COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 116 (2014) 68–80

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

# Hand gesture guided robot-assisted surgery based on a direct augmented reality interface

*Rong Wen*[a], *Wei-Liang Tay*[b], *Binh P. Nguyen*[a,*], *Chin-Boon Chng*[a], *Chee-Kong Chui*[a]

[a] *Department of Mechanical Engineering, National University of Singapore, 9 Engineering Drive 1, Singapore 117576, Singapore*

[b] *Department of Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117583, Singapore*

ARTICLE INFO

*Article history:*
Received 29 July 2013
Received in revised form
17 December 2013
Accepted 22 December 2013

*Keywords:*
Human–robot cooperation
Augmented reality
Augmented interaction
Visual guidance
Image-guided surgery
Projector-camera system

ABSTRACT

Radiofrequency (RF) ablation is a good alternative to hepatic resection for treatment of liver tumors. However, accurate needle insertion requires precise hand-eye coordination and is also affected by the difficulty of RF needle navigation. This paper proposes a cooperative surgical robot system, guided by hand gestures and supported by an augmented reality (AR)-based surgical field, for robot-assisted percutaneous treatment. It establishes a robot-assisted natural AR guidance mechanism that incorporates the advantages of the following three aspects: AR visual guidance information, surgeon's experiences and accuracy of robotic surgery. A projector-based AR environment is directly overlaid on a patient to display preoperative and intraoperative information, while a mobile surgical robot system implements specified RF needle insertion plans. Natural hand gestures are used as an intuitive and robust method to interact with both the AR system and surgical robot. The proposed system was evaluated on a mannequin model. Experimental results demonstrated that hand gesture guidance was able to effectively guide the surgical robot, and the robot-assisted implementation was found to improve the accuracy of needle insertion. This human–robot cooperative mechanism is a promising approach for precise transcutaneous ablation therapy.



## 1. Introduction

Hepatic resection or liver transplation offers the greatest potential for curing primary and metastatic liver tumors. However, only about 10–20% of patients are suitable for such procedures due to surgical and health constraints. A promising alternative treatment is the use of ablative therapies; among which, radiofrequency ablation (RFA) is most common. RFA is minimally invasive and results in reduced blood loss during the liver surgery. Current RFA is typically used to treat small tumors due to the small ablation zone created by the needle electrode. For the treatment of large tumors (tumor's diameter >5 mm), multiple needle insertions are necessary in order to thoroughly destroy the tumor. However, this process is subject to uncertainties in tumor imaging and needle placement, compromising the effect and predictability of RFA treatment [12]. A further difficulty in current RFA operation is that ultrasound guidance is typically used for RFA. Ultrasound lacks depth perception and may create transient

* *Corresponding author*. Tel.: +65 9699 3579.
E-mail addresses: mpenp@nus.edu.sg, ngphubinh@yahoo.com (B.P. Nguyen).
0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2013.12.018

micro-bubbles in ablated tissue, obscuring the target sites and increasing the difficulty of subsequent needle insertions [24]. To overcome these limitations, the use of surgical robots has been proposed for RFA of large tumors requiring multiple insertions [22], as robotic surgery offers increased speed, accuracy and consistency.

Current contact-based interfaces can provide haptic feedback for the operation of surgical robots, but pose other problems for surgery. Contact-based interfaces increase the difficulty of maintaining sterility in the operating theatre and increase the risk of cross-contamination, hindering surgeons from performing other surgical steps without changing gloves and interrupting the work flow. Conversely, a contactless human–robot interaction interface operating by gestures and pointing cues can be used without causing sterility issues. Recently, augmented reality (AR), which is a visualization technology that superimposes computer-generated objects on the real-world scene, has drawn considerable interest in the field of image-guided surgery (IGS) [5,4].

In current RFA surgical procedures, limited visual guidance and hand-eye coordination restrict surgical efficiency. Similarly, accurate RF needle navigation is critical to RFA of the whole tumor and efficiency of the surgery. In addition, a pre-planned surgical scheme of RF needle placement and its intraoperative implementation guidance are necessary.

