence vector. We encounter a problem, however; for example, let $\hat{q}^{(1)} = [0.5 \ 0.49 \ 0.01]^T$ and $\hat{q}^{(2)} = [0.5 \ 0.25 \ 0.25]^T$ be confidence vectors for two signals. While we can label $\hat{q}^{(2)}$ as Class 1 easily, it is clearly hard to make a decision for $\hat{q}^{(1)}$. This maximum confidence measure would assign both signals to Class 1, however, because it does not take into account the rest of the confidence in the confidence vector. A way to rectify that would be to use Shannon entropy; if the entropy is small (less uncertainty, high confidence), it is easy to assign a label to the signal, and vice versa. We still face a problem in the above case because the Shannon entropy tells us that we can label $\hat{q}^{(1)}$ with higher confidence (less uncertainty) as its entropy, $H(\hat{q}^{(1)}) = 1.0707$, is lower than that for $\hat{q}^{(2)}$, $H(\hat{q}^{(2)}) = 1.5$. To resolve this issue, we modify the Shannon entropy measure as,

$$M(\hat{q}) = H(\hat{q}) (\chi_{d>T} + \lambda(d) \chi_{d\leq T}), \quad (5)$$

where χ_I is the indicator function of an interval I , $d = |\hat{q}_{(1)} - \hat{q}_{(2)}|$ with $\hat{q}_{(1)}, \hat{q}_{(2)}$ the first and second largest element in \hat{q} , respectively, T is the threshold, and $\lambda(d)$ is a penalty function that is large when the first and second largest elements are close. In other words, when the difference d between the two largest elements is large, the first term in (5) takes over and $M(\hat{q}) = H(\hat{q})$; when, on the other hand, the difference d between the two largest elements is small, the second term in (5) takes over and the uncertainty $M(\hat{q})$ is large.

We can now use this new uncertainty measure to say that the uncertainty of the s th subband in labeling the i th signal is $M(\hat{q}_s^{(i)})$. Assuming that each signal contributes equally to its subband, we define the uncertainty of each subband as the mean of the uncertainty of all the confidence vectors,

$$M_s = \frac{1}{U} \sum_{i=L+1}^N M(\hat{q}_s^{(i)}).$$

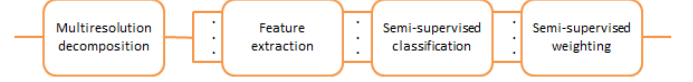


Fig. 3: Semi-supervised multiresolution classification. Supervised classification and weighting algorithm in Figure 2 are replaced with their semi-supervised counterparts so that unlabeled data can contribute to classification.

We now define the discriminative power of the s th subband to be the confidence

$$g_s = \frac{e^{-\beta M_s}}{\sum_{j=1}^S e^{-\beta M_j}}, \quad (6)$$

where β is the decay coefficient that controls the distribution of the discriminant power from all the subbands. When the uncertainty of a subband is large, the confidence is small and the subband gets assigned a low weight, and vice versa. Confidences from all the subbands are collected into a vector g ; note that $\|g\|_1 = 1$.

We now find the weight vector by optimizing a *semi-supervised weighting objective function*,

$$w = \arg \min_{\omega} \left\{ \alpha \sum_{i=1}^L \|\hat{q}^{(i)} - \hat{Q}^{(i)} \omega\| + (1 - \alpha) \|\omega - g\| \right\}, \quad (7)$$

with the constraint $\|\omega\|_1 = 1$, and where $\alpha = L/(L + U)$ is the *labeling ratio*. The first term in (7) represents the contribution from all labeled signals and is a scaled version of (1). The second term in (7) represents the contribution from all unlabeled signals; to obtain it, we fit weights to subbands' confidences. We use the labeling ratio to balance these two terms. Since this is a convex optimization problem, it is numerically efficient to obtain the global optimum by any standard convex optimization package. As in (2), after weighting, we compute the global decision as

$$\hat{y}^{(i)} = \arg \max_c \hat{q}_c^{(i)}, \quad (8)$$

where $\hat{q}^{(i)} = \text{SSW}(\hat{Q}^{(i)}, Q) = \hat{Q}^{(i)} w$ (see Algorithm 3).

IV. ADAPTIVE GRAPH FILTER

The idea of using a graph filter as a binary classifier was first proposed in [13]. For practical applications, however, its use is limited: First, it can only perform binary classification; then, it trains the filter coefficients by choosing a local optimum; and finally, it cannot classify unseen signals. Here, we propose an *adaptive graph filter* as a semi-supervised classifier in (4) that resolves these problems; we also connect this new adaptive graph filter to diffusion maps and diffusion wavelets.

In (4), for the signal $x^{(i)}$ and subband s , the inputs to the semi-supervised classifier are the feature vector $f_s^{(i)}$ and the ground-truth matrix Q , and the output is the confidence vector $\hat{q}_s^{(i)}$. For simplicity, in this section we omit the subband index s and write $f^{(i)}, \hat{q}^{(i)}$; it should be understood, however, that an adaptive graph filter is applied in each subband.

Algorithm 3 Semi-supervised multiresolution classification

Input	\mathcal{X}	input dataset
	\mathcal{Y}	ground-truth labels for \mathcal{L}
Output	$\hat{\mathcal{Y}}$	estimated labels for \mathcal{U}
Parameters	per subband s	$s = 1, \dots, S$
	D_s	multiresolution function
	$a_s^{(i)}$	multiresolution coefficients
	F_s	feature extraction function
	$f_s^{(i)}$	feature vector
	SSC_s	semi-supervised classification function
	$\hat{q}_s^{(i)}$	confidence vector
	for all subbands	
	SSW	semi-supervised weighting function
	$\hat{Q}^{(i)}$	confidence matrix
	w	weight vector, $\ w\ _1 = 1$
	$\hat{q}^{(i)}$	final confidence vector
Function	SSMRC(\mathcal{X})	
	$a_s^{(i)} = D_s(x^{(i)})$	multiresolution decomposition
	$f_s^{(i)} = F_s(a_s^{(i)})$	feature extraction
	$\hat{q}_s^{(i)} = SSC_s(f_s^{(i)}, Q)$	semi-supervised classification
	$\hat{q}^{(i)} = SSW(\hat{Q}^{(i)}, Q)$	semi-supervised weighting
	$\hat{y}^{(i)} = \arg \max_c \hat{q}_c^{(i)}$	
	return $\hat{\mathcal{Y}}$	

A. Graph Filtering as Semi-Supervised Classification

We start by outlining the basic idea, followed by detailed developments. Let the input graph signal be a prior confidence matrix formed from the ground-truth matrix and the graph shift be the Hermitian transpose of the transition matrix. An adaptive graph filter is then built by combining a series of graph shifts. The filter coefficients of the graph filter are trained by fitting the estimated results to the known labels and minimizing the labeling uncertainty. The output graph signal after filtering is the posterior confidence matrix whose i th row will be the desired confidence vector $\hat{q}^{(i)}$.

Let $\mathcal{G} = (\mathcal{F}, P)$ be a graph with $\mathcal{F} = \{f^{(i)}\}_{i=1}^N$ a set of feature vectors in the given subband for the entire dataset. Because we want to be able to represent directed graphs, we propose the graph shift $P \in \mathbb{R}^{N \times N}$ to be

$$P_{i,j} = \frac{\exp\left(\frac{-\rho(f^{(i)}, f^{(j)})}{\sigma}\right)}{\sum_{k=1}^N \exp\left(\frac{-\rho(f^{(i)}, f^{(k)})}{\sigma}\right)}, \quad (9)$$

where ρ is a local distance measurement, such as the ℓ^2 norm or the cosine distance, and σ is a scaling coefficient, which controls the bandwidth. The graph shift we defined here is the Hermitian transpose of the *transition matrix* of the graph. The graph shift thus has a probabilistic interpretation: $P_{i,j}$ gives the probability that the j th node jumps to the i th node in one step [51].

