Surgical Video Motion Magnification with Suppression of Instrument Artefacts

Mirek Janatka[1], Hani J. Marcus[1,2], Neil L. Dorward[2], and Danail Stoyanov[1]

[1] Wellcome/EPSRC Centre for Interventional and Surgical Sciences, University
College London, London, UK
[2] Department of Neurosurgery, The National Hospital for Neurology and
Neurosurgery, Queen Square, London, UK
mirek.janatka@ucl.ac.uk

Abstract. Video motion magnification can make blood vessels in surgical video more apparent by exaggerating their pulsatile motion and could prevent inadvertent damage and bleeding due to their increased prominence. It could also indicate the success of restricting blood supply to an organ when using a vessel clamp. However, the direct application to surgical video could result in aberration artefacts caused by its sensitivity to residual motion from the surgical instruments and would impede its practical usage in the operating theatre. By storing the previously obtained jerk filter response of each spatial component of each image frame - both prior to surgical instrument introduction and adhering to a Eulerian frame of reference - it is possible to prevent such aberrations from occurring. The comparison of the current readings to the prior readings of a single cardiac cycle at the corresponding cycle point, are used to determine if motion magnification should be active for each spatial component of the surgical video at that given point in time. In this paper, we demonstrate this technique and incorporate a scaling variable to loosen the effect which accounts for variabilities and misalignments in the temporal domain. We present promising results on endoscopic transnasal transsphenoidal pituitary surgery with a quantitative comparison to recent methods using Structural Similarity (SSIM), as well as qualitative analysis by comparing spatio-temporal cross sections of the videos and individual frames.

Keywords: Motion Magnification · Surgical Visualisation · Augmented Reality · Computer Assisted Interventions · Image Guided Surgery

1   Introduction

In endoscopic surgery visualising blood vessels is a common challenge as they are often beneath the tissue surface or indistinctive from the surface texture. Major complications can result from instruments causing inadvertent damage and hence bleeding [1] due to vessel imperceptibility that leads to surgical error. In severe cases, such bleeding can place the patient at risk of death if it cannot be controlled or potentially lead to post-operative problems, that would

require additional surgery to address [2]. The major challenge and cause for such problems in endoscopic procedures is that subsurface vessels cannot be visualised directly or detected through touch and palpation [3]. Various schemes for detecting and avoiding subsurface vasculature have been explored including Augmented Reality (AR) that fuses preoperative patient imaging with the surgical video, as well as novel optical imaging [4] and intraoperative ultrasound [5]. Similar schemes have also been used to determine if blood flow continues to pulse into a region, after being obstructed by a clamping mechanism [6]. How-
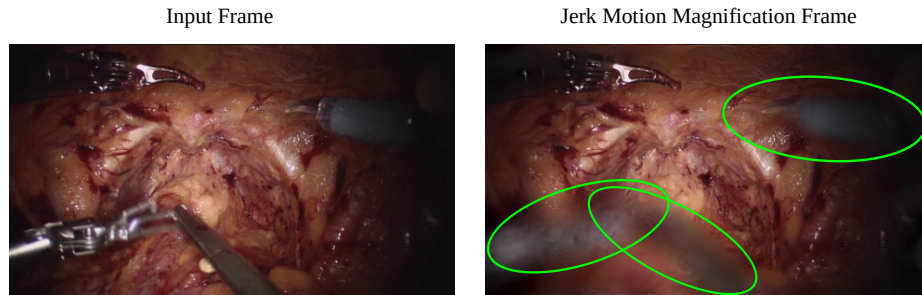
Input Frame   Jerk Motion Magnification Frame