In this paper, we present a surgeon–robot cooperative system for RF treatment using the guidance of both natural hand gestures and a direct AR interface on patient body. The projector-based AR is used to enable surgeons to view anatomical models and surgical data directly from the patient's (skin) surface. Hand gestures are used to intuitively control a robotic manipulator fitted with a RFA device, while contactless pointing allows surgeons to create an in situ surgical plan on the patient body. The hand gesture recognition includes palm recognition for user authorization and an improved dynamic gesture recognition algorithm for interaction with both robot control and AR display. A state machine design is adopted to prevent dangerous state transitions and improve system safety.

The contribution of this study lies in establishment of a robot-assisted natural AR guidance mechanism that incorporates the advantages of the following three aspects: AR visual guidance information, surgeon's experiences, accuracy and consistency of robotic surgery. With this cooperative surgical robot system, the hand-eye limitations in image-guided surgeries can be overcome, assisting surgeons in accurately reaching the target surgical areas inside the patient body.

The paper is structured as follows. Section 2 describes related work as well as our previous research work. In Section 3, an overview of the system design is introduced. With details of the methods presented in Section 4, the experiments and results are described and discussed in Section 5. Finally, the contributions of this paper is summarized in Section 6, concluding with possible future areas of refinement.

## 2. Related work

In image-guided surgeries, hand-eye coordination and experience play key roles in procedures that require surgeons to manipulate surgical tools to accurately reach targeted locations. However, this is a challenging task due to the limited visual information provided by the traditional medical images used in the most of cases. Virtual reality (VR), a technology that can create real-time virtual guidance information, has been applied to surgical practice to improve operation accuracy. Cai et al. [2] developed a VR simulation for the enhancement of hand-eye coordination for percutaneous transluminal coronary angioplasty (PTCA) operation which requires physicians to have good hand-eye coordination while performing the operation. In their system, a tactile interface and a visual interface were designed to provide resemblance of the angiography suite. A backend engine was implemented to support various computational tasks from vascular network modeling to interaction and visualization. Their experiments verified that the artificially created surgical information could enhance hand-eye coordination in surgical training procedure and lead to a short learning curve. Another study utilizing virtual reality for surgical guidance information was the development of a surgical training system for laparoscopic cholecystectomy (LC) operation [1]. This system incorporates modules of knowledge-based learning and a stepwise technical-skill pathway with ongoing feedback, enabling a trainee to acquire knowledge and skills to improve his or her surgical hand-eye cooperation. However, the investment required for this system was significant in terms of simulator cost, maintenance and training space. In addition, the virtual environment could not completely replace the traditional training procedure due to its lack of feedback of real information.

In order to solve the problem that real objects are isolated from the artificially constructed environment, augmented reality has been explored to overlay artificially guidance information on the real organs or patients. Augmented reality and human–computer interaction (HCI) can be used in various capacities in surgery, such as to explore medical data or to control instruments and medical systems [13]. Among the earliest objectives of gesture control in medicine was to provide contact-free navigation of medical data in an aseptic surgical environment, reducing the need to sterilize interaction interfaces and the risk of cross-contamination. In Wachs et al. [16], the Gestix system was proposed to track hand gestures to perform different actions by using a state machine to switch between different classes of actions. Similarly, Kipshagen et al. [10] described a stereo camera-based system to control a medical image viewer where hand gestures performed simple image manipulations such as rotating or zooming, proposing instead the use of remote pointing to perform location sensitive operations such as triggering a mouse click at a specific screen position. Instead of relying on vision-based interfaces, Gallo et al. [6] used Microsoft Kinect to capture the gesture data for remote medical image exploration.