We now build a graph filter as in (3), except that, because of the dependencies on the data in (9), this is an *adaptive graph filter*,

$$H = h(P) = \sum_{k=1}^K h_k P^k. \quad (10)$$

Note that we omit the 0th term since, as we will see, it does not contribute to classification. The graph filter thus represents

the relational dependencies among signals represented via their feature vectors.

Let the graph signal be the confidence matrix of *all* the signals on the graph, called *prior confidence matrix*, that is, the following map:

$$s : \mathcal{F} \rightarrow \mathbb{R}^{N \times C},$$

defined as

$$(\hat{Q}_{\text{pr}})_{j,c} = \begin{cases} 1, & \text{when } y^{(j)} = c; \\ 0, & \text{otherwise,} \end{cases}$$

or,

$$\hat{Q}_{\text{pr}} = \begin{bmatrix} Q \\ \mathbf{0}_{U \times C} \end{bmatrix}.$$

In other words, the first L rows of \hat{Q}_{pr} are the confidence matrix Q representing the labeled dataset, while the other U rows are all zeros representing the unlabeled dataset. The prior confidence matrix thus starts with the prior knowledge on the labeled dataset (the ground-truth matrix) and without any knowledge on the unlabeled dataset.

By applying adaptive graph filtering now, the ground truth propagates from the labeled dataset to unlabeled dataset; the output graph signal, or, the *posterior confidence matrix*, is obtained as

$$\hat{Q}_{\text{ps}} = H \hat{Q}_{\text{pr}}, \quad (11)$$

where the i th row of the posterior confidence matrix \hat{Q}_{ps} is the desired confidence vector $\hat{q}^{(i)}$. Note that graph filtering propagates the labeling information not only from the labeled signals to the unlabeled signals, but among the labeled signals as well, giving the mislabeled signals a chance to be corrected, and consequently providing robustness in classification. In other words, we do not fully trust the initial labels and use the graph structure to help us arrive at the final labels.

One issue left to address is how to choose the filter coefficients h_1, h_2, \dots, h_K . We rewrite (11) as

$$\hat{Q}_{\text{ps}} \stackrel{(a)}{=} \sum_{k=1}^K h_k P^k \hat{Q}_{\text{pr}} = \sum_{k=1}^K h_k \hat{Q}_k, \quad (12)$$

where (a) follows from (10) and \hat{Q}_k is the $N \times C$ confidence matrix for the k th graph shift defined as $\hat{Q}_k = P^k \hat{Q}_{\text{pr}}$. For each i , the desired confidence vector $\hat{q}^{(i)}$ (i th row of \hat{Q}_{ps}) is thus a weighted linear combination of corresponding rows from each graph shift \hat{Q}_k . Since a confidence vector may not sum to 1 after graph filtering, we normalize each confidence vector as

$$\hat{q}^{(i)} \leftarrow \frac{\hat{q}^{(i)}}{\sum_j \hat{q}_j^{(i)}}.$$

This normalization does not influence the classification result because we assign the label of the largest component in the confidence vector; the normalization does help calculate the uncertainty measure in (5). It is now clear why we omitted the 0th term in (10); since $\hat{Q}_0 = \hat{Q}_{\text{pr}}$, it does not contribute to propagating information from the labeled to the unlabeled data. Fitting these estimates to the known labels and minimizing the labeling uncertainty is identical to the semi-supervised

weighting we performed in the last section. We can thus use the same minimization as in (7),

$$h^* = \arg \min_h \left\{ \alpha \sum_{i=1}^L \|q^{(i)} - \hat{Q}'^{(i)} h\| + (1 - \alpha) \|h - \gamma\| \right\}, \quad (13)$$

where $h = [h_1 \dots h_K]^T$ with the constraint $\|h\|_1 = 1$, α is the labeling ratio as in (7), $\hat{Q}'^{(i)} = [\hat{q}_1^{(i)} \dots \hat{q}_K^{(i)}]$ is a $C \times K$ confidence matrix of the i th signal that collects the graph shifts' individual confidence vectors, and $\gamma = [\gamma_1 \dots \gamma_K]^T$ collects the discriminative powers of each graph shift,

$$\gamma_k = \frac{e^{-(\beta/U) \sum_{i=L+1}^N M(\hat{q}_k^{(i)})}}{\sum_{j=1}^K e^{-(\beta/U) \sum_{i=L+1}^N M(\hat{q}_j^{(i)})}},$$

where M is the uncertainty measurement as in (5) and β is the decay coefficient as in (6). Note that, to solve (13), we use the same method as we did to solve (7).

In the first term of (13), we fit the estimated confidence vectors to the ground-truth vectors by changing the filter coefficients. This is another reason we omit the 0th term in (10).

B. Regression: Handling Unseen Data

As defined, adaptive graph filtering can only handle signals in the given dataset; should an unseen signal appear, the graph would need to be rebuilt and the filter coefficients retrained, at a significant computational cost. To handle unseen signals, we introduce regression.

We assume that each signal is randomly sampled from some continuous distribution and that the signals with the same label originate from the same distribution (recall that signals here are subband feature vectors). If we use the given signals and their posterior confidence vectors $\hat{q}^{(i)}$ to estimate the distributions, we can label those unseen signals originating from these distributions. The task is thus to design a regression function to map the unseen signals to their posterior confidence vectors.

Given the subband feature vectors $\mathcal{F} = \{f^{(i)}\}_{i=1}^N$ and their posterior confidence vectors \hat{Q}_{ps} , for a batch of unseen signals $\mathcal{F}_{us} = \{f^{(i)}\}_{i=N+1}^{N+M}$, the posterior confidence matrix for the unseen signals is

$$\hat{Q}_{us} = L \hat{Q}_{ps}, \quad (14)$$

where $L \in \mathbb{R}^{M \times N}$ is a generic form of the smoothing matrix determined by some regression technique, such as polynomial regression, spline regression or kernel regression [52]. For the regularized reproducing kernel regression, a valid kernel function $k(f, f')$ is first chosen to measure the inner product of f and f' in a higher-dimensional space, and then, the smoothing matrix is defined as

$$L = K'(K + \lambda I_N)^{-1},$$

where λ is a regularization parameter, I_N is an $N \times N$ identity matrix, $K'_{i,j} = k(f^{(N+i)}, f^{(j)})$, $i = 1, 2, \dots, M$, $j = 1, 2, \dots, N$, and $K_{i,j} = k(f^{(i)}, f^{(j)})$, $i, j = 1, 2, \dots, N$.

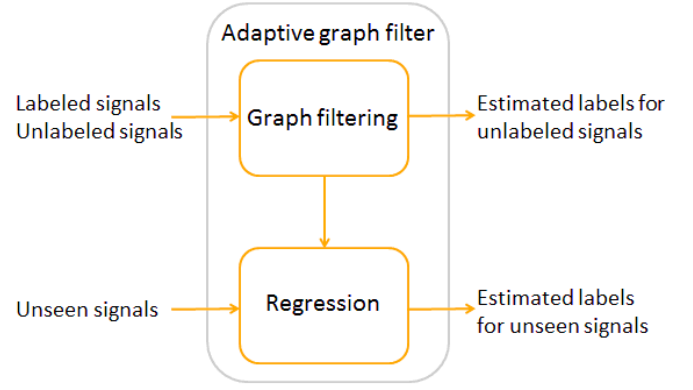


Fig. 4: Adaptive graph filter. Labeled and unlabeled signals are fed into the graph filtering block that outputs the estimated labels for unlabeled signals. Based on existing labeled and unlabeled signals, graph filtering generates the smoothing matrix for the regression block. Unseen signals are fed into the regression block that outputs the estimated labels for unseen signals.

The smoothing matrix calculates the relational dependencies between the unseen and given signals. Closer dependencies lead to higher values in the smoothing matrix. Because of this, the generic regression model predicts the posterior confidence vector of the new signal by weighing the posterior confidence vectors of the given signals locally, which means that the signals close by have similar posterior confidence vectors.

The forms of one step of graph filtering (11) and one step of regression (14) are similar. While both of these assume that similar signals have similar confidence vectors, their goals are different: the graph filtering step builds the relationship between unlabeled and labeled signals and produces labels from limited label information, while the regression step builds the relationship among the signals with the same label and connects unseen signals to their confidence vectors. Therefore, adding regression to graph filtering creates an adaptive graph filter that not only propagates the labeling information within a given dataset, but across unseen signals as well (see Figure 4).

