

A non-invasive subtle pulse rate extraction method based on Eulerian video magnification

Yang Wei¹[0000-0001-6195-8595], Nadezhda Gracheva² and John Tudor²

¹ Nottingham Trent University, Nottingham, NG11 8NS, UK

² University of Southampton, Southampton, SO17 1BJ, UK
yang.wei@ntu.ac.uk

Abstract. Measuring pulse rate by means of video recording of the wrist area is a non-invasive approach. Eulerian Video Linear Magnification (EVLN) is used in this paper to magnify, and make visible, subtle pulse-induced wrist motions. A series of experiments are conducted to investigate the performance of EVLN under various conditions, such as light intensity, background colour and a set of video recording parameters. The results show that a light intensity of around 224 to 229 lx is optimal; excess or inadequate light significantly impairs the success of amplifying the skin movement resulting from the pulse. It is demonstrated that a white background colour enables both the radial and ulnar areas to be clearly visible in the recorded video, thus improving pulse measurement. In addition, it is shown that a female's pulse strength is approximately 40 % weaker than that of a male averaged over the participants.

Keywords: video magnification, non-invasive, pulse detection.

1 Introduction

Video magnification (VM) is a technique that reveals subtle imperceptible motions by selectively amplifying them to be visible to the human eye. Video magnification has developed rapidly in the last few years with applications in visual vibrometry [1], medicine [2] and industrial engineering [3]. In [1], the material properties of visible objects are estimated by analysing subtle and otherwise imperceptible vibrations in a video and the use of VM offers a promising alternative to more specialised tools such as laser vibrometry. He *et.al.* [2] extracted the subtle pulse information from the wrist area based on a video recorded using a digital camera to predict important cardiovascular events. Wadhwa *et.al.* [3] applied VM to estimate the subtle vibrations of large objects such as a crane swaying in the wind. It was found that, in a controlled environment, the estimated vibrations (i.e. acceleration) correlate well with the results from a commercially available accelerometer. A number of papers also report human pulse magnification based on colour and motion magnification. Magnification of subtle colour changes might in addition show evenness of blood flow, which could be useful for early diagnosis of arterial problems [4]. These studies investigate the reliability of the method compared to conventional procedures, such as manual measurement and pulse

oximetry. In a video-based pulse extraction scenario, the wrist [2], the face [5] and the neck [6] are the areas typically exploited.

Eulerian Linear Video Magnification (ELVM) was first presented by Wu *et.al.* [4], where both colour variations in the facial area, and motion magnification in the wrist area, were analysed. It was demonstrated that the algorithm amplifies skin colour changes for both light and dark skin complexions. This paper also proved that the pulse rate extracted using video magnification matches the results obtained using a photoplethysmogram. He *et. al.* [6] measured the pulse transit time from colour changes at both the wrist and neck; an Arduino board and PulseSensors [7] are used with ELVM measurements. 10 subjects were tested but no results or statistics on the accuracy were reported. Miljkovic *et al.* [5] presented pulse rate measurement based on ELVM by amplifying colour changes around the facial area. Two volunteers were involved in the study and the results verified by electrocardiogram. It was demonstrated that ELVM was a reliable method for measuring pulse rate from the face but no data on accuracy was reported.

No study investigated the performance of ELVM under various ambient conditions, such as under different lighting intensities, skin complexions, camera selections and background colours. The contribution of this paper is to study the effect of these parameters on ELVM. This paper also presents an investigation of the environmental, equipment and subject variations in order to optimise ELVM based pulse amplification. Section 2 of this paper describes the background and compares different video magnification techniques. Section 3 details the experimental setup for using video magnification to acquire the pulse rate from the inner wrist area. Section 4 provides measurement results whilst varying the measurement conditions in terms of ambient light, background colour, camera type and participant gender, skin tone and texture. Finally, conclusions are given in Section 5 followed by acknowledgements and references.