Besides enhancement of visual assistance to improve surgical hand-eye coordination, robotic surgery offers increased precision and is especially helpful for minimally invasive surgery [15]. The role of surgical robots is to assist surgeons in performing accurate or sophisticated manual work, enabling surgeons to focus more on the intraoperative analysis and decision making. In our previous research [22], a surgical robot (Fig. 1(a)) was designed to perform RFA needle insertion in large tumor treatment based on a preoperative surgical

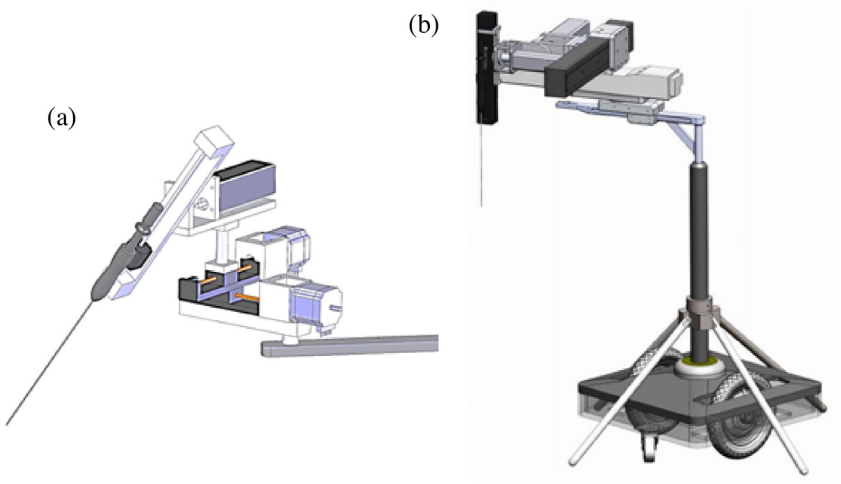



**Fig. 1 – Our surgical robot for RFA needle insertion (a) was upgraded (b) with robotic arms with a higher accuracy, more DOF and a mobile platform for better surgical implementation and surgeon–robot cooperation.**

plan. The developed robotic RFA surgical system consisted of a RFA planning software and a robotic manipulator. The manipulator mechanism consists of a main manipulator and a sub-manipulator system. In the study for a robot-assisted surgery with mixed reality guidance [19], the modular design approach was adopted to facilitate its task-dedicated operation mechanism. We updated the previous robot system with more degrees of freedom (DOF) between the manipulator and the three-axis servo system. A mobile platform was added, enhancing its mobility in order to cooperate with surgeons' operation. The robotic frames were also updated with the ability to move with more stable and precise steps (Fig. 1(b)).

Wong and Matsumoto [21] proposed a cognitive apprenticeship model for cognitive motor learning. The model was derived from the traditional apprenticeship model composed of observation, coaching and practice. This model contributed to the construction of a cooperative surgical system that could sense surgeons' intention and navigate their operation. With the improvements in precision and accuracy of gesture recognition, gesture control has been applied to the direct operation of robots in surgery. Jacob et al. [9] designed Gestonurse, a robotic scrub nurse system, to assist surgeons by passing surgical instruments specified using a gesture command. A state machine model was implemented in order to prevent accidental and hazardous changes between states. Gestonurse was found to reduce the number of movements needed for switching between surgical instruments, thus increasing the efficiency of operations [17]. John Hopkins University has also demonstrated a Kinect-based system for capturing hand gestures to remotely control a surgical robot to perform surturing [14].

## 3. System design and methods overview

As shown in Fig. 2, the cooperative surgical robot system consists of a computational center, a mobile surgical robot module, projector-camera (ProCam) based spatial AR interface and a vision tracking module consisting of a stereo-vision camera and a Kinect device. A rectangular robotic frame (RRF) is constructed to dynamically locate the ProCam system, stereo-vision camera and Kinect device to adapt to the patient and surgeons' positions. The surgical robot module, with its mobile platform, can be flexibly deployed around the operation table to work under the surgeons' hand gesture control. The ProCam system is mounted on the top of the RRF to produce a direct AR interface overlaid on the patient (skin) surface. The direct AR interface constructs an augmented surgical field by integrating virtual medical information (e.g., target organs, vessels and planned needle insertion paths) with the actual patient body to provide surgeons with augmented visual information. With the direct AR interface, hand gesture guidance can be used as a direct and natural method to interact with the augmented surgical information. In order to provide the computational center real-time data, the vision tracking module monitors the display changes in the augmented surgical field and the surgeons' intention delivered by the hand gesture states.