Adaptive graph filter serves as a semi-supervised plugin method for classification. The traditional plugin methods estimate the unknown quantities in the Bayes' rule and plug them in for classification [53]; in the other words, they are equivalent to the regression block of adaptive graph filter. For instance, to label the i th signal, we first estimate the confidence vector $\hat{q}^{(i)}$ by using some regression techniques on the labeled dataset and then plug $\hat{q}^{(i)}$ in (2) to do classification. Without a large number of labeled signals, however, the traditional plugin methods fail to train a robust regression model, which further causes a poor classification performance [54]; this is the reason why we do not use regression in the first place. Adaptive graph filter solves this problem by producing labels for unlabeled signals in the filtering step, such that both labeled and unlabeled signals contribute to the smooth matrix in the regression step.

Algorithm 4 Adaptive graph filter

Input	\mathcal{F}	input dataset
	\hat{Q}_{pr}	prior confidence matrix
Output	\hat{Q}_{ps}	posterior confidence matrix
Parameters	$P_{i,j}$	graph shift
	ρ	local distance
	σ	scaling coefficient
	h_k	filter coefficients
Function	$\text{AGF}(\mathcal{F}, \hat{Q}_{\text{pr}})$	
	$P_{i,j} = (9)$	graph shift construction
	$h = (13)$	filter coefficient optimization
	for $k = 1 : K$ do	diffusion
	$\hat{Q}_k = P^k \hat{Q}_{\text{pr}}$	
	$\hat{Q}_{\text{ps}} = \sum_{k=1}^K h_k \hat{Q}_k$	weighting
	return \hat{Q}_{ps}	

C. Cost Analysis

The adaptive graph filter contains two steps: filtering and regression. In the filtering step, the computation involves the graph shift construction with the cost of $O((N-1)N/2 + N + N^2) = O((3N^2 + N)/2)$, the diffusion operation with the cost of $O(CKN^2)$, and the weighting operation with the cost of $O((K-1)N^2)$, for a total cost of $O(((C+1)K + 1/2)N^2 + N/2)$. In the regression step, the bulk of the cost comes from the construction of the smooth matrix and the inverse it involves with the cost of $O((N-1)N/2 + N + N^3 + MN + MN^2)$ and the matrix multiplication in (14) with the cost of $O(MCN)$, for a total cost of $O(N^3 + (M + 1/2)N^2 + (1/2 + M(C+1))N)$.

D. Relation to Diffusion Maps

We now analyze the adaptive graph filter by connecting it to diffusion maps and show that it reconstructs a robust diffusion map with more flexibility.

1) *Diffusion Maps*: The diffusion maps are coordinates that provide efficient geometric descriptions of signals and are built based on the singular value decomposition of the transition matrix.

Let A be the adjacency matrix of a graph, D the diagonal matrix whose i th element is $D_{i,i} = \sum_j A_{i,j}$ and $T = D^{-1}A$ the transition matrix. Recall that each element of T measures the likelihood of getting from one data point to another in one step; each element of T^k measures the likelihood of getting from one data point to another in k steps. The diffusion distance between two signals x, y in step k is then defined as

$$\begin{aligned} D_k(x, y)^2 &= \|t_k^{(x)} - t_k^{(y)}\|_{D^{-1}}^2 \\ &= (t_k^{(x)} - t_k^{(y)})^T D^{-1} (t_k^{(x)} - t_k^{(y)}), \end{aligned} \quad (15)$$

where $t_k^{(x)}$ and $t_k^{(y)}$ are x th and y th rows of T^k , respectively. Since the diffusion distance takes into account all paths of length k from x to y , it is robust to noise perturbation and outliers. Note that the diffusion distance can also be calculated using the ℓ^2 norm.

Since the transition matrix is asymmetric, we introduce a normalized transition matrix $\hat{T} = D^{\frac{1}{2}} T D^{-\frac{1}{2}}$, which, since

symmetric, can be factored as

$$\hat{T} = V \Lambda V^* = \sum_{i=1}^N v_i \lambda_i v_i^*,$$

where $V = [v_1 \ v_2 \ \dots \ v_N]$ is an orthogonal matrix and λ_i are the singular values; moreover, $1 = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N \geq 0$. We then decompose the transition matrix as

$$T^k = \Psi \Lambda^k \Phi^* = \sum_{i=1}^N \psi_i \lambda_i^k \phi_i^*, \quad (16)$$

where $\Psi = [\psi_1 \ \psi_2 \ \dots \ \psi_N] = D^{-\frac{1}{2}} V$, and $\Phi = [\phi_1 \ \phi_2 \ \dots \ \phi_N] = D^{\frac{1}{2}} V$. Note that \hat{T} and T share the same eigenvalues. The underlying diffusion map for step k and node x is now defined as

$$Y_k^{(x)} = [\lambda_1^k \psi_{1,x} \ \lambda_2^k \psi_{2,x} \ \dots \ \lambda_N^k \psi_{N,x}]^T. \quad (17)$$

If we define the diffusion space to be the space spanned by the columns of Φ , then, the diffusion map $Y_k^{(x)}$ gives the coordinates of x diffused k times in that space.

Using diffusion maps, the diffusion distance is simply

$$\begin{aligned} D_k(x, y)^2 &\stackrel{(a)}{=} (t_k^{(x)} - t_k^{(y)})^T D^{-1} (t_k^{(x)} - t_k^{(y)}), \\ &\stackrel{(b)}{=} \left[\sum_{i=1}^N \lambda_i^k (\psi_{i,x} - \psi_{i,y}) \phi_i^* \right] D^{-1} \left[\sum_{j=1}^N \lambda_j^k (\psi_{j,x} - \psi_{j,y}) \phi_j \right] \\ &= \left[\sum_{i=1}^N \lambda_i^k (\psi_{i,x} - \psi_{i,y}) v_i^* D^{\frac{1}{2}} \right] D^{-1} \left[\sum_{j=1}^N \lambda_j^k (\psi_{j,x} - \psi_{j,y}) D^{\frac{1}{2}} v_j \right] \\ &\stackrel{(c)}{=} \sum_{i=1}^N [\lambda_i^k (\psi_{i,x} - \psi_{i,y})]^2 \\ &= \|Y_k^{(x)} - Y_k^{(y)}\|_2^2, \end{aligned}$$

where (a) follows from (15); (b) from (16); and (c) from the orthogonality of V . Thus, by varying $k = 0, 1, \dots$, the diffusion maps allow us to find an alternate representation that might better separate the data, for example.

2) *Relation to the Adaptive Graph Filter*: We now show how, by using the adaptive graph filter, we construct diffusion maps that allow for more flexibility. Recall that our adaptive graph filter in (10) uses the graph shift P that is a Hermitian transpose of the transition matrix T . Thus, applying (16) to (10), we get

$$\begin{aligned} h(P) &= \sum_{k=1}^K h_k P^k = \sum_{k=1}^K h_k (T^*)^k = \sum_{k=1}^K h_k (T^k)^* \\ &= \sum_{k=1}^K h_k (\Psi \Lambda^k \Phi^*)^* = \sum_{k=1}^K h_k \Phi \Lambda^k \Psi^* \\ &= \Phi \left(\sum_{k=1}^K h_k \Lambda^k \right) \Psi^* = \Phi h(\Lambda) \Psi^*, \end{aligned} \quad (18)$$

with $h(\Lambda) = \text{diag}(\sum_{k=1}^K h_k \lambda_1^k, \dots, \sum_{k=1}^K h_k \lambda_N^k)$.

If we define the diffusion space to be the space spanned by the columns of Ψ , then, we define a diffusion map to be

$$Y_h^{(x)} = [h(\Lambda)_{1,1}\phi_{1,x} \quad \dots \quad h(\Lambda)_{N,N}\phi_{N,x}]^T. \quad (19)$$

Thus, the construction of an adaptive graph filter allows for continuous change of the coordinates in the diffusion space, providing flexibility in finding the best representation in a data-adapted fashion; this adaptivity is reflected by subscript h in $Y_h^{(x)}$. This is in contrast to only discrete changes allowed by (17).