2 Background

Lagrangian motion magnification was the first reported use of VM which was achieved by tracking the trajectories of pixels over time [8]. However, the computational time is long since it requires complex processing. For example, the magnification processing of a video with a resolution of 866x574 pixels, and a duration less than 30 s, requires 10 hours. In contrast, the Eulerian approach does not track motion explicitly as the input video is spatially decomposed, temporally filtered and then the motion of interest is multiplied by an amplification factor α [6]. In comparison to the Lagrangian approach, Eulerian Video Magnification (EVM) is less computationally demanding and does not use complex feature tracking procedures. In addition, EVM produces a higher quality output video and also additionally magnifies colour variations.

Table 1. Eulerian Video Magnification techniques.

Technique	Spatial decomposition	Colour space	Computational time (sec)
Eulerian Linear VM	Laplacian Pyramids	YIQ	35.6
Phase - Based VM	Riesz pyramids	YIQ	75
Phase - Based VM	Complex steerable pyramids	YIQ	325.9
Dynamic VM	Decomposition into a foreground and background	Not provided	Not provided

In Table 1, four EVM techniques are presented [9-12]. All the EVM techniques employ spatial decomposition followed by temporal filtering. However, the various EVM techniques differ in the spatial decomposition approach and whether amplitude or phase based amplification is used. Computational time grows towards the bottom of the table. ELVM has the shortest computational time, but the output video is the lowest quality with higher noise. For the aim of this research, where the pulse magnification in a wrist area is of interest, ELVM is chosen as it offers the shortest computational time and so provides a fast pulse measurement.

To implement ELVM, the raw input video is first subjected to spatial decomposition. Generally, each video frame is represented by a Laplacian pyramid in which each spatial band is temporally processed to extract the motion of interest. The extracted motion is amplified by a given amplification factor α . Finally, the amplified motion is added back to the original video and the pyramid is collapsed to reconstruct the video. The temporal filtration employs a first-order Taylor series expansion analysis [13]. The relationship between temporal processing and motion magnification is represented by a 2D matrix reported in [6], shown in Eq1.

$$f(x) + (1 + \alpha)\delta(t) \frac{\partial f(x)}{\partial x} \approx f(x + (1 + \alpha)\delta(x)) \quad (1)$$

Where x and t are position and time, $\delta(t)$ is the displacement function and α is the amplification factor. The bounds for amplification factor given motion $\delta(t)$ is therefore:

$$(1 + \alpha)\delta(t) < \frac{\lambda}{8} \quad (2)$$

For pulse magnification, the motion of interest is always a motion at relatively low spatial frequencies as the subtle motion itself is never large; this implies implicitly unconstrained parameters so that the amplification factor α is the only explicitly constrained parameter. More detailed analysis is included in [6].

3 Video magnification setup

Pulse can be measured by palpating arteries in the body. All arteries can be used for pulse measurements but it is easier to palpate the artery at particular places [14]:

- Where an artery is closer to the skin surface.
- When the artery is located just above firm tissue.

Two arteries, radial and ulnar, on the wrist can be used to measure the pulse but the radial artery is normally used for manual measurement as this location satisfies both the bulleted conditions above. However, for video magnification, the second condition is not required and therefore the ulnar artery may also be considered if it is close enough to the skin surface so that the subtle intensities changes can be captured. The original MATLAB code of Eulerian Linear Video Magnification is a free source that is available from MIT [15]. It executes C++ functions using MEX files.

3.1 Video recording parameters

Since microscopic intensity changes are to be detected and amplified in this research, the quality of the video is of great importance. The video clips of the wrist area in MOV format are used as the input due to its uncompressed quality compared to the MP4 format [16] [17].