Fig. 1. Demonstrating aberrations from surgical instrument motions using video motion magnification via frame comparison. Left) Original video frame. Right) Aberrations generated jerk-based motion magnification, outlined by the green elliptical annotation [15]

ever, all have some drawbacks in terms of surface registration accuracy, workflow additions, ergonomic problems and signal sensitivity [7]. Hence practically, the problem persists and can be a significant hurdle to the successful completion of many procedures [8]. Video motion magnification (VMM) [9–11] has previously been proposed as a mechanism to aid vessel localisation in endoscopic video, without the requirement of additional *in situ* hardware, surface registration or contrast agents [12, 14, 15]. VMM uses the existing variations in the endoscopic video stream that are minute and out of the range of the surgeon's perception and creates a synthetic video where such motions are perceivable. The generated video characteristics can be temporally selected, so that reoccurring motion within a certain temporal frequency range can be processed exclusively, which allows for motions of known occurrence to be selected [9, 10]. In the case of vessel localisation, it is possible to isolate and amplify motion due to the heart rate (readily available in the OR), that governs the periodicity of vessel distension from the pressure wave that is generated from the heart. VMM assumes spatial consistency, effectively treating every point on the image as a time series, and requires a static view for initialisation. This creates limitations and while VMM has been demonstrated in endoscopic third ventriculostomy and robotic prostatectomy videos, it has not been effectively clinically translated or adopted [12].

VMM has also been suggested for other clinical applications, such as with video otoscopy [13].

One challenge for VMM in surgery is that other motions within the scene can cause aberrational distortions and can be disorientating to surgical view. An alternative method of using VMM was proposed in robotic partial nephrectomy, where respiratory motion is present and needs be accounted for. Rather than use VMM directly a colour map representation was generated from the raw VMM video of where pulsation was used to located vessels as an aid to assist in registration of preoperative patient data [14]. Attempts to deal with large motion presence in motion magnified videos have been suggested by segmentation which is not practical in an operation [19, 20]. Yet, recent developments in temporal filtering have allowed for different components of motion, such as acceleration and jerk, to be selected based upon a principal oscillation frequency. This allows for VMM to leverage a higher band pass of frequencies than the band under investigation to exaggerate motion within video [15–18]. In our previous work, we used the third order of motion (jerk) was utilised to exclude motions from respiration and transmitted motion from larger arteries, whilst motion magnifying blood vessels in surgical video, with a filter designed around the pulse rate. This approach reduced blur distortion caused by the large motions within the scene, whilst still permitting motion magnification of vessel distension, as the pulsatile waveform contains jerk characteristic [15]. However, this jerk filter is not able to prevent the generation of motion blur aberrations from the presence of instruments moving within the scene, making the synthesised video unsuitable for surgical guidance (as shown in Fig.1). For VMM to be a viable option for surgical intervention it must be usable with instrument motion presence or its function would be limited to just observational usage.

In this paper, we propose a technique that would allow for motion magnification to be left unaffected by the introduction of tool motion to the surgical camera's field of view. To operate, it simply requires the known heart rate and a brief sample of the view that is uninterrupted by instrument motion. It maintains the spatial consistency assumption of a fixed view point. We demonstrate this method on four cases of endoscopic video of transnasal transsphenoidal surgery, providing both qualitative and quantitative comparison to a prior method.

2   Methods

2.1   Motion Magnification

VMM operates by spatially decomposing video frames into local frequency components using Complex Steerable Pyramid (CSP) [21, 22] which uses different Gabor-like wavelets (ψ) to decompose images at varying scales and orientations representations notated by S.

$$\tilde{I}(\mathbf{x}, t) = \sum S(\mathbf{x}, t) * \psi + \epsilon(\mathbf{x}, t) \tag{1}$$

Where $\tilde{I}(\mathbf{x}, t)$ represents a CSP reconstructed video frame at time t, with intensity values at $\mathbf{x} = (x, y)$ pixel value, $*$ notes a convolution operation. As well as the band-pass of S a high-pass and low-pass residual ($\epsilon(\mathbf{x}, t)$) that is unalterable is also required for the reconstruction. As the information held in $S(\cdot)$ are complex conjugates representing local frequency information, local phase can be attained. As the local motion is related to the local phase ($\phi(\mathbf{x}, t)$) via the Fourier shift theorem, motion analysis and modulation can be performed by filtering and manipulating the local phase of the video over time. As shown in our previous work [15], higher order of motion magnification utilises a temporally tuned third order Gaussian derivative to detect jerk motion $D(\mathbf{x}, t)$ within a certain pass-band of temporal frequencies. It exploits the linear relationship of convolution to gather third order of motion from local phase using a third order Gaussian derivative:

$$D(\mathbf{x}, t) := \frac{\partial^3 G_\sigma(t)}{\partial t^3} * \phi(\mathbf{x}, t) = \frac{\partial^3 \phi(\mathbf{x}, t)}{\partial t^3} * G_\sigma(t) \tag{2}$$

