A 3D Augmented Reality Training System for Endoscopic Surgery

Rong Wang[1,2], Zheng Geng[1], Xiangbing Meng[1,2], and Renjing Pei[1,2]
[1]Institute of Automation, Chinese Academy of Sciences, Beijing, China
[2]University of Chinese Academy of Sciences, Beijing, China
Contact Author Email: wangrong2013@ia.ac.cn

Abstract
We propose a 3D augmented reality (AR) system for perception training on the ablation of tumor inside the kidney. The system consists of a region-based visual tracking for localizing kidney and tumor and a 3D display for visualizing AR results. We also integrate a virtual instrument in the system.

Author Keywords
augmented reality; visual tracking; 3D display; minimally invasive surgery.

1. Introduction
Augmented reality (AR) refers to a technique that combines real and virtual objects in a single visualization. It is popular in minimally invasive surgery (MIS) and aims to enhance surgeons' visual perception. In typical MIS, an endoscope is inserted in a patient's body through a small incision, and a surgeon observes the captured endoscopic images from a display monitor. Significant challenges in this process include surgeon's limited visual perception and the fact that the surgeon cannot perceive vital structures (e.g. tumor or vessels) underneath the visible surface. Medical AR is achieved by laying over the endoscopic image with hidden structures which are acquired from various imaging modalities, such as computed tomography (CT) or magnetic resonance imaging (MRI). It aims to overcome the aforementioned challenges.

In general, AR systems mainly include localization and display technologies. Localization is to obtain the accurate position of the visible object with respect to the camera. Since the relationship between the visible surface and the hidden object can be determined from preoperative 3D models, the location of the hidden object can be gotten. Display technology determines the appropriate way to present the hidden object. Various localization and display approaches are adopted by different AR systems.

Optical tracking [1] which requires extra devices is precise and widely used in medical AR. However, it is impractical to track internal endoscope's motion in MIS since this method needs to fix optical markers on the target object. Therefore, visual localization approaches attract more and more researchers' attentions. Su et al. [2] used stereo triangulation and 3D iterative closest point (ICP) registration to determine the location of the tumor. With the development of dense simultaneous localization and mapping (SLAM) systems [3], Chang [4] demonstrated that they have better performance in MIS. Since tracking in SLAM usually obtains the camera's motion, initial aligning between the CT model and the reconstructed model is also needed and can be achieved through 3D surface registration or landmark-based registration. In [5], a feature-based tracking is applied in MIS. In our system, we do tracking by adopting the algorithm from [6], which takes the region-based segmentation as a result of tracking. It proves to be convenient and robust in our application.

Instead of using traditional 2D displays, increasing number of medical AR systems attempt to integrate 3D display to bring improved quality of AR visualization. Most of them have focused on the use of Head Mounted Displays (HMDs) which may not be convenient and comfortable for surgeons in the operating room. Therefore, 3D display without eyeglasses, which is also called autostereoscopic 3D display, is proposed for medical AR. Since surgeons observe surgical scenes through recorded endoscopic images, we believe a multi-view autostereoscopic lenticular LCD is feasible to be utilized in MIS. This 3D display uses existing 2D screen fabrication infrastructure. Its cost is relatively low and its implementation is simple. Visualization mode is another important factor in AR. Good modes can provide surgeons with the correct spatial perception. Although this paper does not discuss this topic in detail, we present final display results with different modes proposed in [7].

The goal of this work is to design and implement a 3D AR training system for MIS on kidney tumor. The system consists of a region-based localization and a lenticular-based 3D display. We also integrate a virtual instrument in it for interaction. Our AR-based training is more conformable to reality compared with traditional virtual reality (VR) based training. The training aims to improve surgeons' visual perception, dexterity, and accuracy of surgical manipulation, thus significantly enhance the surgical outcomes. Our main contributions include:

1. Implemented a simultaneous segmentation and tracking method to locate the target object;
2. Designed and implemented an autostereoscopic 3D display for medical AR system;
3. Integrated a virtual instrument in the AR system for surgical training.

