Automatic Segmentation of Occluded Vasculature via Pulsatile Motion Analysis in Endoscopic Robot-Assisted Partial Nephrectomy Video

Alborz Amir-Khalili^{a,*}, Ghassan Hamarneh^b, Jean-Marc Peyrat^c, Julien Abinahed^c, Osama Al-Alao^d, Abdulla Al-Ansari^{c,d}, Rafeef Abugharbieh^a

^a*BiSICL, University of British Columbia, Vancouver, BC, Canada*

^b*Medical Image Analysis Lab, Simon Fraser University, Burnaby, BC, Canada*

^c*Qatar Robotic Surgery Centre, Qatar Science & Technology Park, Doha, Qatar*

^d*Urology Department, Hamad General Hospital, Hamad Medical Corporation, Doha, Qatar*

Abstract

Hilar dissection is an important and delicate stage in partial nephrectomy, during which surgeons remove connective tissue surrounding renal vasculature. Serious complications arise when the occluded blood vessels, concealed by fat, are missed in the endoscopic view and as a result are not appropriately clamped. Such complications may include catastrophic blood loss from internal bleeding and associated occlusion of the surgical view during the excision of the cancerous mass (due to heavy bleeding), both of which may compromise the visibility of surgical margins or even result in a conversion from a minimally invasive to an open intervention. To aid in vessel discovery, we propose a novel automatic method to segment occluded vasculature from labeling minute pulsatile motion that is otherwise imperceptible with the naked eye. Our segmentation technique extracts subtle tissue motions using a technique adapted from phase-based video magnification, in which we measure motion from periodic changes in local phase information albeit for labeling rather than magnification. Based on measuring local phase through spatial decomposition of each frame of the endoscopic video using complex wavelet pairs, our approach assigns segmentation labels by detecting regions exhibiting temporal local phase changes matching the heart rate. We

*Corresponding author

Email address: alborza@ece.ubc.ca (Alborz Amir-Khalili)

Figure 1: Variation of renal artery structure and corresponding percentage of occurrence in 266 kidneys adapted from Sampaio and Passos (1992). In each case all vessels that cross the dotted line must be clamped or ligated to minimize intraoperative hemorrhaging.

demonstrate how our technique is a practical solution for time-critical surgical applications by presenting quantitative and qualitative performance evaluations of our vessel detection algorithms with a retrospective study of fifteen clinical robot-assisted partial nephrectomies.

Keywords: Motion Analysis, Image-guided Surgery, Robot-Assisted Partial Nephrectomy, Vessel Segmentation, Hilar Dissection

1. Introduction

Approximately 64,000 new cases of kidney cancer, commonly renal cell carcinoma, are projected to occur in the U.S. in 2014 Siegel et al. (2014). This constitutes double the number of cases reported in 2005. Kidney resection, also known as a nephrectomy, remains the only known effective treatment for this type of localized cancer Drucker (2005). Robot-assisted partial nephrectomy (RAPN) refers to nephron-sparing techniques performed with surgical robots in which only the cancerous cells are excised and the kidney is reconstructed to retain functionality.

The RAPN procedure is organized into five main stages according to Gill et al. (2002): 1) Bowel mobilization; 2) Hilar dissection and control; 3) Identification and demarcation of tumor margins; 4) Resection of tumor; and 5) Reconstruction of the kidney (renorrhaphy). Hilar dissection stands out as a daunting stage requiring significant expertise since improper clamping due to overlooked accessory renal vessels can cause significant bleeding during resection Singh (2009).

Hilar dissection is a delicate procedure during which the surgeon dissects

through the Gerota’s fascia and removes the connective tissue that surrounds the renal artery (RA) and renal vein (RV). This task is complex due to substantial natural variability in patient vasculature (Fig. 1) and the amount of perinephric fat surrounding the kidney. Access to the hilum grants the surgeon control over the flow of blood into and out of the kidney, which is critical as warm ischemia is required during the excision of the tumor to minimize internal hemorrhaging. In some cases, accessory vessels that branch off from the RA or the abdominal aorta (AA) are accidentally missed as they lie hidden behind a thick layer of perinephric fat. In one study of 200 laparoscopic partial nephrectomy cases by world leading surgeons Ramani et al. (2005), seven incidents of intraoperative bleeding were reported as a result of inadequate hilar control, two of which were directly caused by missed accessory vessels. Although the number of incidents is relatively low, other studies by Urban et al. (2001); Sampaio and Passos (1992) observed the existence of accessory vessels in more than 35% of patients. These accessory vessels also prolong the hilar dissection stage as the surgeon must locate them prior to resection. If the surgeon’s level of experience is limited, the incidence of bleeding and overall dissection time may be much higher. The implications are many, aside from obvious complications that would arise from internal hemorrhaging, as bleeding may also jeopardize the surgical outcome by occluding the surgeon’s view while the tumor is being resected.