The video clips are recorded in the laboratory using: i) a Nikon D5500 camera, and ii) an iPhone 6 camera. When using a professional camera, there are often two video standards available: i) NTSC (60Hz) [9], ii) PAL (50Hz) [18]. This difference is due to the differences in the alternating current (AC) frequency at different geographical areas in the world. This is important to note because light flickering will be clearly visible in the video if a camera is not synchronised with the AC frequency of the area where the video is recorded. Automatic flicker reduction is an option when a camera standard cannot meet the flicker requirements of the area. Since the video is recorded in the UK which is a PAL area, the flicker reduction option was used while using the NTSC video standard. The reasons for working with the NTSC standard are explained in the next paragraph.

The choice of the frame rate is based on the motion to be magnified. If more frequent motions occur, more frames per second are required in order to capture the desired signal with sufficient accuracy to produce an accurate magnification. The videos reported in [6] are recorded at 30 frames per second (fps) which is the option offered with the NTSC standard. A variety of frame rates were investigated in this research: 25 fps (PAL), 30 fps (NTSC), 50 fps (PAL) and 60 fps (NTSC); more frames require greater computational effort. This heavily affects the computational time without giving advantages in quality, so 25/30 fps are preferred over 50/60 fps. Nevertheless, 30 fps produces a result of higher quality compared to 25 fps. As a result, 30 fps NTSC was selected for this research.

A higher resolution provides more pixels and hence more detail in the area of interest. However, the computational time is much longer than with a lower resolution. For a

resolution of 640x424, a full Laplacian pyramid contains 361, 852 elements for each colour layer, whereas a full Laplacian pyramid of a frame with a resolution of 1920x1080 has 2, 774, 885 elements. It is clear that computational effort grows dramatically for a greater resolution as processing of a video with a resolution of 1920x1080 takes up to 14 times longer compared to 640x424 (from 70 to 80 seconds to 20 minutes). Hence, a resolution of 640x424 at 30 fps was chosen as an efficient video specification offering satisfactory quality.

3.2 Algorithm parameters

It is important to note that all the processing is achieved in the YIQ colour space and not RGB. The Y component represents the luminance information, and I and Q are the chrominance component on the orange-blue and purple-green axes, respectively. This colour space is a rotated version of RGB [19]. YIQ colour space is chosen to reduce the colour artefacts by attenuating the chrominance components.

The Laplacian pyramid is used for spatial decomposition as it provides access to different spatial frequencies without performing direct transformations in the frequency domain [20]. To construct a Laplacian pyramid, a Gaussian pyramid is required. A recursive process of building a Gaussian pyramid generally comprises two steps: i) low-pass filtering an image and ii) down sampling the result by a factor of 2. To obtain a Laplacian pyramid, a blurred version of the original image needs to be subtracted at each level of a Gaussian pyramid. In ELVM, a binomial filter is used to construct the Laplacian pyramid [21].

Temporal filtering is required to eliminate amplification of the motion at undesired frequencies. Such motion might be noise due to non-ideal camera sensors or involuntary subtle movements of a hand. For the pulse magnification, the most suitable filter is a relatively wide band pass at low frequencies to ensure that all the pulse frequencies fall within the pass band. A second-order infinite impulse response (IIR) band pass filter is then used to extract the pulse. In general, subtraction of two low pass filters produces a second-order band pass filter [8] and this is employed in the implementation of the ELVM. In practice, the ELVM would become extremely computationally demanding if filtering full Laplacian pyramids for every intensity of a pixel at all frames in a video clip. Therefore, an approximate IIR filter is used to minimise computational effort.

$$\text{lowpass1} = (1 - r1) * \text{pyramid} + r1 * \text{lowpass1} ; \quad (3)$$

$$\text{lowpass2} = (1 - r2) * \text{pyramid} + r2 * \text{lowpass2} ; \quad (4)$$

$$\text{filtered} = \text{lowpass1} - \text{lowpass2}; \quad (5)$$

In the above approximation, $r1$ is the parameter corresponding to an upper cut-off frequency $\omega1$ and $r2$ is the lower cut-off frequency $\omega2$ of the passband. Such an approximation imposes constraints on the frequency parameters, which should take values greater than 0 but less than 1. Under such conditions, the frequency band of 0.4 Hz to 5 Hz, used for pulse extraction, is approximated with a pair of parameters $r2 = 0.04$ and