Where σ is the standard derivation of the Gaussian (G) derivative and is defined $\sigma = \frac{fr}{4\omega}$ [23], where ω is the temporal frequency of interest, in this case the heart rate from the electrocardiogram. fr denotes the sampling rate of the video. The Gaussian derivative convolution is applied to the time series for all scales and orientations of each pixel ($\mathbf{X}$) from the CSP representation. By detecting local jerk motion in a video, it can be exaggerated by an amplification factor α to generate $\hat{S}(\cdot)$.

$$\hat{S}(\mathbf{x}, t) = A(\mathbf{x}, t) e^{i(\phi(\mathbf{x}, t) + \alpha D(\mathbf{x}, t))} \tag{3}$$

Where A is the amplitude and ϕ is the phase of that particular local frequency at $\mathbf{x}$ at time t with respect to the S band's orientation and scale. The summations of which from the various scales and orientations can reconstruct the motion magnified frame $\hat{I}(\mathbf{x}, t)$.

2.2   Tool Motion Artefact Suppression Filter

Assuming the endoscope is statically positioned and there are no surgical instruments within the scene, the response of the jerk filter applied to the video feed can be anticipated as:

$$D(\mathbf{x}, t) = D(\mathbf{x}, t \bmod \frac{1}{\omega_c}) \tag{4}$$

Where mod is the modulo function and $\frac{1}{\omega_c}$ is the time period of the cardiac cycle, reported from the electrocardiogram. ω_c is also the ω value used to determine $G(t)$ from Eq. 2. However, in reality, due to sampling quantization and subtle variations in the heart rate, perfectly aligned repetition rarely occurs. Therefore these values are used as a guide for creating the filter with an offset range, which can then be loosened by a scaling factor. Taking these stored tool free scene readings to be $TL(\mathbf{x})$:

$$TL(\mathbf{x}) = D(\mathbf{x}, 0), D(\mathbf{x}, 1), \ldots D(\mathbf{x}, \frac{1}{\omega_c}) \tag{5}$$

To generate the offset range for the filter, the variability of $TL(\mathbf{x})$ can be found as $R(\mathbf{x})$:

$$R(\mathbf{x}) = \frac{max(TL(\mathbf{x})) - min(TL(\mathbf{x}))}{2} \tag{6}$$

This allows the comparator filter to be:

$$D(\mathbf{x}, t \bmod \frac{1}{\omega_c}) \pm \beta R(\mathbf{x}) \tag{7}$$

Where β is a scaling factor that can widen the filter further. To operate within the VMM, the comparator has to act as a switch so that the motion magnification effect can be deactivated. Therefore we define the state of $\chi(\mathbf{x}, t)$ as:

$$1 = [D(\mathbf{x}, t) < D(\mathbf{x}, t \bmod \frac{1}{\omega_c}) + \beta R(\mathbf{x})] \wedge [D(\mathbf{x}, t) > D(\mathbf{x}, t \bmod \frac{1}{\omega_c}) - \beta R(\mathbf{x})]$$
$$0 = else \tag{8}$$

Where $\wedge$ is a logic *and* function. This consideration can be shown in the CSP magnified band S as $\hat{S}(\cdot)^*$, with the threshold value β being pre-assigned with χ_β.

$$\hat{S}(\mathbf{x}, t)^* = A(\mathbf{x}, t)e^{i(\phi(\mathbf{x},t) + \alpha\chi_\beta(\mathbf{x},t)D(\mathbf{x},t))} \tag{9}$$

Which can be used to reconstruct $\hat{I}(\mathbf{x}, t)^*$, a motion magnified video without blur distortion from tool motion.