The work flow of the AR–surgeon–robot cooperation is illustrated in Fig. 3. The hardware layer of the system is encircled in blue. The computer workstation is the control and computation center to support robot control, generation of AR interface and augmented interaction including human–robot and human–AR interface interaction. Above the hardware layer are the subjects involved in this interactive surgical procedure which includes the surgeon (decision making), surgical robot motion/execution, patient and AR based surgical field. Two modes of surgical procedures, manual and semi-automatic, are proposed for this robot-assisted RFA treatment with an AR interface which provides necessary visual guidance for an in situ surgical planning. The manual operation mode requires surgeons to select needle insertion sites using only image guidance, while the semi-automatic mode can automatically generate the ablation points based on a ablation model planning [22] which is specifically designed for the surgical robot implementation. In the semi-automatic mode, surgeons can use hand gestures to revise the automatically generated surgical planning data directly on the patient body.

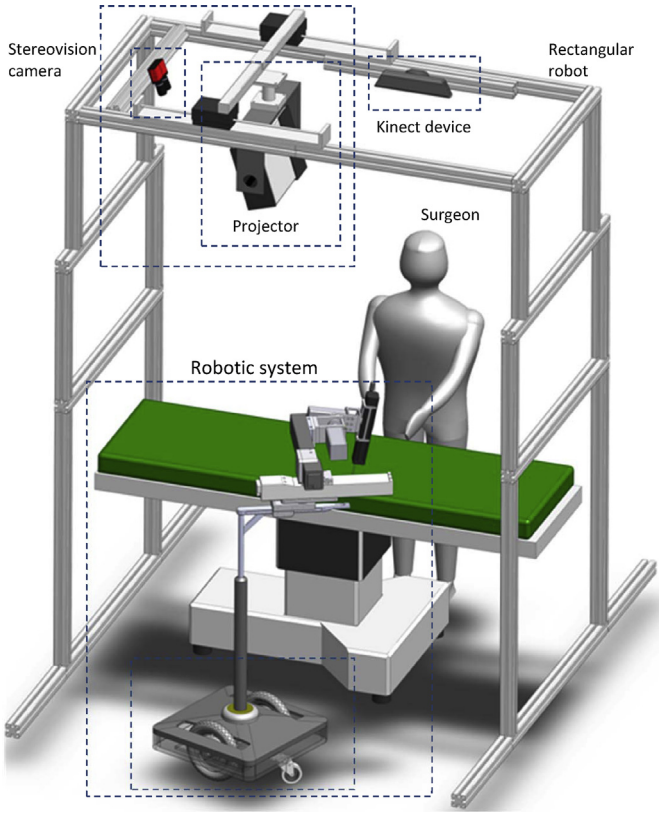



**Fig. 2 – Overview of the cooperative surgical robot system.**

This includes reassignment of the needle insertion ports, RFA model planning and insertion trajectory planning. The data flow, represented by black arrow lines, includes control and feedback flows between the subjects and the hardware modules. The yellow lines representing the data flow among the subjects is a user (surgeon) centered interaction flow.

In order to guarantee a robust human–robot and human–AR interaction in surgical environments, a state