$r1 = 0.5$. This type of approximation restricts the range of frequencies that might be extracted but is sufficient for pulse magnification, while speeding up calculations. For different sub-bands, the amplification factor might vary due to its dependency on spatial wavelength (λ) which controls the cut-off spatial frequency. The spatial wavelength is calculated for each frame and depends only on the resolution. It is clear that, with each down sampling step, λ becomes smaller by a factor of two. Hence, for each iteration, the current amplification factor α_{cur} is recalculated considering the spatial wavelength λ and calculated displacement δ . The amplification factor decreases as the size of a sub-band becomes smaller.

4 Pulse extraction results

Lighting is crucial for recording video for the purposes of pulse magnification. Inadequate or excess light will impair the ability of ELVM to capture subtle intensity changes. The experiments were all conducted with the same lighting conditions arising from the ceiling lamps in the laboratory. However, to investigate how lighting affects the performance of the algorithm, two additional devices were used: a bench fluorescent lamp LC8076 (LightCraft Magnifier Lamp) providing a consistent shadow free light source and Light meter LX - 8809A.

The additional fluorescent lamp was fixed to the bench and positioned with its tube towards the plane on which a wrist is located. The experiments investigated the following light conditions:

- No additional light source.
- An additional light source at three different distances from a wrist.
- With artificial shadow which is introduced by a solid cover above the wrist area.

The detection performance is quantified using a Successful Pulse Magnification Rate (SPMR) which is defined as follows:

$$SPMR = \frac{N_s}{N_T} \quad (6)$$

Where N_s is the number of successful detection and N_t is the total number of participants. In every experiment, the illuminance was measured using the light meter, of which the sensor was positioned in the same plane as the wrist. 10 participants took part in the experiments and the illuminance measurements with the corresponding successful pulse magnification rate (SPMR) taken as an average over all participants, are shown in the Table 3.

Table 2. Light intensity measurement at various conditions.

Lighting condition	Illuminance (lx)	SPMR _{average} (%)
With extra shadow	74 – 76	20
No extra light	224 – 229	86

Lamp at \approx 1 meter away	387 – 391	40
Lamp at \approx 50 cm away	496 – 498	20
Lamp at \approx 20 cm away	1179 – 1180	10

The results show that lighting condition is a crucial parameter and it has a great effect on the performance of ELVM. It is seen that ELVM produces 86 % correct results within a range of 224 to 229 lx, while for lighting conditions around 74 to 76 lx, the SPMR drops below 20 %. With the additional light source at less than 20 cm away, the SPMR was 0 and there is no pulse seen in the output video.

4.1 Recording background variation

Colours reflect light of different wavelengths so that the background may create distinct lighting conditions in a frame. In [6], a higher contrast background produces a higher SPMR. The hypothesis is therefore that the colour reflecting most light would produce a better environment for video magnification. It is anticipated that a white-coloured background is the most suitable one. To test this seven colours were investigated as the background as shown in the frames from each video clip in Figure 1.

