A REPORT

ON

EULERIAN VIDEO MAGNIFICATION FOR REMOTE PHOTOPLETHYSMOGRAPHY

BY:

ANISH GHULE	2020A7PS0129P
RISHAV MUKHERJI	2020AAPS0330G
RYA SANOVAR	2020AAPS0306H

AT:

CENTRAL ELECTRONICS ENGINEERING RESEARCH INSTITUTE

(CEERI), CHENNAI

A Practice School-I Station of

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

JULY 2022

A REPORT

ON

EULERIAN VIDEO MAGNIFICATION FOR REMOTE PHOTOPLETHYSMOGRAPHY

BY:

Anish Ghule	2020A7PS0129P	Computer Science
Rishav Mukherji	2020AAPS0330G	Electronics and Communication
Rya Sanovar	2020AAPS0306H	Electronics and Communication

Prepared in partial fulfillment of the

Practice School-I Course Nos.

BITS C221/BITS C231/BITS C241

AT:

CENTRAL ELECTRONICS ENGINEERING RESEARCH INSTITUTE

(CEERI), CHENNAI

A Practice School-I Station of

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE

JULY 2022

Acknowledgement

We would like to thank the Vice Chancellor of Birla Institute of Technology and Science Pilani, Dr. Souvik Bhattacharyya and Practice School Division, BITS Pilani, who gave us the opportunity of drafting this report on 'Eulerian Video Magnification for remote photoplethysmography'.

We express our special gratitude towards our mentors, Dr. A. Amalin Prince, from BITS Pilani, and Dr. Madan Kumar Lakshmanan, from CEERI Chennai, for the invaluable guidance throughout. They have taught us the methodology to carry out the research and present it clearly. Apart from team-work, we learned many new things about the current scenario of Machine Learning techniques for image magnification, and its application to remote photoplethysmography (rPPG) for our report.

Our sincere thanks to our family for their support and motivation. A big thanks to our friends, and other contributors, who have helped us with providing access to their datasets and/or open-source code. We could finalize the report within limited time, by analysis of the data, partly available publicly, and the dataset received from CEERI Chennai even during the online Practice School-I. Last but not the least, we thank Google (Docs, Forms, Spreadsheets, Groups), Microsoft (MS Office), arXiv, as these platforms provided us convenient and handy methods of preparing our report.

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI (RAJASTHAN)

Practice School Division

Station: Central Electronics Engineering Research Institute

Centre: Chennai

Duration: Till 22nd July 2022

Date of Start: 30th May 2022

Date of submission: 22nd July 2022

Title of Project: Eulerian Video Magnification for Remote Photoplethysmography

ID/ Names/ Disciplines of students	Anish Ghule	2020A7PS0129P	Computer Science
	Rishav Mukherji	2020AAPS0330G	Electronics and Communication
	Rya Sanovar	2020AAPS0306H	Electronics and Communication

Name of Expert Dr. Madan Kumar Lakshmanan

Name of PS Faculty Dr. A Amalin Prince

Keywords: Computer Vision (CV), Remote Photoplethysmography (rPPG), Eulerian Video Magnification (EVM)

Project Areas: Computer Vision

Abstract: The goal is to reveal temporal variations in videos that are difficult or impossible to see with the unaided eye and display them in an indicative manner. The method - Eulerian Video Magnification, takes a standard video sequence as input, and applies spatial decomposition, followed by temporal filtering to the frames, for color and motion magnification. We experimented and found out the varying frequency bands and amplification factors required for different magnification tasks.

Signatures of students

Anish Ghule A.S. Ghule

Rishav Mukherji Rishav

Rya Sanovar RyaSanovar

Date: 22nd July 2022

Signature of PS Faculty

Date:

TABLE OF CONTENTS

PAGE NO.

LIST OF ILLUSTRATIONS

PAGE NO.

Abstract

The microscopic environmental changes in the surroundings are imperceptible to humans. What is impossible to perceive with the naked eye includes minor changes, such as color variations due to blood flow to the face, or motion variations, like the subtle movement of veins beneath human skin. For detecting these variations, an optical microscope aid is required to the human eye. As an alternative, cheaper and newer technologies like high-speed imagery and image processing help to detect such subtle variations. Eulerian video magnification (EVM) is the most popular technique for achieving this. This method, takes a standard video sequence as input, and applies spatial decomposition, followed by temporal filtering to the sequence of frames. Amplification of the resulting signal reveals the hidden information, resulting in color and motion magnified videos. Using this method, we can successfully visualize the flow of blood as it fills the face and also to amplify and reveal invisible motions.