Surgeons often make use of preoperative medical images for identifying troublesome accessory vessels Mottrie et al. (2010). Even with high-resolution scans and segmented preoperative plans available to them, surgeons are still burdened with the complex and error-prone task of mentally transferring these abstractions onto the surgical site during the operation. Reducing the difficulty of navigation has been attempted by various approaches that rely on multi-modal registration to align the preoperative surgical map of the vessels onto the surgeon’s endoscopic view, e.g. Su et al. (2009); Teber et al. (2009); Pratt et al. (2012); Amir-Khalili et al. (2013); Hamarneh et al. (2014); Nosrati et al. (2014); Estépar and Vosburgh (2014); Puerto-Souza et al. (2014). Registering and augmenting preoperative segmentations into intraoperative video is an excellent

idea. However, such techniques have major limitations including selection of
 50 sensitive parameters Nosrati et al. (2014), manual alignments Su et al. (2009);
 Pratt et al. (2012), use of invasive fiducials Teber et al. (2009), and high com-
 putational complexity that prohibits practical real-time operation Estépar and
 Vosburgh (2014); Puerto-Souza et al. (2014); Hamarneh et al. (2014); Amir-
 Khalili et al. (2013). These limitations stem from the difficulty of registering
 55 intraoperative video data with 3D anatomy that deforms and changes due to
 factors such as cutting, retraction, and breathing. Furthermore, these methods
 do not specifically focus on augmenting the location of vasculature.

Recent methods that focus more specifically on the detection of vascula-
 ture include the use of hardware solutions such as near infrared fluorescence
 60 imaging Tobis et al. (2011) or algorithmic methods that only use color inten-
 sity information from the endoscope to highlight vasculature based on perfusion
 models Crane et al. (2010). Solutions that use near infrared fluorescence are not
 widely accessible as they are cost restrictive, requiring additional equipment and
 expert clinicians to control the dosage of fluorescent agents. On the other hand,
 65 simple algorithmic methods fail to identify vessels that are hidden under a layer
 of fat.

Intraoperative ultrasound (US) imaging is another hardware solution em-
 ployed during the tumor demarcation and excision stages of RAPN; mainly to
 resolve uncertainties in the location of tumor boundary Gill et al. (2002). Re-
 70 cent advancements in the field of US guidance, i.e. ‘pick-up’ transducers Schnei-
 der et al. (2011), motivate the use of US during hilar dissection, but such US
 based guidance techniques also incur additional costs in terms of an increase
 in required personnel (as some robotic surgeons are not trained to operate and
 interpret US), operating time, and equipment upkeep. Even with Doppler US
 75 imaging, the localization of complex vascular structures is further ameliorated
 by the fact that the laparoscopic US probes currently available in the market
 can only acquire 2D images and, depending on the availability of picture-in-
 picture visualization, the surgeon may have to look at a separate screen to view
 the US images.

80 Our work is motivated by the need for a hardware-free non-invasive auto-
 mated guidance system that can reduce the complications and the time required
 to perform hilar dissection by assisting the surgeon in localizing hidden acces-
 sory vessels. Our proposed system highlights occluded vessels by analyzing the
 temporal motion characteristics of the scene captured in the endoscopic video.
 85 These pulsatile motion characteristics are novel features for vessel segmentation
 and may be incorporated alongside other complimentary static visual features
 such as color, intensity, and texture. Our method is inspired by video magni-
 fication techniques applied to natural scenes Wu et al. (2012); Wadhwa et al.
 (2013), where an Eulerian approach to analyzing flow within a video sequence
 90 can be used to magnify periodic motions that are invisible to the human eye.
 An extension of Wu et al. (2012) was very recently implemented in the context
 of robot-assisted surgery by McLeod et al. (2014) to enhance the appearance of
 vasculature by magnifying their pulsatile motions. Replacing the actual surgical
 scene by a processed unnatural and exaggerated motion may be a nuisance to
 95 the surgeons and may cause confusion and confound critical visual information
 in the edited video regions. Furthermore, analysis in Wu et al. (2012) demon-
 strated that the linear, first order Taylor series and other variants such as the
 one adopted by McLeod et al. (2014), approximations of motion is susceptible
 to error and magnification of noise when the motion is large. Linear Eulerian
 100 motion analysis is therefore not a good option as a typical surgical scene, in addi-
 tion to containing small pulsatile motions, also contains gross motion of organs
 caused by breathing and peristalsis. Since a first order approximation is not
 suitable for our application, in our current work¹, we adapt phased-based video
 analysis to map detected subtle motion patterns into segmentations rather than
 105 magnified motions. To demonstrate the potential utility of our method to sur-
 geons, we evaluated our method with a retrospective study of fifteen challenging
 RAPN clinical cases (selection criteria is detailed in Section 3.2).