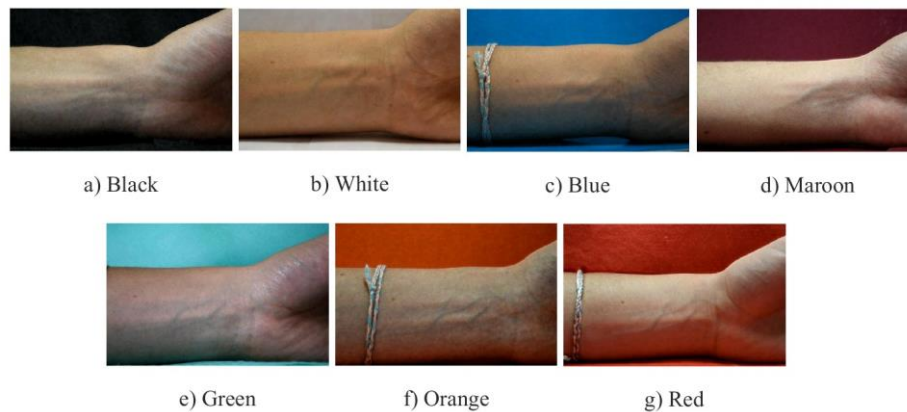


Fig. 1. Comparisons between backgrounds with different colours.

It is clear that each frame has a slightly different tone. Apart from the white background, the bottom side of the arm is covered in shadow and so is less visible and clear because of the different reflective properties of the different background colours. White is the only colour to provide a clear full wrist frame image. In Figure 1, the radial artery is located at the upper part of the wrist and clearly visible in all frames. However, the ulnar artery is located at the bottom and is in shadow for all background colours except white. For some participants, the pulse occurs only in the area of the ulnar artery and therefore using the non-white backgrounds increases the risk of missing important pulse information.

4.2 Recording camera selection

Two subjects are tested with a Nikon digital single-lens reflex (DSLR) D5500 and an iPhone 6 camera to compare processing results. The first participant, P1, has a strong pulse, which is clearly magnified under different conditions in various experiments. The second participant, P2, has a moderate pulse, which might not be magnified if the conditions are not chosen carefully.



Fig. 2. Output video frames recorded using (a) an iPhone 6 built-in camera and (b) a Nikon D5500.

Processing the iPhone 6 video with the same parameters as were successful for the Nikon does not reveal the magnified pulse. This indicates that the sensors of the phone camera do not sufficiently capture the tiny intensity changes. In addition, there is a greater noise level in the iPhone output video when compared with that of the professional Nikon camera.

4.3 Final test parameters and experiments

10 male and 5 female participants with different physical skin textures and tones took part the experiments to optimise the ELVM parameters. All the participants were tested under the conditions of: a white background, an illuminance of around 225 Lux and a distance to the camera of approximately 27 cm. The ELVM processing parameters, however, are tuned in order to investigate the performance in detail.

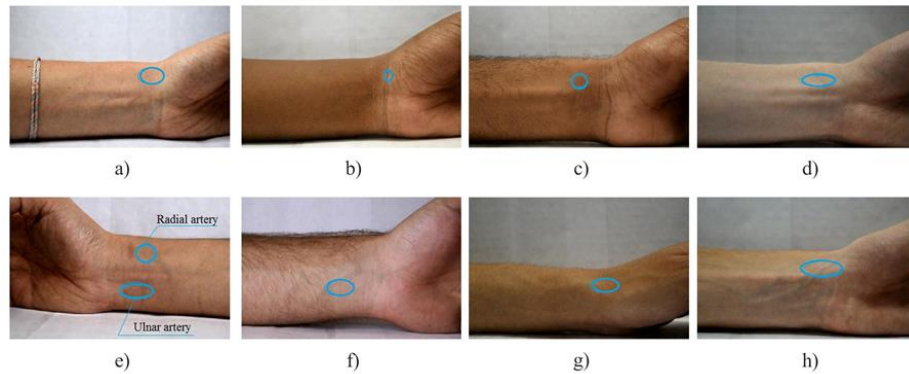


Fig. 3. Examples of areas where pulses appear to be magnified on different participants.

The pulses have different strengths and generally female participants have a weaker pulse than males. Two magnified pulses are demonstrated in Figure 4.

