Live Demonstration: Event-based Visual Microphone

Ryogo Niwa[1], Tatsuki Fushimi[1], Kenta Yamamoto[1], and Yoichi Ochiai[1,2]
[1] University of Tsukuba, [2] R&D Center for Digital Nature
{niwa.ryogo, tfushimi, kenta.yam,wizard}@digitalnature.slis.tsukuba.ac.jp

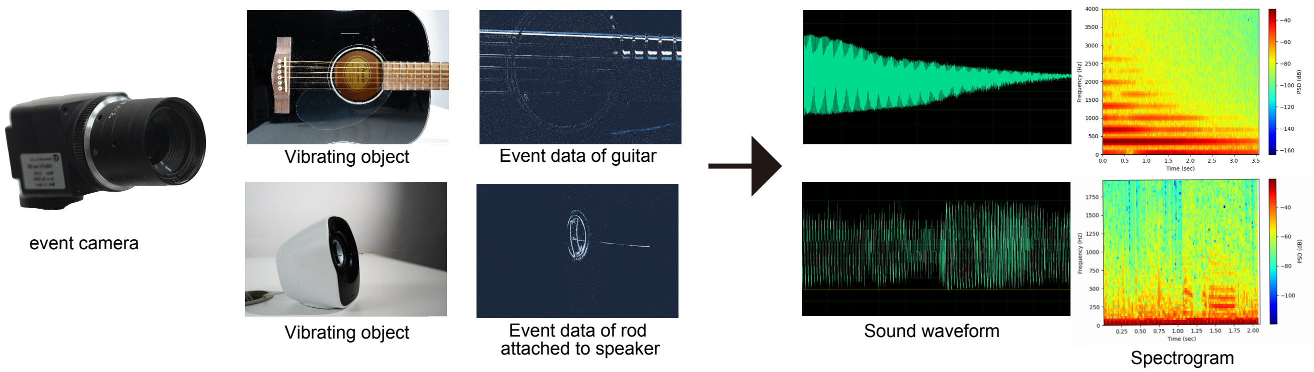

Figure 1. Sound recovery from event data. Left: The event camera efficiently captures the vibrating objects. Right: Sound waveform and spectrogram reconstructed from recorded event data.

Abstract

Non-contact measurement devices, such as laser Doppler vibrometers(LDV) and high-speed cameras, have been developed to measure various forms of vibrations. LDVs are expensive, and high-speed cameras have a significant trade off between resolutions and sampling frequency. Herein, we propose a non-contact, simple, and inexpensive approach to measure vibration using an event camera that records only changes in brightness. We demonstrated the reconstruction of audible sounds using the event camera; to the best of our knowledge, this study could be considered the first to successfully apply this method. This study addresses the challenges in measuring and reconstructing vibrations without requiring additional lighting and using a cost-effective method that can recover audible sounds and human voice. This method provides a new avenue for scientists and engineers to develop cost-effective and non-contact methods for measuring and reconstructing vibrations and sounds.

1. Introduction

From the quantum scale to the universe scale, the vibration of an object can encode various types of information, such as faults in an engineering structure, or conversations in the next room. Therefore, a wide range of contact and non-contact methods have been developed to measure various forms of vibration.

For example, a laser Doppler vibrometer (LDV) can measure vibrations through the frequency shift of light and high-speed cameras can capture minute vibrations with high sensitivity up to MHz for both LDV and high-speed cameras. Devices with such high capability are expensive and complex. In addition, owing to the limits imposed by the Nyquist frequency, it is challenging even for high-end cameras to recover audible sounds from their recordings [1].

Herein, we propose "Event-based Visual Microphone" a non-contact, simple, and inexpensive approach for measuring vibration using an event camera. An event camera is a special type of camera that only records changes in brightness at each pixel, unlike an RGB camera. As vibrating parts are accompanied by changes in brightness, event cameras can capture a wide field of vision at a high speed (up to 1 MHz) without significant trade off. In addition, high-frequency vibrations can theoretically be reconstructed without extra lighting.

