Vision-based Sensor
Case Study: Verrazano-Narrows Bridge

Khawla Sidaoui – HNTB Corporation

The advent of video analytics and non-contact monitoring methods offers the ability to look at the larger potential of SHM. A vision-based sensor is introduced with methods for noise filters, camera motion compensation and target detection to optimize it for practical use on long-span bridges. The sensor was implemented on Verrazano-Narrows Bridge to study its global behavior using traffic camera feed in a recent wind induced event.

1. Trackers & sampling frequency

The proposed vision sensor system was developed using several trackers available in Python's OpenCV library that detect a rigid body and estimate its motion between consecutive frames in a video sequence. The main trackers used are Median-Flow and KCF, the choice depends on how well they perform given the limitations in each video, such as occlusion, contrast and light. For accurate results and to capture the required frequency bandwidth, the video's minimum *fps* must be at least double the highest frequency that needs to be captured. For applications where the frequency could be higher than 15 Hz, a slow-motion video with *fps* > 30 is required.

2. Trace & camera displacement compensation

The algorithm uses the trackers to calculate coordinates of a user-defined bounding box centered around a target of interest on the bridge, in this case the main cable (Fig.1). The coordinates are extracted in each frame *i* and assigned to a list across *n* frames.

$$[X]_{Target} = [x_{ci}, x_{ci+1}, \dots, x_{cn}]$$

$$[Y]_{Target} = [y_{ci}, y_{ci+1}, \dots, y_{cn}]$$

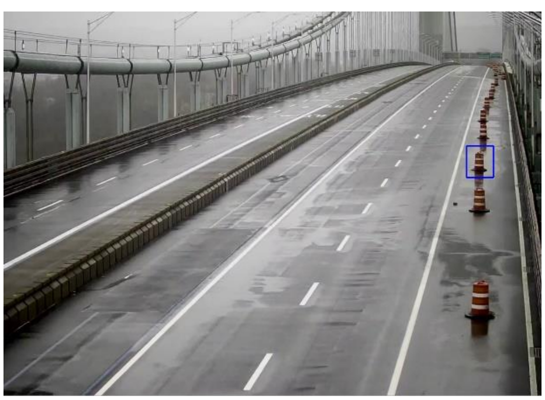


Figure 1- Bounding box around target on bridge

Figure 2- Bounding box around fixed target

Given that the traffic camera is on the bridge and is likely moving, the same process is repeated for a stationary target to compensate for camera shake. The light in the distance (Fig.2) served as a frame of reference to get the camera motion in the x and y directions. Subtracting the 2 coordinates generates the bridge's displacement trace with the amplitude in pixels.

$$[X]_{Compensated} = [X]_{Target} - [X]_{Fixed}$$

$$[Y]_{Compensated} = [Y]_{Target} - [Y]_{Fixed}$$

3. Signal Post-Processing & Mode Identification:

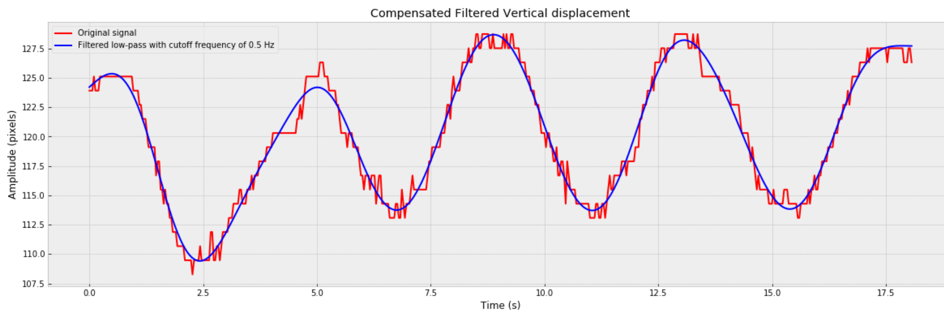


Figure 3 – Original and filtered signals.

A low-pass filer (Butterworth Filter) is applied to clean the signal and remove high unwanted frequencies. Figure 3 shows the original signal in red, and the filtered trace after setting the threshold at 0.5 Hz.