¹An earlier version of this appeared in Amir-Khalili et al. (2014)

2. Methodology

As motivated earlier, hilar dissection prior to clamping is a critical stage in RAPN that would be greatly simplified if occluded blood vessels near the renal hilum are automatically highlighted to surgeons. Blood vessels at the hilum pulsate due to the large amount of blood flowing into and out of the kidney. These periodic pulsations are within a narrow temporal passband centered around the heart rate of the patient. With a high definition surgical endoscope, one can observe the pulsations of the vessels as faint movements on the surface of the connective tissue that covers them. Our goal is to automatically process every frame in the surgical video and label pixels that exhibit this characteristic motion. We denote our labels as:

$$L(x, y, t) : \mathbb{R}^2 \times \mathbb{R}^+ \rightarrow l \in [0, 1], \quad (1)$$

where l is a normalized fuzzy value that is proportional to the magnitude of pulsatile motion measured at the pixel $(x, y) \in \mathbb{R}^2$ at a given point in time $t \in \mathbb{R}^+$. In the remainder of this section, we elaborate on our methodology for generating these labels from local phase measurements, starting with the relationship between phase and motion.

The shift property of Fourier transform states that motion is directly related to the change in phase, however the explicit computation of motion vectors from phase with techniques like Gautama and Van Hulle (2002) is computationally expensive. To avoid this cost, we aim to exploit the relationship between phase and motion to segment the pulsating regions of interest in the video sequence. The overview of our method is shown in Fig. 2 and it is described in the following 1D intuitive manner, without loss of generality. First, consider a video, a simplified version of the one shown in Fig. 2a, denoted as $I(x, t)$ representing a function mapping given pixel $x \in \mathbb{R}$ at time t to an intensity value. Suppose that this video is a sequence composed of the 1D image $f(x)$ that is translated by a displacement vector $d_x(t)$ along the x direction as a function of time t such that $I(x, t) = f(x + d_x(t))$. To extract this motion, we decompose each frame

Figure 2: Overview of our proposed method: (a) a synthetic input video composed of one circle that pulsates (top) and another that remains stationary (bottom). (b) Steerable filter bank with illustrated impulse responses decompose the information inside each frame into (c) magnitude and local phase at different scales and orientations. (d) The phase information of all frames of the video is temporally filtered using an ideal bandpass filter centered on the frequency of the pulsating circle. (e) A median filter is applied to (f) the magnitude weighted response of the filtered phases to remove phase noise. (g) The results are then combined to generate the fuzzy labels L and (h) added back to the input video as an overlay. A spatiotemporal cross section of the video (lower right) illustrates four pulsations across 30 frames of the synthetic video. High resolution images are available in the digital copy.

of the video into spatial sub-bands

$$I(x, t) = f(x + d_x(t)) = \sum_{\omega=-\infty}^{\infty} A_{\omega} e^{i\omega(x+d_x(t))} \quad (2)$$

with each sub-band representing a complex sinusoid $S_{\omega}(x, t) = A_{\omega} e^{i\omega(x+d_x(t))}$ at spatial frequency ω . The phase of each sub-band is defined as $\Phi_{\omega}(x, t) = \arg(S_{\omega}) = \omega(x + d_x(t))$. Since $\omega d_x(t)$ is the only motion related component of the phase that varies with time, we can isolate it from the DC component ωx by applying a DC-balanced temporal bandpass filter with a wide enough passband to capture all temporal variations in $d_x(t)$.

Generally, motions in a video are not merely a simple global translation. The displacement vector $d_x(t)$ is, in fact, $d_x(x, t)$ since it varies as a function of both time and space. To measure these spatially localized variations, rather than using a Fourier series expansion, we decompose each frame of the video using a spatial filter bank (pyramid) consisting of a cascade of Gabor odd and even symmetric filter pairs (analogous to Hilbert transform filter pairs in 1D)

with limited spatial support (note that the impulse responses of these wavelets have been enlarged for clarity in Fig. 2b). In the steerable pyramid, the spatial extent of each filter is determined by the scale or spatial passband of the Gabor wavelets and, at each scale, the filters are designed to measure motion along a certain direction or orientation in 2D space (note that in Fig. 2b a pyramid consisting of two scales and two orientations is used). In the 1D example, if the local motion is from a single sinusoid with spatial frequency ω along the x direction, we would only need a single pair of Gabor wavelets to extract the motion from the change in local phase. Local, or instantaneous, phase is calculated from the argument of the response to the wavelet pair (Fig. 2c). We then estimate local motion $d_x(x, t)$ from the change in local phase by applying a DC-balanced temporal bandpass filter to the obtained local phase values. We filter the local phase measurements using an ideal bandpass filter:

$$H(x, t) = 2B_H \text{sinc}(2B_H t) - 2B_L \text{sinc}(2B_L t), \quad (3)$$

120 where B_L is the low frequency cut-off and B_H is the high frequency cutoff and the two **sinc** functions are the spatial domain representations of two **rect** functions in the frequency domain that construct an ideal bandpass. The response of the temporal bandpass filter is denoted by $B_\omega(x, t) = \Phi_\omega * H = \omega d_x(x, t)$. We tune the passband of the filter to the typical heart rate of a patient so that we
125 then can simply and effectively isolate components of the local motion that are synchronous with the heart rate and hence to vascular pulsation. In this work we have set the passband of the temporal filter wide enough such that it can separate pulsatile motion from breathing motion in all of the fifteen cases. Future development should involve a tighter estimate of the patients heart rate to
130 improve the results. Such estimates may be recorded directly from the patients heart rate monitor or obtained from the anesthetist.

