System Designs for Augmented Reality Based Ablation Probe Tracking

Hao Bo Yu and Harvey Ho(✉)

Auckland Bioengineering Institute, The University of Auckland,
70 Symonds Street, Auckland 1010, New Zealand
harvey.ho@auckland.ac.nz

Abstract. In this paper we present two Augmented Reality (AR) systems and associated algorithms to track and visualise a surgical ablation probe. The first system is based on the Kinect sensor while the second system makes use of a stereo-vision camera (OvrVision Pro) and a Head Mounted Display (HMD) device. Both systems utilise the fiducial markers on a custom-built rig attached to the ablation probe. We applied the first AR system to the navigation of a virtual liver, and the second AR system to the prediction of the 3D position of probe tip. The predication error for the tip was about 5–10 mm, with a computational speed of 10 FPS. In conclusion, two AR systems were designed and implemented with potential for further improvements to be applied in an actual clinical context.

Keywords: Augmented reality · Ablation probe · Computer vision Virtual model

1 Introduction

Ablation is used to destroy liver tumours in-situ. It can be a primary curative option for small liver cancers (<3 cm) such as hepatocellular carcinomas and colorectal liver metastases, and is ideal for those that cannot undergo surgery. It can be performed percutaneously (under radiological guidance) or surgically (either laparoscopically or at open operation). The procedure involves the insertion of a gauge applicator into the tumour centre to destroy tumour cells using radiofrequency or microwave energy to generate heat. The placement accuracy of the probe tip is critical to ensure the destruction of the entire tumour whilst limiting damage to healthy cells. Current surgical ablation procedures typically use 2D ultrasound (US) images displayed on a monitor to guide the advancement of ablation probes. The US images are then compared with pre-surgical 3D CT, MRI images [1,2]. This practice poses challenges to a surgeon in interpreting the information from 2D images to guide the 3D positioning of the ablation probe [1,10], thus introduces errors due to *ad hoc* judgements of the location of the needle tip. Adding to the error are respiration and cardiac motion induced organ displacements and needle deflection [3,4].

© Springer International Publishing AG, part of Springer Nature 2018
M. Paul et al. (Eds.): PSIVT 2017, LNCS 10749, pp. 87–99, 2018.
https://doi.org/10.1007/978-3-319-75786-5_8

To address this problem, surgical navigation systems have been proposed to aid the prediction of 3D positions of the ablation probe. Some proposed solutions involve the use of robotic assistance [5,7]. Others use US images registered with pre-operational CT/MRI images [2]. In particular, CT images and intra-operative video are used to create an augmented reality (AR) guidance system [9]. While these all address some of the issues of the probe navigation problem, they each possess their own problems. These include the operator having to look away at a separate screen and/or the visualisation is only in 2D and still requires interpretation and mental processing to visualise the 3D position. This raises questions whether a head mounted displays (HMD) device could be utlised in this context. Indeed it was shown, as early as 2001 that AR system with HMD produced better results than traditional ablation ultrasound guided methods [10]. However, most HMD devices and their associated vision algorithms used in the surgical context are proprietary or in-house built, not based on consumer grade products.

The aim of this work is therefore to design and build an AR system using inexpensive, commercially available products. After testing the systems in simulated surgical scenes and phantoms we hope to gain insights of more sophisticated AR systems that are capable of yielding accurate and fast predication of probe positions to surgeons.

2 Methods

2.1 Camera Model

Regardless of camera systems to be used, the core of computer vision (CV) is to relate 2D image points and their corresponding 3D object points. This correspondence is built at first through the process of camera calibration, where the intrinsic, extrinsic and distortion parameters of a vision camera are found. Specifically, an AR scene setup requires the *projection* matrix P, which comprises of the *intrinsics* matrix (M), the *extrinsic* parameters (R and $\overrightarrow{T}$) as:

$$P = M[R|\overrightarrow{T}]$$

The pixel coordinates can be related to the 3D coordinates by the extrinsic parameters as:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = [P] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

where u, v are the pixel coordinates, w is associated with depth, X, Y, Z are the 3D coordinates of an object point. The problem we need to solve is the Perspective-n-Points (PnP) problem [8], and can be stated as: given a set of object points (3D) to image points (2D) pairs how to find a projection matrix, P, such that it satisfies Eq. (1) for all 3D to 2D pairs.

