

MULTIRESOLUTION CLASSIFICATION WITH SEMI-SUPERVISED LEARNING FOR INDIRECT BRIDGE STRUCTURAL HEALTH MONITORING

*Siheng Chen^{1,2}, Fernando Cerda^{3,5}, Jia Guo², Joel B. Harley¹, Qing Shi¹,
Piervincenzo Rizzo⁴, Jacobo Bielak³, James H. Garrett³ and Jelena Kovačević^{1,2}*

¹Dept. of ECE, ²Dept. of BME and Center for Bioimage Informatics

³Dept. of CEE, Carnegie Mellon University, Pittsburgh, PA, USA

⁴Dept. of CEE, University of Pittsburgh, Pittsburgh, PA, USA

⁵Dept. Ingeniería Civil, Universidad de Concepción, Concepción, Chile

ABSTRACT

We present a multiresolution classification framework with semi-supervised learning for the indirect structural health monitoring of bridges. The monitoring approach envisions a sensing system embedded into a moving vehicle traveling across the bridge of interest to measure the modal characteristics of the bridge. To enhance the reliability of the sensing system, we use a semi-supervised learning algorithm and a semi-supervised weighting algorithm within a multiresolution classification framework. We show that the proposed algorithm performs significantly better than supervised multiresolution classification.

Index Terms— multiresolution classification, semi-supervised learning, bridge structural health monitoring

1. INTRODUCTION

Bridge structural health monitoring (SHM) has been an intense research area for some time. Traditional, direct approaches, are to collect acceleration signals by installing sensors on a bridge. The drawback of such direct approaches is that they require a sophisticated and expensive electronic infrastructure with installation, maintenance and power support. Recently, indirect approaches have been proposed [1, 2, 3], suggesting the use of moving vehicles to collect data from accelerometers inside the vehicles, a far less expensive and complex solution (see Figure 1).

In indirect approaches less data is collected and the data is noisier (as it is farther from the source); thus, data analysis plays a crucial role. Moreover, although it is easy to get a large number of data samples, it is expensive to label them (which involves physically inspecting the bridge and determining its health); thus, very few data samples are actually

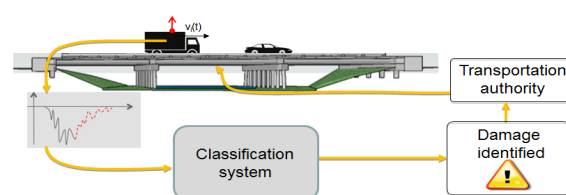


Fig. 1: Indirect bridge SHM system. Acceleration signals are collected from a moving vehicle and sent to a classification system, which identifies the bridge status and reports it to a transportation authority.

labeled. This real-world constraint turns the indirect bridge SHM into a semi-supervised classification problem.

We propose a novel semi-supervised classification framework that takes advantage of supervised multiresolution classification (MRC) [4], which extracts hidden features in localized time-frequency regions (subbands), and a semi-supervised learning algorithm [5], which uses both labeled and unlabeled samples. This is followed by a semi-supervised weighting algorithm that combines information from all the subbands of all the signals to make a global decision.

The outline of the paper as follows: Section 2 states the problem and gives a brief overview of MRC and semi-supervised learning; Section 3 describes our proposed algorithm, which is validated in Section 4 on acceleration signals collected from a lab-scale bridge-vehicle dynamic system. Section 5 concludes with pointers to future directions.

2. BACKGROUND AND PROBLEM FORMULATION

Classification. Our task in bridge SHM is to label acceleration signals as belonging to different classes of structural change or damage, a task known as classification [6]. Let $\mathcal{X} = \{x^{(i)} \in \mathbb{R}^N\}_{i=1}^n$ be the given dataset with n signals, l labeled and u unlabeled; $\mathcal{Y} = \{y^{(i)} \in \{1, 2, \dots, C\}\}_{i=1}^\ell$

The authors gratefully acknowledge support from the NSF through awards 1130616 and 1017278, as well as CMU Carnegie Institute of Technology Infrastructure Award.