INTRODUCTION

Eulerian Video Magnification is a procedure developed by MIT (Massachusetts Institute of Technology) that allows us to measure biological signals which are otherwise imperceivable to the unaided human eye. While EVM is certainly useful for unobtrusively measuring biological signals, it has a major drawback - in the fact that it can also significantly amplify noise as the magnification factor increases. The noise amplification can lead to serious enhancement issues. When using EVM with a narrow bandpass filter, we can extract any frequency within that particular passband. This can lead to the amplification of noise and the false attribution of this signal to blood flow and heart rate.

DISCUSSION

Remote photoplethysmography (rPPG)

Utility of rPPG

Remote photoplethysmography (rPPG) is a low-cost, non-contact heart rate measuring technology that can be used for telemedicine purposes. Photoplethysmography (PPG) is non-invasive technique that can be used to detect blood volume changes in the microvascular bed of tissue, often, skin surface.

rPPG uses images of a person from video feed. The technique measures the variance of RBG spectrum reflection changes from the skin, in the form of contrast between specular reflection and diffused reflection. Specular reflection is the light reflection caused by the outer skin, on the other hand, diffused reflection is the reflection that remains from the absorption and scattering in skin tissue, and it varies along with blood volume changes.

Convenience and Viability

The potential applications of rPPG include:

- Contactless vital signs monitoring
- Improved medical accessibility to remote places and low-income families
- Increased remote access to medical services

Challenges in rPPG

One of the key problems faced for carrying out rPPG is video and image quality. Issues arises from low lighting conditions, motion blur, lack of pixels in region of interest (ROI), etc. This can be countered by carrying out necessary image and video enhancement using a variety of techniques.

Image and Video Enhancement techniques

The Cartoon Animation Filter can be used create perceptually appealing motion exaggeration. This filter follows a Lagrangian perspective, which is in reference to fluid dynamics, where particles' trajectory is tracked over time. However, since they rely on highly accurate motion estimation, it is computationally expensive and difficult to make artifact-free, especially at regions of occlusion boundaries and complicated motions. [Hao-Yu W et al. 2012]

We are looking into Eulerian Video Magnification (EVM). In contrast to Lagrangian perspective; in which the frame of reference moves with the object, we utilize an Eulerian perspective which uses a fixed frame of reference to characterize fluid properties over time at each fixed location. [Neal Wadhwa et al. 2017]

