

Higher Order of Motion Magnification for Vessel Localisation in Surgical Video

Mirek Janatka¹, Ashwin Sridhar, John Kelly, and Danail Stoyanov¹

Wellcome / EPSRC Centre for Interventional and Surgical Sciences, University College London, United Kingdom

¹Department of Computer Science, University College London, United Kingdom
`mirek.janatka@ucl.ac.uk`

Abstract. Locating vessels during surgery is critical for avoiding inadvertent damage, yet vasculature can be difficult to identify. Video motion magnification can potentially highlight vessels by exaggerating subtle motion embedded within the video to become perceivable to the surgeon. In this paper, we explore a physiological model of artery distension to extend motion magnification to incorporate higher orders of motion, leveraging the difference in acceleration over time (jerk) in pulsatile motion to highlight the vascular pulse wave. Our method is compared to first and second order motion based Eulerian video magnification algorithms. Using data from a surgical video retrieved during a robotic prostatectomy, we show that our method can accentuate cardio-physiological features and produce a more succinct and clearer video for motion magnification, with more similarities in areas without motion to the source video at large magnifications. We validate the approach with a Structure Similarity (SSIM) and Peak Signal to Noise Ratio (PSNR) assessment of three videos at an increasing working distance, using three different levels of optical magnification. Spatio-temporal cross sections are presented to show the effectiveness of our proposal and video samples are provided to demonstrate qualitatively our results.

Keywords: Video Motion Magnification, Vessel Localisation, Augmented Reality, Computer Assisted Interventions

1 Introduction

One of the most common surgical complications is due to inadvertent damage to blood vessels. Avoiding vascular structures is particularly challenging in minimally invasive surgery (MIS) and robotic MIS (RMIS) where the tactile senses are inhibited and cannot be used to detect pulsatile motion. Vessels can be detected by using interventional imaging modalities like fluorescence or ultrasound (US) but these do not always produce a sufficient signal, or are difficult to use in practice [1]. Using video information directly is appealing because it is inherently available, but processing is required to reveal any vessel information hidden within the video and is not apparent to the surgeon, as can be seen in the right image of Fig.1.

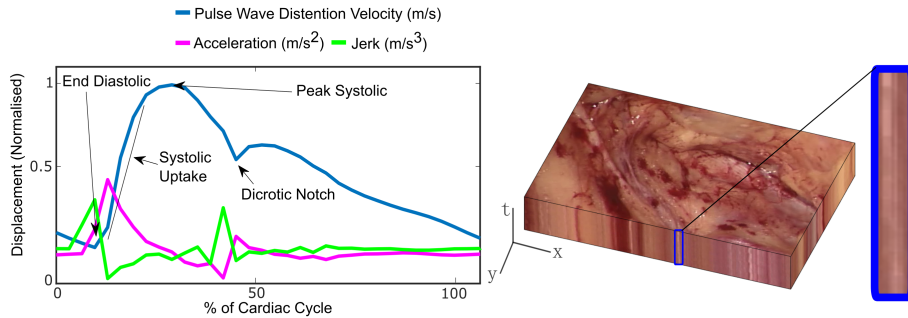


Fig. 1: (Left): The vessel distension-displacement from the pulse wave, with the higher order derivatives along with annotation of the corresponding cardio-physiological stages. Down sampled to 30 data points to reflect endoscope frame rate acquisition. (1-D Virtual Model of arterial behaviour [2]) (Right): Endoscopic video image stack. The blue box surrounds an artery with no perceivable motion, shown by the vertical white line in the cross section

The cardiovascular system creates a pressure wave that propagates through the entire body and causes an equivalent distension-displacement profile in the arteries and veins [3]. This periodic motion has intricate characteristics, shown in Fig. 1 (left), that can be highlighted by differentiating the distension-displacement signal. The second order derivative outlines where the systolic uptake is located, whilst the third derivative highlights the end diastolic phase and the dicrotic notch. This information can be present as spatio-temporal variation between image frames and amplified using Eulerian video magnification (EVM). EVM could be applied to endoscopic video for vessel localisation by using an adaptation of an EVM algorithm and showing the output video directly to the surgeon [4]. Similarly, EVM can aid vessel segmentation for registration and overlay of pre-operative data [5], as existing linear based forms of the raw magnified video can be abstract and noisy to use directly within a dynamic scene. Magnifying the underlying video motion can exacerbate unwanted artifacts and unsought motions, and in this case regarding surgical video, of those which are not the blood vessels but due to respiration, endoscope motion or other physiological movement within the scene.