i	sample index	
$\mathcal{X} = \{x^{(i)}\}$	input dataset	$i = 1, \dots, n$
$\mathcal{Y} = \{y^{(i)}\}$	ground-truth labels for $\mathcal{L}$	$i = 1, \dots, \ell$
$\mathcal{L} = \{(x^{(i)}, y^{(i)})\}$	labeled dataset	$i = 1, \dots, \ell$
$\mathcal{U} = \{x^{(i)}\}$	unlabeled dataset	$i = \ell + 1, \dots, n$
$\mathcal{F}$	feature extraction function	
$f^{(i)}$	feature vector	$i = 1, \dots, n$
$q^{(i)}$	ground-truth vector	$i = 1, \dots, \ell$
$\hat{y}^{(i)}$	estimated label	$i = \ell + 1, \dots, n$
$\hat{q}^{(i)}$	confidence vector	$i = \ell + 1, \dots, n$
$\hat{\mathcal{Y}} = \{\hat{y}^{(i)}\}$	estimated labels for $\mathcal{U}$	$i = \ell + 1, \dots, n$

Table 1: Parameters used in a generic classification system.

the *ground-truth labels* for the *labeled dataset* $\mathcal{L} = \{(x^{(i)} \in \mathcal{X}, y^{(i)} \in \mathcal{Y})\}_{i=1}^{\ell}$; and $\mathcal{U} = \{x^{(i)} \in \mathcal{X}\}_{i=\ell+1}^n$ the *unlabeled dataset*. Then, the problem can be formulated as designing a map that associates an input signal to a class label with a certain probability. Typically, a generic classification system will have an intermediate block between the two, a *feature extractor* $\mathcal{F}$ (FE), aimed at reducing the dimensionality of the problem; this is followed by a *classifier* $\mathcal{C}$. The outputs of the classifier are the *estimated labels* $\hat{\mathcal{Y}} = \{\hat{y}^{(i)} \in \{1, 2, \dots, C\}\}_{i=\ell+1}^n$ for the unlabeled dataset $\mathcal{U}$. If the classifier is supervised, we denote the block by SC (see Figure 2).

Note that a label can also be viewed as a posterior probability vector $\hat{q}$ of size $C \times 1$, where the c th component of the vector, $\hat{q}(c)$, is the probability that a sample belongs to the c th class. Since $\hat{q}$ gives a confidence to an assigned label, we name it a *confidence vector*. The confidence vector for a labeled sample is called the *ground-truth vector*, q .

MRC. MRC is a supervised classification framework (see Figure 2), originally proposed for bioimaging applications [4, 7, 8]. It decomposes images into S localized space-frequency subbands using wavelet packets, a data-adaptive MR technique [9]. In each subband, MRC extracts features, classifies them, and produces a local classification decision. A supervised weighting algorithm combines all local decisions into a global decision (see Algorithm 1).

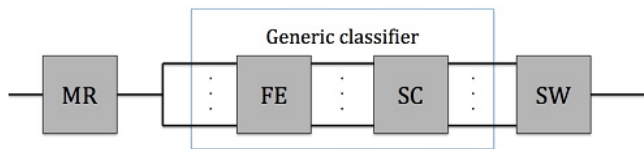


Fig. 2: Supervised MRC decomposes images into localized space-frequency subbands using wavelet packets (MR), followed by feature extraction (FE) and supervised classification (SC) in each subband, yielding a local classification decision. A supervised weighting algorithm (SW) combines all local decisions into a global decision.