The rest of the paper is organized as follows. After this introduction, the principle and implementation of localization, 3D display, and interaction are introduced in Section 2. Next, some experimental results are presented and evaluated in Section 3. Finally, conclusions are drawn in Section 4.

2. Technical Methods
The overview of our medical AR training system is shown in **Figure 1**. Pre-processing, which is not listed in the overview, only has to be performed once. It includes camera calibration and CT segmentation. From CT segmentation, we can get models of kidney and tumor. The main process consists of tracking, autostereoscopic 3D display and virtual instrument interaction. We use a 3D printed kidney model in our simulated experiment. It is segmented from CT images. We add some artificial textures on its surface to imitate real scenario. We will describe technical details in the following parts.

Localization: In our system, we adopted the tracking algorithm first proposed in [6] to obtain the motion of the kidney. For simplicity, we assume that the kidney is a non-deformable rigid body, thus the tumor's motion is exactly the same as that of the kidney. The spatial relationship between the kidney and the tumor is obtained from pre-operative CT. In the algorithm, 2D image segmentation and 3D pose tracking are simultaneously processed.

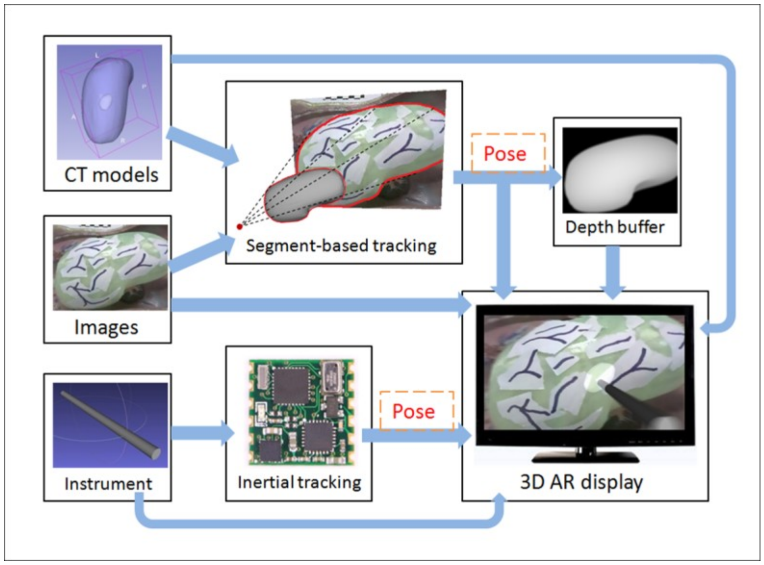

Figure 1. The overview of our medical AR training system.

This method requires the to-be-tracked object's 3D model first. Our system satisfies this condition because we could obtain the kidney's 3D model from pre-operative CT images. This method takes the region-based segmentation as a result of 3D tracking. Compared with other visual tracking techniques, the algorithm adopted in our system has its own advantages. It will still work when the feature-based tracking fails due to insufficient textures. It is easier than existing SLAM which is lack of scale with a single camera and also needs initial registration. We will briefly introduce the tracking algorithm in the following part. For more details, please refer to [6].

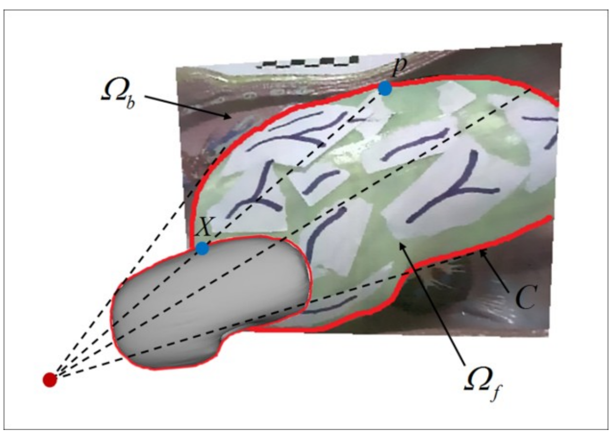

Figure 2. An illustration of the simultaneous segmentation and tracking algorithm.