In this paper, we propose to utilise features that are apparent in the cardiac pulse wave, particularly the non-linear motion components that are emphasised by the third order of displacement, known as jerk (Green plot Fig. 1, left). We devise a custom temporal filter and use an existing technique for spatial decomposition of complex steerable pyramids [6]. The result is a more coherent magnified video compared to existing lower order of motion approaches [7,8], as the high magnitudes of jerk are prominently exclusive to the pulse wave in the surgical scene, as our method avoids amplification of residual motions due to respiration or other periodic scene activities. Quantitative results are difficult for

such approaches but we report a comparison to previous work using Structure Similarity [9] and Peak Signal to Noise Ratio (PSNR) of three robotic assisted surgical videos at separate optical zoom. We provide a qualitative example of how our method achieves isolation of two cardio-physiological features over existing methods. A supplementary video of the magnifications is provided that further illustrates the results.

2 Methods

Building on previous work in video motion magnification [8] [7] [10] we set out to highlight the third order motion characteristics created by the cardiac cycle. In an Eulerian frame of reference, the input image signal function is taken as $I(\mathbf{x}, t)$ at position \mathbf{x} ($\mathbf{x} = (x, y)$) and at time t [10]. With the linear magnification methods, $\delta(t)$ is taken as a displacement function with respect to time, giving the expression $I(\mathbf{x}, t) = f(\mathbf{x} + \delta(t))$ and is equivalent to the first-order term in the Taylor expansion:

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

This Taylor series expansion appropriation can be continued into higher orders of motion, as shown in [8]. Taking it to the third order, where $\hat{I}(\mathbf{x}, t)$ is the magnified pixel at point \mathbf{x} and time t in the video.

$$\hat{I}(\mathbf{x}, t) \approx f(\mathbf{x}) + (1+\beta)\delta(t) \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} + (1+\beta)^2\delta(t)^2 \frac{1}{2} \frac{\partial^2 f(\mathbf{x})}{\partial^2 \mathbf{x}} + (1+\beta)^3\delta(t)^3 \frac{1}{6} \frac{\partial^3 f(\mathbf{x})}{\partial^3 \mathbf{x}} \quad (2)$$

In a similar vein to [8], we equate a component of the expansion to an order of motion and isolate these by subtraction of the lower orders

$$I(\mathbf{x}, t) - I(\mathbf{x}, t)_{non-linear(2^{nd} order)} - I(\mathbf{x}, t)_{linear} \approx (1+\beta)^3\delta(t)^3 \frac{1}{6} \frac{\partial^3 f(\mathbf{x})}{\partial^3 \mathbf{x}} \quad (3)$$

assuming $(1+\beta)^3 = \alpha$, $\alpha > 0$.

$$D(\mathbf{x}, t) = \delta(t)^3 \frac{1}{6} \frac{\partial^3 f(\mathbf{x})}{\partial^3 \mathbf{x}} \quad (4)$$

$$\hat{I}_{non-linear(3^{rd} order)}(\mathbf{x}, t) = I(\mathbf{x}, t) + \alpha D(\mathbf{x}, t) \quad (5)$$

This produces an approximation for for the input signal and a term that can be attenuated in order to present an augmented reality (AR) view of the original video.

2.1 Temporal Filtering

As jerk is the third temporal derivative of the signal $\hat{I}(\mathbf{x}, t)$, a filter has to be derived to reflect this. To achieve acceleration magnification, the Difference of Gaussian (DoG) filter was used [8]. This allowed for a temporal bandpass to be

assigned, by subtracting two Gaussian filters, using $\sigma = \frac{r}{4\omega\sqrt{2}}$ [11] to calculate the standard deviations of them both, where r is the frame rate of the video and ω is the frequency under investigation. Taking the derivative of the second order DoG we create an approximation of the third order, which follows Hermitian polynomials [12]. Due to the linearity of the operators, the relationship between the the jerk in the signal and the third order DoG as:

$$\frac{\partial^3 I(\mathbf{x}, t)}{\partial t^3} \otimes G_\sigma(t) = I(\mathbf{x}, t) \otimes \frac{\partial^3 G_\sigma(t)}{\partial t^3} \quad (6)$$

2.2 Phase-based Magnification

In the classical EVM approach, the intensity change over time is used in a pixel-wise manner [10] where a second order IIR filter detects the intensity change caused by the human pulse. An extension of this uses the difference in phase w.r.t spatial frequency [7] for linear motion, as subtle difference in phase can be detected between frames where minute motion is present. Recently, phase-based acceleration magnification has been proposed [8]. It is this methodology we utilise and amend for jerk magnification. By describing motion as phase shift, a decomposition of the signal $f(x)$ with displacement $\delta(t)$ at time t , the sum of all frequencies (ω) can be shown as:

$$f(\mathbf{x} + \delta(t)) = \sum_{\omega=-\infty}^{\infty} A_\omega e^{i\omega(\mathbf{x} + \delta(t))} \quad (7)$$

where the global phase for frequency ω for displacement $\delta(t)$ is $\phi_\omega = \omega(\mathbf{x} + \delta(t))$.