To generate fuzzy segmentation labels from the computed local motion, B_ω is first attenuated in regions where the magnitude response A_ω of the spatial sub-band is weak. This is done by computing the product between the

bandpassed phases and the normalized magnitude of the spatial filter response vectors $\hat{A}$ to obtain $Q_\omega = \hat{A}_\omega B_\omega$ (Fig. 2e). Since local phase measurements Φ_ω are wrapped between the interval $(-\pi, \pi]$, and since H in (3) acts as a derivative, the jumps in wrapped phase become impulse noise in Q_ω . We remove this noise from the product Q_ω using a spatiotemporal median filter (Fig. 2f). For faster performance, the spatiotemporal median filter is replaced with a spatial pseudo-median filtering process using efficient 2D morphological opening $\circ$ and closing $\bullet$ operations as follows

$$\tilde{Q}_\omega = Q_\omega \circ E + Q_\omega \bullet E - Q_\omega, \quad (4)$$

where E is a 2×2 square structuring element. This denoised product $\tilde{Q}_\omega$ is averaged across all spatial sub-bands (Fig. 2g) and in the 2D case, across all filter orientations to obtain our final fuzzy labels

$$L = \frac{1}{N} \sum_{\forall \omega} \frac{|\tilde{Q}_\omega|}{2\pi\omega}, \quad (5)$$

where N is a normalizing factor. The resulting sequence of fuzzy labels L may be displayed as an overlay or separately to highlight this pulsatile motion (Fig. 2h).

In a real-time application, the ideal temporal filter H may be replaced with an
135 infinite impulse response filter.

3. Experiments and Results

In this section, we detail the *in vivo* dataset and the parameters used for the qualitative, quantitative, and clinical evaluation of our method. Results are illustrated in-situ following the description of the experiments in each subsection.
140 tion. Extended discussions of all experimental findings are carried out in the next section.

3.1. Experimental Setup and Data Acquisition

Video sequences from fifteen clinical RAPN interventions were used for validation. All endoscopic video data were acquired by a da Vinci Si surgical system
145 (Intuitive Surgical, California, USA). HD (1080i) videos were downsampled to 480×270 pixels, to reduce computation times and memory requirements of the overcomplete wavelet representation. The publicly available code from Portilla and Simoncelli (2000) was used to generate sixteen complex steerable filter pairs (four orientations at four scales). The number of scales was set to four
150 since, at the downsampled resolution, this number restricts the spatial extent of the wavelets such that most of the large structures in the videos are detected without blurring the motion of the smaller structures. In Section 4, we point to cases where the detection can be improved with a higher number of scales. The number of orientations was chosen such that diagonal motion is detected
155 more accurately without increasing the computational and memory complexity of the algorithm. Increasing the orientations would improve the detection but it would increase the over completeness of the pyramid representation at a faster rate than an increase in number of scales. Refer to Wadhwa et al. (2013) for more details regarding over completeness.

160 The passband of the temporal filter was set between $B_L = 60$ to $B_H = 120$ beats per minute. Average runtime of our unoptimized MATLAB code to process these four second clips (120 frames) was 45 seconds (35 seconds with the pseudo-median filter). All results shown were obtained using the spatio-temporal median filter with a window size of $3 \times 3 \times 3$.

165 To provide an objective framework for validation, we compared the segmentations obtained through our guidance system against the manually obtained ground truth vasculature. To generate the ground truth, we segmented the kidney, tumor/cyst, AA, inferior vena cava (IVC), RA, RV, and accessory vessels (Fig. 3) using the publicly available ITK-SNAP semi-automatic segmentation
170 algorithm Yushkevich et al. (2006). The resulting meshes were then manually aligned onto the first frame of each endoscopic scene (Fig. 4a) by rigidly transforming the models using a 6-degrees of freedom (DOF) 3D mouse to adjust

Figure 3: Manual segmentation of each surgical case with the number of accessory vessels included in parenthesis, showing kidney (brown), tumor/cyst (green), veins (cyan), and arteries (red).

all DOF contemporaneously. Anatomical landmarks such as the contour of the kidney, visible parts of the vessels, tumor, liver, spleen, ribs, and the surgical
175 dissection planes were used to guide the registration process. Initial alignments were made by a graduate student with extensive knowledge of renal anatomy, and finalized by a junior urologist to ensure correctness. The segmentations and alignments were done prior to performing the vascular motion segmentation. Examples of the registered ground truths are presented in Fig. 4b. Small
180 observable discrepancies between the aligned ground truth segmentation and the endoscopic view are attributed to non-rigid deformations of the organs and vasculature caused by deformation during insufflation, retraction, or the mobilization of organs during the dissection, which our rigid registration does not take into account. By comparing the observable discrepancies against the known
185 size of visible structures obtained from the CT images, we estimate an average of 1-3mm of alignment error in most cases and a maximum of 4-7mm in cases where organs have been significantly retracted by the surgical instruments or mobilization of other organs such as in Cases 2, 3, 4, 5, 9, and 14.

