

# Hybrid Tracking for Improved Registration of Laparoscopic Ultrasound and Laparoscopic Video for Augmented Reality

William Plishker<sup>1</sup>(✉), Xinyang Liu<sup>2</sup>, and Raj Shekhar<sup>1,2</sup>

<sup>1</sup> IGI Technologies, Inc., College Park, MD, USA

will@igitechnologies.com, RShekhar@childrensnational.org

<sup>2</sup> Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Health System, Washington, DC, USA

**Abstract.** Laparoscopic *augmented reality* (AR), improves the surgeon's experience of using multimodal visual data during a procedure by fusion of medical image data (e.g., ultrasound images) onto live laparoscopic video. The majority of AR studies are based on either computer vision-based or hardware-based (e.g., optical and electromagnetic tracking) approaches. However, both approaches introduce registration errors because of variable operating conditions. To alleviate this problem, we propose a novel approach of hybrid tracking which comprises of both hardware-based and computer vision-based approaches. It consists of the registration of an ultrasound image with a time-matched video frame using electromagnetic tracking followed by a computer vision-based refinement of the registration and subsequent fusion. Experimental results demonstrate not only the feasibility of the proposed concept but also improved tracking accuracy that it provides and the potential for its integration into a future clinical AR system.

**Keywords:** Augmented reality · Laparoscopic ultrasound · Electromagnetic tracking · Hybrid tracking

## 1 Introduction

Laparoscopic surgery is an increasingly accepted mode of surgery because it is minimally invasive and leads to much faster recovery and improved outcomes. In a typical laparoscopic surgery, the primary means of intraoperative visualization is a real-time video of the surgical field acquired by a laparoscopic camera. Compared to open surgery, laparoscopic surgery lacks tactile feedback. Moreover, laparoscopic video is capable of providing only a surface view of the organs and cannot show anatomical structures beneath the exposed organ surfaces. One solution to this problem is augmented reality (AR), which is a method of overlaying imaging data—laparoscopic ultrasound (LUS) images in the present work—onto live laparoscopic video. Potential benefits of AR are improved procedure planning, improved surgical tool navigation and reduced procedure times. A typical AR approach consists of registration of real-time LUS images on live laparoscopic video followed by their superimposition.

Image-to-video registration methods can be divided into two broad categories: (1) computer vision (CV)-based and (2) hardware-based methods. The first category uses CV techniques to track in real time natural anatomical landmarks and/or user-introduced patterns within the field of view of the camera used. When ultrasound is the augmenting imaging modality, tracking the ultrasound transducer in the video is the goal in these approaches. For example, some earlier methods [1, 2] attached user-defined patterns on the ultrasound transducer and tracked those patterns in the video. Feuerstein et al. [3], on the other hand, directly tracked the LUS transducer in the video by detecting lines describing the outer contours of the probe. However, the CV-based approaches may fail or degrade in the presence of occlusion and variable lighting conditions [4].

The second category concerns the use of external tracking hardware devices. The most established method at present is optical tracking, which uses infrared cameras to track optical markers affixed rigidly on the desired tools and imaging devices. The method has been employed in many AR applications [5–7]. AR systems based on electromagnetic (EM) tracking have also been proposed [8, 9]. Tracking hardware is susceptible to two types of errors: system and calibration. The system-based errors in EM tracking often stem from ferrous metals and conductive materials in tools that are close enough to the field generator [10]. Optical markers frequently face the line-of-sight problem. Calibration-based registration errors could be associated with experimental errors from system calibration, which includes ultrasound calibration [11] and laparoscopic camera calibration [12].