Different methods can be used to solve the PnP problem, including but not limited to, Levenberg-Marquart optimization, P3P, and EPnP methods [13]. The open source CV library OpenCV includes the cv::solvePnP function that solves for the projection matrix P. Our methodology requires *intrinsic* values to be known beforehand, but methods such as Direct Linear Transformation (DLT) [8] solve for P without knowing the *intrinsic* parameters. However, knowing the *intrinsic* properties can reduce the number of correspondences required, for instance, in our probe navigation implementation only four correspondences were used.

2.2 Ablation Probe and Custom-Built Rig

The AR setup is based on an actual microwave ablation probe (Acculis Accu2i), which is used in the Auckland City Hospital, New Zealand for tissue ablation (Fig. 1a). The ablation probe's applicator, or the shaft, is the part inserted into

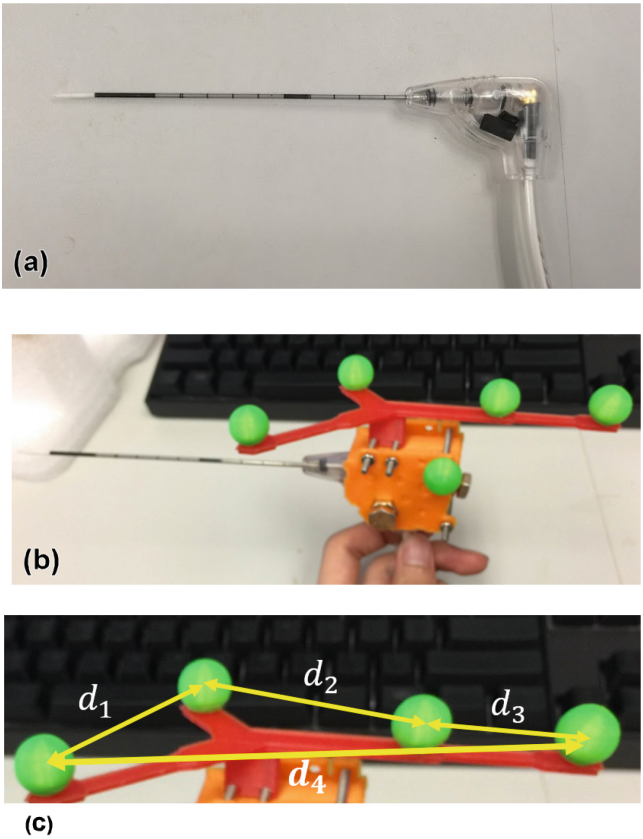

Fig. 1. (a) Accu2i ablation probe; (b) Removable rig designed for the probe; (c) Distances between individual fiducials for error estimation.

an organ, and heat released from the leading section of the probe serves the purpose of thermal coagulation. The probe is capable of ablating a tissue size of 4.5 cm × 5.5 cm in 6 min (per Accu2i manual). Hence it is critical to accurately predict the location of the applicator, even when it is hidden, e.g., when it is inside a soft phantom. This is one of the challenges faced by surgeons, and the motivation of our work.

In order to monitor the movements of the probe, a rig system with fiducial markers was custom-designed and built (Fig. 1b). The rig was first designed in SolidWorks (Dassault Systemes) by taking measurements from the Accu2i ablation probe. The rig is designed to be removable, and has three screws that holds the probe firmly. The rig has three parts: a base in orange colour, a frame in red colour and fiducial markers (Fig. 1b). The red frame functions as a holder for the fiducial markers, and can be made in different shapes depending on users' need.

In terms of fiducial markers, two kinds of markers were used. Firstly RGB markers were used to complement the RGB stereo camera used. Secondly, the markers can be switched to retro-reflective (or self reflecting) markers suitable for an infra-red camera. RGB markers were used in this work for their simplicity and availability; it is compatible with the two vision system used, i.e. an OvrVision Pro based system and a Kinect-based system, to be introduced later.

Lastly, all rig parts including the base, frame, and markers were 3D printed except infra-red reflective markers, which are commercially available.