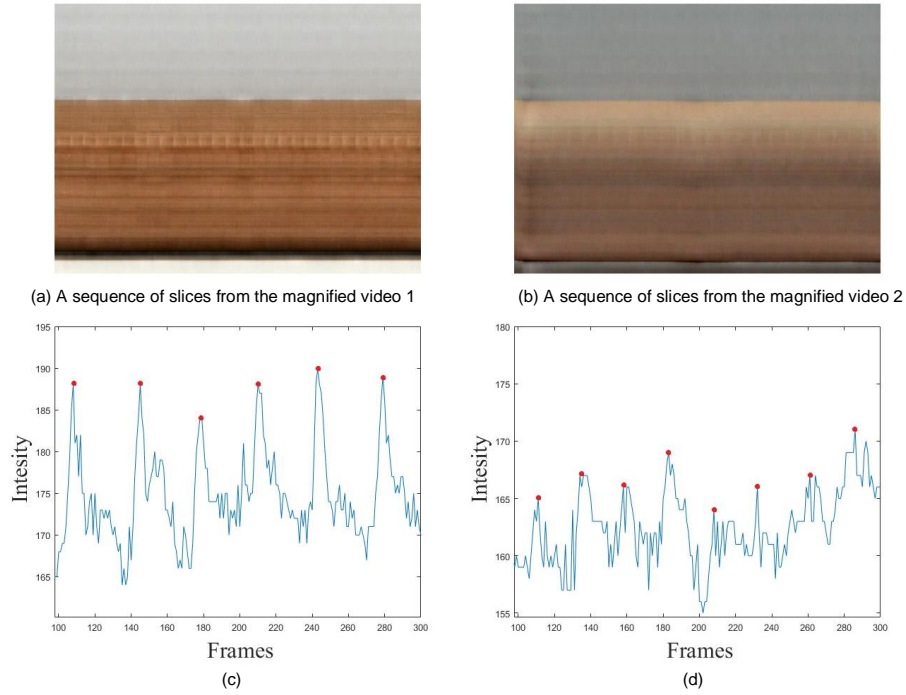


Fig. 4. Magnified pulse of a male participant (a,c) and a female participant (b,d).

A magnified pulse of a male is shown in Figure 4 a) and the pulse as a pixel intensity series of a single pixel is given in Figure 4 c). Peaks that correspond to the pulse are clearly visible in Figure 4 c). Such a pulse is defined as strong as it is clearly magnified under various conditions with different processing parameters. The maximum-to-minimum difference for this intensity series is 25. A pulse of a female participant is demonstrated in b) and d). It is seen, in a sequence of slices and in the intensity series, that this pulse is weaker. The maximum-to-minimum difference for this series is 15, which is 40% less than that for the male pulse in c). These graphs demonstrate the generally observed difference in the strength of male and female participants. To validate this information, the whole group of participants is summarised in Table 3.

Table 3. Summary of pulse strength.

Category	Participants		Percentage (%)	
	Male	Female	Male	Female
Weak	0	2	0	40
Moderate	3	2	30	40

Strong	7	1	70	20
--------	---	---	----	----

The performance of the algorithm did not depend on complexion but only on the strength of the pulse. The use of ELVM is proven to work for both light and dark skin tones. The magnification pulse is visible at one of the arteries or at both.

The results also showed that Body Weight Index (BMI) also has an influence on the success of the ELVM algorithm as the pulse rates of those participants with a BMI higher than 25 were difficult to detect which is believed to be due to both the radial and ulnar arteries being covered by excess body fat.

5 Conclusions

The Eulerian Linear Video Magnification (ELVM) algorithm was evaluated by implementing it in MATLAB with various operation parameters and conditions for measuring the pulse rate from the inner wrist area of a human subject. A video clip from a camera was used as the source and the pulse at the radial artery was successfully amplified to reveal the subtle pulse variation. For operation parameters, environmental light intensity and the background colour were evaluated as well as the type of camera. The results show that a light intensity between 224 and 229 lx, with a white colour background, provide the highest magnification rate when the video is taken using a professional DSLR camera. It is also shown that the successful magnification rate on male subjects is much higher than on female subjects due to the weaker pulse strength of females.