Eulerian Video Magnification

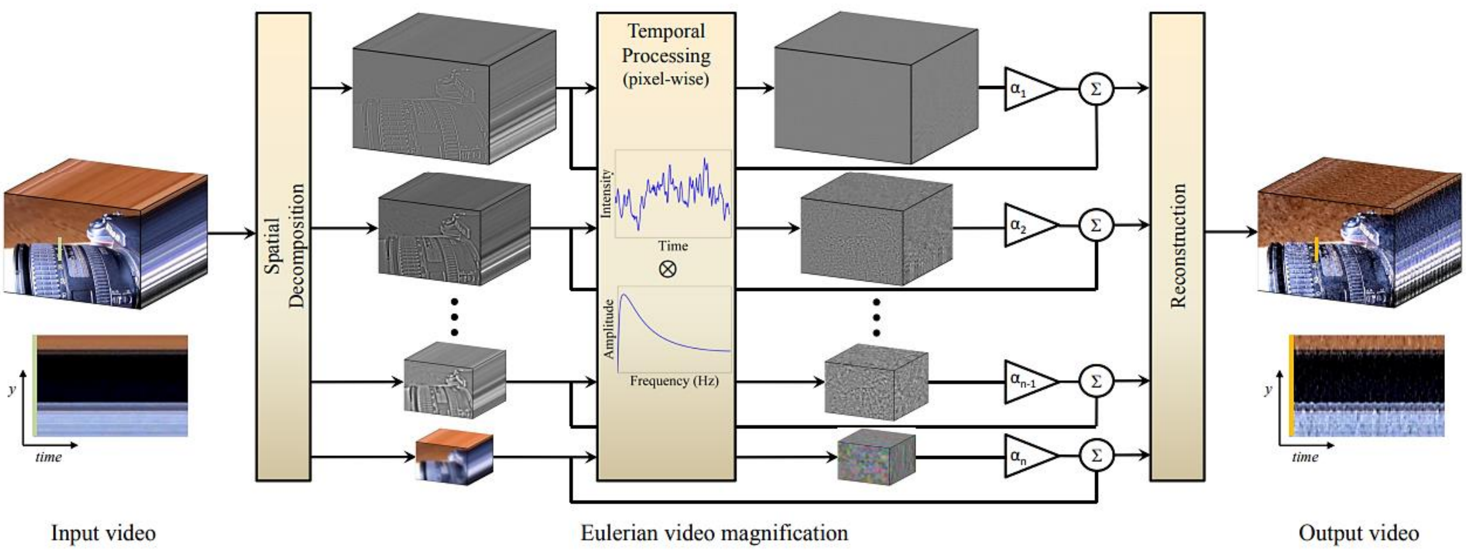

Figure 1: Overview of Eulerian Video Magnification (EVM) framework

The EVM model is described in Figure 1.the system first decomposes the input video sequence into different spatial frequency bands, and applies the same temporal filter to all bands. The filtered spatial bands are then amplified by a given factor α, added back to the original signal, and collapsed to generate the output video. The choice of temporal filter and amplification factors can be tuned to support different applications [Hao-Yu W et al. 2012].

The core idea of Eulerian video magnification is to independently process the time series of the colour values at each pixel. It is done by applying standard 1D temporal signal processing to each time series for amplifying a band of interesting temporal frequencies, for example, around 1 Hz (60 beats per minute) for colour changes and motions related to heart-rate. The new resultant time series at each pixel gives an output video. Tiny changes which were impossible to see in the input, for example, the reddening of a person's face with each heart beat get magnified and are thus clearly visible [Neal Wadhwa et al. 2017]. This is visible in the example shown in Figure 2.

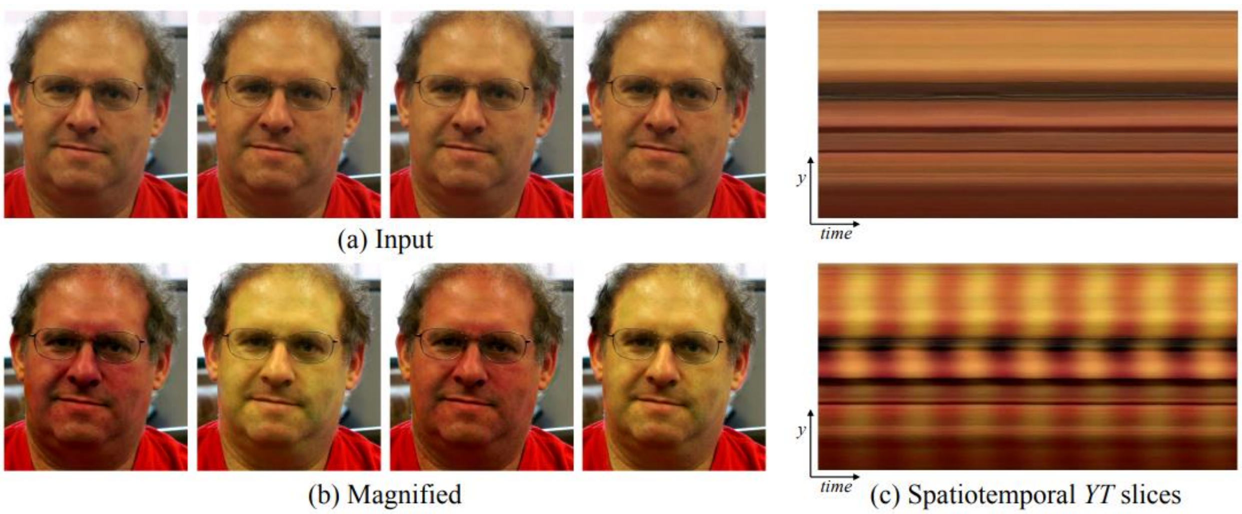

Figure 2: An example of using our Eulerian Video Magnification framework for visualizing the human heart rate. In the input sequence the signal is imperceptible, but in the magnified sequence the variation is clear.