We illustrate the above discussion with an example. Let $\lambda_1 = 1$, $\lambda_2 = 0.5$, $Y_0^{(x)} = [1 \ 1]^T$, $Y_0^{(y)} = [2 \ 1]^T$, and $Y_0^{(z)} = [1 \ 2.4]^T$. The distance between x and y in this step is $D_0(x, y) = 1$, smaller than the distance between x and z , $D_0(x, z) = 1.4$. Diffusing once, we get $Y_1^{(x)} = [1 \ 0.5]^T$, $Y_1^{(y)} = [2 \ 0.5]^T$, and $Y_1^{(z)} = [1 \ 1.2]^T$. The distance between x and y in this step is $D_1(x, y) = 1$, larger than the distance between x and z , $D_1(x, z) = 0.7$. In other words, by changing the power k , distances change in the diffusion space. Since k can only be an integer, it is not possible, for example, to make the distances between x and y , and x and z be the same. By using the adaptive graph filter, however, we can find the optimal filter coefficients to match such a requirement using (13).

E. Relation to Diffusion Wavelets

We now analyze the adaptive graph filter by connecting it to diffusion wavelets and show that it performs multiresolution classification on graphs.

1) *Diffusion Wavelets*: Diffusion wavelets are a multiscale framework to analyze signals with complex structures [16]. They can be seen as an extension of the classical wavelet theory, where, the diffusion wavelet basis is learned from the geometry of the signal structure in a data-adapted fashion. The diffusion wavelet basis is constructed by dilation using the dyadic powers of the transition matrix, the idea being that they propagate local relationships to global relationships throughout the graph.

Given a graph, at the j th resolution level, we have T^{2^j} , as the transition matrix, $j = 1, 2, \dots$. Since the second singular value of the transition matrix is less than 1 to keep the graph connected, only the first singular value is 1. Consequently, if the transition matrix is raised to a high power, all the singular values disappear except for the first one,

$$\lambda_i^{2^j} \rightarrow 0, \quad i = 2, 3, \dots, N \quad \text{as } j \rightarrow \infty.$$

When $j = 1$, the transition matrix T measures local pairwise similarities; increasing the power j gradually decreases the rank of the transition matrix T^{2^j} and causes local information of the graph to be missed since the resolution on the graph changes from finest to the coarsest. Thus, by changing j , we can both perform a multiresolution analysis as well as do it in a computationally efficient manner.

2) *Relation to the Adaptive Graph Filter*: The adaptive graph filter in (10) is formed as a linear combination of graph shifts P raised to power k . When k is large, P^k become a low-rank matrix describing the global information of the graph, just as T^{2^j} does for diffusion wavelets. Each k corresponds to a different resolution on the graph, and thus, adaptive graph filter actually performs multiresolution classification on the graph. It weighs the classification results from each resolution to produce the global result. The filter coefficients represent the discriminative power of each resolution. This also explains why the objective functions to optimize the filter coefficients in (7) and the weights of each subband in (13) are the same.

V. EXPERIMENTAL RESULTS

In this section, we validate the proposed framework on indirect bridge structural health monitoring, and show that it performs remarkably better than previous approaches.

A. Dataset

To study the bridge behavior under various conditions comprehensively, a lab-scale bridge-vehicle dynamic system was built; see Figure 5. Accelerometers were installed on a vehicle that moves across the bridge; acceleration signals were then collected from those accelerometers. We collected 30 acceleration signals for each of 13 different bridge conditions, 8 different speeds, from 1 m/s to 2.75 m/s with increments of 0.25 m/s, and 2 vehicles with different weights, 4.8 kg and 5.2 kg respectively, for a total of 6240 acceleration signals. The 13 bridge conditions included one pristine condition and 4 different damage severities for each of 3 different damage proxy scenarios. These damage proxy scenarios simulate possible damage on the bridge. For example, varying rotational restraints simulates rubber bearings becoming stiffer in time or corrosion of rocker supports; there were four of these, at one corner, two, three, and four. Another four scenarios were modeled by adding one or two dampers at quarter span of the bridge, two at quarter span and two at midspan, and finally, two each at quarter, mid, and three-quarter span. The final four scenarios were modeled by adding mass at midspan, of 50 g, 100 g, 200 g, and 300 g. For more details, see [21].

B. Experimental Setup

Given a specific vehicle driven at a specific speed, we want to classify 13 bridge conditions, in particular, with a low labeling ratio. We consider 16 vehicle-speed combinations for each of which there are 30 acceleration signals for each of the 13 scenarios; the final accuracy is the average over the 13 scenarios; the baseline accuracy is thus $100/13 = 7.7\%$.

We choose a Coiflet filter bank [55] with 4 levels (for a total of 15 subbands) in the multiresolution block [24], principal component analysis [23] in the feature extraction block, radius kernel support vector machine [56] in the supervised classification block, and adaptive graph filtering in the semi-supervised classification block. In each subband, we use the coefficients that preserve roughly 95% of energy after principal component



Fig. 5: The lab-scale bridge (from [21] with permission).

analysis as the feature vector for each signal. We construct the graph by choosing ρ as the cosine distance, defined as [57]

$$\rho(v_1, v_2) = 1 - \frac{\langle v_1, v_2 \rangle}{\|v_1\| \|v_2\|},$$

and $\sigma = (1/N^2) \sum_{i,j} \rho(f^{(i)}, f^{(j)})$ in (9). The length of the adaptive graph filter is $K = 30$. In the semi-supervised weighting function, we choose the penalty threshold $T = 0.02$ and the penalty function $\lambda(d) = 1 + 5(d/T - 1)^2$ in (5). The decay coefficient $\beta = 1$ is chosen to minimize (7) and (13). To solve (7) and (13), we used CVX, a package for specifying and solving convex programs [58], [59]. We performed a 30-fold cross-validation. Table II summarizes all the parameters at a glance. The details about the setting of parameters see a reproducible research page [17]. Note that in semi-supervised classification, we train the model based on both labeled and a large number of unlabeled signals without considering unseen signals, we thus do not have to worry about data snooping and overfitting. We compare our method, semi-supervised multiresolution classification with adaptive graph filtering, against:

- *Generic classification* with kernel support vector machine, diffusion functions, harmonic functions, and adaptive graph filtering.
- *Supervised multiresolution classification* with kernel support vector machine.
- *Semi-supervised multiresolution classification* with diffusion functions, harmonic functions, and adaptive graph filtering.

We use the following shorthands in figures and tables: S for supervised (if with another acronym) or speed (if without another acronym), SS for semi-supervised, no MRC for generic classification, MRC for multiresolution classification, SVM for kernel support vector machine, DF for diffusion functions, HF for harmonic functions, AGF for adaptive graph filtering, and V for vehicle.

C. Classification Results

We validate our method from three standpoints: (1) the performance of the semi-supervised multiresolution classification framework; (2) the ability of adaptive graph filtering

Experimental setup

dataset		
V	vehicles	2
S	speeds	8
	damage scenarios	13
	signals/scenario	30
multiresolution decomposition		
D	Coiflet filter bank	4 levels
S	number of subbands	15
feature extractions		
F	principal component analysis	
classification		
SC	kernel support vector machine	
SSC	adaptive graph filter	
ρ	local distance	cosine distance
σ	scaling coefficient	$(1/N^2) \sum_{i,j} \rho(f^{(i)}, f^{(j)})$
β	decay coefficient	1
K	length of graph filter	30
weighting		
T	threshold	0.02
$\lambda(d)$	penalty function	$1 + 5((d/T) - 1)^2$

TABLE II: Parameters used in the experiments.

to handle mislabeled signals; and (3) the ability of adaptive graph filtering to handle unseen signals.