2.3 Pose Estimation for the Probe

The alignment of two sets of data cloud is a common problem in computer science. Three methods are commonly used, namely Principle Component Analysis (PCA), Singular Value Decomposition (SVD), and the Iterative Closest Point (ICP) methods [12]. In particular, ICP is a method that does not require correspondence between the initial and target data points [1]. However ICP is used for large data clouds, thus not suitable for our application. SVD and PCA, on the other hand, require correspondence between the two sets of data clouds. We use an exhaustive search method, similar to that of [6], to find the correspondence between a set of initial points, markers on the rig, and the corresponding set of target points, which are detected 3D points from depth sensors or stereo vision cameras.

Figure 2 explains the algorithm. On the left we have the *template* geometry ($\mathcal{V}$), with known relative distances between each of the nodes. On the right side, we have a set of nodes ($\mathcal{W}$) detected with vision systems. **Each *template* node in $\mathcal{V}$ needs to correspond to a *world* marker in $\mathcal{W}$**, and this is the problem we need to solve.

The *template* points are the *initial* points, and they correspond to the actual markers on the surgical probe (Fig. 1). A closed loop was used to connect template points, for instance $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$, and the distance between each node i.e. the edge length in the loop can be determined. If $n = 3$ we have:

$$\mathcal{L} = \{l_1 = ||\boldsymbol{t_1} - \boldsymbol{t_2}||, l_2 = ||\boldsymbol{t_2} - \boldsymbol{t_3}||, l_3 = ||\boldsymbol{t_3} - \boldsymbol{t_1}||\} \tag{2}$$

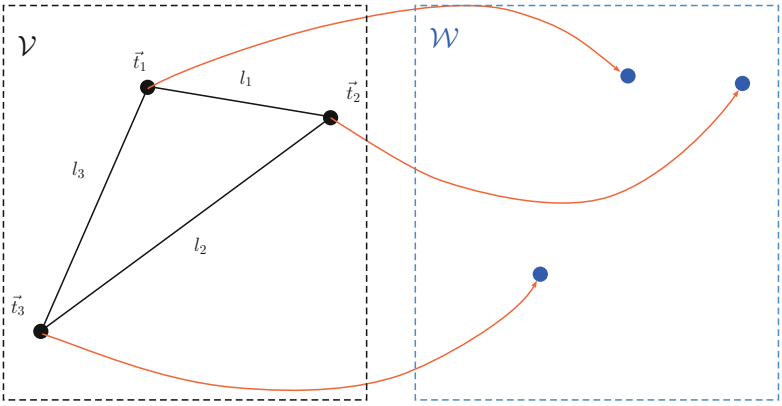

Fig. 2. This figure illustrates the correspondence problem. The left figure shows the template points $\mathcal{V} = \{\boldsymbol{t}_1, \boldsymbol{t}_2, \boldsymbol{t}_3\}$. On the right, the three world points $\mathcal{W}$ detected by vision systems are represented by blue dots. (Color figure online)

After establishing the set of edge lengths $\mathcal{L}$, we move on to retrieve the 3D positions of the vertices from depth sensors.

Assuming the raw 3D position is unordered, $\mathcal{W}_{\text{unordered}} = \{\boldsymbol{p}_1, \boldsymbol{p}_3, \boldsymbol{p}_2\}$, all possible loops that can be formed by these points are listed in Eq. 3. For each scenario, $d_1, \ldots, d_3$ represents the distances between each node pair. Our goal now is to find which set of distances is the closest to the template distances $\mathcal{L}$.

$$\begin{aligned}
loop_1 &:= p_1 \xrightarrow{d_1} p_2 \xrightarrow{d_2} p_3 \xrightarrow{d_3} p_1 \\
loop_2 &:= p_1 \xrightarrow{d_1} p_3 \xrightarrow{d_2} p_2 \xrightarrow{d_3} p_1 \\
loop_3 &:= p_2 \xrightarrow{d_1} p_1 \xrightarrow{d_2} p_3 \xrightarrow{d_3} p_2 \\
loop_4 &:= p_2 \xrightarrow{d_1} p_3 \xrightarrow{d_2} p_1 \xrightarrow{d_3} p_2 \\
loop_5 &:= p_3 \xrightarrow{d_1} p_1 \xrightarrow{d_2} p_2 \xrightarrow{d_3} p_3 \\
loop_6 &:= p_3 \xrightarrow{d_1} p_2 \xrightarrow{d_2} p_1 \xrightarrow{d_3} p_3
\end{aligned} \tag{3}$$