2. Method

We used a Silky Evcam HD camera with a 1300 Mega

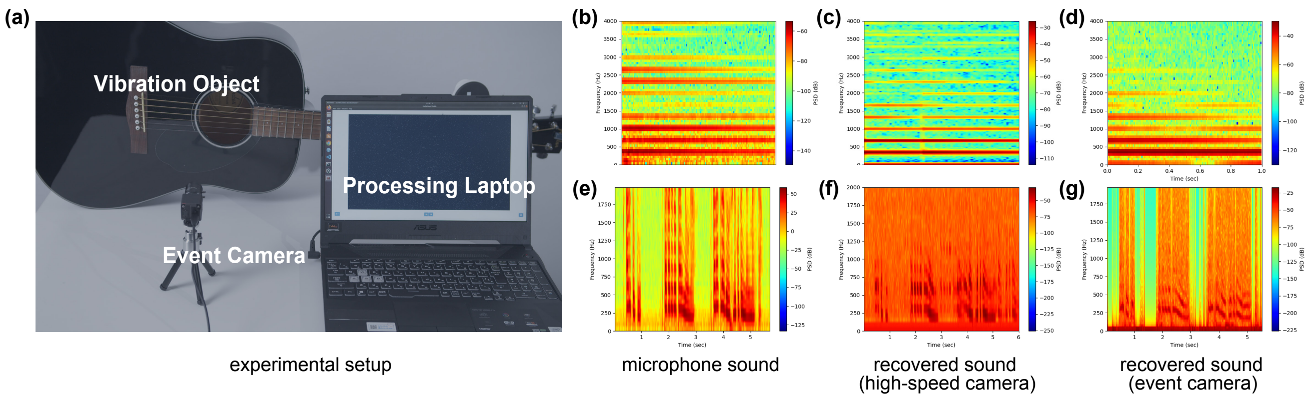

Figure 2. Sound recovery from event data. Left: The event camera efficiently captures the vibrating objects. Right: Sound waveform and spectrogram reconstructed from recorded event data.

events/sec event rate, a TAMRON AF180mm F/3.5 Di lens for object capture, and a 2x teleconverter (TC-2001) and SIGMA 150-600mm F6.3 lens for distant objects. We employed the principle analysis approach developed by Dorn et al. [2] to extract vibration from the event camera data.

Events are recorded as $e(x, y, t) = 1$ when the logarithmic change in brightness exceeds a threshold. An event indicates that the edge or pattern of an object vibrates at that location. Chronological changes in the local phase of the event data represent the movement of edges and patterns. Event cameras do not have frames, so as in Dorn et al., we define a pseudo-frame, $\hat{S}$, as an $n \times n$ two-dimensional array centered on the pixel where the event occurs, with one at the center and 0 elsewhere. Then, we compute the local response $R = \hat{S} * G_{\omega_0}$ using a single-scale complex Gabor filter for $\hat{S}$. By performing this calculation, the local phase values are computed and assigned only to the pixels in the θ direction near the event occurrence, where θ represents the parameter of Gabor filter. We perform this operation for all event data. To reduce the effects of noise, an amplitude-weighted Gaussian blur should be performed on the local response, R. This spatially blurred phase signal can be represented as

$$\hat{\phi} = \frac{A\phi * K_\rho}{A * K_\rho} \tag{1}$$

where A is the local amplitude $A = |R|$, ϕ is the raw local phase $\phi = angle(R)$, and K_ρ is the Gaussian kernel. Then, by arranging the local phase values of each pixel in a chronological order and performing linear interpolation, we can restore the vibration of the all pixel.

If we want to reconstruct sounds with higher accuracy, it can be can be achieved by recovering the sound from multiple pixels and performing signal processing on them. For more details, please refer to the supplementary information.

3. Application

When an object is the source of the sound, the sound produced by the object or the attached object can be reconstructed. For instance, our method can restore the sound of guitar strings and rod attached to speaker by capturing their events. In addition, our method can restore only the sound that we want to record even in a noisy environment, and by using a telephoto lens, we can restore the sound of a sound source several tens of meters away. Fig. 2 compares the microphone-recorded sound and the sound reconstructed by our method.

4. Visitors Experience

In the demo, visitors can experience sound reconstruction using our method, even in noisy environments and from distant sources like a guitar 10 m away. For more details, please see the video on https://youtu.be/PRXId6gMqR0.

Acknowledgement

This work was supported by JST CREST Grant Number JPMJCR19F2, including the AIP Challenge Program.

References

[1] Abe Davis, Michael Rubinstein, Neal Wadhwa, Gautham J. Mysore, Frédo Durand, and William T. Freeman. The visual microphone: Passive recovery of sound from video. *ACM Transactions on Graphics*, 33(4):1–10, July 2014. 1

[2] Charles Dorn, Sudeep Dasari, Yongchao Yang, Charles Farrar, Garrett Kenyon, Paul Welch, and David Mascareñas. Efficient Full-Field Vibration Measurements and Operational Modal Analysis Using Neuromorphic Event-Based Imaging. *Journal of Engineering Mechanics*, 144(7):04018054, July 2018. 2