Algorithm 1 MRC

Input	$\mathcal{X} = \{x^{(i)}\}$	input dataset
Output	$\hat{\mathcal{Y}} = \{\hat{y}^{(i)}\}$	estimated labels for $\mathcal{X}$
	s	subband index $s = 1, 2, \dots, S$
	$\mathcal{D}_s$	MR function
	$a_s^{(i)}$	MR coefficients
	$f_s^{(i)}$	feature vector
	$\mathcal{C}_s$	supervised classification function
	$\hat{q}_s^{(i)}$	confidence vector
	$\mathcal{W}$	supervised weighting function
	w	weighting vector $\ w\ _1 = 1$
	$\hat{q}^{(i)}$	confidence vector after weighting

MRC($\mathcal{X}$)

MR	$a_s^{(i)} = \mathcal{D}_s(x^{(i)})$
FE	$f_s^{(i)} = \mathcal{F}(a_s^{(i)})$
SC	$\hat{q}_s^{(i)} = \mathcal{C}_s(f_s^{(i)})$
SW	$\hat{q}^{(i)} = \mathcal{W}([\hat{q}_1^{(i)}, \hat{q}_2^{(i)}, \dots, \hat{q}_S^{(i)}])$
	$\hat{y}^{(i)} = \arg \max_c \{\hat{q}^{(i)}(c)\}$
return	$\hat{\mathcal{Y}}$

(Only parameters different from Table 1 are listed.)

MRC provides various options: the choice of the filter bank $\mathcal{D}$ used in the MR block [10], the feature extraction method $\mathcal{F}$ used in the FE block, and the supervised classifier $\mathcal{C}$ used in the SC block. In the s th subband, the function producing the filter bank output is denoted by $\mathcal{D}_s$ and the supervised classifier by $\mathcal{C}_s$ (different classification boundaries in different subbands). 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 $\mathcal{W}$ as

$$\hat{q}^{(i)} = \mathcal{W}(\hat{Q}^{(i)}) = \hat{Q}^{(i)}w,$$

where the weighting vector w , which assigns weight to each subband according to its discriminative power, is chosen by optimizing a supervised weighting objective function

$$w = \arg \min_w \left\{ \sum_{i=1}^{\ell} \|q^{(i)} - \hat{Q}^{(i)}w\| \right\}. \quad (1)$$

The optimization is performed over all labeled data samples with the constraint that $\|w\|_1 = 1$.

Semi-Supervised Learning. Semi-supervised learning is a technique for training classifiers with both labeled and unlabeled data that assumes that unlabeled data can provide distribution information to build a stronger classifier. It includes generative mixture models with expectation maximization, co-training, transductive support vector machine and graph-based approaches [5]; we focus here on label propagation [11, 12], one of graph-based approaches. It assumes that, while the measured samples exist in a high-dimensional space, they are distributed in a low-dimensional manifold. Based on this, a graph is constructed to analyze the distribution of all samples; by understanding how labels propagate on this graph, classification can be achieved (see Algorithm 2).

Algorithm 2 Label propagation

Input	$\mathcal{X} = \{x^{(i)}\}$	input dataset
Output	$\hat{\mathcal{Y}} = \{\hat{y}^{(i)}\}$	estimated labels for $\mathcal{X}$
	A	adjacency matrix
	ρ	local measurement
	σ	scaling coefficient
	P	transition matrix
	t	transition time

Label propagation($\mathcal{X}$)		
Construction	$A_{ij} = \exp(-\rho(x^{(i)}, x^{(j)})/\sigma)$	
Normalization	$P = D^{-1}A, \quad D_{ii} = \sum_j A_{ij}$	
Initialization	$\hat{Q}$, with $\hat{Q}_{y^{(i)}, i} = 1, \quad i = 1, \dots, \ell$	
Diffusion	$\hat{Q} \leftarrow \hat{Q}P^t$	
Labeling	$\hat{y}^{(i)} = \arg \max_c \{\hat{q}^{(i)}(c)\}$	
return	$\hat{\mathcal{Y}}$	

(Only parameters different from Algorithm 1 are listed.)

3. PROPOSED ALGORITHM

MRC analyzes data to uncover hidden information; in its original form, it uses supervised classification, and can thus train on labeled samples only. When the labeled set is small or contains improperly labeled samples, 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 frameworks and gain the best of both worlds: a semi-supervised MRC.

Semi-Supervised Classification. The MR and FE blocks from Figure 2 work as before. The first change is that the supervised classifier block, SC, is replaced by a semi-supervised one, SSC (see Figure 3), using both labeled and unlabeled samples to make a labeling decision in each subband.