2.3 Synthetic Example

To better visualise how the tool motion artefact suppression filter (TMASF) works, Fig. 2 shows a demonstration on a synthetic arterial displacement profile [24], shown as a phase reading that could be taken from a single pixel from an arbitrary $S(\cdot)$, where vessel motion exists. After just under two cardiac cycles, a tool passes across the point, denoted by "tool motion" that alters the phase reading. The reading then returns to that of the vessel motion as before. The response of this displacement profile from the jerk filter [15] is shown in the middle plot of Fig. 2 in green. After a single cardiac cycle, the TMASF can be constructed, the bounds of which are shown in red. So long as the jerk filter response stays within these bounds, $\chi(\cdot)$ is equal to 1 and the amplification is performed.

However, if the jerk filter response moves outside these bounds, $\chi(\cdot)$ is equal to 0 for that particular pixel around that moment and the amplification effect is nullified. This can be seen around the time the tool motion is present. The bottom plot of Fig. 2 shows the resultant amplified signal, using both the TMASF filter (dashed black curve) and without (magenta curve). As shown, the amplification generates tool motion artefacts on prior method where tool motion is present. The TMASF reduces such artefacts, but is not immune to them. The extent to which they are created can be reduced by decreasing the β value, which would essentially bring the red lines in Fig. 2 closer together. Yet, this could also be detrimental to the amplification effect, due to the imperfections in regularity and sampling.
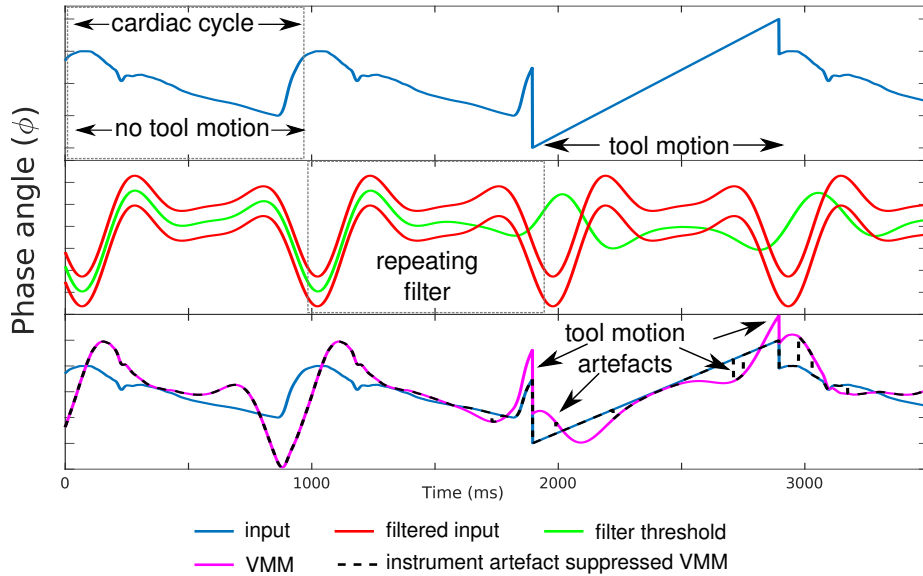
Fig. 2. Synthetic one dimensional explanation of the instrument motion suppression filter. The top plot shows the observed cardiac motion with a tool passing over it. The middle plot shows the filtered component of the motion signal and the generation of the suppression filter. The bottom plot shows the difference the suppression has on the motion magnification output signal.

3   Results

3.1   Experimental Setup

To verify if the TMASF works in surgical videos, we performed a proof of concept study using retrospective data (IDEAL Stage 0) [25], applying the filter to a series of patient cases (n=4) that underwent endoscopic transnasal transsphenoidal surgery. The study was registered with the local institutional review board, patients provided their written consent, and videos were fully anonymised. Each video was processed with the TMASF at three β threshold scalings ($\beta = 1$, 3 and 5) and without, using the previous jerk motion magnification without the TMASF [15], using 8 orientations and quarter octave CSP, from hereafter referred to as VMM.

All samples have a brief few seconds before surgical instruments are visible in the scene for the filter to initialise. This would be a reasonable prerequisite requirement and perhaps automatically possible with a surgical robot as it is aware when the system is stationary. For all videos the magnification factor was set to $\alpha = 10$. The resolution of the capture videos were 1280 x 720, however were cropped to a 720 x 720 pixel square to account for the visible area of the video. The videos were scaled down by a half for quicker offline processing.