6 Acknowledgement

We acknowledge support for this project as part of the partnership resource of the EPSRC Interdisciplinary Research Collaboration 'SPHERE' - a Sensor Platform for Healthcare in a Residential Environment. EPSRC grant number: EP/K031910/1.

7 References

1. A. Davis, K. Bouman, J. Chen, M. Rubinstein, O. Buyukozturk, F. Durand and W. Freeman, "Visual vibrometry: estimating material properties from small motions in video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 732-745, 2017.
2. X. He, R. Goubran and X. Liu, "Wrist pulse measurement and analysis using Eulerian video magnification," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, Las Vegas, US, 2016.
3. N. Wadhwa, M. Rubinstein, F. Durand and W. Freeman, "Phase based video motion processing," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 80, 2013.
4. N. Miljkovic and D. Trifunovic, "Pulse rate assessment: Eulerian video magnification vs. electrocardiography recordings," in *12th Symposium on Neural Network Applications in Electrical Engineering (NEUREL)*, 2014.

5. X. He, R. Goubran and X. Liu, "Using Eulerian video magnification framework to measure pulse transit time," in IEEE International Symposium on Medical Measurements and Applications (MeMeA), 2014.
6. H. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," ACM Transactions on Graphics (TOG), vol. 31, no. 4, 2012.
7. "Pulsesensor," [Online]. Available: <https://pulsesensor.com>. [Accessed 14.12.2018].
8. C. Liu, A. Torralba, W. Freeman, F. Durand and E. Adelson, "Motion magnification," ACM Transactions on Graphics (TOG), vol. 24, pp. 519-526, 2005.
9. M. Elgharib, M. Hefeeda, F. Durand and W. Freeman, "Video magnification in presence of large motions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
10. L. Liu, L. Lu, J. Luo, J. Zhang and X. Chen, "Enhanced Eulerian video magnification," in Image and Signal Processing (CISP), 2014.
11. N. Wadhwa, M. Rubinstein, F. Durand and W. Freeman, "Riesz pyramids for fast phase-based video magnification," in IEEE International Conference on Computational Photography (ICCP), 2014.
12. N. Wadhwa, H. Wu, A. David, M. Rubinstein, E. Shih, G. Mysore, J. Chen, O. Buyukozturk, J. Guttag and W. Freeman, "Eulerian video magnification and analysis," Communications of the ACM, vol. 60, pp. 87-85, 2016.
13. B. Horn and B. Schunck, "Determining optical flow: a retrospective," Artificial Intelligence, vol. 59, pp. 81-87, 1993.
14. W. Nichols, M. Rourke and C. Vlachopoulos, McDonald's blood flow in arteries: theoretical, experimental and clinical principles, CRC Press, 2011.
15. H. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand and W. Freeman, "Video magnification," [Online]. Available: URL <http://people.csail.mit.edu/mrub/vidmag>. [Accessed 03.05.2018].
16. K. Rao, D. Kim and J. Hwang, Video coding standards, Netherlands: Springer, 2014.
17. T. Sikora, "MPEG digital video-coding standards," IEEE Signal Processing Magazine, pp. 82-100, 1997.
18. B. Evans, Satellite communication systems, IET, 1999.
19. A. Ford and A. Roberts, "Colour space conversions," Westminster University, 1998. [Online]. Available: <http://www.photo-lovers.org/pdf/coloureq.pdf>. [Accessed 03.05.18].
20. P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," IEEE Transactions on Communications, vol. 31, pp. 532-540, 1983.
21. E. Kupce and R. Freeman, "Binomial filters," Journal of Magnetic Resonance, vol. 99, pp. 644-651, 1992.
22. J. Suh and Y. Kim, Accelerating MATLAB with GPU Computing: A Primer with Examples, Elsevier Inc., 2014.