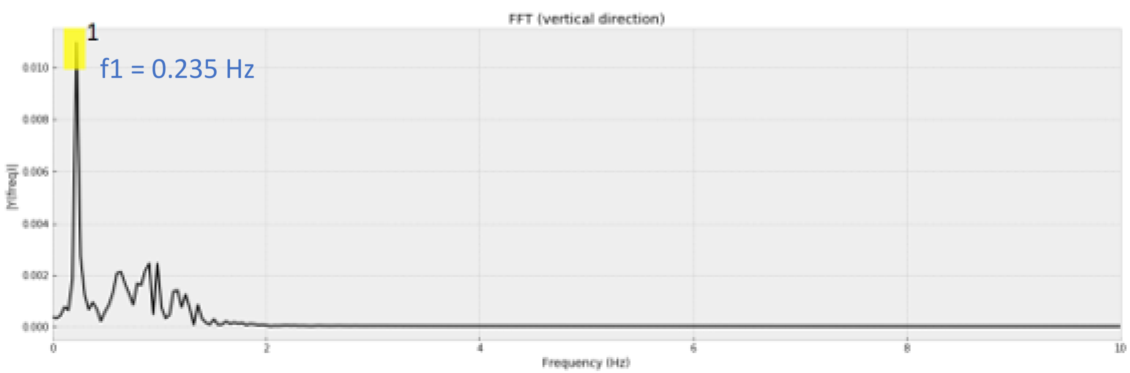


Figure 4 – FFT

The algorithm then produces the FFT of the compensated filtered signal to identify the modes and natural frequencies. Based on a study done by RDWI, the frequency of mode 13 is 0.233 Hz. This correlates with the frequency from the video analysis f1=0.235 Hz with a 1% error.

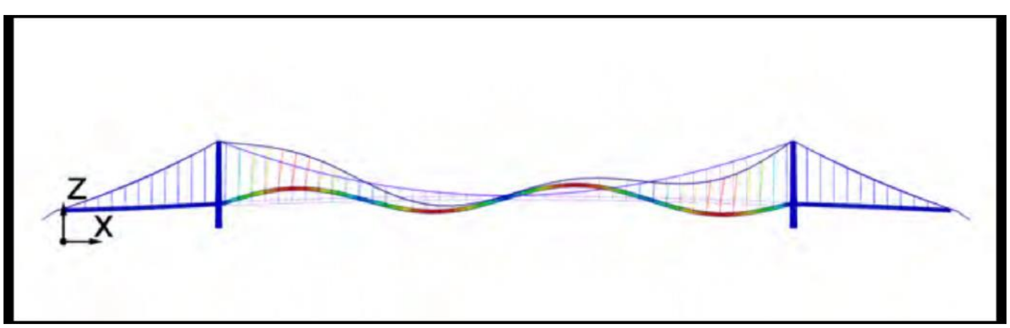


Figure 5 – VNB mode shape 13 (study by RDWI).

4. Scaling & Global Behavior

To get a better sense of the deflections measured, the movement is scaled using a target's known dimension *d* and its size in pixels *f*. The amplitude is normalized (Fig. 7) using the following scaling factor:

$$SF = \frac{d}{f}\ inches/pixel.$$

On VNB, traffic cylinders were used as targets to estimate displacement at multiple points along the deck, and the movement was scaled using their height (H=39.7"). Assuming that the distance between the cylinders is approximately 50 ft, and that the traffic camera is at midspan, the displacement of 5 targets at 50, 150, 200, 300 and 500 ft from midspan was measured (Fig. 6)

Figure 6 – Targets deflection and distance from midspan.