We propose a novel hybrid tracking method comprising of both hardware-based and vision-based methods, which may provide consistent, more accurate and reliable image fusion for an AR system. In this work, we focus on applying our method to EM tracking that is capable of tracking the LUS transducer with a flexible imaging tip. The same framework can also be applied to optical tracking. After an ultrasound image is registered with and overlaid on a time-matched video frame using EM tracking, a vision-based algorithm is used to refine the registration and subsequent fusion. Such a rectified calibration method can be accomplished in two stages by: (1) computing a correction transformation which when applied to a 3D Computer Aided Design (CAD) model of the LUS probe improves the alignment of its projection with the actual LUS probe visible in the camera image and (2) incorporating the calculated correction transformation in the overall calibration system.

## 2 Methods

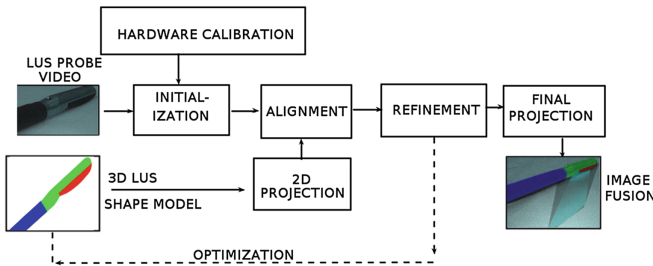
Our AR system in this study includes a clinical vision system (Image 1 Hub, KARL STORZ, Tuttlingen, Germany) with a 10-mm  $0^\circ$  laparoscopic camera (Image 1 HD), an ultrasound scanner (Flex Focus 700, BK Ultrasound, Analogic Corporation, Peabody, MA, USA) with a 9-mm LUS transducer with a flexible imaging tip (Model 8836-RF), an EM tracking system with a tabletop field generator (Aurora, Northern Digital Inc., Waterloo, ON, Canada), and a graphics processing unit (GPU)-accelerated laptop computer that runs the image fusion software. As shown in Fig. 1, we designed and 3D-printed a wedge-like mount to hold the EM sensor (Aurora 6DOF Flex Tube, Type 2, 1.3 mm diameter) using

an existing biopsy needle introducer track in the LUS transducer [9]. The mount was made as thin as possible so that the integrated transducer can still go through a 12-mm trocar, a typical-sized trocar for use with the original transducer.



**Fig. 1.** Custom-designed EM tracking mount on the LUS transducer.

The outline of our hybrid tracking framework is illustrated in Fig. 2. It has two main stages. The first stage consists of two parts: (1) computing calibration of AR system components: laparoscope and LUS transducer; (2) registration of LUS image and the projection of the 3D LUS transducer model on the camera image using the calibration results. In the second stage, the 2D projection of the 3D LUS transducer model is fitted to the actual transducer seen in the camera image. To achieve this, the position and pose parameters of the 3D LUS transducer model are optimized to determine the best fit of its projection to the camera image. Such a correction transformation matrix is fed back to Stage 1, and thus the registration of the LUS image to video is refined.



**Fig. 2.** The outline of the proposed framework.

## 2.1 System Calibration for AR

We first briefly describe the method for our hardware-based AR visualization. Let  $p_{US} = [x\ y\ 0\ 1]^T$  denote a point in the LUS image in homogeneous coordinates, in which the  $z$  coordinate is 0. Let  $p_{Lap}^U = [u\ v\ 1]^T$  denote the point that  $p_{US}$  corresponds to in the undistorted camera image. If we denote  $T_A^B$  as the  $4 \times 4$  transformation matrix from the coordinate system of A to that of B, the registration of  $p_{US}$  on the undistorted camera image can be expressed as

$$p_{Lap}^U = C \cdot [I_3\ 0] \cdot T_{Mark-Lap}^{Cam} \cdot T_{Tracker}^{Mark-Lap} \cdot T_{Mark-US}^{Tracker} \cdot T_{US}^{Mark-US} \cdot p_{US} \quad (1)$$

where US refers to the LUS image; Mark-US refers to the EM sensor attached on the LUS transducer; Tracker refers to the EM tracker; Mark-Lap refers to the EM sensor attached on the laparoscope; Cam refers to the laparoscopic camera;  $I_3$  is an identity matrix of size 3; and  $C$  is the camera matrix.  $T_{US}^{\text{Mark-US}}$  can be obtained from ultrasound calibration;  $T_{\text{Mark-US}}^{\text{Tracker}}$  and  $T_{\text{Tracker}}^{\text{Mark-Lap}}$  can be obtained from tracking system;  $T_{\text{Mark-Lap}}^{\text{Cam}}$  and  $C$  can be obtained from laparoscope calibration [12].

