

# Eulerian Video Magnification in Surveillance Systems

D. Ricanek<sup>1</sup>, P. Honec<sup>2</sup>

<sup>1</sup>Master's Degree Programme (2), FEEC, Brno University of Technology

<sup>2</sup>Department of Control and Instrumentation, FEEC, Brno University of Technology

E-mail: [xrican02@vutbr.cz](mailto:xrican02@vutbr.cz), [honecp@vutbr.cz](mailto:honecp@vutbr.cz)

**Abstract**—In this article we look at the research behind Eulerian Video Magnification, its beginnings and innovations that lead to its potential use in real time heart rate measurement. The spatial amplitude changes are exaggerated using EVM in conjuncture with face detection using the Viola-Jones algorithm to allow measurement of moving subjects. Fourier transform on the temporal signal of pooled skin pixels then yields heart rate. The algorithm is tested against movement and at different distances.

**Keywords**—Riesz transform, pyramid, phase, Eulerian video magnification, non-contact pulse measurement

## 1. INTRODUCTION

Having a network composed of millions of microscopic sensors, able to cover the surface of an object sounds sensational, yet modern digital cameras could be compared to exactly such a network.

The human eye is not well adapted to recognizing subtle spatio-temporal changes. Even though we're surrounded by an entire invisible world full of useful information. For example, the human head wobbles slightly in rhythm with the beating of the heart. Exaggerating these subtle changes is the domain of a family of algorithms called Eulerian Video Magnification (EVM) invented by Wu et al. in [1] in 2012.

Despite the ten years since its inception, EVM still suffers from major problems with accuracy when large motions occur and as such, isn't suitable to replace traditional monitoring systems e.g. in hospital environments. This work combines the Viola-Jones face detection algorithm and EVM to test its usability in surveillance system for heart rate (HR) detection and measurement in moving subjects.

## 2. PHASE BASED EULERIAN VIDEO MAGNIFICATION

The original linear EVM - which results in the color amplification seen in Fig. 1 - suffers from a low signal to noise ratio (SNR) due to its linear amplification of noise. This can be mitigated by using differences in phases of the image instead of just intensity. This way we can amplify the motion in an image up to four times as much as in the linear method while keeping the noise largely unchanged. Apart from that it also becomes possible to attenuate unwanted small motions such as jittering or movements due to breathing or incorrect stabilization of the measured patch of skin.



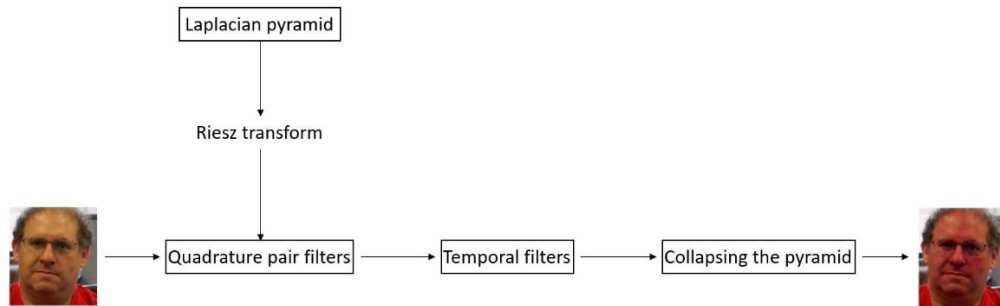
**Figure 1:** Pulsing skin color after exaggeration. Green line shows the section which has been expanded onto the time axis below the picture

However, due to the necessity of calculating quadrature pairs, the overall computational time goes up by a multiple of eight which makes the algorithm unusable in real time applications.

This problem may be dealt with by using a special type of a complex steerable pyramid filter called the Riesz pyramid.

This pyramid is composed of spatial sub bands gained from an invertible version of the Laplacian pyramid used in the original study from Wu et al. in [1], and subsequent approximation of the Riesz transform for all these sub bands.

Since Riesz transform is a special two-dimensional case of the Hilbert transform, its output is a quadrature pair of filters with a phase difference of  $90^\circ$ . The algorithm is visualized in Fig. 2.



**Figure 2:** Individual steps of the Riesz EVM

### 3. SOLUTION

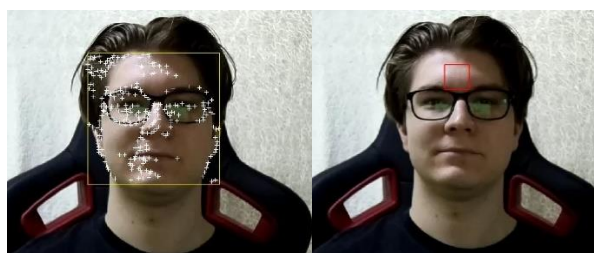
The results from Ramya a Anudev in [2] would suggest that camera choice is largely irrelevant. However, since we're testing the accuracy of measurements on subjects at larger distances, higher resolution had been chosen with Raspberry Pi HQ camera which uses a Sony IMX477 sensor that has a resolution of 12 Mpx. Additionally, the IR filter on this camera can be removed to allow measurements at night.

An oximeter Beurer PO 80 gathering timestamped data was used to obtain ground truth for comparison with HR measured by the algorithm.

The algorithm itself runs in Matlab due to the ease-of-use functions for computer vision. At the beginning the CascadeObjectDetector, which uses Viola-Jones, is applied on the video frame to detect the subjects face. The resulting significant points within the area where a face had been detected are being tracked by the PointTracker method which yields translation and rotation matrices as the frame sequence progresses. This unfortunately takes a significant amount of time and together with separating the frames from the codec makes the algorithm run far from real time.

