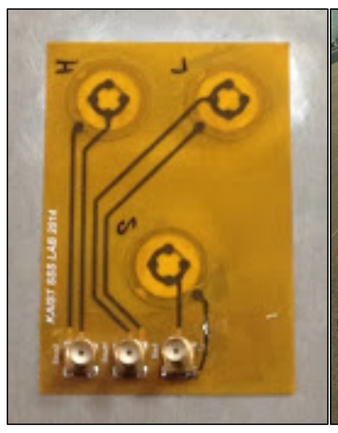
Data-driven Fatigue Crack Evaluation based on Wave Propagation Data

Seongwoon Jeong (swjeong3@stanford.edu)

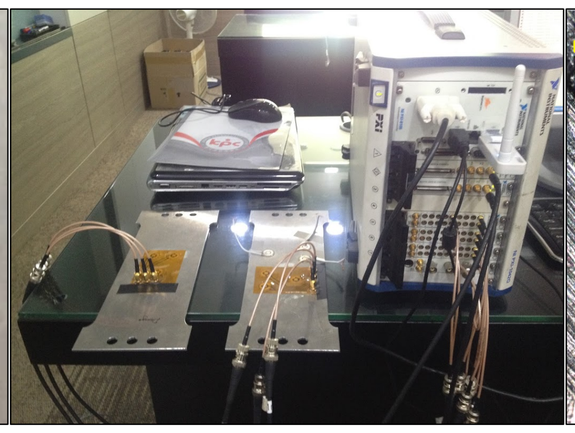
Overview

The goal of this project is to propose a data-driven crack evaluation method based on the machine learning techniques to better support the decision making process of bridge owners. To achieve this goal, different input feature combinations based on sensor data are defined and tested, and different classification methods are utilized to determine a specimen is intact or damaged. The sensor data is acquired from steel specimen using a high-frequency fatigue crack sensor. The raw sensor data is pre-processed so that several features representing meaningful information of sensor data can be extracted. In addition, I defined two target classes: one is intact and another is damaged. Based on these feature-target data, binary classification analysis is conducted. The results shows that the proposed feature-target data has certain trend that can distinguish damaged specimen from intact specimen, and the proposed classification scheme performs well for the training data set. However, the proposed scheme seems to have an over-fitting issue so far and does not show the expected performance when a test data set is applied. In the future work, I will try to figure out how to resolve this problem so that the proposed method can successfully evaluate the fatigue crack based on the sensor data.

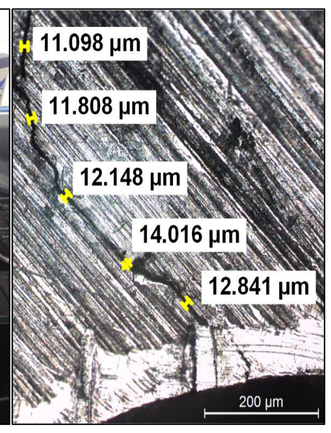
Fatigue Crack Sensor



Fatigue crack sensor (Exciter + Sensor)



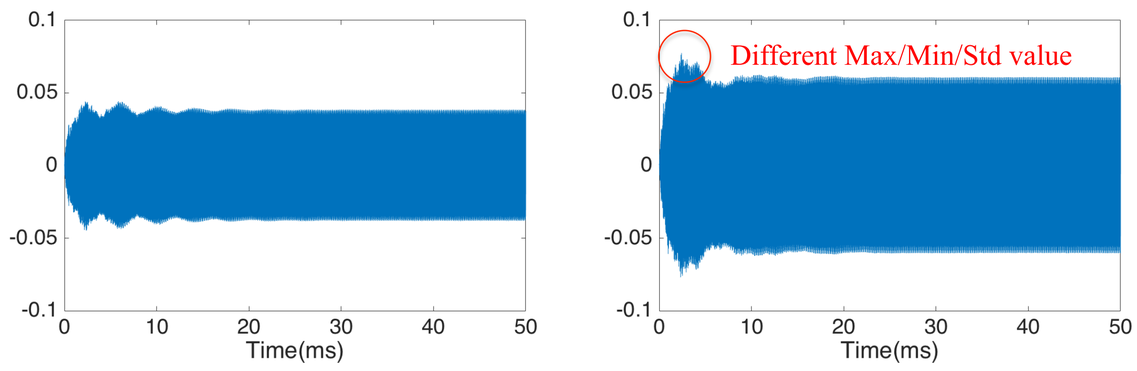
Fatigue crack sensors installed on specimens (Intact & Damaged)