Semi-Supervised Weighting. We now explain how to weigh decisions from all the subbands to get a global decision in a semi-supervised manner. Labeled samples contribute to weighting directly by fitting their confidence vectors from all the subbands to the ground truth; unlabeled samples cannot do the same as they do not have the ground truth. We could use Shannon entropy to measure the confidence of labeling an unlabeled sample; if the entropy is small (less uncertainty, high confidence), it is easy to assign a label to the sample, and vice versa. In the label propagation algorithm, Algorithm 2, we label each sample by finding the largest element in its confidence vector; we could thus normalize each confidence vector to sum to 1 to measure its entropy. We encounter a problem, however; for example, let $\hat{q}^{(1)} = [0.5 \ 0.5 \ 0]^T$ and $\hat{q}^{(2)} = [0.5 \ 0.25 \ 0.25]^T$ be confidence vectors. While we can label $\hat{q}^{(2)}$ as Class 1 but cannot make a decision for $\hat{q}^{(1)}$, the entropy measure tells us that we can label $\hat{q}^{(1)}$ with higher confidence (less uncertainty) because its entropy is lower. To resolve this issue, we define a new uncertainty measure,

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

Algorithm 3 Semi-supervised MRC

Input	$\mathcal{X} = \{x^{(i)}\}$	input dataset
Output	$\hat{\mathcal{Y}} = \{\hat{y}^{(i)}\}$	estimated labels for $\mathcal{X}$
	$\mathcal{C}$	semi-supervised classification function
	$\mathcal{W}$	semi-supervised weighting function

SSMRC($\mathcal{X}$)		
MR	$a_s^{(i)} = \mathcal{D}_s(x^{(i)})$	
FE	$f_s^{(i)} = \mathcal{F}(a_s^{(i)})$	
SSC	$\hat{q}_s^{(i)} = \mathcal{C}_s(f_s^{(i)})$	
SSW	$\hat{q}^{(i)} = \mathcal{W}([\hat{q}_1^{(i)}, \hat{q}_2^{(i)}, \dots, \hat{q}_s^{(i)}])$	
	$\hat{y}^{(i)} = \arg \max_c \{\hat{q}^{(i)}(c)\}$	
return	$\hat{\mathcal{Y}}$	

(Only parameters different from Algorithm 1 are listed.)

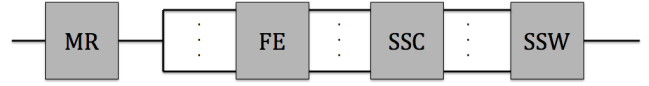


Fig. 3: Semi-supervised MRC. Classification and weighting algorithm in Figure 2 are replaced with their semi-supervised versions so that unlabeled data can contribute to classification.

where $H(\hat{q})$ is the entropy of confidence vector $\hat{q}$, 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. Let $M_s^{(i)}$ be the uncertainty of the s th subband to label the i th sample. Since entropy is additive, the total uncertainty of a subband when classifying unlabeled samples is the mean uncertainty over all the unlabeled samples in this subband. Thus, the total uncertainty of the s th subband is

$$m_s = \frac{1}{u} \sum_{i=l+1}^{l+u} M_s^{(i)}.$$

Define the normalized confidence of the s th subband as

$$g_s = \frac{e^{-\beta m_s}}{\sum_{k=1}^S e^{-\beta m_k}},$$

where β is the decaying coefficient. When the uncertainty of a subband is large, the confidence is small and the subband gets assigned a low weight, and vice versa. This confidence is the discriminative power of a given subband.