In **Figure 2**, the image is denoted as I, and a pixel in the image is p whose RGB value is represented as y. A 3D point X is in the camera coordinate and it is the rigid transformation of a point X_0 in the object coordinate. The transformation has 6 degrees of freedom (3 for rotation and 3 for translation). The goal is to find the optimal 3D pose to segment the image into foreground Ω_f and background Ω_b whose color appearance models are $P(y \mid M_i), i \in \{f,b\}$. The energy function which needs to be minimized is defined in Eq. (1), where $\phi(p)$ is the level-set function and $H(\bullet)$ is the Heaviside step function. P_f and P_b, which are computed by Eq. (2), are posteriori probabilities of each pixel belonging to foreground and background. Then we differentiate the objective function with respect to each motion parameter and use gradient descent to solve the optimization problem.

$$E(\phi) = -\sum_{p\in\Omega} \log\left(H(\phi)P_f + \left(1-H(\phi)\right)P_b\right) \tag{1}$$

$$\begin{gathered} P_i = \frac{P(y \mid M_i)}{\eta_f P(y \mid M_f) + \eta_b P(y \mid M_b)}, i \in \{f,b\} \\ \eta_f = \sum_{p\in\Omega} H(\phi),\ \eta_b = \sum_{p\in\Omega}\left(1-H(\phi)\right) \end{gathered} \tag{2}$$

Autostereoscopic 3D Display: Considering the characteristics of MIS, the most feasible way to achieve 3D display is based on lenticular sheet. In our system, we use a 9-view autostereoscopic lenticular LCD to display our medical AR contents. The general pipeline of this process is shown in **Figure 3**.

Each 3D image shown on this 3D display requires inputs of 9 different images from 9 views. These 9 views can be obtained through rendering from a reference RGB image and its corresponding depth map. Since we have recovered the motion of the kidney through visual tracking, the depth value of the projected kidney model can be read from the depth buffer. We set the depth of the other part of the scene to zeros. Thus we can obtain the depth map of the current scene. It should be noted that the obtained depth map is smoothed by a Gaussian filter before used to render 9 different views in order to decrease the effect of edge artifacts. We then do the rendering process. Each 3D point which corresponds to one pixel in the endoscopic image is transformed according to the view transformation and the transformed point is projected on an image to form a new view. There will be some holes in the newly rendered view, because of discrete re-projection. We then fill each hole by the weighted average value of its neighboring pixels which also have similar depth values.

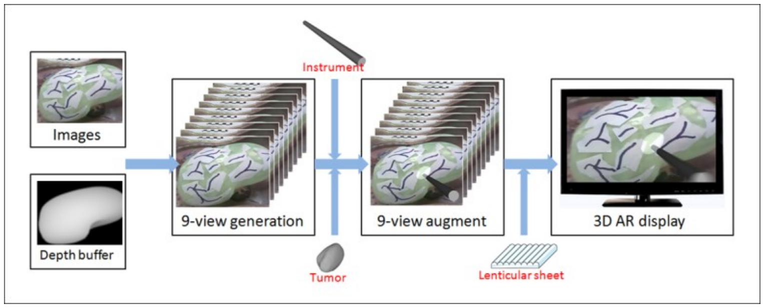

Figure 3. The general pipeline to achieve autostereoscopic 3D AR display.

After obtaining these 9 different views, we then render 3D models (including tumor and instrument) with corresponding poses on each of them. Finally, 9 rendered 2D images are used to generate a synthetic image for autostereoscopic 3D display. Visualization mode (e.g. transparency or window) is also important in AR for spatial perception. We implement several modes in our system but do not describe them in detail in this paper. For more knowledge about them, please refer to our previous work [7].

Interaction: A virtual instrument is incorporated in our medical AR training system for interaction. Surgeons can control a real-world instrument just like the real surgical tool. We use an inertial measurement unit (IMU) which is attached to the real instrument to get its pose. The virtual instrument is rendered according to this pose. There is a depth test between the virtual instrument and the kidney surface. When the instrument penetrates the kidney surface, the appearance of the instrument's inside part changes according to the adopted visualization mode. In such a way we provide a cue for the relationship between the kidney surface and the instrument. We also produce a hint when the instrument touches the tumor. We hope that with our training system's settings, surgeons can improve their visual perception abilities and manipulation dexterities in real medical AR.