Acquisition of the endoscope was at 25 fps. For quantitative analysis and comparison Structure Similarity Image Matrix index (SSIM) was used to compare the magnified frames to the corresponding input frames. The closer to 1 the similar the frame is to the input, meaning the less noise or motion magnification effect has been generated. To verify that the motion magnification effect is still operational, comparative spatio-temporal cross sections from places of interest are compared across all videos. Additionally, all videos are supplied in the supplementary material.

3.2   Results and Analysis

As a comparative example of how the TMASF functions, the frame-wise SSIM result for each scaling factor of the TMASF and VMM can be seen in Fig. 3, that depicts the result from case four, and is a common representation for all cases. The "tool motion" region shows how the various β values of the TMASF
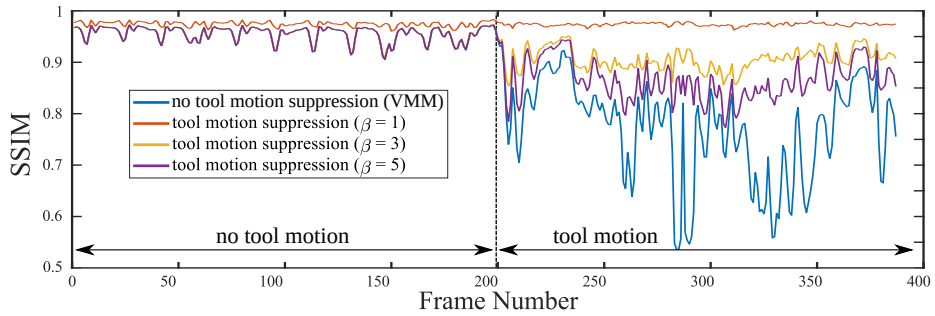

Fig. 3. Frame-wise comparison of SSIM for all four motion magnified videos (Case 4)

perform. The VMM curve shows the impact tool movement has on the generated video, with SSIM values dropping to as low as 0.55 and are volatile for this duration, with rarely reaching the lowest reading from the "no tool motion" region. Similarly to the prior "no tool motion" section, TMASF $\beta = 1$ shows that the motion magnification is impinged by the suppression filter at this β level. The other two β values show a drop in SSIM readings, compared to the "no tool motion" region, however are not as severe as the VMM curve, with $\beta = 5$ reaching an offset of -0.05 compared to $\beta = 3$ at most. These results suggest that the TMASF works, on β levels 3 and higher. To report a fuller quantitative performance of the TMASF, box plots from the SSIM reading of the entire videos are shown on the right bottom corner of Fig.4. A running trend of the order of the medians from all cases for the different scaling values can be seen, with VMM (N/A) being lowest, followed by $\beta = 5$, then $\beta = 3$ and highest being $\beta = 1$. The images in Fig.4 shows qualitative comparisons of each of the four cases with select frames from VMM and TMASF $\beta = 3$ to the input video. Next to each
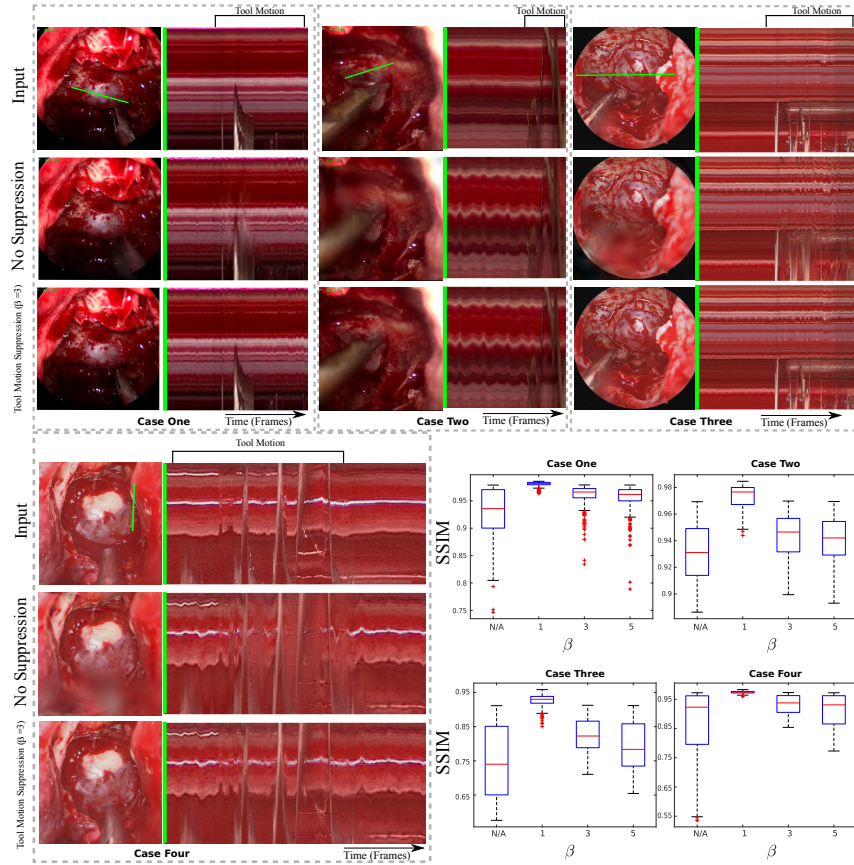
Fig. 4. Qualitative (input, VMM and TMASF $\beta = 3$ - green line on image frame indicates sample site for spatio-temporal cross section) and quantitative (using SSIM: VMM and TMASF $\beta = 1,3$ and 5) comparisons of all four cases.