This is achieved by computing Eq. 4 for every loop from 1 to 6, where l_1, l_2, l_3 represents the template lengths. The loop that gives us the minimum value can be used to reorder $\mathcal{W}_{\text{unordered}}$.

$$||l_1 - d_1|| + ||l_2 - d_2|| + ||l_3 - d_3|| \tag{4}$$

For example, if $loop_5$ is the minimum to the above equation, then the ordered 3D point will be $\mathcal{W}_{\text{ordered}} = \{\boldsymbol{p}_3, \boldsymbol{p}_1, \boldsymbol{p}_2\}$. The above exhaustive algorithm was implemented for the case when $n = 4$. Once correspondence is established, SVD (in the PCL library [11]) was used to find the rotation $[R]$ and translation $[T]$ between the *template* points and the ordered 3D points $\mathcal{W}_{\text{ordered}}$, as detected by vision systems which were used in the work, as introduced below.

The idea behind the algorithm is explained by taking the number of data points $n = 3$, however the same concept applies to more data points.

2.4 Kinect-Based AR System

Kinect Sensors. The Kinect V2 sensor (Microsoft) was primarily aimed at the virtual 3D gaming community for the 3D gaming console XBox, however it is also used in scientific research, in particular in those areas that require depth information. The Kinect has a RGB *colour camera* with a resolution of 1920×1080 pixels, and an infra-red camera with a resolution of 512×424 pixels to capture depth information. The refresh rate for both cameras is 30 frames per second (FPS).

In a 3D virtual surgical guidance system, all elements important to surgeons during an ablation surgery need to be embodied, namely the organ, the probe, and the surgical world. In this system we used the depth sensor for the detection of the world or the environment, and the RGB sensor to detect fiducials and markers.

This process was achieved in three main steps. In the first step, 3D fiducials are detected by the Kinect. Next, the above exhaustive method and SVD are used to find the relative rotation $[R]$ and translation T between the *template* points and the 3D points detected from the Kinect. Thirdly the applicator is transformed with the same transformations (Fig. 3).

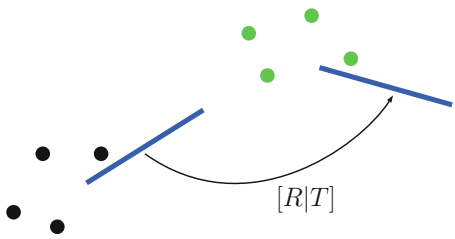

Fig. 3. Black dots represents the *template* points, and the green dots represents the 3D points retrieved from the Kinect. The blue line represents the applicator. (Color figure online)

Virtual Liver. Instead of applying the system in an actual surgical sense, we used a virtual liver model digitised from the Visible Human dataset, which contains the liver surface and the portal and hepatic veins (to be shown in Results). The virtual liver mesh was anchored to the AR scene by two ArUco markers shown in Fig. 4. All relevant geometries were then visualised in an OpenGL viewer environment allowing the user to rotate, pan, and zoom in the environment.

In the Kinect based system, the Kinect sensor was mounted on a tripod, and its location was fixed. The ArUco markers on the desk were also static, though they could be moved anywhere in the scene (Fig. 4). The moving element of the

system is the probe and the rig attached to it. Next we introduced a HMD-based system where both the cameras and the probe are in motion in the scene, thus poses more challenges.

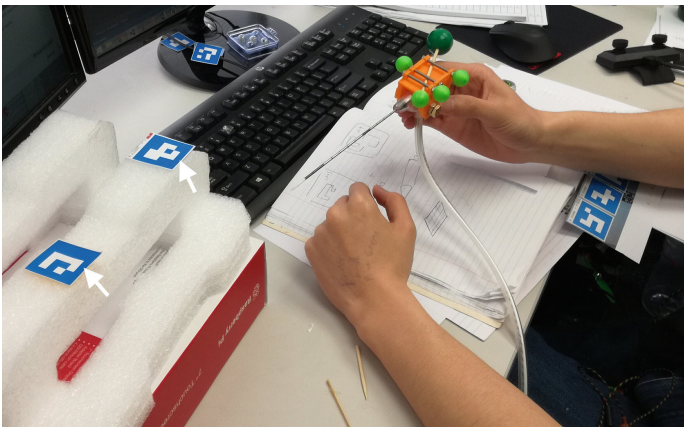

Fig. 4. Setup of Kinect-based system: The kinect was mounted on a tripod, from where this image was taken. Two ArUco markers, indicated by arrows, were used as fiducials to locate the virtual liver mesh in an AR scene. Note, the ArUco markers were different from the fiducials on the rig and they serve different purposes.