## 2.2 Improved System Calibration for AR

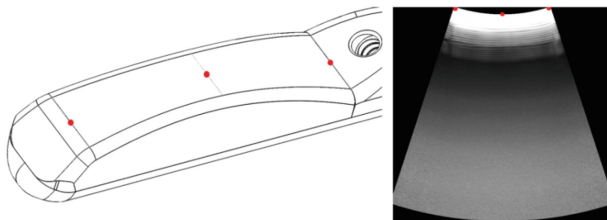
To refine the registration of the LUS image, we first project a 3D LUS transducer model on the camera image using the standard calibration results. We then apply a vision-based algorithm to register the projected 3D transducer model with the actual LUS transducer shown in the video. This yields a correction matrix  $T_{\text{Corr}}$  as a rigid transformation. Since there is a fixed geometric relationship between the LUS transducer and the LUS image, the same  $T_{\text{Corr}}$  can be used to refine the location of the LUS image overlaid on the video. As an update to Eq. 1, a summary of our general approach can be expressed as

$$p_{\text{Lap}}^U = C \cdot [I_3 0] \cdot T_{\text{Corr}} \cdot T_{\text{Mark-Lap}}^{\text{Cam}} \cdot T_{\text{Tracker}}^{\text{Mark-Lap}} \cdot T_{\text{Mark-US}}^{\text{Tracker}} \cdot T_{US}^{\text{Mark-US}} \cdot T_{\text{Model}}^{\text{US}} \cdot p_{\text{Model}} \quad (2)$$

where points of the 3D LUS transducer model are first transferred to the LUS image coordinate system through  $T_{\text{Model}}^{\text{US}}$ , which is described in the next section.

## 2.3 LUS Probe Model and Calibrations

We obtained a CAD model of the LUS probe used in this study from the manufacturer. Because the exact mechanical relationship between the imaging tip of the LUS transducer and the LUS image is proprietary information and not known to the research community, we developed a simple registration step to transfer the coordinate system of the CAD model to that of the LUS image (supposing the LUS image space is 3D with  $z = 0$ ). As illustrated in Fig. 3, we selected three characteristic points on the CAD model and their corresponding points on the LUS image plane. Without loss of generality, we fixed the scan depth of the LUS image to 6.4 cm, a commonly used depth setting for ultrasound imaging during abdominal procedures. A simple three-point rigid registration was then performed to obtain  $T_{\text{Model}}^{\text{US}}$  in Eq. 2.



**Fig. 3.** The three points selected on the LUS CAD model (left) and on the LUS image (right).

We performed ultrasound calibration using the tools provided in the PLUS library [11]. Laparoscope calibration was performed using the fast approach of [12], which requires only a single image of the calibration pattern.

### 2.4 Model Projection and Alignment

To compare the pose and position of the rendered virtual model and the probe in the camera image, we propose the workflow of the CV-based refinement algorithm as presented in Fig. 4. First a region of interest (ROI) is generated for each frame of the laparoscopic video using fast visual tracking based on robust discriminative correlation filters [13] such that subsequent processing focuses on the imaging tip. Based on this coarse estimate of the probe’s location, the bounding box surrounding the imaging tip is intended to include at least some portion of the top, middle, and tip of the probe as seen by the camera. To find the straight edges of these features of the probe, the camera image is first converted to a gray scale image based on brightness, followed by Canny edge detection. We used the Probabilistic Hough Transform (PHT) to extract a set of lines from the edge detection result within the ROI, an example of which is shown in Fig. 5. The line set was filtered by creating a coarse grain 2D histogram with the axes defined by PHT parameters  $(r, \theta)$  and values of histogram defined by the sum of the lengths of lines in the bin. All lines not contained within the highest peak present in the 2D histogram are removed to produce a set of lines that corresponds with the long edges, parallel or close to parallel present in the probe. From this smaller set of lines, a fine grain 2D histogram based on the PHT parameters  $(r, \theta)$  is created. The two highest peaks in this histogram represent the top and middle of the probe. The indices of the peaks are then used in the cost function for the optimization of the virtual probe location.