3. Results

In this part we show some experimental results of our medical AR training system.

First, we compare tracking accuracy in our system against two SLAM systems which use visual tracking as well. They are parallel tracking and mapping (PTAM) [8] using sparse feature points and dense tracking and mapping (DTAM) [3] using all the pixels in the image. Because SLAM with a single camera cannot recover the real size of the scene, we capture a checker board pattern to calculate the scale. The comparison results using a video sequence are shown in **Figure 4**. We also present qualitative results of the 100th frame overlaying the tracked kidney model with 3 methods in **Figure 5**. Among the 3 methods in our case, PTAM is the worst because our experimental scene is lack of sufficient features. When X and Y axis are defined as the out-of-screen rotations, we find that there are relatively large differences between DTAM and ours in the X and Y directions. These are caused by the fact that our adopted algorithm mainly focuses on the contour. Since the object was viewed from the frontal view, a tilt in the X or Y direction had little effect on the change of the contour. Despite the limitations of our adopted algorithm, the comparison results demonstrate that it has a satisfying performance.

After obtaining the poses of tumor and the instrument from visual tracking and inertial tracking respectively, we render our final AR images for 3D display. Display results achieved by different visualization modes are shown in **Figure 6**. Transparent overlay is to overlay the tumor onto the endoscopic image by transparency. Virtual window is to open a window to see the tumor behind the visible kidney surface. Random-dot mask is to create many small holes on the visible surface. Transparent mask modulates the transparency depending on the distance to the interested area. The ghosting method assigns

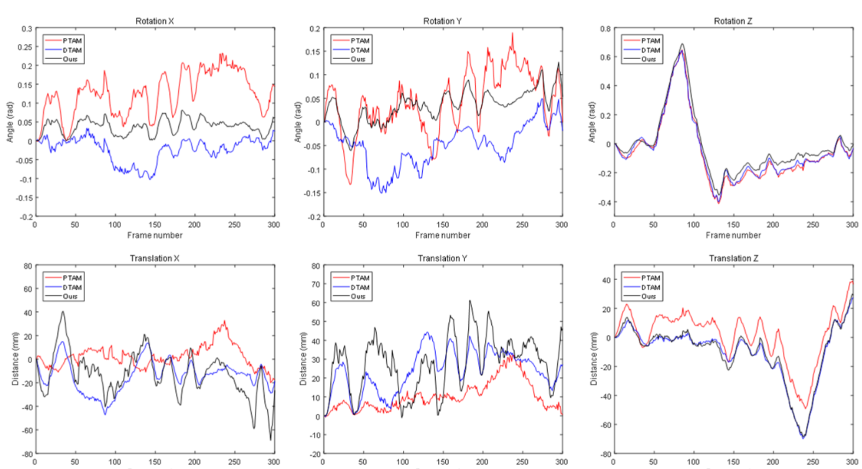

Figure 4. Tracking results with PTAM, DTAM and ours.

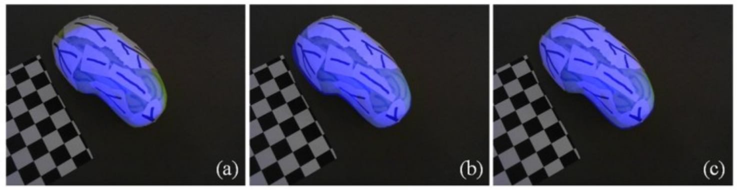

Figure 5. Qualitative tracking results of the 100th frame with: (a) PTAM; (b) DTAM; (c) ours.