2.5 HMD and Stereo Vision-Based AR System

Stereo Vision Camera. The *OvrVision Pro* stereo camera system, shown in Fig. 5(a) is the product of the Wizapply company in Japan (www.ovrvision.com). It has two 1.8 MP cameras, with *optical* centres about 6 cm apart. This camera was originally designed for HMD-based AR applications. It possesses a large *barrel* distortion, and is corrected based on the principle that the corrected image is a function of the radial distance between the distorted points and the undistorted image centre r. This is achieved in the camera calibration process, which is explained in many texts and is not repeated here. We refer the interested reader to [13] for further reading. The result of the correction matrix for the OvrVision Pro camera is:

$$[M]_{\text{Left}} = \begin{bmatrix} 706.07 & 0 & 441.39 \\ 0 & 706.4 & 492.67 \\ 0 & 0 & 1 \end{bmatrix} [M]_{\text{Right}} = \begin{bmatrix} 703.31 & 0 & 482.2 \\ 0 & 703.19 & 477.62 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

The comparison with barrel effects of the OvrVision cameras before and after corrected is shown in Fig. 6.

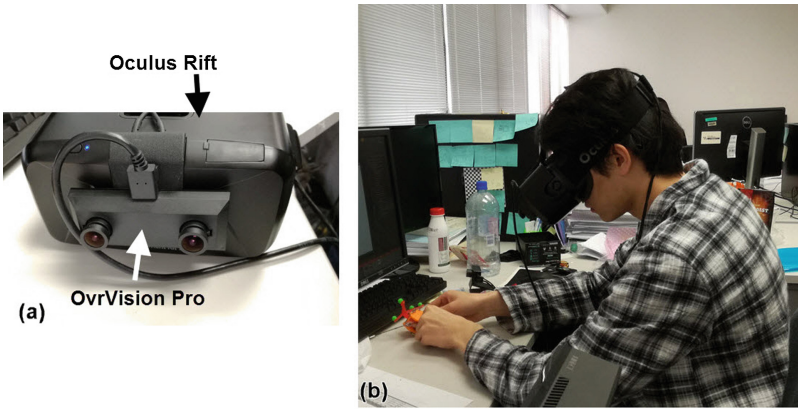

Fig. 5. Setup of the HMD-based AR environment: (a) An OvrVision Pro stereo vision camera is mounted on a Oculus Rift; (b) The AR scene is viewed from the Oculus goggles with real-time video stream captured by OvrVision. See Sect. 3.2 for an actual AR scene viewed from the goggles.

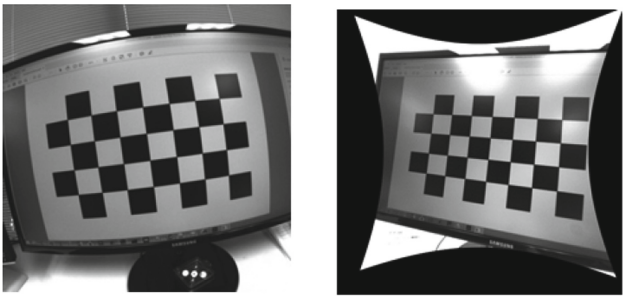

Fig. 6. Barrel effects of the stereo vision system is corrected.

Oculus Rift and OvrVision. To create an HMD-based AR environment, the OvrVision sensor was mounted on the Oculus Rift DK2, which is a virtual reality (VR) device used to visualise computer generated virtual environments by using stereoscopic displays. Oculus Rift itself does not have an interface to view the physical world. Rather, OvrVision was used to provide stereo-images and overlay them on the Oculus image frames. Indeed, the OvrVision Pro was designed for such purposes, and it has a mechanism for easy attachment on to the front of the Oculus DK2 (Fig. 5a).

Image fusion between the Oculus Rift and OvrVision begins by locating the ablation markers in 3D by using stereo-triangulation. Then, the exhaustive search method outlined above and the SVD method were used to find the transformation matrix $[R|T]$. This matrix is used to transform the applicator from its previous state to its new orientation.