3.2. Qualitative Results

190 The sequences that constitute our dataset, represent patients with varying visceral adiposity, RENAL scores (e.g. cyst or tumor, endophytic or exophytic), and vasculature. To explore the advantages and limitations of our method, we included challenging conditions, such as heavy presence of specular noise (Cases 1, 2, 5, 9, 10, and 12), the endoscope being close (Cases 3, 10, and 14) or far
195 (Cases 1, 5, and 6) from the tissue surface, vessel occlusion by other organs (Cases 4, 6, 8, and 10), retraction of blood vessels (Cases 3, 5, 9, 13, and 14), and tool motion (Cases 7 and 12). Though not all the clinical examples were successful (Cases 7 and 8), the cases cover a wide range of typical scenes and potential problems that can arise.

200 In Fig. 4 we illustrate the qualitative performance of our proposed method. In our experiments, we observed that although venous and arterial structures pulsate at the same frequency, their pulsations are not always in-phase. In fact,

Figure 4: Exemplar video frames with the proposed automatic localization of veins and arteries in the scene. (a) The first frame of the sequence, (b) manually localized venous (cyan) and arterial (red) structures, (c) the binary mask used for quantitative evaluation, (d) temporal mean of the segmentations, and exemplar frames at the time of (e) venous and (f) arterial pulsation. High resolution images are available in the digital copy.

the temporal resolution of the surgical endoscope (30 frames per second) was able to observe the motion of the IVC and RV an average of six frames ahead of the AA and RA. To present this temporal phenomenon in Fig. 4, we manually extracted two frames of the segmented video that highlight venous vs. arterial pulsation.

The two exemplar frames were manually extracted from the same cardiac cycle. The fifteen four-second clips contain between four to eight cardiac cycles; this number varies depending on the patient’s heart rate during acquisition. Each motion label frame within a cardiac cycle, containing 15-30 frames, was visually compared to the ground truth to identify a pair of frames that best represent the venous and arterial structures. The ground truth was not altered during this process. Compared to the reference in Fig. 4b, the motions highlighted in Fig. 4e correspond to the cyan structures (venous) and Fig. 4f corresponds to the red structures (arterial).

3.3. Quantitative Evaluation

Currently our method is not able to automatically differentiate between the locations of veins and arteries. Therefore, to quantify a measure of detection for such pulsating structures, the automatic segmentations were first binarized (at a fixed threshold throughout the sequence) and combined across the frames of the video to generate a single binary image containing the union of all pulsatile motion regions in the video. The resulting image was then compared to a binary mask (Fig. 4c) of the reference manual segmentation, combining all vasculature into a single mask. Fig. 5 illustrates the segmentation performance of all cases, at different threshold values, via their receiver operating characteristics (ROC). The areas under these ROC curves are presented in the legend of in Fig. 5.

3.4. Clinical User Study

We performed an initial user-centric assessment of the proposed technique by analyzing feedback from different surgeons at the Hamad Medical Corporation. We recruited four surgeons (two junior surgeons, and two senior surgeons) and

Figure 5: Quantitative results on fifteen RAPN cases. Left: ROC of all cases with the associated area under curve in parenthesis. Right: median (red) and mean (green) of all ROC. Mean area under all ROC is 0.72 with a standard deviation of 0.05.

asked them to watch short surgical scenes (video clips) of hilar dissection from six different patients. Each surgeon was presented two types of video clips for each patient and was asked to identify the vasculature in the scene. The first set of videos consisted of clips from the original unprocessed surgical scenes, while the second set of videos consisted of the same clips but augmented using our proposed method. To prevent the surgeon’s performance from being biased by their viewing of the original and augmented videos of the same patient, we separated the viewing dates by one week and randomized the order of viewing for both original and augmented scenes. In addition, the surgeons were given the option of viewing the history video, defined as a 5 to 10 minutes video clip representing the history of the surgery preceding the selected short scene. This was required by the surgeons as it mimics real surgical scenarios. During a real operation, the events and actions leading to the current scene provide context for the short (4 second) video clips, thus implicitly providing the surgeon with vital information such as where the vessels may be based on how the organs have been retracted and positioned. In each experiment, we measured the time it took for the surgeon to annotate the clip and saved the surgeon’s annotations on the first frame of each scene. We then compared the annotations with the ground truth data of that first frame (binary mask in Fig. 4b) using the Dice

similarity coefficient (DSC).