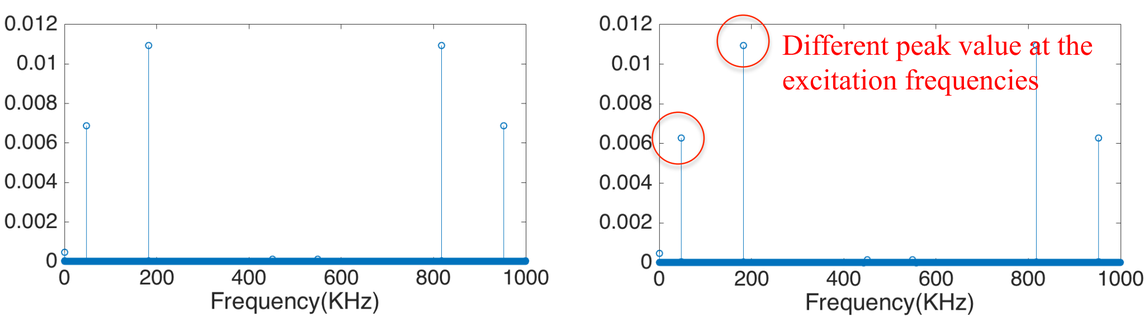
Fatigue crack on the damaged specimen

- Sampling rate: 1.0 MHz
- Sampling duration: 0.5 second per excitation
- Output data: Time-series wave amplitude
- High frequency(HF) excitation: 180 ~ 185 KHz
- Low frequency(LF) excitation: 50 KHz
- Number of sensors: 2 per specimen
- Number of specimens: 5 intact, 4 damaged specimens
- Number of data sets:
  6(HF) × 1(LF) × 2(sensors per specimen) × 9(number of specimen) = 108

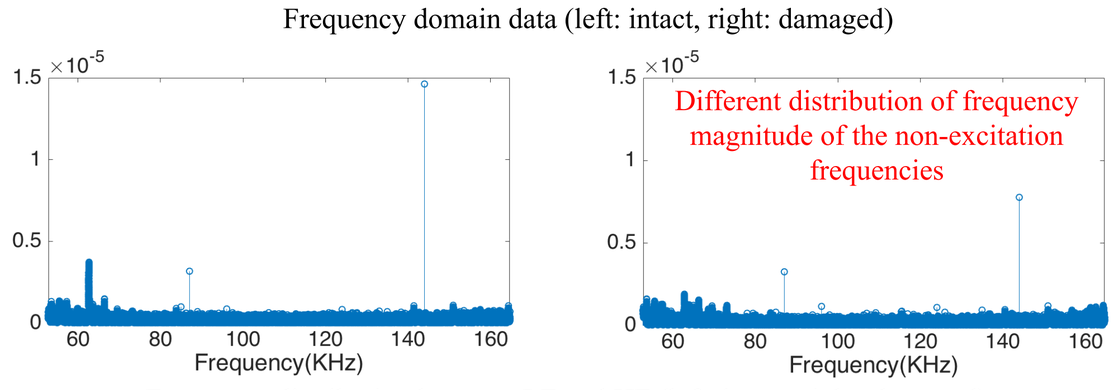
Sensor Data



Time-series data (left: intact, right: damaged)



Frequency domain data (left: intact, right: damaged)



