

Technical Perspective

Magnifying Motions the Right Way

By Richard Szeliski

THE ABILITY TO reliably amplify subtle motions in a video is a wonderful tool for investigating a wide range of phenomena we see in the natural world. Such techniques enable us to visualize the subtle blood flow in a person's face, the rise and fall of a sleeping infant's chest, the vibrations of a bridge swaying in the wind, and even the almost imperceptible trembling of leaves due to musical notes.

The development of image processing techniques to amplify such small motions is one of the recent breakthroughs in the computational photography field, which applies algorithmic enhancement techniques to photos and videos in order to create images that could not be captured with regular photography. Some of the earlier work on this topic (originating from the same research group at MIT) used motion estimation (optical flow) techniques to recover small motions, amplify them, and then digitally warp the images. Unfortunately, optical flow techniques are very sensitive to noise, lack of texture, and discontinuities, which make this approach very brittle.

More recently, the idea of adding scaled amounts of temporal intensity differences, which the authors call the *Eulerian approach* because of its connection to fluid dynamics (which also models motion), has produced a simpler—and in many cases more robust—approach. However, this technique also amplifies noise, and it breaks down for larger amplification factors.

To see why this is the case, think of a thin line (say a telephone wire) swaying slightly in the wind. The main difference between two adjacent video frames is a darkening of the sky along one edge (where the wire is moving to) and a brightening of the pixels at the opposite edge (where the wire has moved away, revealing the brighter sky). Simply adding scaled versions of this temporal difference results in intensity clipping artifacts for large magnification factors, such as the 75x magnification the authors apply to a video of a construction

crane (which we would rightly assume to be quite rigid) swaying imperceptibly in the wind. Mathematically speaking, the phenomenon is due to the breakdown of a Taylor Series approximation of the signal for larger motions.

The solution to this dilemma, as detailed in the following paper, is to think about amplifying the various *phases* inherent in a multi-scale decomposition of the image. Each phase difference at a given frequency band, which is due to the small motion, can be independently amplified and added back into the original signal. The authors demonstrate that this results in a perfect shift for pure sinusoids.

For a multi-scale decomposition, which groups adjacent frequencies into related sub-bands, the approximation of a shift through the addition of phase-shifted signals results in much better results than the simpler linear (all-scale) difference amplification.

While this analysis is valid for amplifying the motion seen in a single pair of video frames, improved results can be obtained by combining this analysis with selective *temporal* filtering to only amplify particular vibration frequencies. The video signal is decomposed into “three-dimensional” spatio-temporal bands, and only those

bands corresponding to the particular phenomenon of interest (vibration, swaying, breathing, and so on) are amplified, which both highlights the motions being studied and drastically reduces the amplification of video noise.

The resulting spatiotemporal motion magnification algorithms can be applied to a wide range of phenomena, including blood flow and breathing, the small motions of rigid man-made structures, and even biological (inner ear) membrane vibration.

The most surprising real-world result, however, is probably the ability to recover simple audio signals (musical notes or human speech) from the visual vibrations of a plant or bag placed in the same room as the audio source. The authors call this setup the *visual microphone*. While this may sound similar to the kinds of optical microphones used to recover sound from vibrations of windowpanes, these latter approaches use optical interferometry, while the visual microphone processes regular videos. A related approach can also be used to measure physical properties of other materials such as fabrics. Details on this and many of the other techniques discussed in the paper are provided in the ample citations.

Overall, Eulerian Motion Magnification and Analysis is a delightful tour through one of the most surprising and useful developments in computational videography in the last decade. The ability to both magnify and quantify subtle visual motions from video sequences is both a testament to the mathematical sophistication of today's multi-scale video processing algorithms and to the tremendous potential of computational photography to bring us a deeper and richer understanding of real-world phenomena. 

The following paper is a delightful tour through one of the most surprising and useful developments in computational videography in the last decade.

Richard Szeliski (szeliski@fb.com) is the director and a founding member of the Computational Photography group at Facebook, Seattle, WA.

Copyright held by author.