METHODOLOGY

Mathematics of EVM

Let $I(x, t)$ be denote the image intensity at position x and time t. We can express the observed intensities with respect to a displacement function $\delta(t)$, such that $I(x, t) = f(x + \delta(t))$ and $I(x, 0) = f(x)$. The aim of EVM is to synthesize a signal which amplifies magnitude for some amplification factor α which is written as

$$\hat{I}(x, t) = f\big(x + (1 + \alpha)\delta(t)\big) \quad (1)$$

Applying a first order Taylor series expansion on $f(x + \delta(t))$ about x, we can write

$$I(x, t) \approx f(x) + \delta(t)\frac{\partial f(x)}{\partial x} \quad (2)$$

Let $B(x, t)$ be the result of applying a broadband temporal bandpass filter to $I(x, t)$ at every position x.

$$B(x, t) = \delta(t)\frac{\partial f(x)}{\partial x} \quad (3)$$

We amplify that bandpass signal by α and add it back to $I(x, t)$, resulting in the magnified signal

$$\tilde{I}(x, t) = I(x, t) + \alpha B(x, t) \quad (4)$$

Combining equations 2,3,4 we get

$$\tilde{I}(x, t) \approx f(x) + (1 + \delta)(t)\frac{\partial f(x)}{\partial x} \quad (5)$$

Assuming the first-order Taylor expansion holds for the amplified larger perturbation, the processed output is simply

$$\tilde{I}(x, t) \approx f(x + (1 + \alpha)\delta(t)). \quad (6)$$

A visual representation of the process on a single sinusoidal curve is shown in Figure 3 [Hao-Yu W et al. 2012].

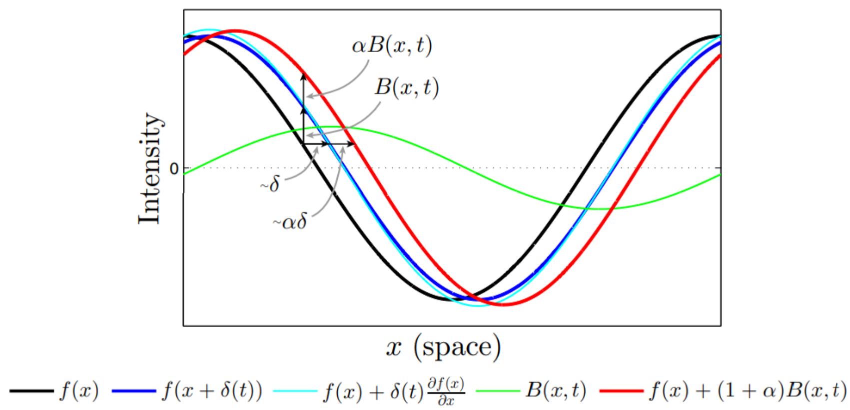

Figure 3: The input signal is shown at two time instances: $I(x,t) = f(x)$ at time t and $I(x,t+1) = f(x+\delta)$ at time $t+1$. The first-order Taylor series expansion of $I(x,t+1)$ about x approximates well the translated signal. The temporal bandpass is amplified and added to the original signal to generate a larger translation. In this example $\alpha = 1$, magnifying the motion by 100%, and the temporal filter is a finite difference filter, subtracting the two curves.

Limitations on Amplification

When the motions are small, this approach to motion magnification is robust and fast. However, for large motions, this processing can result undesired output. But it is possible to detect when this happens and suppress magnification in this case stabilizing the input video first. Limitation to how well spatio-temporal filtering can remove noise and amplified noise can cause image structures to move incoherently.

Linear amplification relies on a first-order Taylor expansion, which breaks down when either the amplification factor or the input motion is too large. For overly large amplification factors, the magnified video overshoots and undershoots the video's white and black levels causing clipping artifacts near edges where the second derivative is non-negligible thus making the initial Taylor expansion inaccurate [Neal Wadhwa et al. 2017]. Figure 4 visualizes affect of magnitude of amplification on a sinusoidal signal.