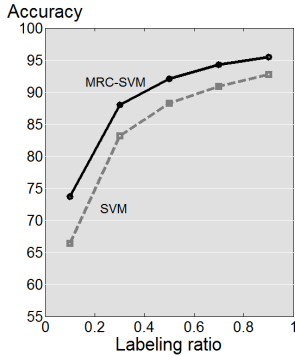
1) *Semi-Supervised Multiresolution Classification*: We validate the proposed framework, semi-supervised multiresolution classification, by comparing it to generic classification and supervised multiresolution classification with a low labeling ratio of 10% (see Table III).

We detect three trends, the first two of which validate each component of the framework (multiresolution classification and semi-supervised learning), while the third validates the entire framework: (1) Multiresolution framework improves classification accuracy (darker columns 7-10 versus lighter columns 3-6): supervised multiresolution classifier (MRC SVM, dark red column 6) performs better than the corresponding supervised generic classifier (SVM, light red column 2) and each semi-supervised multiresolution classifier (MRC DF, HF, and AGF, dark blue columns 8-10) performs better than the corresponding semi-supervised generic classifier (DF, HF, and AGF, light blue columns 4-6). (2) Semi-supervised learning improves classification accuracy (blue columns versus red columns): each semi-supervised generic classifier (DF, HF, and AGF, light blue columns 4-6) performs better than the supervised generic classifier (SVM, light red column 3) and each semi-supervised multiresolution classifier (MRC DF, HF, and AGF, dark blue columns 8-10) performs better than the supervised multiresolution classifier (MRC SVM, dark red column 7). (3) Multiresolution framework with semi-supervised learning (MRC DF, HF, and AGF, dark blue columns 8-10) improves classification accuracy over the supervised generic classifier (SVM, light red column 3) by $\sim 30\%$.

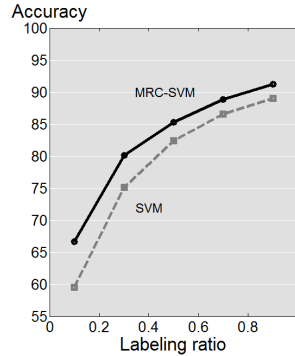
We further validate these trends under different labeling ratios. Figures 6–9 show the dependence of classification accuracy on the labeling ratio for 2 vehicles averaged across 8 speeds. Figures 6 and 7 validate Trend 1 for supervised (SVM) and semi-supervised (AGF) classifiers. In each case and for both vehicles, multiresolution framework improves

V	S	No MRC				MRC			
		S SVM	DF	SS HF	AGF	S SVM	DF	SS HF	AGF
1	1	57.97	83.70	88.92	87.98	81.19	99.63	99.80	99.83
	2	70.42	85.78	90.38	89.55	84.71	99.83	99.98	99.99
	3	74.29	86.25	90.67	89.91	84.59	99.11	99.40	99.38
	4	74.82	88.08	94.79	93.87	80.29	99.89	99.93	99.98
	5	70.68	74.87	78.92	77.64	72.35	93.76	96.13	94.52
	6	67.13	82.43	86.00	85.24	69.72	91.54	93.47	93.43
	7	59.48	65.31	66.05	66.23	59.75	77.44	78.79	78.74
	8	56.53	66.52	67.58	67.11	56.63	75.37	77.93	77.23
2	1	49.75	78.17	82.28	80.99	71.69	85.36	84.63	85.12
	2	53.64	67.30	71.38	70.32	60.69	80.19	80.51	80.59
	3	67.96	82.50	87.84	86.09	74.94	95.12	94.62	94.81
	4	61.52	79.57	82.85	82.86	65.27	86.65	86.44	87.34
	5	62.75	77.92	82.15	81.22	66.42	88.86	88.34	88.72
	6	66.89	80.17	81.65	81.73	69.75	84.08	83.61	84.18
	7	65.09	82.87	85.19	85.48	70.59	89.15	89.51	89.66
	8	48.57	80.32	83.01	82.57	53.67	92.03	93.62	93.31

TABLE III: Accuracy comparison of Vehicles (V) 1 and 2, with Speeds (S) 1, 2, ..., 8, and labeling ratio of 10%.

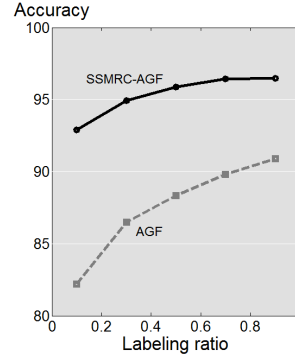


(a) Vehicle 1.

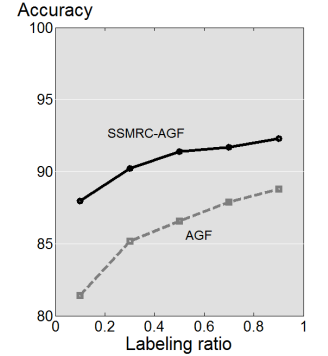


(b) Vehicle 2.

Fig. 6: Trend 1: Multiresolution framework improves classification accuracy. Comparison of supervised classifiers.



(a) Vehicle 1.



(b) Vehicle 2.

Fig. 7: Trend 1: Multiresolution framework improves classification accuracy. Comparison of semi-supervised classifiers.

classification accuracy across all labeling ratios.

Figure 8 validates Trend 2 for supervised (SVM) and semi-supervised (AGF) multiresolution classifiers. For both vehicles, semi-supervised learning improves classification accuracy across all labeling ratios. Moreover, as the labeling ratio decreases, accuracy drops sharply for multiresolution-based SVM; performance of the multiresolution-based adaptive graph filter stays relatively flat, however, even at very low labeling ratios.

Figure 9 validates Trend 3 for semi-supervised (AGF) multiresolution classifier and generic supervised classifier (SVM). For both vehicles, semi-supervised multiresolution classifier with adaptive graph filtering dramatically improves classification accuracy across all labeling ratios.

2) *Ability of Adaptive Graph Filtering to Handle Mis-labeled Signals:* In real-world problems, some of the labeled signals could be unreliable for different reasons, for example, negligence or uncertainty. As mentioned in Section IV-A, one of the advantages of using adaptive graph filtering is to provide robustness to mislabeling. To validate that, we randomly mislabel a fraction of labeled signals, feed them

into the classifiers together with correctly labeled signals, and compare the fault tolerances of the three semi-supervised classifiers. Tables IV and V show results where 20% of signals are labeled, with 15.38% and 33.33% of these labeled signals mislabeled, respectively. The trends from before still hold: Multiresolution framework improves classification accuracy (darker columns 6-8 versus lighter columns 3-5). Among the semi-supervised multiresolution classifiers, adaptive graph filtering (AGF, dark columns 8) performs the best in each case. Moreover, as the ratio of mislabeled signals increases from 15.38% to 33.33%, the performance of adaptive graph filtering is relatively unaffected, while the performance of the other two semi-supervised classifiers, diffusion functions and harmonic functions, decreases dramatically. We thus conclude that the semi-supervised multiresolution classification using adaptive graph filtering is robust to mislabeled signals.

3) *Ability of Adaptive Graph Filtering to Handle Unseen Signals:* Finally, to validate the claim from Section IV-B that using regression allows us to handle unseen signals, we keep a portion of signals as unseen signals. For each vehicle and speed, we have 13 damage scenarios with 30 signals for each

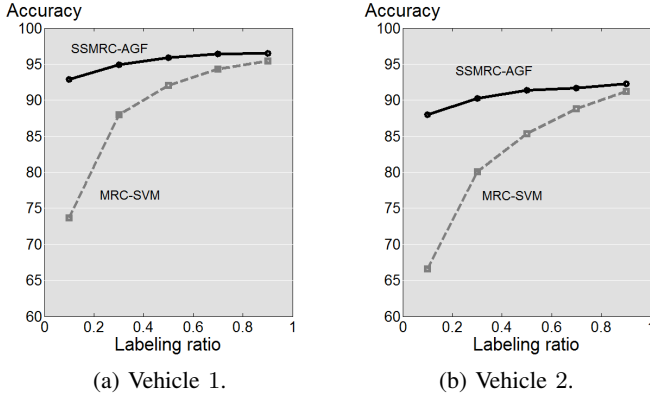


Fig. 8: Trend 2: Semi-supervised learning improves classification accuracy. Comparison of multiresolution classifiers.