In general, the proposed method was found to improve vessel detection, mainly by reducing detection-time for junior surgeons. More specifically, the proposed method reduced vessel detection time by 22% on average for junior
255 surgeons, though it did not seem to affect detection-time for expert surgeons. Our proposed overlay only increased the average DSC for all users by a marginal 0.024. Upon debriefing, both junior and senior surgeons confirmed that they were relying on visual cues and prior knowledge for locating vessels and that the overlays were mainly used to confirm their own localizations, which in and of
260 itself was reported as an added benefit by the surgeons. The feedback from the surgeons was generally positive, with the exception of one junior surgeon who stated that the visualization of the segmentations were difficult to interpret. Nevertheless, the average DSC for all cases (with and without augmentation) performed by all participants was 0.13 with a standard deviation of 0.11, whereas
265 our method performed with a DSC of 0.50 with a standard deviation of 0.24. Note that to compute the DSC for our method, we thresholded our fuzzy labels at 0.10 for all cases, this threshold was chosen as it gave the highest average DSC across all cases. Our interpretation of these findings is reported in the following section.

270 4. Discussion

In this section, we analyze the significance and implications of our experimental findings. We were careful in the selection of our cases and in ensuring that the alignment of the ground truth is, to the best of our ability, without error. The reported misalignment errors in ground truth (up to 4-7mm in cases
275 with visible non-rigid deformations) may appear to be large to some readers, but it is rather small compared to the size of the structures (up to 30mm in diameter for the IVC) that we are trying to detect. This small error does not have a significant impact on our interpretation of the results.

Our quantitative assessment indicates a mean area under the ROC curve of

280 0.72, midway between the noise baseline of 0.5 and the highest possible value
of 1.0 (indicating perfect detection). Our detection rate may be slightly higher
than the presented value due to the aforementioned misalignment of the ground
truth. A conservative value of 0.72 suggests that our generated labels are dis-
criminative and suitable for integration (as an additional data term) into exist-
285 ing vessel segmentation techniques that use other visible features such as color,
shape and texture.

To put the reported detection rates into perspective and to identify oppor-
tunities for achieving higher accuracy, we manually chose two exemplar frames
(depicting venous and arterial structures) from the resulting segmentations and
290 compared the results to our ground truth. The following observations were
made for each of the fifteen cases.

Case 1 The RV and IVC are correctly labelled under the heavy occlusion.

The small RA is correctly identified at the hilum, even though some false
positives were detected on the tools. We attribute the false positives to
295 the fact that the vascular pulsations cause the end effectors of the surgical
tools to vibrate through the abdominal wall.

Case 2 The ground truth location of the RA (Fig. 4b) is slightly misaligned
due to retraction by the surgical instrument. The detected locations of
the AA, RA, IVC and RV are correct despite the retraction.

300 **Case 3** All structures are correctly identified, including the portion of the IVC
that is occluded by the surgical gauze and the small accessory RA to the
left of RV.

Case 4 All structures are identified but it is difficult to differentiate between
arterial and venous structures; possibly because of the heavy occlusion by
305 the perinephric fat. Possible misalignment of the ground truth (suprarenal
vein) due to mobilization of the spleen.

Case 5 All structures are identified. Some false positives are present during
arterial pulsation (Fig. 4f). There may be a misalignment in the ground

truth as retraction has shifted the abdominal aorta up.

310 **Case 6** Branching of the RA is detected on both sides of RV and the pulsation
of heavy vascular region has caused the tumor to pulsate in the centre of
the frame.

Case 7 The RV is correctly identified but significant amount of false positives
are observed due to the motion of the tools in the scene. Ideally, surgical
315 instruments should remain motionless during the acquisition of the video.

Case 8 Like Case 7, Case 8 also posed a big challenge as the vasculature is
heavily occluded by the bowel and many false positives are detected in
fluid filled cavities to the left of the cyst.

Case 9 All structures are roughly identified. A specific patch of specular re-
320 flections on top of the RA in the left side of the frame (Fig. 4d) has signif-
icantly skewed the normalization of the labels. Specular highlights pose
a great challenge to endoscopic video analysis, and although the spec-
ular patches are not large at the hilum compared to other organs such
as the liver, their presence does have noticeable effects, i.e., regions with
325 highlights are emphasized more (stronger response) in comparison to their
neighbors. Cases 10 and 12 are also notable examples of this.

Case 10 The RA and RV are correctly identified. A large part of the bowel is
occluding the large (30mm) IVC making it difficult to detect.

Case 11 The RA is correctly identified but the pulsations of the RV are missed.
330 We attribute this to the large (20mm) size of the structure. In such cases,
the number of scales of the spatial steerable pyramid may be increased to
five levels.

Case 12 Both RA and RV are clearly detected in Case 12; small motion arte-
facts are present on the suction tool.

335 **Case 13** All structures are identified. This case provides a good visualization
of the phase difference between arterial and venous pulsations.

Case 14 In addition to identifying the RV our method was also capable of localizing the gonadal vein present on the right side of the image (Fig. 4e).

Case 15 All structures are identified. There is a Heavy presence of false positives caused by specular highlights on the top right corner.