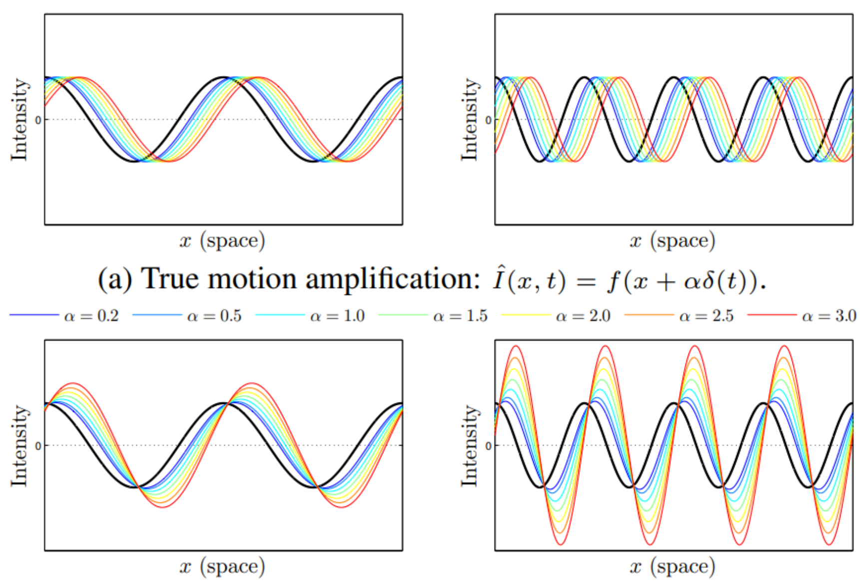

Figure 4: Illustration of motion amplification on a 1D signal for different spatial frequencies and α values.

Furthermore, another limitation on EVM is that any noise that is present in the video is also amplified. This was evident when we implemented EVM on a custom video.

Implementing EVM

As part of our research, we have implemented EVM using python (jupyter notebook) on Google Colab. The main challenge was to make sure the RAM usage was within compute limits. The official MATLAB implementation required 32GB of RAM while we only had access to 16GB. Our initial attempts constantly led to the session crashing.

We went through existing repositories of EVM implementations and tried to make it more efficient. After trying different implementations we were able to make it run and carry out colour magnification for the frequency bandwidth required for rPPG. We were able to emulate the official video as used in the actual paper but also carried out the magnification on a video taken using our laptop webcam.

We selected a temporal bandpass filter to pull out the motion or signals that needs to be amplified and the choice of the filter is dependent on the application. A narrow passband produces a more noise-free output for colour amplification of blood flow. Therefore, because they feature passbands with sharp cut off frequencies, ideal bandpass filters are employed for colour amplification.

Low order IIR filter can be useful for both colour amplification and motion magnification and are also convenient for a real-time implementation. In general, we use two first order lowpass IIR filter with cut off frequencies ω_l and ω_h to construct an IIR bandpass filter. After that desired magnification value α and spatial cut-off frequency λ_c is selected. Using higher value of α violates the bound to exaggerate specific motions or colour changes at the cost of increasing noise. This approach achieves the chrominance component of each frame by doing all the processing in the YIQ space; Y represents luma information and I and Q chrominance. The chrominance component I and Q can be attenuated before conversion to the original colour space.

In order to boost the power of the specific signal, spatial characteristics of the signal can be used to estimate the spatial filter size. For human pulse color amplification, where we seek to emphasize low spatial frequency changes, we may force $\alpha = 0$ for spatial wavelengths below λ_c. For motion magnification videos, we can choose to use a linear ramp transition for α.

Different magnification tasks require a different frequency band for spatio-temporal filter. For magnifying motion, the cutoff frequencies are set wider apart. For example, in the case of magnifying the string of a guitar ω_l and ω_h are taken to be 175Hz and 225Hz respectively. On the other hand, to magnify colour changes such as for rPPG applications we use a narrower frequency band with ω_l and ω_h equal to be 0.8Hz and 1Hz. The two filters are shown in Fig. 5

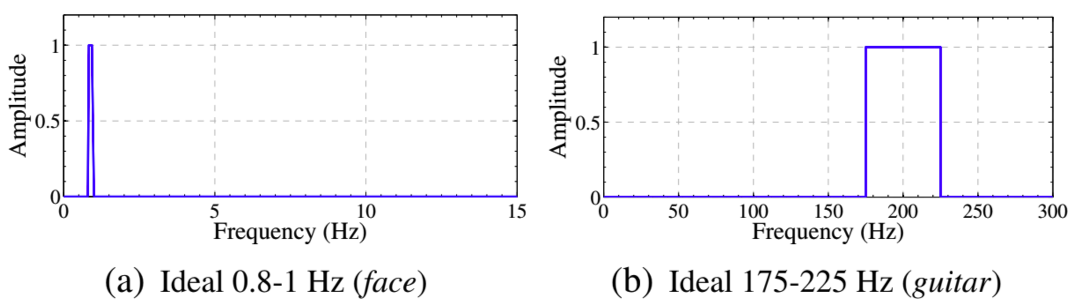

Figure 5: Different temporal filters used for varying magnification purposes.

