Seminar Paper Critical Review

Computer Integrated Surgery II

Yvonne Jiang

Paper citation:

Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, William T. Freeman
Eulerian Video Magnification for Revealing Subtle Changes in the World
ACM Transactions on Graphics, Volume 31, Number 4 (Proc. SIGGRAPH), 2012

The paper that I chose for my seminar presentation and critical review is "Eulerian Video Magnification for Revealing Subtle Changes in the World"—which describes an Eulerian approach to video magnification. I chose this paper due to its potential applications to my team's project, which is Mobile Perfusion Analysis. We seek a low-cost, portable alternative to perfusion assessment, the gold standard of which is currently Laser Doppler Imaging (LDI), a technique that is not only expensive, but also highly inaccessible. The wait times to use an LDI system often result in delayed treatment for chronic wounds, which can result in amputation, and in the worst case, death. Our alternative would leverage the color amplification feature of the EVM algorithm to assess wound perfusion levels in videos taken by a mobile phone, using the subtle changes in skin color as indicators of blood flow. These subtle changes would normally be invisible to the naked eye, but with EVM, the changes are amplified and have previously been proven to be able to be used for similar applications such as measuring human pulse.

First, some background information regarding the article. To begin with, one might wonder what significance of the title "*Eulerian*" Video Magnification is. In this particular case, Eulerian refers to the Eulerian specification of a flow field in fluid dynamics. This is a frame of reference in which fluid motion is observed over time by focusing on specific locations, such as if you were to observe air currents by studying the variation of measurements recorded at fixed vane anemometers. This is as opposed to the Lagrangian specification of the flow field, in which the observer follows the fluid particle as it moves through space and time. This would be analogous to attaching a tracker to free-flying kite and observing its direction and velocity over time (this is a purely hypothetical measurement technique for the purposes of illustration and would likely be inaccurate or fail according to the laws of physics).

In terms of video analysis, Eulerian essentially corresponds to studying the variation of individual pixel values over time whereas a Lagrangian approach would utilize feature tracking to detect motion. According to the paper, the majority of previous work in motion magnification has usually involved the Lagrangian method. However, the Lagrangian approach requires explicit motion estimation, which is very computationally intensive and furthermore is not applicable to color amplification, which is the main application that my team's project would need to assess perfusion. On the other hand, the authors of the paper claim that the Eulerian approach *does* incorporate this color amplification feature in addition to being less computationally intensive to the point of being able to be run in real-time, but still remains robust.

The algorithmic approach itself is actually quite simplistic, and consists of four main steps. The first step in the process is to decompose the input video into different spatial frequencies using a Laplacian Pyramid, which is a stack of increasingly downscaled versions of the input video where a pixel in a lower level is the average of the corresponding pixels in the Laplacian of the next highest level. This allows changes over time to be observed on multiple spatial scales, which is useful when trying to separate motion amplification from color amplification, and for reducing the effects of noise (which would generally manifest as high spatial frequency motion). The algorithm facilitates this spatial filtering by taking a frequency cutoff as a parameter with which to apply a low pass spatial filter. For the purposes of human pulse color amplification, the paper suggests that we emphasize low spatial frequency changes using this filter.

The next step is to apply the same temporal bandpass filter to all levels of the decomposition, which is another algorithm of the parameter (the frequency band of interest is user specified). The paper suggests narrower bands for applications in which the temporal frequency of motion can be closely estimated, like the range of normal human heart rates, or for the vibration of guitar strings. Narrower bands are also best for color amplification, so as to reduce noise. Broadband on the other hand is better for motion amplification, especially for when the motions being observed are wider, but still subtle.

Then the resulting videos are amplified by a factor α, which is another program parameter, which is simply the magnification factor. I.e., the larger the α, the more exaggerated the motion/color amplification. This magnified motion signal is then added back to the original signal and the resulting pyramid is then collapsed back to form the output video.

The reason as to why this simple temporal processing produces motion magnification even without the explicit motion estimation used in the Lagrangian method lies in using an analysis that relies on the first-order Taylor series expansions which are, according to the paper, common in optical flow analyses. The basis for this is explained theoretically in the paper (and applied visually to a simple sinusoidal wave function for verification). The proof involves the use of first order Taylor-series approximations of the original color intensity signal at each pixel, and the Taylor-series approximation of this signal after the EVM process (when the signal has been displaced by a displacement function (representing motion/color change) at time *t*, had a bandpass filter applied to it, then amplified by the α factor and added back to the original signal). If these two approximations are compared, the functions can be rearranged to show that the processed signal simplifies to the original signal plus the displacement function multiplied by the quantity (1+α). That is, by applying the bandpass filter to the displaced signal and then amplifying, the displacement function is amplified by exactly α%.