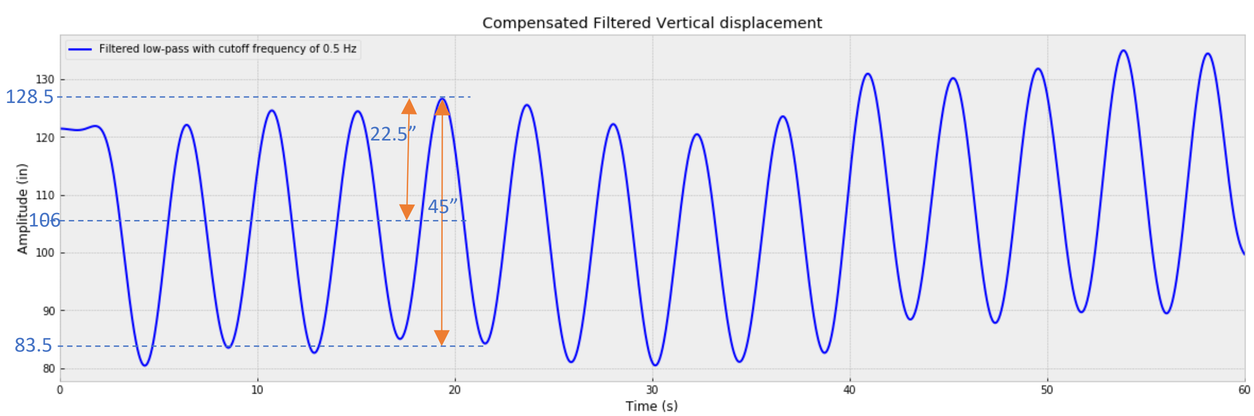
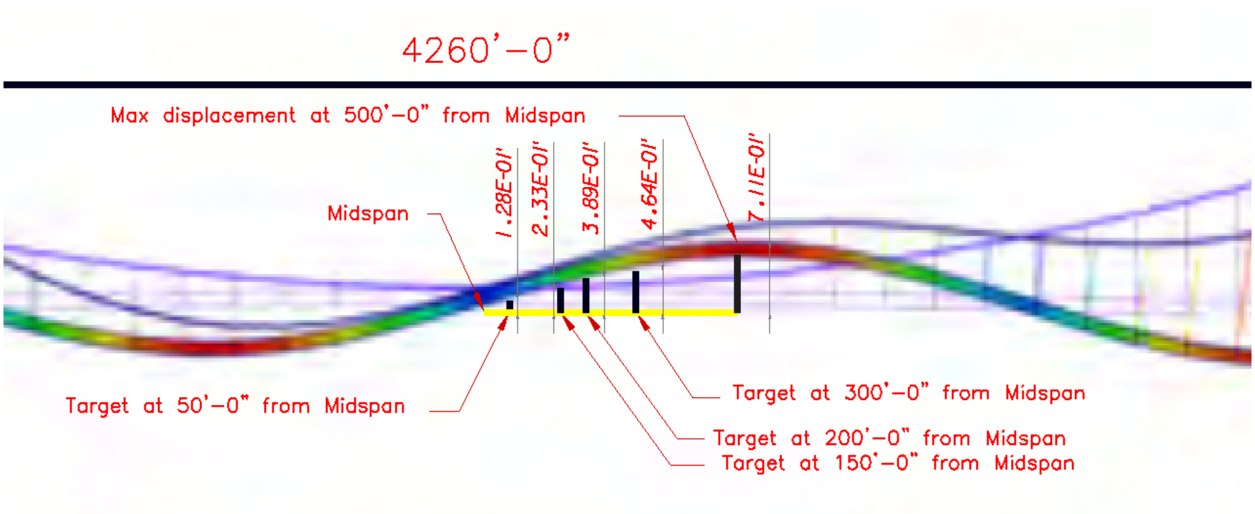


Figure 7- Target 3 normalized displacement.

According to the scaled deflected shape of mode 13, the maximum deflection of the bridge is at 500 ft from midspan. The deflection was measured at the same 5 locations using a random scale (Fig. 8). The gradual increase in motion from the video analysis was compared to the deflected shape to confirm that the deck's behavior is consistent with mode 13 (see Table).



Target	Approx. Distance from midspan (ft)	Displacement based on video analysis (in)	Vid ratio di/d0	Mode 13 deflected shape (random scale)	Deflected shape ratio di/d0	% Error = 1-[ratio(vid)/ratio(mode)] x100
1	50	**4** (=d0)	1	1.28 (=d0)	1	
2	150	7.5	1.875	2.33	1.82	2.9
3	200	12.5	3.125	3.89	3.04	2.8
4	300	15	3.75	4.64	3.63	3.3
At peak	500	**22.5**	5.625	7.11	5.55	1.3

Max amplitude at 500 ft = 22.5" x 2 = **3.75 ft**
Max displacement from baseline to peak at 500 ft = 22.5" = **1.875 ft**

Conclusion

Using only traffic camera feed, the frequency, mode, and maximum displacement of the bridge were estimated:

- Frequency from multiple videos and various targets was consistent with Mode 13.
- Deck motion was consistent with deflected shape of mode 13.
- Maximum amplitude was 3.75 ft at 500 ft from midspan, and maximum deflection was 1.875 ft.