V	S	No MRC			MRC		
		DF	HF	AGF	DF	HF	AGF
1	1	83.88	85.15	88.53	95.92	96.60	99.39
	2	87.15	87.76	90.84	97.72	98.09	99.79
	3	87.28	88.77	91.03	97.47	97.71	99.40
	4	86.97	88.99	93.45	96.93	96.99	99.52
	5	74.52	73.99	77.72	94.41	95.28	95.74
	6	83.05	84.08	87.54	93.21	93.94	95.75
	7	63.99	63.45	66.08	78.37	79.21	80.06
	8	66.92	67.15	68.14	78.38	79.59	80.44
2	1	77.38	79.28	81.88	83.53	83.39	85.68
	2	66.78	67.65	69.98	77.97	78.32	79.86
	3	82.09	83.80	86.72	94.13	93.61	95.07
	4	79.99	81.00	84.56	86.67	86.61	88.31
	5	79.42	80.40	83.20	87.39	87.61	89.01
	6	80.40	80.97	83.34	84.25	83.43	84.69
	7	81.96	82.34	86.03	88.43	89.01	89.85
	8	79.59	81.47	83.31	92.31	92.94	94.40

TABLE IV: Robustness to mislabeled signals: accuracy comparison of Vehicles (V) 1 and 2, with Speeds (S) 1, 2, ..., 8, with labeling ratio of 20% and misslabeling ratio of 15.38%.

for a total of 390 available signals. We assign $M = 65$ of these to be unseen, leaving us with $N = 325$ total signals, out of which we label $L = 13$. We thus have 4% (13/325) labeled and 20% (65/325) unseen signals. Table VI shows results for both the generic classifier (lighter columns 3-4) and semi-supervised multiresolution classifiers with adaptive graph filtering (darker columns 5-6). In each case, we compare accuracies of $U = 312$ unlabeled signals as well as of $M = 65$ unseen signals. Although the unseen signals never appear in the filtering stage, classification accuracy is close to that of unlabeled signals. If the signal set is sufficiently large, the adaptive graph filter learns the distribution of signals from both labeled and unlabeled signals, which it then uses for the unseen signals.

VI. CONCLUSIONS AND FUTURE WORK

We presented a novel classification framework that combines multiresolution classification with semi-supervised learning; adaptive graph filtering for semi-supervised classification that allows for classifying unlabeled as well as unseen

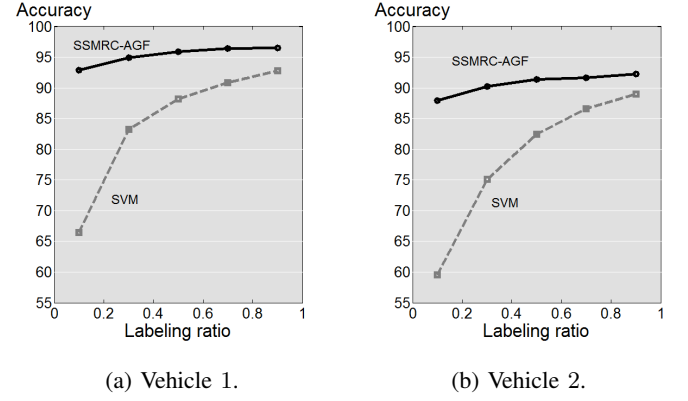


Fig. 9: Trend 3: Multiresolution framework with semi-supervised learning improves classification accuracy.

V	S	No MRC			MRC		
		DF	HF	AGF	DF	HF	AGF
1	1	72.55	73.41	81.99	87.16	87.43	97.20
	2	75.84	76.18	85.16	89.37	89.96	98.14
	3	79.36	79.40	87.17	90.33	90.73	97.44
	4	77.47	77.03	89.15	88.50	88.83	97.90
	5	66.32	62.75	72.87	86.78	85.67	90.08
	6	74.44	72.98	81.92	86.45	85.35	90.98
	7	57.70	55.89	62.02	72.43	72.69	76.06
	8	60.14	59.12	63.88	71.29	71.64	75.26
2	1	66.76	68.00	76.29	74.46	74.50	81.90
	2	57.62	57.50	63.41	71.08	70.85	75.94
	3	73.54	73.69	82.47	87.25	87.18	93.19
	4	71.27	70.05	79.78	78.49	78.70	84.78
	5	70.76	69.87	78.66	79.95	79.67	85.80
	6	72.39	71.84	80.41	77.18	77.13	82.55
	7	74.39	73.82	82.53	81.45	82.01	87.04
	8	72.88	71.12	80.04	85.82	86.29	92.19

TABLE V: Robustness to mislabeled signals: accuracy comparison of Vehicles (V) 1 and 2, with Speeds (S) 1, 2, ..., 8, with labeling ratio of 20% and misslabeling ratio of 33.33%.

signals and for correcting mislabeled signals; and solution to indirect bridge structural health monitoring.

The proposed framework builds upon supervised multiresolution classification, which extracts hidden features in localized time-frequency subbands, and semi-supervised learning, which uses both labeled and unlabeled signals. We link the two via a novel weighting algorithm that combines information from all the subbands of all the signals to make a global decision in a semi-supervised fashion. We propose a novel semi-supervised classifier, adaptive graph filter; also, the first real application of signal processing on graphs. We further connect it to diffusion maps and diffusion wavelets and show that it performs multiresolution classification on graphs.

We validate the proposed framework on the task of indirect bridge structural health monitoring and show that: (1) multiresolution framework on its own, (2) semi-supervised learning on its own, and (3) the two together, all improve classification accuracy. Furthermore, we show that adaptive graph filtering has the ability to handle unlabeled, mislabeled as well as unseen signals. Applications to different physical

V	S	No MRC-AGF		MRC-AGF	
		Unlabeled	Unseen	Unlabeled	Unseen
1	1	76.79	74.21	99.19	96.92
	2	80.34	74.72	99.26	96.77
	3	83.84	79.18	97.38	94.41
	4	85.67	84.51	99.70	98.77
	5	67.44	73.85	82.81	80.92
	6	75.20	74.97	85.35	80.97
	7	62.02	62.72	72.97	71.13
	8	57.97	58.05	67.39	66.97
2	1	75.63	75.38	81.99	82.41
	2	56.77	57.64	76.05	76.41
	3	77.13	77.08	93.25	92.00
	4	71.57	72.62	80.49	81.13
	5	69.31	68.77	83.73	82.62
	6	73.27	74.05	78.13	78.46
	7	77.98	78.10	83.60	83.59
	8	74.69	77.64	86.82	87.64

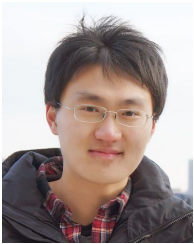
TABLE VI: Robustness to unseen signals: accuracy comparison of Vehicles (V) 1 and 2, with Speeds (S) 1, 2, . . . , 8, labeling ratio of 4% and ratio of unseen signals of 20%.

situations are illustrated in [17].

Some near-future tasks are to use more features in each time-frequency subband, prune wavelet packet tree to achieve faster implementation and test the framework on real-world bridge-vehicle dynamic systems.

VII. ACKNOWLEDGMENT

The authors gratefully acknowledge support from the NSF through awards 1130616 and 1017278, as well as CMU Carnegie Institute of Technology Infrastructure Award. They also would like to thank Dr. Aliaksei Sandryhaila, Mr. Jia Guo, Mr. Joel B. Harley, Mr. Qing Shi for helpful discussions, and an editor and four reviewers for comments that led to improvements in the manuscript. We follow the principles of reproducible research. To that end, we created a reproducible research page available to readers [17]. Initial parts of this work were presented at ICASSP 2013 [11] and GlobalSIP 2013 [14].



Siheng Chen (S'13) received his B. Eng degree in optoelectronics engineering in 2011 from Beijing Institute of Technology and the M.S degree in electrical and computer engineering in 2012 from Carnegie Mellon University. He is currently a Ph.D student in the Department of Electrical and Computer Engineering and the Center for Bioimage Informatics at Carnegie Mellon University. His research interests include graph signal processing, indirect bridge health monitoring and biomedical image analysis.