Soft Phantom. A soft phantom was made from a gelatine powder (DAVIS Gelatine) combined with water at a concentration of 12% weight of gelatine. The size of the phantom is 150 mm × 150 mm × 100 mm. The phantom is placed inside a plastic box. Furthermore, a CharUco pattern (or chessboard + ArUco markers) [13] was placed over the gelatine phantom to create a virtual tissue block in the AR scene (to be shown in Results).

3 Results

The AR systems were run on a Intel i7-4790 CPU @ 3.60 GHz, RAM 32 GB, GPU 2 GB, with a NVIDIA 745 (OEM) GPU. The programming language is C++ in Visual Studio 2013.

3.1 Kinect-Based AR System

Figure 7 shows four screenshots of the 3D AR system in action. The purple cylinder represents the ablation applicator (the rig is currently not included). Figure 7(a) shows the relative position between the applicator and the virtual liver. The data cloud generated from the RF sensor is also visible in the background. Figure 7(b) is the close-up view, where the four white points represent the virtual fiducials whose transformation applies to the applicator as well. The blue points are the actual fiduals in the rig detected by the Kinect RGB camera.

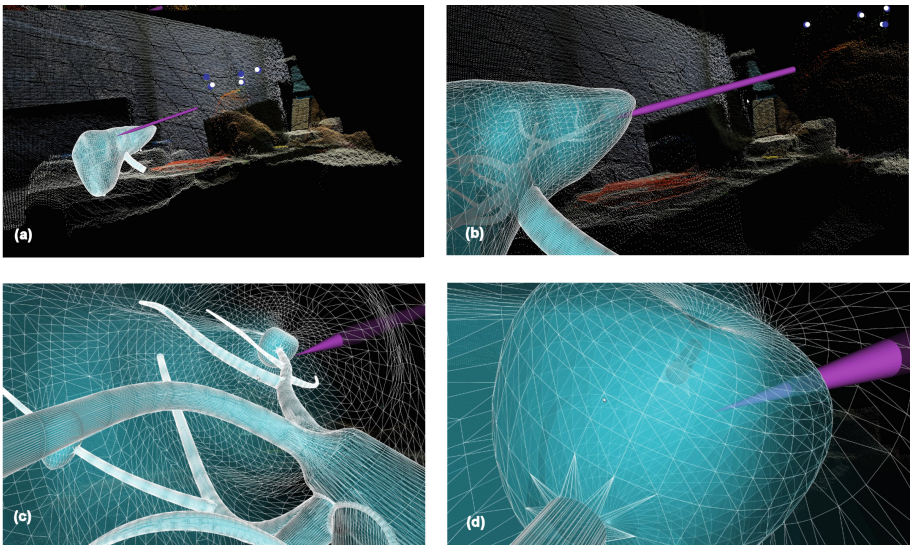

Fig. 7. Snapshots of the ablation visualisation system: (a) and (b) Relative position of the probe and the liver, and the environment as captured by Kinect; (c) The user is able to view inside of the liver, where the internal vessels geometry and tumour are visualised; (d) The virtual probe applicator approaches the tumour. (Color figure online)

The user is able to rotate, pan and zoom the applicator by using the mouse and keyboard. By doing so, the user is able to navigate within the liver mesh and visualise the geometries internally. In Fig. 7(d) the ablation tip can be seen inside of a tumour. This ablation visualiser system performed at 8–15 FPS. The tip of the applicator probe had an error of less than 10 mm.

3.2 HMD-Based AR System

The result of the HMD-based AR system is shown in Fig. 8. The left and right images are the video images viewed from the left and right lenses of the OvrVision stereo vision camera. The white line projected on the Oculus frames represents the ablation probe applicator, and the applicator tip is represented by a yellow circle. The virtual applicator follows the movements of the actual probe applicator (Fig. 8a). The CharUco described in Sect. 2.5 acts as landmarks for the virtual phantom, currently represented by a square wire frame (Fig. 8b).