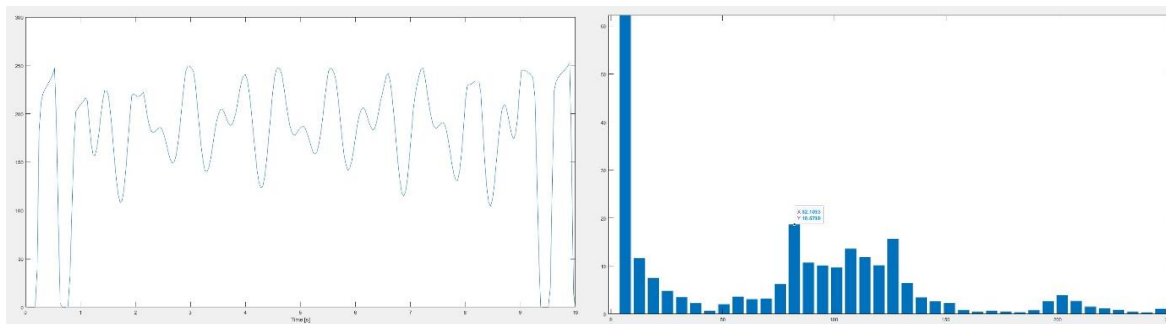
We can then apply these matrices on an arbitrary point on the face – in this case the middle of the forehead – and create a separate video of just this area; effectively stabilizing the video on the area we want to measure as shown in Fig. 3.

Of course, it is nearly impossible to create a perfectly centered video this way due to several reasons. But because the heart's pulse wave travels slow enough we can get away with just applying EVM to this cut out section. The downside is that SNR suffers tremendously as can be seen in Fig. 4 but performing EVM on a smaller video speeds up this part of the algorithm.



**Figure 3:** On the left: Face detected by the Viola-Jones algorithm and the significant points used to

track the face as it moves. On the right: Square area that is to be cut out and composed into another video. The pose of the red square is calculated together with the other significant points



**Figure 4:** On the left: Amplitude of the green channel of the pooled pixels on the subject's forehead over ten seconds. On the right: Fourier transform of this signal with the most significant harmonic being the measured HR

#### 4. RESULTS

The data set was collected under LED lighting powered by a DC power supply to prevent mains frequencies from seeping into the signal via conventional light bulbs. A total of eight measurements had been made to compare the algorithms performance under different frame rates, motion, and different distances. Additionally, two special measurements were made to see if resolution would have any effects and to compare using EVM alone on a mostly still subject.

Unless stated otherwise, the analyzed videos were 640x480 pixels and shot close to the camera (about 1 meter away) at 30 frames per second (fps).

**Table I:** EVM results compared to readings from an oximeter under various conditions

Conditions	Oximeter [BPM]	EVM [BPM]
1280x720 resolution, still subject	75.0	75.8
Still subject, 30 fps	82.0	82.4
Moving subject, 30 fps	81.0	82.7
Still subject, 60 fps	80.0	80.7
Moving subject, 60 fps	68.0	65.0
Still subject, farther away	82.0	63.0
Moving subject, farther away	81.0	65.3
Still subject, EVM applied alone	82	88.7

#### 5. CONCLUSION

Contactless measuring has made a huge step forward since the publication of the pioneering study from MIT in 2012. From experimental biology through defect detection on bridges to monitoring of breathing in sleeping infants, EVM continues to find many use cases.

Thanks to switching from linear EVM to phase based it became possible to exaggerate much smaller movements and with the use of Riesz pyramids this can be done in real time. This has opened the possibility of using EVM in surveillance systems for detecting pathologies and monitoring patients.

Such application of EVM faces some challenges when the subject moves. This article deals with enabling measurements on moving subjects by stabilizing an area of interest on the subject's forehead. This is accomplished by detecting the subject's face using the Viola-Jones object detector in Matlab and passing the extracted, smaller video of just the area on the subject's forehead to EVM for color magnification.

This results in surprisingly accurate readings at close distances even while the subject is moving their head, which wasn't possible when using just EVM on the whole video. In fact, focusing on just a part of the subject's forehead even improves the accuracy and speed of the algorithm when the subject is still.

As was expected there was no difference between the two main frame rates used by today's cameras when measuring HR. At farther distances, the algorithm begins to falter as can be seen in table I. This could be amended by using a higher resolution camera, focusing on skin detection rather than face detection or measuring HR based on the changes in the shape of the subject's head.

As for the real time application. Both EVM and Viola-Jones can be run in real time separately up to a certain resolution. However, at farther distances higher resolution may be required and further testing in this area will be necessary.

We've seen that it's possible to use EVM even on a moving subject, provided they're not too far away. To truly make this algorithm viable in surveillance systems it will be necessary to come up with ways of measuring at farther distances and preventing the signal from being drowned by noise as is almost the case in Fig. 4. Developing a solution for those two problems will be the focus of the future work.

## REFERENCES

- [1] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. 2012. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.* 31, 4, Article 65 (July 2012), 8 pages. doi:<https://doi.org/10.1145/2185520.2185561>
- [2] RAMYA MARIE, S a J ANUDEV. Response Analysis of Eulerian Video Magnification. *Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB)*. Cham: Springer International Publishing, 2019, s. 671-678. ISBN 9783030006648. ISSN 2212-9391. doi:[10.1007/978-3-030-00665-5\\_66](https://doi.org/10.1007/978-3-030-00665-5_66)
- [3] Wadhwa, Neal, et al. "Riesz pyramids for fast phase-based video magnification." 2014 IEEE International Conference on Computational Photography (ICCP). IEEE, 2014. <http://dx.doi.org/10.1109/ICCPHOT.2014.6831820>