The Eulerian magnification process is straight forward:

1. Spatial decomposition: An image sequence is decomposed into different spatial frequency bands using Laplacian pyramids
2. Band-pass filter: The time series corresponding to the value of a pixel on all levels of the pyramid are band-pass filtered to extract frequency bands of interest
3. Magnification: The extracted band-passed signals are multiplied by a magnification factor and this result is added to the original signals
4. Reconstruction: The magnified signals that compose the spatial pyramid are collapsed to obtain the final output

If the input video has multiple channels (e.g., each frame is a colour image in RGB colour space), then we can process each channel independently. The YIQ colour space is particularly suggested for Eulerian magnification since it allows to easily amplify intensity and chromaticity independently of each other.

```
def rgb2yiq(src):
    [rows,cols]=src.shape[:2]
    dst=np.zeros((rows,cols,3),dtype=np.float64)
    T = np.array([[0.299, 0.587, 0.114], [0.5959, -0.2746, -0.3213],
                  [0.2115, -0.5227, 0.3112]])
    for i in range(rows):
        for j in range(cols):
            dst[i, j]=np.dot(T,src[i,j])
    return dst
```

The first step to augment a video is to compute a Laplacian pyramid for every single frame. The Laplacian pyramid was originally proposed by Burt and Adelson in their 1983 paper The Laplacian pyramid as a compact image code, where they suggested to sample the image with Laplacian operators of many scales. This pyramid is constructed by taking the difference between adjacent levels of a Gaussian pyramid, and approximates the second derivative of the image, highlighting regions of rapid intensity change.

```
def build_laplacian_pyramid(src,levels=3):
    gaussianPyramid = build_gaussian_pyramid(src, levels)
    pyramid=[]
    for i in range(levels,0,-1):
        GE=cv2.pyrUp(gaussianPyramid[i])
        L=cv2.subtract(gaussianPyramid[i-1],GE)
        pyramid.append(L)
    return pyramid
```

We consider the time series corresponding to the value of a pixel on all spatial levels of the Laplacian pyramid. We convert this time series to the frequency domain using the Fast Fourier Transform, and apply a band pass filter to this signal. The choice of the band-pass filter is crucial. The `butter_bandpass_filter` function is used to generate a Butterworth band-pass filter of a particular order.

```
def butter_bandpass_filter(data, lowcut, highcut, fs, order=5):
    omega = 0.5 * fs
    low = lowcut / omega
    high = highcut / omega
    b, a = signal.butter(order, [low, high], btype='band')
    y = signal.lfilter(b, a, data, axis=0)
    return y
```

```
def temporal_ideal_filter(tensor,low,high,fps,axis=0):
    fft=fftpack.fft(tensor,axis=axis)
    frequencies = fftpack.fftfreq(tensor.shape[0], d=1.0 / fps)
    bound_low = (np.abs(frequencies - low)).argmin()
    bound_high = (np.abs(frequencies - high)).argmin()
    fft[:bound_low] = 0
    fft[bound_high:-bound_high] = 0
    fft[-bound_low:] = 0
    iff=fftpack.ifft(fft, axis=axis)
    return np.abs(iff)
```

After extracting the frequency band of interest, we need to amplify it and add the result back to the original signal. After amplifying the signals, all that is left is to collapse the Laplacian pyramids into a single image per frame. Notice that we can attenuate the amplification to obtain different results, or we can low-pass filter the amplified signal to reduce effects on high frequency components of the images, such as borders.

```
def reconstract_video(amp_video,origin_video,levels=3):
    final_video=np.zeros(origin_video.shape)
    for i in range(0,amp_video.shape[0]):
        img = amp_video[i]
        for x in range(levels):
            img=cv2.pyrUp(img)
        img=img+origin_video[i]
        final_video[i]=img
    return final_video
```

```
def magnify_color(video_name,low,high,levels=3,amplification=20):
    t,f=load_video(video_name)
    gau_video=gaussian_video(t,levels=levels)
    filtered_tensor=temporal_ideal_filter(gau_video,low,high,f)
    amplified_video=amplify_video(filtered_tensor,
                                  amplification=amplification)
    final=reconstract_video(amplified_video,t,levels=3)
    save_video(final)
```

RESULTS

The Eulerian Video Magnification has been implemented in Python, in the Colab environment. We used our custom test video, recorded using laptop webcam at 30 fps, with a duration of 10 seconds. Each frame in the video is of width 1280px and height 780px. Fig. 6 illustrates the results of `color_magnification` on the video with well-lighting conditions, and chrominance attenuation across 4 layers in the Gaussian blur-stack.