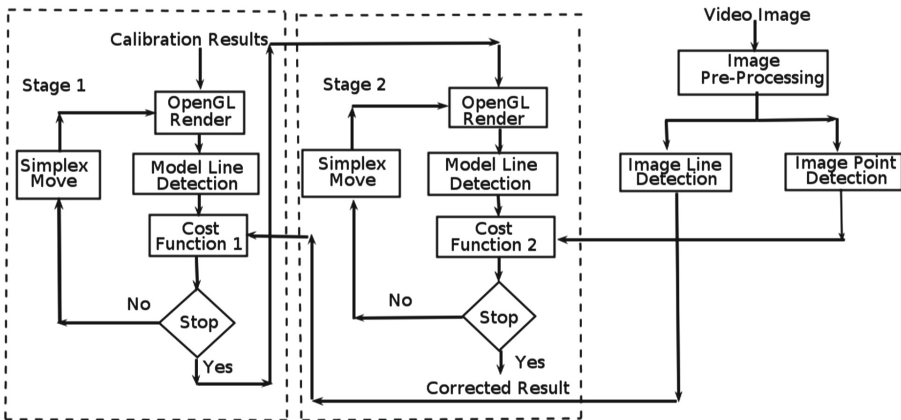
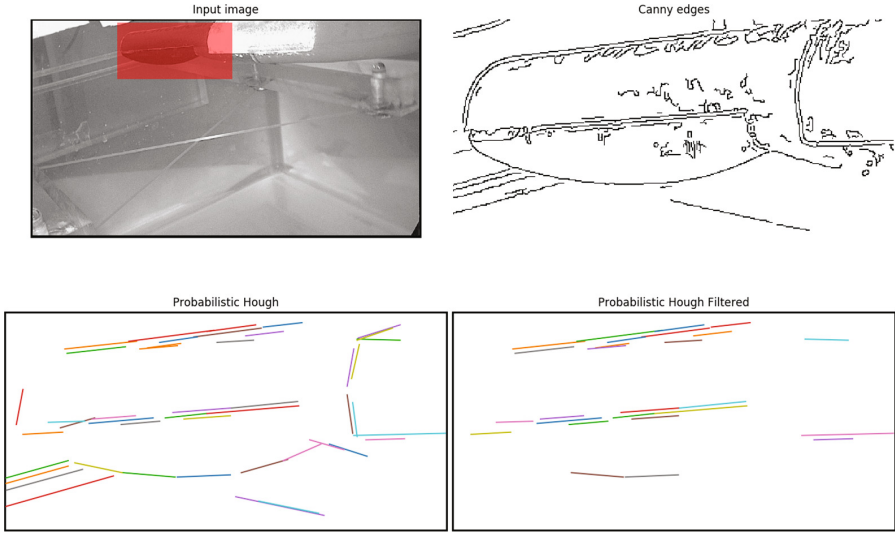


Fig. 4. The proposed refinement algorithm. Dotted areas depict iterative processes.