The theoretical background checks out, but the real world performance is the more significant factor. Unfortunately there isn't much empirical evidence provided in the paper as to the EVM technique's overall performance, but they do state the set of input videos that were tested, along with the parameters used to magnify each of the videos (which are available at the MIT CSAIL website, here: http://people.csail.mit.edu/mrub/vidmag/). In general, all the paper had to say about the results of these tests were that the target motion was able to be amplified, and that to refer to the supplementary videos online to view the effects. Even in the cases where they could compare to a ground truth such as in the videos where pulse could be detected, the researchers included no observations beyond that the EVM output matched the simultaneously measured ground truth data, without describing the means by which they reached that conclusion. They do however offer some graphical evidence as to how EVM fares at detecting a pulse signal from a video of a human face under noisy conditions with and without a sufficient spatial low pass filter applied, showing that the pulse signal is only visible in the former case. Along with this they include an equation to estimate the size of the spatial filter needed to reveal the signal at a certain noise power level.

As for the Eulerian vs Lagrangian comparison, according to the paper, because the two methods take different approaches to motion—Lagrangian approaches explicitly track motions, while EVM does not—they can be used for complementary motion domains. Lagrangian approaches work better to enhance motions of fine point features and support larger amplification factors, while EVM is better suited to smoother structures and small amplifications. The paper includes a graphical representation of the results of an experiment in which EVM and a Lagrangian algorithm were run on a synthetic video with varying amounts of noise (synthetic so as to allow true motion and thus amplified motion to be objectively known, and thus each algorithm's deviation from the true magnification could be quantified). The results of their experiment objectively showed that the Eulerian-based approach fares worse as the amplification factor increases, while the Lagrangian approach fares significantly worse as pure spatial noise increases.

To summarize, the paper introduces a straightforward method that takes a video as input and exaggerates subtle color changes and imperceptible motions without performing feature tracking or optical flow computation, but merely spatio-temporal processing. This Eulerian based method, which temporally processes pixels in a fixed spatial region, presents a single framework that can be used to amplify both spatial motion and purely temporal changes, which is a feature which is not supported by previous Lagrangian methods. The color amplification of the EVM technique, which has been previously shown to amplify skin color fluctuations enough to enable pulse detection, is instrumental to our project, potentially enabling better chronic wound treatments, and more accessible perfusion assessment in general.

Before launching into criticisms, I will first mention the qualities that I found to be quite admirable about the paper. Firstly, the theoretical basis provided was solid and well explained, and worded in a way that was not laborious to parse either. All major aspects of the algorithm were deconstructed mathematically and thoroughly described in plaintext as well, along with the expected performance of the Eulerian approach against the existing Lagrangian approach. Furthermore their description of the experimental setup was very thorough, and those who wish to replicate the same tests would likely find it very straightforward to do so.

Unfortunately, this brings me to my first point of discontent: while the videos they tested with certainly produce the results they described when input to EVM, there was surely a preprocessing step before the algorithm was run. Preprocessing would be needed to take care of real world imperfections such as changing ambient light conditions, a shaky camera, or other non-ideal environments. The paper does not however discuss any of these potential variables and only touches upon preprocessing very briefly. This being the case, I felt that there was certainly not enough information to comfortably intuit what preprocessing steps should be applied to real world videos before EVM could be run to reach the desired effect.

Furthermore, as I hinted at before, the lack of concrete, numeric results describing the algorithm's performance on non-synthetic video was somewhat disconcerting. Of course, it is hard to define what a ground truth for the magnification of a real world video would be, but the fact that they failed to include even an error analysis of their pulse detection method against the ground truth pulse on those videos where they measured pulse seems to be an indication of poor form, as far as results sections go. That they included a graphical comparison of the pulse detection performance under noisy conditions with and without proper spatial filtering was somewhat mollifying, but only caused me to desire to see more comparisons of the effects of changing the program parameters. All the experiment descriptions included what parameters were used and usually an explanation as to why, but never any description, visual or textual, of what would occur if the parameters had been otherwise varied. Of course, maybe the effects can be inferred by the purpose of each parameter, but I believe that it's always more effective (and interesting) to see the effects graphically.

Lastly, I felt that, despite reading the paper several times, I failed to get a good sense of the level of impact this novel technique would really have on the scientific community, and the world at large. It of course included many mentions of the pulse measurement functionality, but there was no indication as to why using this technique to measure pulse signified an improvement over existing products, though I suppose to those with any medical background the comparison might be obvious. For me as well, since I went in with an expectation of what we would want to do with this technique, the lack of more listed applications and their significance was not as noticeable at first, though it immediately became concerning when I finally started wondering who else would ever want or need to use the technique. Clearly a need exists, or else the Lagrangian predecessor wouldn't have been invented either, but the paper sheds little light on how wide of a domain those needs might span. However significant or insignificant the method may prove to be in the future, though, I found it to be a fascinating and novel approach to an imperfectly solved existing problem, with a definite impact on my team's project goals.