It has been shown that spatially localised phase information of a series of image over time is related to local motion [13] and has been leveraged for linear magnification [7]. This is performed by using complex steerable pyramids [14] to separate the image signal into multi-frequency bands and orientations. These pyramids contain a set of filters $\Psi_{\omega_s, \theta}$ at multiple scales, ω_s and orientations θ . The local phase information of a single 2D image $I(\mathbf{x})$ is

$$(I(\mathbf{x})) \otimes \Psi_{\omega_s, \theta}(\mathbf{x}) = A_{\omega, \theta}(\mathbf{x}) e^{i\phi_{\omega_s, \theta}(\mathbf{x})} \quad (8)$$

Where $A_{\omega, \theta}(\mathbf{x})$ is the amplitude at frequency ω and orientation θ , and where $\phi_{\omega_s, \theta}$ is the corresponding phase at scale (pyramid level) ω_s . The phase information is extracted ($\phi_{\omega_s, \theta}(\mathbf{x}, t)$) at a given frequency ω , orientation θ and frame t . The jerk constituent part of the motion is filtered out with our third order Gaussian filter and can then be magnified and reinstated into the video ($\hat{\phi}_{\omega, \theta}(\mathbf{x}, t)$) to accentuate the desired state changes in the cardiac cycle, such as the dicrotic notch and end diastolic point, shown in Fig. 1 (left).

$$D_\sigma(\phi_{\omega, \theta}(\mathbf{x}, t)) = \phi_{\omega, \theta}(\mathbf{x}, t) \otimes \frac{\partial^3 G_\sigma(t)}{\partial t^3} \quad (9)$$

$$\hat{\phi}_{\omega, \theta}(\mathbf{x}, t) = \phi_{\omega, \theta}(\mathbf{x}, t) + \alpha D_\sigma \phi_{\omega, \theta}(\mathbf{x}, t) \quad (10)$$

Phase unwrapping is applied as with the acceleration methodology in order to create the full composite signal [8][15].

3 Results

To demonstrate the proposed approach, endoscopic video was captured from robotic prostatectomy using the da Vinci surgical system (Intuitive Surgical Inc, CA), where a partially occluded obturator artery could be seen. Despite being identified by the surgical team the vessel produced little perceivable motion in the video. This footage was captured at 1080p resolution at 30Hz. For processing ease, the video was cropped to a third of the original width, which contained the motion of interest, yet still retains the spatial resolution of the endoscope. The video was motion magnified using the phase-based complex steerable pyramid technique described in [7] for first order motion and the video acceleration magnification described in [8] offline for comparison. Our method appended the video acceleration magnification method. All processes use a four level pyramid and half octave pyramid type. For the temporal processing, a bandpass was set at 1Hz \pm 0.1 to account for a pulse around 54 to 66 bpm. From the patient's ECG reading, their pulse was stable at 60 bpm during video acquisition. This was done at three magnification factors (x2, x5, x10). Spatio-temporal slices were then taken at a site along the obturator artery for visual comparison of each temporal filter type. For a quantitative comparison, the Peak Noise to Signal Ratio (PNSR) and Structural Similarity (SSIM) index [9] was calculated on a hundred frame sample, comparing the magnified videos to their original equivalent frame.

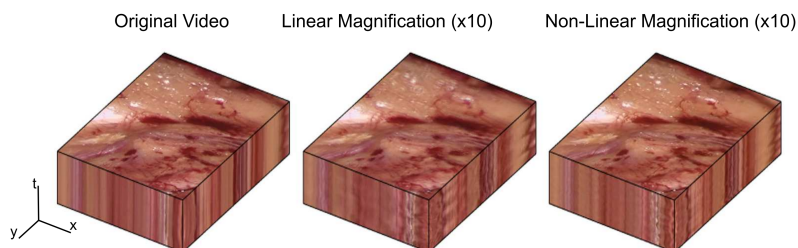


Fig. 2: Volumetric image stacks of an endoscopic scene under different types of magnification.