**Fig. 5.** An example of the hybrid based tracking system with intermediate results displayed: Top Left – gray scale representation of the original camera image and the mask box being used shown in red, Top Right – Canny edges detected within the mask, Lower Left – Probabilistic Hough Transform line set from Canny, Lower Right – line feature based filtering of the line set (Color figure online)

In Stage 1 of optimization, we use the same procedure to detect the same two feature lines both for the rendered 3D LUS transducer model and for the actual transducer shown in the camera image. We compared the alignment of the feature lines using a cost function defined as

$$F_1(x) = \sum_{i=1}^2 \left[ w_r \cdot \left( r_{img}^i - r_{gl}^i(x) \right)^2 + w_\theta \cdot \left( \theta_{img}^i - \theta_{gl}^i(x) \right)^2 \right] \quad (3)$$

where  $w$  is a scalar,  $img$  refers to the camera image, and  $gl$  refers to the OpenGL-rendered 3D LUS transducer model. The optimization used the simplex method [15] to search for the five parameters  $x$  associated with a rigid transformation ( $T_{Corr}$  in Eq. 2). In our current work, we fixed the other parameter, i.e., the one associated with the rotation about the LUS transducer axis. With only two feature lines as constraints, the optimization in Stage 1 may not accurately estimate parameters associated with translation along the LUS transducer axis.

In Stage 2 of optimization, we detect a feature point of the tip of the probe in both images to address inaccuracies along the transducer axis. We used gradient descent-based active contours method [14] to segment the LUS probe from the camera image and identify a feature point  $p$  as the farthest point corresponding with the tip of the transducer. The initialization for segmentation was provided by an ellipse encompassing the ROI. We compared the feature points using another cost function

$$F_2(x) = w_p \cdot d(p_{\text{img}}, p_{\text{gl}}(x))^2 \quad (4)$$

where  $d(\cdot, \cdot)$  is the Euclidean distance in image. In this stage, we restricted the simplex search to focus on only one of the six parameters: the one associated with translation along the LUS transducer axis. The other five parameters are kept fixed to their results from Stage 1. For both stages of optimization, the search terminated according to the tolerances set on both input and cost function deltas.

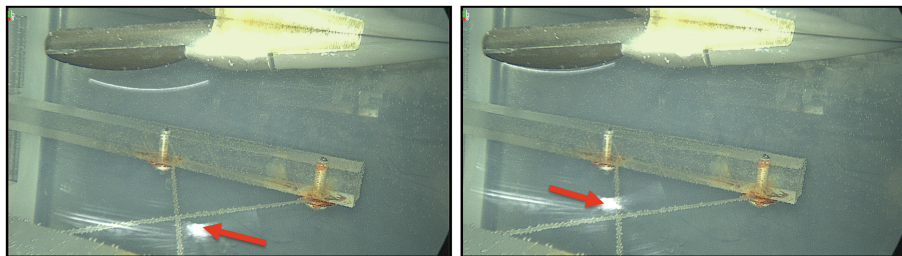
Our hardware-based AR visualization has been implemented using C++ on a GPU-accelerated laptop computer. Currently, the described CV-based refinement algorithm is implemented using OpenCV and scikit-image [17] in Python. The Python based refinement implementation utilizes internal APIs with the C++ based AR code base to transfer images and results between the two.

### 3 Experiments and Results

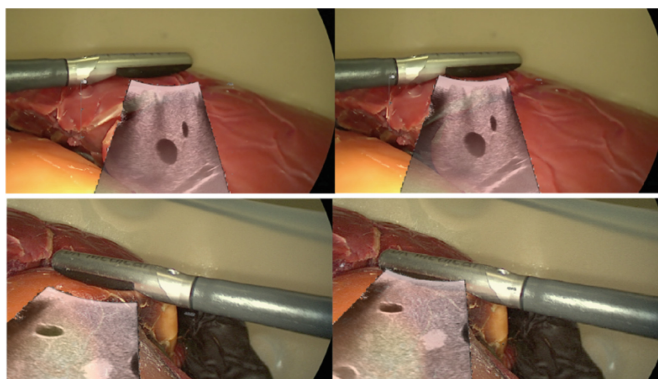
To show the improvement from applying hybrid tracking, we performed experiments to measure and compare target registration error (TRE) between the EM tracking-based approach and the hybrid approach. A target point, the intersection of two cross wires immersed in a water tank, was imaged using the LUS transducer. The target point along with the imaging tip of the LUS transducer was viewed with the laparoscope, whose lens was immersed in water as well. The LUS image was overlaid on the camera image through the EM tracking-based approach (Sect. 2.1) as well as the hybrid approach. The target point in the overlaid LUS image can then be identified and compared with the actual target point shown in the camera image. Their Euclidean distance in the image plane is the TRE.