frame is a spatio-temporal cross section, that is taken from each respective video (indicated by a green line) and shows how the pixels change in that location over time. By looking at the frames from VMM and TMASF $\beta = 3$ it can be seen that blur distortion is reduced and that the TMASF makes the video clearer to view. The spatio-temporal cross sections show that physiological motion magnification is present in all cases, however for the TMASF structure can still be seen where tool motion is present (similar to the image in the input cross section), whilst is lost in the VMM cross section. This suggests that TMASF successfully reduces aberrations from VMM videos whilst retaining the desired motion magnification effect of exaggerating the motion of the physiology. In general, the results suggest that there is a trade-off between impinging the motion magnification effect and permitting aberration occurring from instrumentation motion, depending on the β value used with TMASF.

4   Discussion

In this paper, we have proposed a filter constructed from local phase information collected prior to the insertion of surgical tools into the surgical field of view for motion magnification in endoscopic procedures. This approach can prevent aberration from being created due to instrument motion and hence allow more clinically usable surgical motion magnification augmentation, such as critical structure avoidance and assistance in vessel clamping.

We have shown that the proposed filter can reduce amplification of motion due to tools on example videos from endoscopic neurosurgery where the camera and surgical site are confined and do not move too much. This is an important consideration because our filter would need re-initialising if the endoscope is moved or after large changes to the surgical scene. Yet this could be performed quickly, as the initialisation period is the length of a heart beat but more work is needed to detect and automate any re-initialisation strategy.

The application of the TMASF is causal and is possible to combine with a real-time system as it is only reliant on past information. Further work is needed to consider user studies investigating how augmented visualisation of motion magnification can be presented to the clinical team to access the risk of cognitive overload and inattention blindness [26]. Both the motion magnification factor α and TMASF β value are variables that can be tuned to create the optimal synthesized video, however, it is not essential that the raw video is used alone. TMASF could assist in existing visualisation approaches to only show physiological motion [14]. Additionally, the ability to segregate physiological and non-physiological motion from a surgical scene could assist in surgical instrument tracking and anatomical segmentation tasks, as well as in other modalities of medical imaging that utilise motion estimation.

4.1   Acknowledgements

References

1. Stanley A Feldman and Vincent Marks. The problem of haemorrhage during anaesthesia and surgery. *Anaesthesia*, 16(4):410–431, 1961.
2. M Marietta, L Facchini, P Pedrazzi, S Busani, and G Torelli. Pathophysiology of bleeding in surgery. In *Transplantation proceedings*, 38, pages 812–814. Elsevier, 2006.
3. Danail Stoyanov Surgical Vision *Annals of biomedical engineering*, 40(2):332–345. Springer, 2012.