Fig. 2 shows an apprehensible overview of our video magnification investigation. The pulse from the external iliac artery can be seen in the right corner and the obturator artery on the front face. Large distortion and blur can be observed on the linear magnification example, particularly in the front right corner, where as this is not present on the non-linear example, as change in velocity is exaggerated, where as any velocity is exaggerated in the linear case. Figure 3 displays a magnification comparison of spatio-temporal slices taken from three different for mentioned magnification methods. E and G in this figure, demonstrates the improvement in pulse wave motion granularity using jerk has in temporal processing, compared to the lower orders. The magenta in E shows a periodic saw

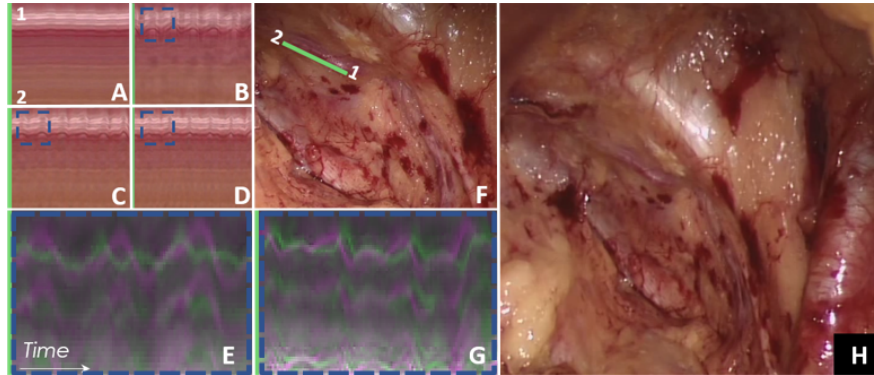


Fig. 3: Motion Magnification of the obturator artery (x10). (a) Unmagnified spatio-temporal slice (STS); (b) Linear magnification [7]; (c) Acceleration magnification [8]; (d) Jerk Magnification (our proposal); (e),(g) Comparative STS, blue box from (d) (jerk) in green, with (b) in magenta in (e) and (c) in magenta in (g); (f) Sample site (zoomed); (h) Overview of the surgical scene.

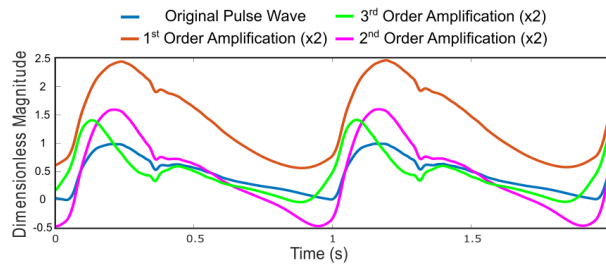


Fig. 4: 1D distension-displacement pulse wave signal amplification, using virtual data [2]. The jerk magnification shown in green creates two distinct peaks that is not present in the other two methods of lower order.

wave, with no discerning features relating to the underlying pulse wave signal. The magenta in G that depicts the use of acceleration shows a more bipolar triangle wave. The green in both E and G shows a consistent periodic twin peak, with the second being more diminished, which suggests that our hypothesis of a jerk temporal filter being able to detect the dicrotic notch as correct and comparable to our model analysis shown in Fig. 4. Table 1 shows a comparison of a surgical scene at three separate working distances. This was arranged to diminish the spatial resolution with the same objective in the endoscope. All three aforementioned magnification algorithms were used on each at three different motion magnification (α) factors (x2, x5, x10).

As a comparative metric, SSIM and PSNR are used as a quantitative metric, with PSNR being based on mathematical model and SSIM taking into account characteristics of the human visual system [9]. SSIM and PSNR allow for objec-

α	Assessment	Scene Level 1			Scene Level 2			Scene Level 3		
		Linear	Acc.	Jerk	Linear	Acc.	Jerk	Linear	Acc.	Jerk
x_2	<i>PSNR</i>	34.95	34.7	35.65	33.87	34.33	35.31	34.41	35.02	35.31
	<i>SSIM</i>	0.94	0.95	0.96	0.93	0.94	0.96	0.95	0.96	0.96
x_5	<i>PSNR</i>	30.6	31.5	33.18	28.98	31.05	33	30.32	31.88	33
	<i>SSIM</i>	0.88	0.9	0.93	0.85	0.89	0.92	0.9	0.92	0.92
x_{10}	<i>PSNR</i>	27.76	28.94	30.56	25.92	28.41	30.43	27.75	29.36	30.43
	<i>SSIM</i>	0.82	0.85	0.88	0.78	0.83	0.87	0.85	0.86	0.88