We performed experiments with four different poses of the laparoscope and the LUS transducer. The average TRE of the EM tracking-based approach was measured to be  $102.0 \pm 60.5$  pixel ( $8.2 \pm 4.9$  mm), and that of the hybrid approach was  $46.7 \pm 10.1$  pixel ( $3.8 \pm 0.8$  mm) with an image resolution of the camera of  $1920 \times 1280$ . The hybrid approach improved overlay accuracy of the original EM tracking-based approach. The CV-based refinement process took on average 52 s, the major bottleneck being the C++ API interface required to read in a new candidate correction matrix. The total number of iteration steps in optimization was fewer than 110 steps for examples tried. Figure 6 shows an example of the refinement results.

We also tested our approach with a more realistic camera and ultrasound images by testing images from phantom. While we did not have a quantitative evaluation of such images, we confirmed that the image processing and subsequent optimization qualitatively worked as well as in the wire phantom. Figure 7 shows examples of this evaluation.



**Fig. 6.** Example of vision-based refinement showing the initial AR visualization using the EM tracking approach (left), and corrected AR visualization using the hybrid tracking approach (right). The arrow indicates the target point shown in the overlaid LUS image.



**Fig. 7.** Two examples of vision-based refinement on an abdominal phantom. The initial AR visualization (left) and corrected AR visualization using the hybrid tracking approach (right).

## 4 Discussion and Conclusion

In this work, we developed a computer vision-based refinement method to correct registration error in hardware-based AR visualization. Initial hardware-based registration is essential to our approach because it provides an ROI for robust feature line detection, as well as a relatively close initialization for simplex-based optimization. We have developed a vision-based solution to refine image-to-video registration obtained using hardware-based tracking. A 3D LUS transducer model was first projected on the camera image based on calibration results and tracking data. The model was then registered with the actual LUS transducer using image processing and simplex optimization. Finally, the resulting correction matrix was applied to the ultrasound image. The method is promising as evidenced by our preliminary results included in this work. After further refinement, the proposed hybrid framework could greatly improve the accuracy and robustness of a laparoscopic AR system for clinical use.



Although the current computational time is relatively lengthy even for periodic correction, we can more tightly integrate the algorithm with our C++ and GPU-accelerated AR system in the future. If implemented on GPU, the Hough Transform can be achieved in 3 ms [16], and the entire refinement process could take less than 1 s. Currently, Stage 1 of the optimization algorithm only used five of the six parameters associated with a rigid transformation. The rotation about the LUS probe axis is not refined. In the future, we will include the refinement of this parameter in our algorithm. In addition, determining how often the vision-based refinement should be repeated during AR visualization will also be one of our areas of investigation.

**Acknowledgement.** This work was supported by the National Institutes of Health/National Cancer Institute under Grant CA192504.