In summary, our method was qualitatively successful in all cases, except for Cases 7 and 8, both of which are associated with the lowest area under curve values. Camera and tool movement may seem to be a big challenge but coping with them is rather simple in the context of hilar dissection during RAPN. Hilar dissection stage is not time-critical in comparison to the following resection and reconstruction stages. Our surgeons have confirmed that they can afford a four second pause during the operation while our system acquires the required information for processing. The surgical instruments can also be moved out of sight during the acquisition. The most challenging sources of false positives are the specular highlights and pulsatile vibrations in background structures. All organs inside a live human exhibit minute pulsatile motion in sync with the heart rate. These minute vibrations are very small compared to the motion of major vasculature, yet their apparent motion is magnified with the presence of specular highlights or fluids. In the future, we plan to focus mainly on an adaptive estimation for noise (to boost its robustness to specular reflections) and automating the process of differentiation between veins and arteries.

Future improvements to the methodology will be carried out in tandem with more clinical user studies. Our initial user study was insightful as we observed a significant difference between the performance of the surgeons (DSC of 0.13) compared to our segmentations (DSC of 0.50). This difference does not indicate that the ground truth is inaccurate or that the surgeons disagree with it. The performance of the surgeons appears to be poor since surgeons were not able to locate large segments of the vasculature that were hidden/occluded. This difference in DSC and the fact that the performance of the surgeons only improved marginally imply that, perhaps with more training and a better visualization technique, our method has the potential to improve the surgeon's performance

even further. We plan to address this need by developing new (clinically appropriate) visualizations, user inputs that grant the surgeon the ability to control the opacity of the displayed overlay, and a spotlight option that enables the surgeon to choose the region within which the segmentations are overlaid. Although our initial trial is promising, further studies with more participants and data are required to quantify the clinical impact and effectiveness of our method in finding accessory vessels.

5. Conclusions

We have proposed a novel method for localizing and labeling regions in endoscopic video that contain occluded vessels. Our method extends Eulerian phase-based video motion processing techniques to detect and label small motions that are barely visible on the surface of the perinephric fat. To the best of our knowledge, we are the first to attempt the challenging task of localizing occluded vasculature in endoscopic video without the use of additional hardware or preoperative scans. We validated our novel method qualitatively in a retrospective *in vivo* study to verify its application in a clinical setting. Using manually aligned preoperative models of the *in vivo* patient data as ground truth, we performed conservative quantitative validation of our method to report well known measures of detection, i.e., the area under the ROC curve. Furthermore, we conducted a preliminary clinical study, and received very enthusiastic feedback from the surveyed urologists.

The results from our evaluation and the user study demonstrated that our method is suitable for integration alongside existing techniques (e.g. as an additional cue) that use other visible features such as color, shape and texture. We plan to extend our method by developing more clinically oriented visualizations, an adaptive estimation for noise (to boost its robustness to specular reflections), and optimizing the code to operate in real-time. With each iteration of improvement, we plan to perform more user studies to assess the practicality of our method and to judge whether the method is ready for translation into

widespread clinical practice. We are also exploring the addition of our motion segmentations as an additional data term for guiding automatic non-rigid registration of preoperative surgical models with endoscopic video in the context of RAPN. In the current work, the alignment of the pre-segmented 3D model was
400 manual and did not utilize a calibrated stereo video stream. In future extensions to our method, we plan to use automatic alignment leveraging both left and right frames, as recently proposed by Nosrati et al. (2014).

Acknowledgement

This publication was made possible by NPRP Grant #4-161-2-056 from
405 the Qatar National Research Fund (a member of the Qatar Foundation). The statements made herein are solely the responsibility of the authors.

References

- Amir-Khalili, A., Nosrati, M.S., Peyrat, J.M., Hamarneh, G., Abugharbieh, R., 2013. Uncertainty-encoded augmented reality for robot-assisted partial
410 nephrectomy: A phantom study, in: *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*. Springer, pp. 182–191.
- Amir-Khalili, A., Peyrat, J.M., Abinahed, J., Al-Alao, O., Al-Ansari, A., Hamarneh, G., Abugharbieh, R., 2014. Auto localization and segmentation of
415 occluded vessels in robot-assisted partial nephrectomy, in: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (Eds.), *Medical Image Computing and Computer-Assisted Intervention*. Springer International Publishing, volume 8673 of *Lecture Notes in Computer Science*, pp. 407–414.
- Crane, N.J., Gillern, S.M., Tajkarimi, K., Levin, I.W., Pinto, P.A., Elster, E.A.,
420 2010. Visual enhancement of laparoscopic partial nephrectomy with 3-charge coupled device camera: assessing intraoperative tissue perfusion and vascular anatomy by visible hemoglobin spectral response. *The Journal of Urology* 184, 1279–1285.