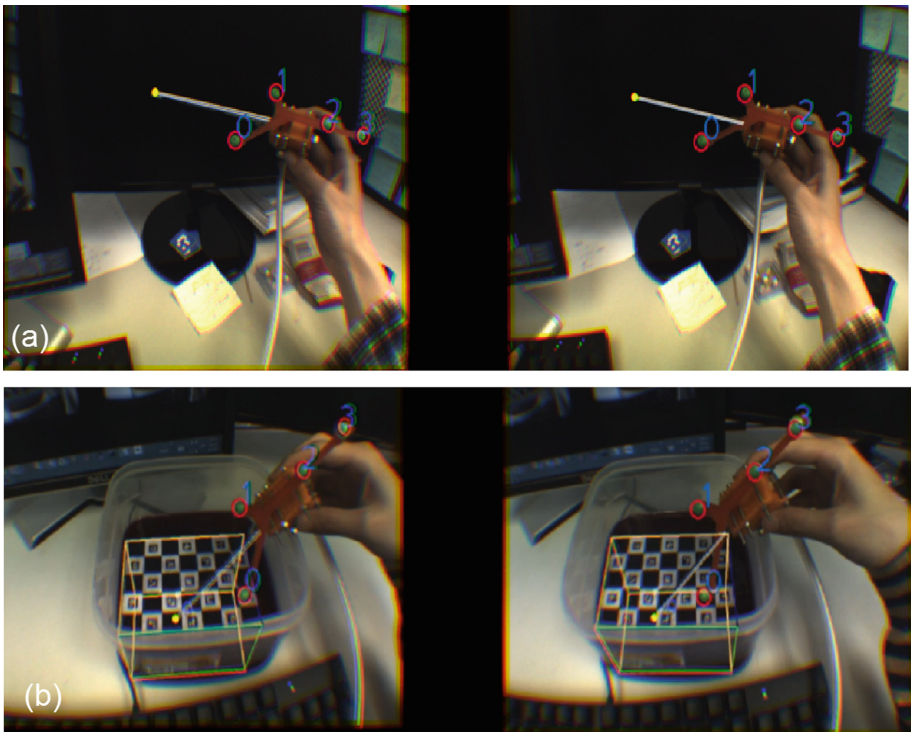

Fig. 8. Results of the HMD-based AR system: (a) The applicator is represented by a white line, the tip by a yellow circle; (b) The tip of the applicator and parts of the applicator is inserted into the phantom and is not visible, but by projecting the lines in Oculus HMD the 3D position of the tip becomes known to the user. (Color figure online)

Note that in Fig. 8(b) the probe was actually inserted into the phantom, i.e. the probe tip was invisible to cameras but was visible in the virtual AR scene to indicate the position of the applicator tip. This is exactly what we aimed to achieve, to inform surgeons where the tip reaches inside an organ.

3.3 Error Evaluation for the System

The error in the HMD-based system was evaluated in the following experiment. The distances between markers, $d_1, \ldots, d_4$ in Fig. 1(c), was compared with that computed from the stereo camera when the ablation probe was rotated and translated in space. The RMS errors (mm) and standard deviation (mm) are summarised in Table 1. These errors are slightly larger than literature [14], where they have achieved RMS values between -2 and 2 mm. Possible reasons include the OvrVision camera having a large lens distortion, and also possibly more camera calibrations are required. Future work should use stereo-vision systems with less *barrel* distortions.

Table 1. RMS error and standard deviation for measured distances.

	Dimension (mm)	RMS (mm)	Standard deviation (mm)
d_1	59	2.17	0.82
d_2	60	2.76	1.45
d_3	39	1.81	1.69
d_4	135	3.52	0.93

4 Discussion

The use of virtual reality and augmented reality in surgery is a growing area of research. They have been implemented in various ways including the projection of patient specific information from pre-operative scans onto the patient, or overlaying such information on live video footage which is displayed on a screen (for a brief review see [9]). Compared to VR applications, an AR application has an additional computation layer, which is to analyse a real-time video stream where a virtual model can be overlaid. This can cause computational bottlenecks if the AR system is not designed and implemented properly.

In this work, two different systems for the navigation and visualisation of ablation probes are presented. Four optical markers were used for both systems, with an exhaustive method to find correspondences between template points and corresponding 3D points. In the first system, the user was able to navigate in a 3D virtual organ. Through this system, a surgeon can have control over the 3D AR scene by using the keyboard and mouse to navigate to regions of interest, as illustrated in Fig. 7. However, such a 3D visualisation approach has

its own drawbacks, including the operator has to look away at a separate screen. Moreover, several aspects need addressing. Firstly, the prediction error (∼10 mm) of the applicator tip is rather large for such applications. This error was likely a result of the accumulation of errors from the individual markers because the tip is far from the marker arrangement, any small error from the markers is multiplied and passed to the tip. Secondly, the computation speed of 8 FPS is too slow for an actual clinical application.