To combine confidences of all subbands into a global decision, we choose the weighting vector by optimizing a semi-supervised weighting objective function,

$$\begin{aligned} w &= \arg \min_{\omega} \left\{ \frac{\alpha}{\ell} \sum_{i=1}^{\ell} \|q^{(i)} - \hat{Q}^{(i)}\omega\| \right. \\ &\quad \left. + (1 - \alpha)\|\omega - g\| \right\}, \end{aligned} \quad (2)$$

where w is as defined before and α is the labeling ratio defined as $\alpha = \ell/(\ell+u)$. The first term in (2) is exactly (1), representing the contribution from labeled samples. In the second term, we fit weights to subbands' confidences, representing the contribution from the unlabeled samples. We use the labeling ratio to balance these two terms; when we have a large number of labeled samples, the first term dominates, otherwise, the second one does. Since this is a convex optimization problem, it is numerically efficient to solve. After getting the weights, we can get the global decision as $\hat{y}^{(i)} = \arg \max_c \hat{q}^{(i)}(c)$, where $\hat{q}^{(i)} = \hat{Q}^{(i)}w$ (see Algorithm 3).

4. EXPERIMENTAL RESULTS

Dataset. We built a lab-scale bridge-vehicle dynamic system, put a sensor on a vehicle, and let it move across the bridge. We collected 30 samples for each of 13 different bridge damage scenarios, 8 different speeds and 2 different vehicles [13].

Experimental Setup. Given a specific vehicle driven at a specific speed, we want to classify 13 scenarios, in particular with a low labeling ratio. We consider 16 vehicle-speed cases for each of which there are 30 samples per 13 of the scenarios, and vary the labeling ratio as 10%, 30%, 50%, 70% and 90%; the final accuracy is the average over the 13 scenarios. We compare the performance of our proposed algorithm to the supervised MRC as well as the label propagation algorithm. We choose a Coiflet filter bank [14] with 4 levels in the MR block, principal component analysis in the FE block, naive Bayes, logistic regression and radius kernel SVM in the SC block [15], and label propagation in the SSC block. For label propagation, we choose the local measurement ρ to be the cosine distance, scaling coefficient $\sigma = 1$, transition time $t = 8$. For computational efficiency, we construct a $k = 4$ regular graph (each vertex connects to 4 neighbors). In semi-supervised weighting function, we choose penalty threshold $T = 0.02$ and the penalty term $\lambda(d) = 1 + 5(d/T - 1)^2$. We performed a 30-fold cross-validation and found that parameters do not influence the results too much.

Results. Table 2 compares the performance of different classifiers with the low labeling ratio of 10%. We use V for vehicle, S for speed, SMRC for supervised MRC, LR for logistic regression, NB for naive Bayes, KSVM for kernel SVM, LP for label propagation, SSMRC for semi-supervised MRC and LP-W for label propagation with a semi-supervised weighting algorithm in (2). We see that when the labeling ratio is low, supervised MRC performs poorly, label propagation works well, and semi-supervised MRC works the best.

Figure 4 shows the dependence of classification accuracy on the labeling ratio for 2 vehicles averaged across 8 speeds. Both figures show similar trends; as the labeling ratio decreases, accuracy drops sharply for all algorithms except for semi-supervised MRC, which consistently outperforms them all and whose performance stays relatively flat even at very low labeling ratios.

V	S	LR	SMRC	KSVM	LP	SSMRC
		NB	NB		LP-W	LP-W
1	1	47.8	62.4	84.1	81.3	99.8
	2	57.0	62.9	84.5	86.2	99.9
	3	52.4	60.9	84.7	86.0	99.4
	4	66.2	63.2	89.1	90.8	99.9
	5	48.0	46.5	81.8	85.2	94.5
	6	34.2	45.7	74.8	86.0	93.5
	7	37.4	46.8	66.0	69.0	72.2
	8	38.8	43.4	59.0	75.9	82.5
2	1	40.9	58.5	76.2	75.6	85.9
	2	37.7	57.7	61.6	68.3	80.5
	3	58.1	65.6	81.3	81.7	94.7
	4	46.7	56.2	73.1	80.0	87.3
	5	47.9	59.0	72.5	76.5	88.1
	6	44.4	54.5	73.3	78.6	83.8
	7	48.3	63.7	76.9	83.3	88.2
	8	54.2	62.3	79.1	90.4	93.8