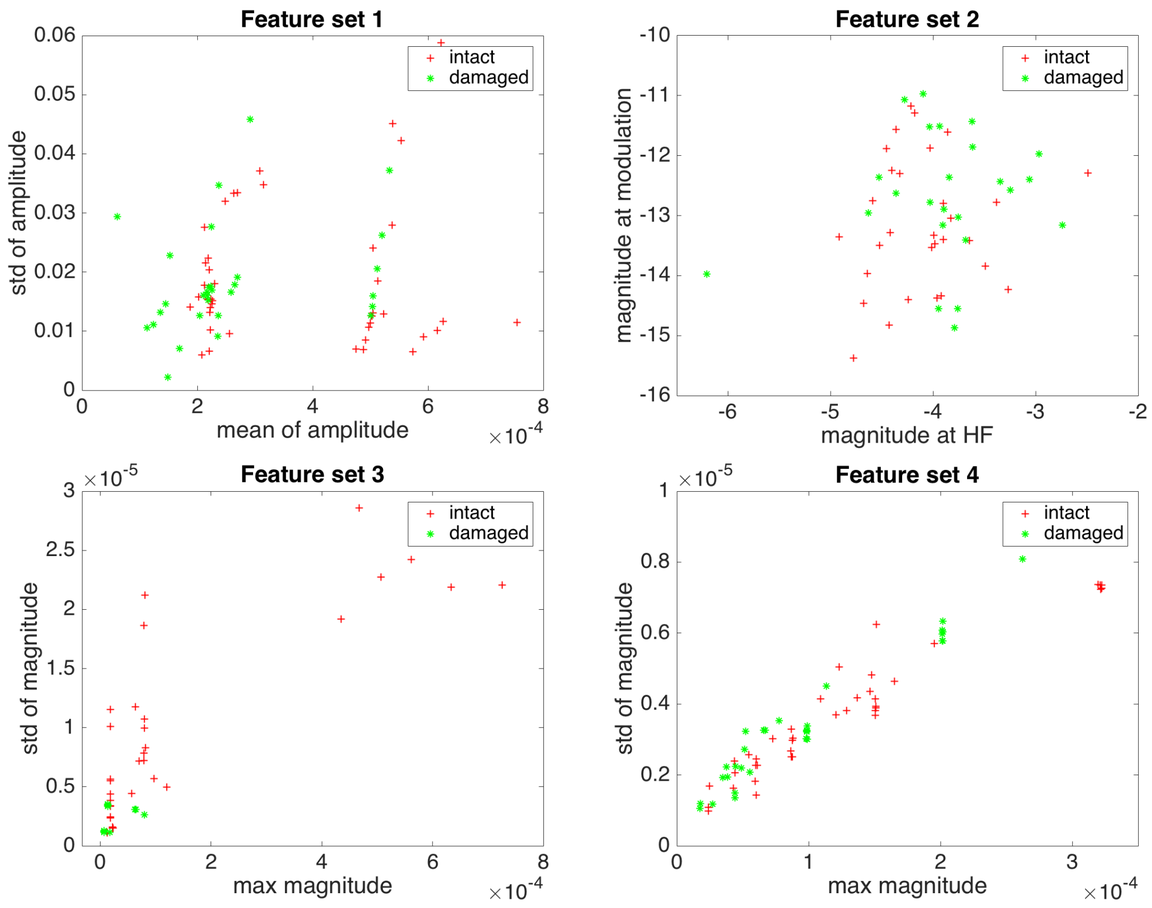
Frequency distribution between LF and HF (left: intact, right: damaged)

- The raw sensor data is a short but very dense time-series data, which consists of 0.5 million consecutive double values.
- Since it is difficult to utilize the raw data, statistically meaningful information such as maximum, minimum and standard deviation values of the raw data are calculated and used as input features.
- In addition, the raw time series data is converted to frequency domain data using Discrete Fourier Transform (DFT), so that the contribution of specific frequencies and distribution of amplitude can also be used as input features.