- Drucker, B.J., 2005. Renal cell carcinoma: current status and future prospects.
 425 Cancer treatment reviews 31, 536–545.
- Estépar, R.S.J., Vosburgh, K.G., 2014. Multimodality guidance in endoscopic and laparoscopic abdominal procedures, in: Intraoperative Imaging and Image-Guided Therapy. Springer, pp. 767–778.
- Gautama, T., Van Hulle, M.M., 2002. A phase-based approach to the estimation of the optical flow field using spatial filtering. Neural Networks, IEEE Transactions on 13, 1127–1136.
 430
- Gill, I.S., Desai, M.M., Kaouk, J.H., Meraney, A.M., Murphy, D.P., Sung, G.T., Novick, A.C., 2002. Laparoscopic partial nephrectomy for renal tumor: Duplicating open surgical techniques. The Journal of Urology 167, 469 – 476.
- 435 Hamarneh, G., Amir-Khalili, A., Nosrati, M.S., Figueroa, I., Kawahara, J., Al-Alao, O., Peyrat, J.M., Abi-Nahed, J., Al-Ansari, A., Abugharbieh, R., 2014. Towards multi-modal image-guided tumour identification in robot-assisted partial nephrectomy, in: Biomedical Engineering (MECBME), 2014 Middle East Conference on, IEEE. pp. 159–162.
- 440 McLeod, A.J., Baxter, J.S., de Ribaupierre, S., Peters, T.M., 2014. Motion magnification for endoscopic surgery, in: SPIE: Medical Imaging, pp. 9036–11.
- Mottrie, A., De Naeyer, G., Schattelman, P., Carpentier, P., Sangalli, M., Ficarra, V., 2010. Impact of the learning curve on perioperative outcomes in
 445 patients who underwent robotic partial nephrectomy for parenchymal renal tumours. European Urology 58, 127–133.
- Nosrati, M.S., Peyrat, J.M., Abinahed, J., Al-Alao, O., Al-Ansari, A., Abugharbieh, R., Hamarneh, G., 2014. Efficient multi-organ segmentation in multi-view endoscopic videos using pre-operative priors, in: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014, Springer. pp. 324–
 450 331.

- Portilla, J., Simoncelli, E.P., 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. *IJCV* 40, 49–70.
- Pratt, P., Mayer, E., Vale, J., Cohen, D., Edwards, E., Darzi, A., Yang, G.Z.,
455 2012. An effective visualisation and registration system for image-guided robotic partial nephrectomy. *Journal of Robotic Surgery* , 1–9.
- Puerto-Souza, G., Cadeddu, J., Mariottini, G.L., 2014. Toward long-term and accurate augmented-reality for monocular endoscopic videos. *Biomedical Engineering, IEEE Transactions on* 61, 2609–2620. doi:10.1109/TBME.2014.
460 2323999.
- Ramani, A.P., Desai, M.M., Steinberg, A.P., Ng, C.S., Abreu, S.C., Kaouk, J.H., Finelli, A., Novick, A.C., Gill, I.S., 2005. Complications of laparoscopic partial nephrectomy in 200 cases. *The Journal of urology* 173, 42–47.
- Sampaio, F., Passos, M., 1992. Renal arteries: anatomic study for surgical and
465 radiological practice. *Surgical and Radiologic Anatomy* 14, 113–117.
- Schneider, C., Guerrero, J., Ngan, C., Rohling, R., Salcudean, S., 2011. Intra-operative pick-up ultrasound for robot assisted surgery with vessel extraction and registration: A feasibility study, in: *Information Processing in Computer-Assisted Interventions*. Springer, pp. 122–132.
- 470 Siegel, R., Ma, J., Zou, Z., Jemal, A., 2014. Cancer statistics, 2014. *CA: a cancer journal for clinicians* 64, 9–29.
- Singh, I., 2009. Robot-assisted laparoscopic partial nephrectomy: Current review of the technique and literature. *Journal of minimal access surgery* 5, 87.
- 475 Su, L.M., Vagvolgyi, B.P., Agarwal, R., Reiley, C.E., Taylor, R.H., Hager, G.D., 2009. Augmented reality during robot-assisted laparoscopic partial nephrectomy: Toward real-time 3d-ct to stereoscopic video registration. *Urology* 73, 896–900.

- Teber, D., Guven, S., Simpfendörfer, T., Baumhauer, M., Güven, E.O., Yencilek, F., Gözen, A.S., Rassweiler, J., 2009. Augmented reality: a new tool to improve surgical accuracy during laparoscopic partial nephrectomy? Preliminary in vitro and in vivo results. *European Urology* 56, 332–338.
- Tobis, S., Knopf, J., Silvers, C., Yao, J., Rashid, H., Wu, G., Golijanin, D., 2011. Near infrared fluorescence imaging with robotic assisted laparoscopic partial nephrectomy: initial clinical experience for renal cortical tumors. *The Journal of Urology* 186, 47–52.
- Urban, B.A., Ratner, L.E., Fishman, E.K., 2001. Three-dimensional volume-rendered CT angiography of the renal arteries and veins: Normal anatomy, variants, and clinical applications. *RadioGraphics* 21, 373–386.
- Wadhwa, N., Rubinstein, M., Durand, F., Freeman, W.T., 2013. Phase-based video motion processing. *ACM Transactions on Graphics* 32, 80.
- Wu, H.Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., Freeman, W.T., 2012. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics* 31, 65.
- Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures. *Neuroimage* 31, 1116–1128.