Fernando Cerda received the Structural Engineering Diploma, Eng. degree and Master of Science from the Structural and Geotechnical Engineering Department, Pontificia Universidad Católica de Chile, Santiago Chile, in 2006. He joined the Civil Engineering Department at Universidad de Concepción, Concepción, Chile, in 2006. He received a Master of Science in Advanced infrastructure Systems and Ph.D. degrees from the Civil and Environmental Engineering Department at Carnegie Mellon University. His dissertation topic was on Indirect Structural Health Monitoring of Bridges. His research interests include structural health monitoring of infrastructure, structural diagnostics, risk mitigation, infrastructure management, machine learning, structural dynamics, numerical modeling and seismic response. His research approach involves problem driven collaboration with interdisciplinary teams



Piervincenzo Rizzo is an associate professor in the Department of Civil and Environmental Engineering at the University of Pittsburgh. In 1998, he received a Laurea (5 years degree) in Aeronautical Engineering at the University of Palermo, Italy, and a M.S. and a Ph.D. in Structural Engineering at the University of California, San Diego, in 2002 and 2004, respectively. His research interests are in the area of active and passive wave propagation and signal processing for nondestructive testing and structural health monitoring applications. To date, he has authored 60 refereed papers, over 100 conference papers, 8 book chapters, and he holds 2 patents. In 2012 he was awarded the Achenbach Medal, which has been created to recognize a young individual who has made an outstanding contribution to the advancement of the field of structural health monitoring. Other honors include the 2007 Faculty Grant Award and the 2002 Fellowship Research Award both from the American Society for Nondestructive Testing.



Jacobo Bielak received his Civil Engineer's degree from the National University of Mexico (UNAM), MS from Rice University, and PhD from Caltech. He joined Carnegie Mellon University in 1978, where he is now the Paul Christiano University Professor. His research is in the areas of earthquake engineering and engineering seismology, and structural health monitoring. He was a member of the original Applied Technology Council (ATC) committee that drafted the first tentative seismic provisions for soil-structure interaction in the US based mainly on his work. These provisions are now, in modified form, part of the NEHRP seismic provisions. Recognition for his work includes the Gordon Bell Prize for Special Accomplishment Based on Innovation. He is a Distinguished Member of the American Society of Civil Engineers, a Fellow of the US Association for Computational Mechanics, and a member of the National Academy of Engineering.



James H. Garrett, Jr. is currently Dean of the College of Engineering (since 2013) at Carnegie Mellon University and the Thomas Lord Professor of Civil and Environmental Engineering at Carnegie Mellon. He was Head of the Department of Civil and Environmental Engineering from June 2006 to December 2012. Until 2013, he was the founding co-director of the Pennsylvania Smarter Infrastructure Incubator, which was a research center aimed at creating and evaluating sensing, data analytics and intelligent decision support for improving the construction, management and operation of infrastructure systems. Garrett also served as the Co-Chief Editor of the ASCE Journal of Computing in Civil Engineering from 2008-2013. Garrett's research and teaching interests are oriented toward applications of sensors and sensor systems to civil infrastructure condition assessment; application of data mining and machine learning techniques for infrastructure management problems in civil and environmental engineering; mobile hardware/software systems for field applications; representations and processing strategies to support the usage of engineering codes, standards, and specifications; knowledge-based decision support systems. Garrett has published his research in over 60 refereed journal articles, over 80 refereed conference papers, over 90 other conference papers and 10 sections or chapters in books or monographs.



Jelena Kovačević (S'88–M'91–SM'96–F'02) received the Dipl. Electr. Eng. degree from the EE Department, University of Belgrade, Yugoslavia, in 1986, and the M.S. and Ph.D. degrees from Columbia University, New York, in 1988 and 1991, respectively. From 1991–2002, she was with Bell Labs, Murray Hill, NJ. She was a co-founder and Technical VP of xWaveforms, based in New York City and an Adjunct Professor at Columbia University. In 2003, she joined Carnegie Mellon University, where she is a Professor and Head of Electrical

and Computer Engineering, Professor of Biomedical Engineering, and the Director of the Center for Bioimage Informatics. Her research interests include bioimaging as well as multiresolution techniques such as wavelets and frames. Dr. Kovačević coauthored the books *Wavelets and Subband Coding* (Prentice Hall, 1995) and *Foundations of Signal Processing* (Cambridge University Press, 2014), a top-10 cited paper in the Journal of Applied and Computational Harmonic Analysis, and the paper for which A. Mojsilović received the Young Author Best Paper Award. Her paper on multidimensional filter banks and wavelets was selected as one of the Fundamental Papers in Wavelet Theory. She received the Belgrade October Prize in 1986 and the E.I. Jury Award at Columbia University in 1991. She is a past Editor-in-Chief of the IEEE Transactions on Image Processing, served as a guest co-editor on a number of special issues and is/was on the editorial boards of several journals. She was a regular member of the NIH Microscopic Imaging Study Section and served as a Member-at-Large of the IEEE Signal Processing Society Board of Governors. She is a past Chair of the IEEE Signal Processing Society Bio Imaging and Signal Processing Technical Committee. She has been involved in organizing numerous conferences. She was a plenary/keynote speaker at a number of international conferences and meetings.

REFERENCES

- [1] N. Saito, "Classification of geophysical acoustic waveforms using time-frequency atoms," *Proc. Am. Stat. Assoc. Stat. Comput.*, pp. 322–327, 1996.
- [2] T. McConaghy, H. Leung, E. Bossé, and V. Varadan, "Classification of audio radar signals using radial basis function neural networks," *IEEE Trans. Instrum. Meas.*, vol. 52, no. 4, pp. 1771–1779, Dec. 2003.
- [3] H. Sohn, C. R. Farrar, N. F. Hunter, and K. Worden, "Structural health monitoring using statistical pattern recognition techniques," *J. Dyn. Sys., Meas., Control*, vol. 123, no. 4, pp. 1519–1524, Dec. 2001.
- [4] A. Chebira, Y. Barbotin, C. Jackson, T. E. Merryman, G. Srinivasa, R. F. Murphy, and J. Kovačević, "A multiresolution approach to automated classification of protein subcellular location images," *BMC Bioinformatics*, vol. 8, no. 210, 2007.
- [5] R. A. Kellogg, A. Chebira, A. Goyal, P. A. Cuadra, S. F. Zappe, J. S. Minden, and J. Kovačević, "Towards an image analysis toolbox for high-throughput Drosophila embryo RNAi screens," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Arlington, VA, Apr. 2007, pp. 288–291.
- [6] A. Chebira, J. A. Ozolek, C. A. Castro, W. G. Jenkinson, M. Gore, R. Bhagavatula, I. Khaimovich, S. E. Ormon, C. S. Navara, M. Sukhwani, K. E. Orwig, A. Ben-Yehudah, G. Schatten, G. K. Rohde, and J. Kovačević, "Multiresolution identification of germ layer components in teratomas derived from human and nonhuman primate embryonic stem cells," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Paris, May 2008, pp. 979–982.
- [7] T. Chen, W. Yin, X. Sean, Z. Dorin, C. Thomas, and S. Huang, "Total variation models for variable lighting face recognition and uneven background correction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 1519–1524, May 2006.
- [8] C. Lin and Y.B. Yang, "Use of a passing vehicle to scan the fundamental bridge frequencies: An experimental verification," *Engineering Structures*, vol. 27, pp. 1865–1878, 2005.
- [9] F. Cerda, J. Garrett, J. Bielak, R. Bhagavatula, and J. Kovačević, "Exploring indirect vehicle-bridge interaction for bridge SHM," in *Proc. Int. Conf. Bridge Maint., Safety Manag.*, Philadelphia, PA, Jul. 2010, pp. 696–702.
- [10] F. Cerda, J. Garrett, J. Bielak, P. Rizzo, J. A. Barrera, Z. Zhang, S. Chen, M. T. McCann, and J. Kovačević, "Indirect structural health monitoring in bridges: scale experiments," in *Proc. Int. Conf. Bridge Maint., Safety Manag.*, Lago di Como, Jul. 2012, pp. 346–353.
- [11] S. Chen, F. Cerda, J. Guo, J. B. Harley, Q. Shi, P. Rizzo, J. Bielak, J. H. Garrett, and J. Kovačević, "Multiresolution classification with semi-supervised learning for indirect bridge structure health monitoring," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, May 2013, pp. 3412–3416.
- [12] X. Zhu, "Semi-supervised learning literature survey," Univ. Wisconsin-Madison, Tech. Rep. 1530, 2005.
- [13] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, 2013.
- [14] S. Chen, A. Sandryhaila, J. M. F. Moura, and J. Kovačević, "Adaptive graph filtering: Multiresolution classification on graphs," in *Proc. IEEE Glob. Conf. Signal Information Process.*, Austin, TX, Dec. 2013, pp. 427–430.
- [15] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, pp. 5–30, 2006.
- [16] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmon. Anal.*, pp. 53–94, Jul. 2006.
- [17] S. Chen, F. Cerda, P. Rizzo, J. Bielak, J. H. Garrett, and J. Kovačević, (2013) Semi-supervised multiresolution classification using adaptive graph filtering with application to indirect bridge structural health monitoring. http://jelena.ece.cmu.edu/repository/rtr-13_ChenCRBGK/13_ChenCRBGK.html.
- [18] A. Chebira, L. P. Coelho, A. Sandryhaila, S. Lin, W. G. Jenkinson, J. MacSleyne, C. Hoffman, P. Cuadra, C. Jackson, M. Püschel, and J. Kovačević, "An adaptive multiresolution approach to fingerprint recognition," in *Proc. IEEE Int. Conf. Image Process.*, San Antonio, TX, Sep. 2007, pp. 457–460.
- [19] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Process. Mag.*, vol. 30, pp. 83–98, 2013.
- [20] Y. B. Yang, C. W. Lina, and J. D. Yau, "Extracting bridge frequencies from the dynamic response of a passing vehicle," *J. Sound Vibrat.*, pp. 471–493, May 2004.