Feature Selection

Defined Features

- Feature sets 1: μ, σ, max. and min. of time-series data
- Feature sets 2: peak magnitude at excitation and modulation frequencies
- Feature sets 3: max. and σ of frequency magnitude of the non-excitation frequencies (frequency < LF)
- Feature sets 4: max. and σ of frequency magnitude of the non-excitation frequencies (LF < frequency < HF)

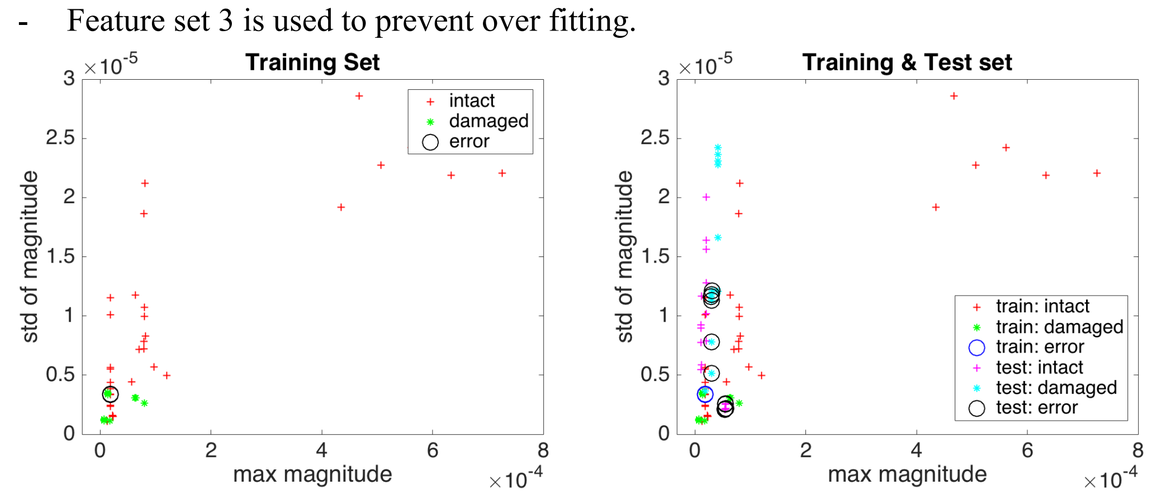


Accuracy of different feature sets

	Linear discriminant	Linear SVM	Quadratic SVM	Gaussian SVM	Max with k-folds (k=5)
Feature 1	73.6%	75.0%	87.5%	**94.4%**	66.7%
Feature 2	66.7%	70.8%	94.4%	**100%**	61.9%
Feature 3	56.9%	70.8%	**97.2%**	88.9%	93.1%
Feature 4	61.1%	62.5%	86.1%	**87.5%**	61.9%

Training and Test

- The sensor data is divided into training set (66%) and test set (33%).
- Quadratic SVM is used with k-folds cross validation (k=5) and box constraint 1.
- Feature set 3 is used to prevent over fitting.



Confusion matrix for training set

True \ Predict	Intact	Damaged
Intact	37 (88.1%)	5 (11.9%)
Damaged	0 (0%)	30 (100%)

Confusion matrix for test set

True \ Predict	Intact	Damaged
Intact	12 (66.7%)	6 (33.3%)
Damaged	6 (33.3%)	12 (66.7%)

Limitation and Future Work

- Although the SVM model performs fine for the training set, the model shows unacceptable errors for the test set. (High variance)
- In order to resolve this problem, it would be good to try getting more training examples.
- Trying smaller set of features usually helps to settle the high variance problem, but the number of features of this model is not very large. In this case, it would be helpful to try extracting different features from raw data.

Acknowledgement

The information, photos, and sensor data of fatigue crack sensor used in this project are provided by Prof. Sohn and Dr. Lim in KAIST.