4. Neil T Clancy, Geoffrey Jones, Lena Maier-Hein, Daniel S Elson and Danail Stoyanov Surgical spectral imaging *Medical Image Analysis* 101699. Elsevier, 2020.
5. Kevin Cleary and Terry M Peters. Image-guided interventions: technology review and clinical applications. *Annual review of biomedical engineering*, 12:119–142, 2010.
6. Pedro Norat, Sauson Soldozy, Mazin Elsarrag, Jennifer Sokolowski, Kaan Yagmurlu, Min S Park, Petr Tvrdik and M Yashar S Kalani Application of Indocyanine Green Videoangiography in Aneurysm Surgery: Evidence, Techniques, Practical Tips. *Fountiers in surgery*, 6:34. Frontiers Media SA, 2019.
7. René Ludwig Bernays. Intraoperative imaging: current trends, technology and future directions. *Transsphenoidal surgery. Philadelphia: Saunders*, pages 56–69, 2010.
8. Aryeh Shander. Financial and clinical outcomes associated with surgical bleeding complications *Surgery*, 142(4):20–25, Elsevier, 2007.
9. Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T. Freeman. Phase-based video motion processing. *ACM Transactions on Graphics*, 32(4):1–10, ACM New York, 2013.
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. *ACM Transactions on Graphics*, 31(4):1–8, ACM New York, 2012.
11. Anh Cat Le Ngo and Raphaél C-W Phan. Seeing the invisible: Survey of video motion magnification and small motion analysis. *ACM Computing Surveys (CSUR)*, 52(6):1–20, 2019.
12. Jonathan A McLeod, John SH Baxter, Sandrine de Ribaupierre, and Terry M Peters. Motion magnification for endoscopic surgery. In *Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling*,9036: 90360C. International Society for Optics and Photonics, 2014.
13. Mirek Janatka, Krishan S Ramdoo, Taran Tatle, Kirttin Pachtrachai, Daniel S Elson and Danail Stoyanov Examining in vivo tympanic membrane mobility using smart phone video-otoscopy and phase-based Eulerian video magnification *Medical Imaging 2017: Computer-Aided Diagnosis*, 10134:101341Y. International Society for Optics and Photonics, 2017.
14. 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.
15. Mirek Janatka, Ashwin Sridhar, John Kelly, and Danail Stoyanov. Higher order of motion magnification for vessel localisation in surgical video. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 307–314. Springer, 2018.
16. Yichao Zhang, Silvia L Pintea, Jan C Van Gemert. Video acceleration magnification In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 529–537, 2017.
17. Shoichiro Takeda, Kazuki Okami, Dan Mikami, Megumi Isogai, Hideaki Kimata. Jerk-aware video acceleration magnification In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1769–1777, 2018.
18. Shoichiro Takeda, Yasunori Akagi, Kazuki Okami, Megumi Isogai, Hideaki Kimata. Video magnification in the wild using fractional anisotropy in temporal distribution In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1614–1622, 2019.

19. Mohamed Elgharib, Mohamed Hefeeda, Fredo Durand, and William T Freeman. Video magnification in presence of large motions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4119–4127, 2015.
20. Julian FP Kooij and Jan C van Gemert. Depth-aware motion magnification. In *European Conference on Computer Vision*, pages 467–482. Springer, 2016.
21. 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.
22. William T Freeman and Edward H Adelson. Steerable filters. In *Optical Society of America Conference on Image Understanding and Machine Vision*, pages 114–117, 1989.
23. Tony Lindeberg. Scale-space for discrete signals. *IEEE transactions on pattern analysis and machine intelligence*, 12(3):234–254, 1990.
24. 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):663–675. American Physiological Society, 2015.
25. Art Sedrakyan, Bruce Campbell, Jose-G Merino, Richard Kuntz, Allison Hirst, Peter McCulloch IDEAL-D: a rational framework for evaluating and regulating the use of medical devices, *Bmj*,353:i2372. British Medical Journal Publishing Group, 2016.
26. Hani J Marcus, Philip Pratt, Archie Hughes-Hallett, Thomas P Cundy, Adam P Marcus,Guang-Zhong Yang, Ara Darzi and Dipankar Nandi Comparative effectiveness and safety of image guidance systems in neurosurgery: a preclinical randomized study, *Journal of neurosurgery*, 123(2):307–313. American Association of Neurological Surgeons, 2015.