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

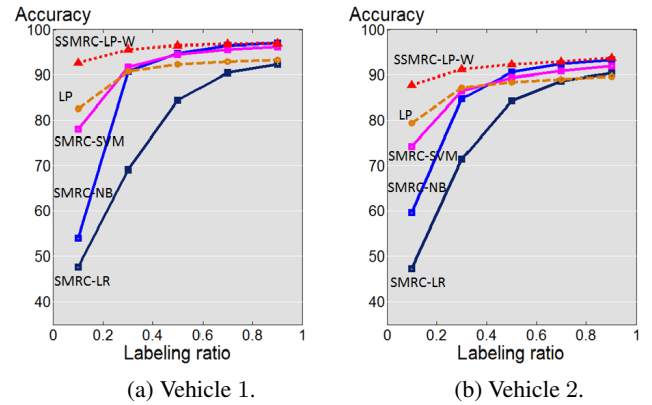


Fig. 4: Accuracy as a function of the labeling ratio.

5. CONCLUSIONS AND FUTURE WORK

We presented a framework and algorithm for indirect bridge SHM. Compared to our previous work [2, 3], we added two new ingredients: (1) We analyze signals in MR spaces, instead of only in frequency domain. (2) We focus on semi-supervised learning setting, instead of supervised learning. Our system combines MR techniques and semi-supervised learning by using a semi-supervised weighting algorithm. The new system performs consistently better than supervised MRC, and significantly better when the labeling ratio is small.

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

6. REFERENCES

- [1] Y. B. Yanga, C. W. Lina, and J. D. Yaub, "Extracting bridge frequencies from the dynamic response of a passing vehicle," *Journ. of Sound and Vibration*, pp. 471–493, May 2004.
- [2] F. Cerda, J. Garrett, J. Bielak, R. Bhagavatula, and J. Kovačević, "Exploring indirect vehicle-bridge interaction for bridge SHM," in *Proc. Int. Conf. Bridge Maintenance, Safety and Management*, Philadelphia, PA, July 2010, pp. 696–702.
- [3] F. Cerda, J. Garrett, J. Bielak, P. Rizzo, J. A. Barrera, Z. Zhang, S. Chen, M. McCann, and J. Kovačević, "Indirect structural health monitoring in bridges: scale experiments," in *Proc. Int. Conf. Bridge Maintenance, Safety and Management*, Lago di Como, Italy, July 2012, pp. 346–353.
- [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] X. Zhu, "Semi-supervised learning literature survey," Tech. Rep. 1530, Univ. Wisconsin-Madison, 2005.
- [6] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, John Wiley & Sons, Englewood Cliffs, NJ, 2001.
- [7] 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. Imaging*, Arlington, VA, Apr. 2007, pp. 288–291.
- [8] 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. Imaging*, Paris, France, May 2008, pp. 979–982.
- [9] R. R. Coifman, Y. Meyer, S. Quake, and M. V. Wickerhauser, "Signal processing and compression with wavelet packets," Tech. Rep., Yale Univ., 1991.
- [10] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Signal Processing. Prentice Hall, Englewood Cliffs, NJ, 1995, <http://waveletsandsubbandcoding.org/>.
- [11] X. Zhu and Z. Ghahramani, "Learning from labeled and unlabeled data with label propagation," Tech. Rep., Carnegie Mellon Univ., 2002.
- [12] A. D. Szlam, R. R. Coifman, and M. Maggioni, "A general framework for adaptive regularization based on diffusion processes," *Journ. Mach. Learn. Res.*, , no. 9, pp. 1711–1739, Aug. 2008.
- [13] 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. Journ. Smart Struct. Syst.*, 2012, Submitted.
- [14] I. Daubechies, "Orthonormal bases of compactly supported wavelets II. Variations on a theme," *SIAM Journ. Math. Anal.*, vol. 24, no. 2, pp. 499–519, Mar. 1993.
- [15] C. M. Bishop, *Pattern Recognition and Machine Learning*, Information Science and Statistics. Springer, 2006.