Table 1: Results from SSIM analysis and PSNR for our surgical videos at three levels of magnification across the different temporal processing approaches.

tive comparisons of a processed image to a reference source, whilst it is expected that a magnified video to be altered, the residual noise generation by the process can be seen by these proposed methods. SSIM is measured in decibels (db), where the higher the number the better the quality is. PSNR is a percentile reading, with 1 being the best possible correspondence to the reference frame. For the all surgical scene, our proposed temporal process of using jerk out performs the other low order motion magnification methods across all magnifications for SSIM and equals or outperforms the acceleration technique, particularly at $\alpha=10$.

4 Conclusion

We have demonstrated that the use of higher order motion magnification can bring out subtle motion features that are exclusive to the pulse wave in arteries. This limits the amplification of residual signals present in surgical scenes. Our method particularly relies on the definitive cardiovascular signature characterized by the twin peaks of the end diastolic point and the dicrotic notch. Additionally, we have shown objective evidence that less noise is generated when used within laparoscopic surgery compared to other magnification technique, however, a wider sample and case specific examples would be needed to verify this claim. Further work will look at a real-time implementation of this approach as well as methods of both ground truth validation and subjective comparison within a clinical setting. Practical clinical use cases are also needed to verify the validity of using such techniques in practice and to identify the bottlenecks to translation.

Acknowledgements

The work was supported by funding from the EPSRC (EP/N013220/1, EP/N027078/1, NS/A000027/1) and Wellcome(NS/A000050/1).

References

1. Ashwin N Sridhar, Archie Hughes-Hallett, Erik K Mayer, Philip J Pratt, Philip J Edwards, Guang-Zhong Yang, Ara W Darzi, and Justin A Vale. Image-guided robotic interventions for prostate cancer. *Nature Reviews Urology*, 10(8):452, 2013.
2. Marie Willemet, Phil Chowienczyk, and Jordi Alastruey. A database of virtual healthy subjects to assess the accuracy of foot-to-foot pulse wave velocities for estimation of aortic stiffness. *American Journal of Physiology-Heart and Circulatory Physiology*, 309(4):H663–H675, 2015.
3. Jordi Alastruey, Kim H Parker, Spencer J Sherwin, et al. Arterial pulse wave haemodynamics. In *11th International Conference on Pressure Surges*, pages 401–442. Virtual PiE Led t/a BHR Group: Lisbon, Portugal, 2012.
4. A Jonathan McLeod, John SH Baxter, Sandrine de Ribaupierre, and Terry M Peters. Motion magnification for endoscopic surgery. 9036:90360C, 2014.
5. Alborz Amir-Khalili, Ghassan Hamarneh, Jean-Marc Peyrat, Julien Abinahed, Osama Al-Alao, Abdulla Al-Ansari, and Rafeef Abugharbieh. Automatic segmentation of occluded vasculature via pulsatile motion analysis in endoscopic robot-assisted partial nephrectomy video. *Medical image analysis*, 25(1):103–110, 2015.
6. Eero P Simoncelli and Edward H Adelson. Subband transforms. In *Subband image coding*, pages 143–192. Springer, 1991.
7. Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman. Phase-based video motion processing. *ACM Transactions on Graphics (TOG)*, 32(4):80, 2013.
8. Yichao Zhang, Silvia L Pinteá, and Jan C van Gemert. Video acceleration magnification. *arXiv preprint arXiv:1704.04186*, 2017.
9. Zhou Wang, Ligang Lu, and Alan C Bovik. Video quality assessment based on structural distortion measurement. *Signal processing: Image communication*, 19(2):121–132, 2004.
10. Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. 2012.
11. Krystian Mikolajczyk and Cordelia Schmid. Indexing based on scale invariant interest points. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 525–531. IEEE, 2001.
12. Bart M Haar Romeny. *Front-end vision and multi-scale image analysis: multi-scale computer vision theory and applications, written in mathematica*, volume 27. Springer Science & Business Media, 2008.
13. David J Fleet and Allan D Jepson. Computation of component image velocity from local phase information. *International journal of computer vision*, 5(1):77–104, 1990.
14. Javier Portilla and Eero P Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, 40(1):49–70, 2000.
15. Daichi Kitahara and Isao Yamada. Algebraic phase unwrapping along the real axis: extensions and stabilizations. *Multidimensional Systems and Signal Processing*, 26(1):3–45, 2015.