## References

1. Leven, J., et al.: DaVinci canvas: a telerobotic surgical system with integrated, robot-assisted, laparoscopic ultrasound capability. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 811–818. Springer, Heidelberg (2005). doi:[10.1007/11566465\\_100](https://doi.org/10.1007/11566465_100)
2. Pratt, P., Jaeger, A., Hughes-Hallett, A., Mayer, E., Vale, J., Darzi, A., Peters, T., Yang, G.Z.: Robust ultrasound probe tracking: initial clinical experiences during robot-assisted partial nephrectomy. *Int. J. Comput. Assist. Radiol. Surg.* **10**(12), 1905–1913 (2015)
3. Feuerstein, M., Reichl, T., Vogel, J., Traub, J., Navab, N.: New approaches to online estimation of electromagnetic tracking errors for laparoscopic ultrasonography. *Comput. Aided Surg.* **13**(5), 311–323 (2008)
4. Bouget, D., Allan, M., Stoyanov, D., Jannin, P.: Vision-based and marker-less surgical tool detection and tracking: a review of the literature. *Med. Image Anal.* **35**, 633–654 (2017)
5. Feuerstein, M., Mussack, T., Heining, S.M., Navab, N.: Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection. *IEEE Trans. Med. Imaging* **27**(3), 355–369 (2008)
6. Shekhar, R., Dandekar, O., Bhat, V., Philip, M., Lei, P., Godinez, C., Sutton, E., George, I., Kavic, S., Mezrich, R., Park, A.: Live augmented reality: a new visualization method for laparoscopic surgery using continuous volumetric computed tomography *Surg. Endosc.* **24**(8), 1976–1985 (2010)
7. Kang, X., Azizian, M., Wilson, E., Wu, K., Martin, A.D., Kane, T.D., Peters, C.A., Cleary, K., Shekhar, R.: Stereoscopic augmented reality for laparoscopic surgery. *Surg. Endosc.* **28**(7), 2227–2235 (2014)
8. Cheung, C.L., Wedlake, C., Moore, J., Pautler, S.E., Peters, T.M.: Fused video and ultrasound images for minimally invasive partial nephrectomy: a phantom study. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6363, pp. 408–415. Springer, Heidelberg (2010). doi:[10.1007/978-3-642-15711-0\\_51](https://doi.org/10.1007/978-3-642-15711-0_51)
9. Liu, X., Kang, S., Plishker, W., Zaki, G., Kane, T.D., Shekhar, R.: Laparoscopic stereoscopic augmented reality: toward a clinically viable electromagnetic tracking solution. *J. Med. Imaging (Bellingham)* **3**(4), 045001 (2016)
10. Franz, A.M., Haidegger, T., Birkfellner, W., Cleary, K., Peters, T.M., Maier-Hein, L.: Electromagnetic tracking in medicine – a review of technology, validation, and applications. *IEEE Trans. Med. Imaging* **33**(8), 1702–1725 (2014)

11. Lasso, A., Heffter, T., Rankin, A., Pinter, C., Ungi, T., Fichtinger, G.: PLUS: open-source toolkit for ultrasound-guided intervention systems. *IEEE Trans. Biomed. Eng.* **61**(10), 2527–2537 (2014)
12. Liu, X., Plishker, W., Zaki, G., Kang, S., Kane, T.D., Shekhar, R.: On-demand calibration and evaluation for electromagnetically tracked laparoscope in augmented reality visualization. *Int. J. Comput. Assist. Radiol. Surg.* **11**(6), 1163–1171 (2016)
13. Danelljan, M., Häger, G., Khan, F., Felsberg, M.: Accurate Scale Estimation for Robust Visual Tracking. *BMCV* (2014)
14. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE Trans. Image Process.* **10**(2), 266–277 (2001)
15. Nelder, J.A., Mead, R.A.: A simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965)
16. van den Braak, G.-J., Nugteren, C., Mesman, B., Corporaal, H.: Fast Hough Transform on GPUs: exploration of algorithm trade-offs. In: Blanc-Talon, J., Kleihorst, R., Philips, W., Popescu, D., Scheunders, P. (eds.) *ACIVS 2011. LNCS*, vol. 6915, pp. 611–622. Springer, Heidelberg (2011). doi:[10.1007/978-3-642-23687-7\\_55](https://doi.org/10.1007/978-3-642-23687-7_55)
17. van der Walt, S., Schönberger, J., Nunez-Iglesias, J., Boulogne, F., Warner, J., Yager, N., Gouillart, E., Yu, T., The scikit-image Contributors: scikit-image scikit-image: Image processing in Python. *PeerJ* **2**, e453 (2014). <http://dx.doi.org/10.7717/peerj.453>