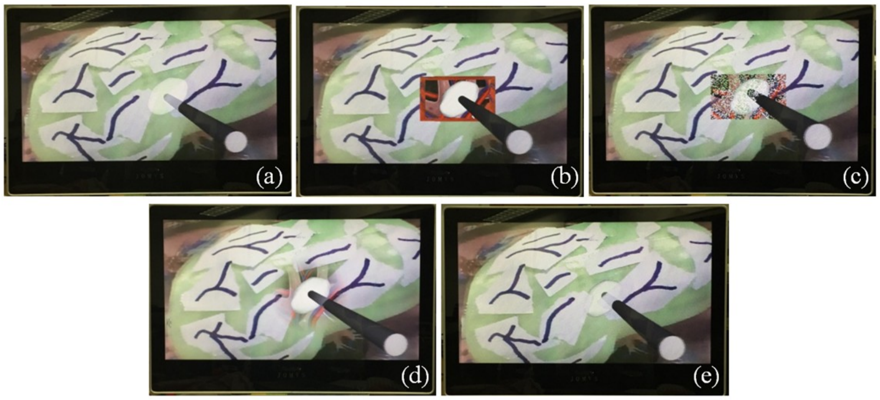

Figure 6. Autostereoscopic 3D AR display results (snapshots of the 3D display monitor's screen) with: (a) transparent overlay; (b) virtual window; (c) random-dot mask; (d) transparent mask; (e) the ghosting method.

transparency according to the saliency analysis of each pixel in the endoscopic image. The display results clearly present the relationships among the visible kidney, the tumor and the virtual instrument. Surgeons can choose visualization modes in which they feel comfortable and easy to understand. Surgeons are trained on this system in order to improve their AR perception and surgical skills with medical AR systems.

4. Conclusion

We propose a 3D AR training system for MIS in this paper. The system consists of a region-based localization, a lenticular-based 3D display, and an interactive instrument. The significance of our work can be concluded as:

- We use simultaneous segmentation and tracking in our AR system. In such a way, we can merge tracking and registration in a unified manner which is convenient and feasible to use. The comparison results with PTAM and DTAM demonstrate that the tracking algorithm adopted in our system has a satisfying performance.
- We design and implement an autostereoscopic 3D display in the medical AR system. We also design several different modes to visualize AR results in order to provide better spatial perception. The display results clearly present the relationships among the visible surface, the hidden object and the virtual instrument.
- We integrate a virtual instrument in the training system. Since many medical training systems are based on VR, our AR-based training is novel. Surgeons can experience the operating scene in a more realistic way.

5. Acknowledgements

This work was supported in part by the National High-tech R&D Program (863 Program) of Institute of Automation, Chinese Academy of Sciences (CASIA) under Grant 2012AA011903 and 2015AA015905. This work was also supported in part by the National Natural Science Foundation of China under Grant 61605240.

6. References

[1] M. Kersten-Oertel et al., “Augmented reality in neurovascular surgery: feasibility and first uses in the operating room,” Int. J. CARS **10(11)**, 1823-1836 (2015).

[2] L.M. Su et al., “Augmented Reality During Robot-assisted Laparoscopic Partial Nephrectomy: Toward Real-Time 3D-CT to Stereoscopic Video Registration,” Urology **73(4)**, 896-900 (2009).

[3] R.A. Newcombe et al., “DTAM: Dense Tracking and Mapping in Real-Time,” in ICCV, 2011, pp. 2320-2327.

[4] P.L. Chang, “Dense Vision in Image-guided Surgery,” Ph.D. dissertation, ICL, London, UK, 2014.

[5] G.A. Puerto-Souza, G.L. Mariottini, “Toward Long-term and Accurate Augmented-Reality Display for Minimally-Invasive Surgery,” in ICRA, 2013, pp. 5384-5389.

[6] V.A. Prisacariu, I.D. Reid, “PWP3D: Real-Time segmentation and tracking of 3D objects,” Int. J. Computer Vision **98(3)**, 335-354 (2012).

[7] R. Wang et al., “Visualization Techniques for Augmented Reality in Endoscopic Surgery,” in MIAR, 2016, pp. 129-138.

[8] G. Klein, D. Murray, “Parallel tracking and mapping for small AR workspace,” in ISMAR, 2007, pp. 225-234.