- [21] F. Cerda, S. Chen, J. Bielak, J. H. Garrett, P. Rizzo, and J. Kovačević, "Indirect structural health monitoring of a simplified laboratory-scale bridge model," *Int. J. Smart Struct. Syst., sp. iss. Challenge on bridge health monitoring utilizing vehicle-induced vibrations*, Jul. 2014.
- [22] G. Lederman, Z. Wang, J. Bielak, H. Noh, J. H. Garrett, S. Chen, J. Kovačević, F. Cerda, and P. Rizzo, "Damage quantification and localization algorithms for indirect SHM of bridges," in *Proc. Int. Conf. Bridge Maint., Safety Manag.*, Shanghai, Jul. 2014, to appear.
- [23] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Englewood Cliffs, NJ: John Wiley & Sons, 2001.
- [24] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, ser. Signal Processing. Englewood Cliffs, NJ: Prentice Hall, 1995, <http://waveletsandsubbandcoding.org/>.
- [25] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of Signal Processing*. Cambridge University Press, 2014, <http://www.fourierandwavelets.org/>.
- [26] J. Kovačević, V. K. Goyal, and M. Vetterli, *Fourier and Wavelet Signal Processing*. Cambridge: Cambridge University Press, 2014, <http://www.fourierandwavelets.org/>.
- [27] J. Kovačević and A. Chebira, "Life beyond bases: The advent of frames (Part I)," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 86–104, Jul. 2007.
- [28] J. Kovačević and A. Chebira, "Life beyond bases: The advent of frames (Part II)," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 115–125, Sep. 2007.
- [29] J. Kovačević and A. Chebira, "An introduction to frames," *Found. Trends Signal Process.*, vol. 2, no. 1, pp. 1–94, 2008.
- [30] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal processing and compression with wavelet packets," Yale Univ., Tech. Rep., 1991.
- [31] R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, pp. 786–804, 1979.
- [32] D. Gabor, "Theory of communication," *J. IEE*, vol. 93, pp. 429–457, 1946.
- [33] J. Jones and L. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *J. Neurophysiol.*, vol. 58, pp. 1233–1258, 1987.
- [34] J. Daugman, "Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, pp. 1169–1179, 1988.
- [35] A. Jain and F. Farrokhnia, "Unsupervised texture segmentation using Gabor filters," *Pattern Recogn.*, vol. 24, pp. 1167–1186, 1991.
- [36] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, pp. 971–987, 2002.
- [37] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, pp. 2037–2041, 2006.
- [38] G. Zhao and M. Pietikainen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, pp. 915–928, 2007.
- [39] R. Bhagavatula, M. C. Fickus, J. W. Kelly, C. Guo, J. A. Ozolek, C. A. Castro, and J. Kovačević, "Automatic identification and delineation of germ layer components in H&E stained images of teratomas derived from human and nonhuman primate embryonic stem cells," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Rotterdam, Apr. 2010, pp. 1041–1044.
- [40] A. Kuruvilla, J. Li, P. Hennings Yeomans, P. Quelhas, N. Shaikh, A. Hoberman, and J. Kovačević, "Otitis media vocabulary and grammar," in *Proc. IEEE Int. Conf. Image Process.*, Orlando, FL, Sep. 2012, pp. 2845–2848.
- [41] M. T. McCann, R. Bhagavatula, M. C. Fickus, J. A. Ozolek, and J. Kovačević, "Automated colitis detection from endoscopic biopsies as a tissue screening tool in diagnostic pathology," in *Proc. IEEE Int. Conf. Image Process.*, Orlando, FL, Sep. 2012, pp. 2809–2812.
- [42] A. Kuruvilla, N. Shaikh, A. Hoberman, and J. Kovačević, "Automated diagnosis of otitis media: A vocabulary and grammar," *Int. J. Biomed. Imag., sp. iss. Computer Vis. Image Process. for Computer-Aided Diagnosis*, Aug. 2013.
- [43] S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, pp. 2786–2799, 2012.
- [44] D. K. Hammond, P. Vandergheynst, and R. Gribonval, "Wavelets on graphs via spectral graph theory," *Appl. Comput. Harmon. Anal.*, vol. 30, pp. 129–150, 2011.
- [45] K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell, "Text classification from labeled and unlabeled documents using EM," *Mach. Learn.*, vol. 39, no. (2/3), pp. 1–32, 2000.
- [46] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. Conf. Computational Learning Theory*, Madison, WI, 1998.
- [47] X. Zhu, "Semi-supervised learning with graphs," Ph.D. dissertation, Carnegie Mellon Univ., 2005.
- [48] A. D. Szlam, R. R. Coifman, and M. Maggioni, "A general framework for adaptive regularization based on diffusion processes," *J. Mach. Learn. Res.*, no. 9, pp. 1711–1739, Aug. 2008.
- [49] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," Carnegie Mellon Univ., Tech. Rep., 2002.
- [50] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *Proc. ICML*, 2003, pp. 912–919.
- [51] F. R. K. Chung, *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*. Am. Math. Soc., 1996.
- [52] L. Wasserman, *All of Nonparametric Statistics*, ser. Springer Texts in Statistics. Springer, 2005.
- [53] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning; Data Mining, Inference, and Prediction*. Springer-Verlag, 2001.
- [54] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*, ser. Springer Texts in Statistics. Springer, 2002.
- [55] I. Daubechies, "Orthonormal bases of compactly supported wavelets II. Variations on a theme," *SIAM J. Math. Anal.*, vol. 24, no. 2, pp. 499–519, Mar. 1993.
- [56] C. M. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. Springer, 2006.
- [57] A. Singhal, "Modern information retrieval: A brief overview," *Bulletin IEEE Computer Soc. Tech. Committee Data Eng.*, vol. 24, no. 4, pp. 35–42, 2001.
- [58] M. Grant and S. Boyd, "Cvx: Matlab software for disciplined convex programming, version 2.0 beta," <http://cvxr.com/cvx>, Sep. 2013.
- [59] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag, 2008, pp. 95–110, <http://stanford.edu/~boyd/graph/dcp.html>.