In the second system, the HMD-based AR system overlay the virtual applicator onto the vision field, creating the illusion that the virtual object was "real" by providing spatial position of the tip to the user. After insertion of the applicator into the phantom, the portion inside of the gelatine becomes obstructed, but the applicator is augmented, allowing the user to see the occluded tip position. This system can be used as a core algorithm for an advanced ablation probe tracking system; for example, an integrated system with both preoperative CT/MRI data, and probe position relative to the patient, in real time. An ongoing work in our group is to use an abdominal biopsy phantom in conjunction with its corresponding MR images to create the AR scene.

Both proposed systems suffer from a high computation time thus further efforts are required to reduce the computation cost. For instance, retro-reflective markers can reduce the search space of the circle detection algorithm and increase the overall speed of the system. Lastly, a Finite Element solver can be added to update tumour positions accurately in order to mitigate for patient movements during surgery.

5 Conclusion

Two Augmented Reality based systems for microwave ablation probe were designed and implemented in this work. With further improvements on computational efficiency and rig design, the system has potential to be used in clinical scenarios for surgical training and probe navigation.

References

1. Jolesz, F. (ed.): Intraoperative Imaging and Image-Guided Therapy. Springer Science & Business Media, New York (2014). https://doi.org/10.1007/978-1-4614-7657-3
2. Ward, T., Goldman, R., Weintraub, J.: Electromagnetic navigation with multimodality image fusion for image-guided percutaneous interventions. Tech. Vasc. Intervent. Radiol. **16**(3), 177–181 (2013)
3. Clifford, M., Banovac, F., Levy, E., Cleary, K.: Assessment of hepatic motion secondary to respiration for computer assisted interventions. Comput. Aided Surg. **7**(5), 291–299 (2002)
4. Abolhassani, N., Patel, R., Moallem, M.: Needle insertion into soft tissue: a survey. Med. Eng. Phy. **29**(4), 413–431 (2007)

5. Loser, M.H., Navab, N.: A new robotic system for visually controlled percutaneous interventions under CT fluoroscopy. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 887–896. Springer, Heidelberg (2000). https://doi.org/10.1007/978-3-540-40899-4_92
6. Srikrishna, B.K., Musti, U., Heikkil, J.: Geometry based exhaustive line correspondence determination. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp. 4341–4348. IEEE (2016)
7. Patriciu, A., Awad, M., Solomon, S.B., Choti, M., Mazilu, D., Kavoussi, L., Stoianovici, D.: Robotic assisted radio-frequency ablation of liver tumors – randomized patient study. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 526–533. Springer, Heidelberg (2005). https://doi.org/10.1007/11566489_65
8. Okuma, T., Sakaue, K., Takemura, H., Yokoya, N.: Real-time camera parameter estimation from images for a mixed reality system. In: Proceedings of the 2000 15th International Conference on Pattern Recognition, vol. 4, pp. 482–486. IEEE (2000)
9. Nicolau, S., Pennec, X., Soler, L., et al.: An augmented reality system for liver thermal ablation: design and evaluation on clinical cases. Med. Image Anal. **13**, 494–506 (2009)
10. Rosenthal, M., State, A., Lee, J., Hirota, G., Ackerman, J., Keller, K., Fuchs, H.: Augmented reality guidance for needle biopsies: an initial randomized, controlled trial in phantoms. Med. Image Anal. **6**(3), 313–320 (2002)
11. Rusu, R.B., Cousins, S.: 3D is here: point cloud library (PCL). In: 2011 IEEE International Conference on Robotics and automation (ICRA), pp. 1–4. IEEE (2011)
12. Kramer, J., Burrus, N., Echtler, F., Daniel, H.C., Parker, M.: Hacking the Kinect, vol. 268. Apress, New York (2012)
13. Forsyth, D., Ponce, J.: Computer Vision: A Modern Approach. Pearson, New York (2015)
14. Ren, H., Liu, W., Lim, A.: Marker-based surgical instrument tracking using dual kinect sensors. IEEE Trans. Autom. Sci. Eng. **11**(3), 921–924 (2014)