Figure 6: Color amplification with chrominance attenuation across 4 layers in the Gaussian pyramid (blur-stack), with bandpass of 0.83-1.0 Hz under well lighting condition

On the other hand, Fig. 7 demonstrates the results in uneven lighting on the face, and does not attenuate chrominance. With the default bandpass of 0.4-3 Hz, the high frequency components such as borders are sensitively affected. The resolution is to use a narrow passband to attenuate the amplification. We have experimentally observed that the bandpass with low 50/60 (0.833) and high 60/60 (1.0) is the best suited for face.

Figure 7: Results of color amplification without chrominance attenuation across 4 layers in the Gaussian pyramid (blur-stack), with bandpass of 0.4-3.0 Hz under uneven lighting conditions on the face

CONCLUSION

1. Eulerian Video Magnification is an experimental technique to extract microscopic signals from videographic material. This can be applied to physiologically attributed datasets to extract magnified contact-based signals such as heart rate which otherwise cannot be acquired.
2. Future work can involve automation of the filter adaptation from a wide to narrow passband, the involvement of motion capture and motion compensation and the examination of physiological signals other than heart rate.
3. Future work can also involve developing an algorithm that looks for skin pixels, to find a good region for magnification that has relatively less noise and that region alone is provided to the EVM algorithm, following which the magnified video is put through the RRPG logic. This logic makes an estimate of the pulse.
4. Elucidation of 3 cases: The above algorithm subjected to the original entire video (control), on only the magnified portion of the chosen pixels, on the entire magnified video.

BIBLIOGRAPHY

References:

1. Hau-Yu, W., Rubinstein, M., Shih, E., Guttag, J., Durand, F., & Freeman, W. (2012). Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Trans. Graph.*, *31*(4). doi:10.1145/2185520.2185561
2. Liu, L., Lu, L., Luo, J., Zhang, J., & Chen, X. (2014). Enhanced Eulerian video magnification. *2014 7th International Congress on Image and Signal Processing*, 50–54. doi:10.1109/CISP.2014.7003748
3. Wadhwa, N., Wu, H.-Y., Davis, A., Rubinstein, M., Shih, E., Mysore, G. J., … Durand, F. (2016). Eulerian Video Magnification and Analysis. *Commun. ACM*, *60*(1), 87–95. doi:10.1145/3015573

GLOSSARY

Term	Definition
Amplify	Increasing the amplitude of a time-varying signal by a given factor
Attenuation	Reduction in the strength of the given signal
Bandpass	A bandpass filter restricts signals at unwanted frequencies from getting through, while allowing the signals within a selected range of frequencies to pass
Chromaticity	An objective specification of the quality of a color, consisting hue and colorfulness (saturation), regardless of its luminance
Chrominance	Chrominance (chroma or C for short) is the signal used in video systems to convey the color information of the picture, separately from the accompanying luma signal (or Y' for short)
Colour Space	A specific organization of colors, which supports reproducible representations of color
Gaussian Pyramid	A technique in image processing that breaks down an image into successively smaller groups of pixels to blur it.
IIR	Infinite Impulse Response - a property applying to linear time-invariant systems that are distinguished by having an impulse response which does not become exactly zero past a certain point, but continues indefinitely.
Laplacian Pyramid	A linear invertible image representation which consists of a set of band-pass images, which are spaced an octave apart, plus a low-frequency residual.
Luma	The brightness of an image
Passband	The range of frequencies or wavelengths that can pass through a filter
Remote Photoplethysmography (rPPG)	A non-invasive, contactless method that monitors the blood volume changes by capturing pixel intensity changes from the skin, based on the facial video captured by a webcam, to measure pulse rate
Sinusoidal	A mathematical curve defined in terms of the sine trigonometric function
Spatio-Temporal	Relating to both space and time
Taylor Series	The Taylor series of a function is an infinite sum of terms that are expressed in terms of the function's derivatives at a single point. The function and the sum of its Taylor series can be approximated equal near this point