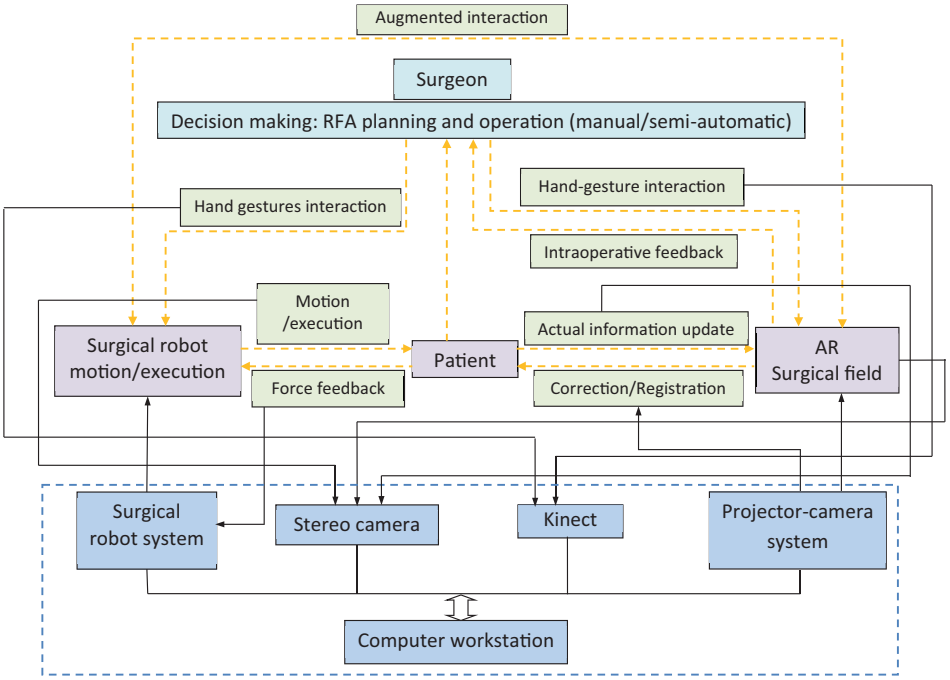



**Fig. 3 – Work flow of the AR based surgeon–robot cooperation.**

machine is used to constrain the surgeon–robot cooperation to the two operation modes. The state machine based system control mechanism explicitly defines robot states that respond to gesture commands. For each hand gesture command, only one state is available for the robot to populate. Whenever the robot motion or execution conflicts with patient safety, the latter takes precedence. For safety consideration in AR display, methods of accurate ProCam calibration and registration of the AR display with an actual patient body were developed. In addition, the safety strategy of accessibility identification by palm recognition enables the system to be controlled only by authorized users.

## 4. Materials and methods

### 4.1. *Surgical AR interface construction*

The projection image is usually distorted by an arbitrary surface. The surfaces' irregular 3D geometry and textures are the main factors in causing distortions such as geometric and radiometric distortion. In order to create an immersive AR surgical field where the computer-generated imagery (e.g., tumors, critical anatomy structures and surgical planning data) appears correctly and seamlessly overlaid on the patient's body, both projection correction and ProCam AR registration are required.

#### 4.1.1. *AR interface correction*

Motivated by calibration of ProCam system which aims to find pixel correspondence between projector and camera images, projection distortion (Fig. 4(a)) can be represented by the pixel correspondence between the source images in the projector and the distorted images captured by the camera. With knowledge of a specific distortion (the pixel correspondence), projection correction can be achieved by pre-warping a source image and then producing a corrected image on the object surface with specific pixel assignment (Fig. 4(b)). Structured-light illumination [20] can be used to find pixel correspondence with a hybrid method combining gray-coded pattern and phase shift. The phase can be generated by Eqs. (1)–(3) using varying intensity values to encode and decode the pixel correspondence between projector and camera images.

$$I_1(x, y) = I_0(x, y) + I_{mod}(x, y)\cos(\varphi(x, y) - \alpha), \tag{1}$$

$$I_2(x, y) = I_0(x, y) + I_{mod}(x, y)\cos(\varphi(x, y)), \tag{2}$$

$$I_3(x, y) = I_0(x, y) + I_{mod}(x, y)\cos(\varphi(x, y) + \alpha), \tag{3}$$

where $I_1, I_2, I_3$ are the intensity of the fringe patterns produced by the projector images varying with the phase,

$$\varphi(x, y) = \arctan\left[\sqrt{3}\frac{I_1(x, y) - I_3(x, y)}{2I_2(x, y) - I_1(x, y) - I_3(x, y)}\right], \tag{4}$$

and the phase shift $\alpha$. $I_0$ is the intensity obtained by averaging the white and black illumination, and $I_{mod}$ is the modulation amplitude.

Once the pixel correspondence is acquired, a texture mapping based method can be used to generate the pre-warped projector image. Assume the coordinates of the projector and the camera image are $(x, y)$ and $(u, v)$, respectively. The number of their pixel correspondence is $n$ and denoted by $f$: $(u_k, v_k) \leftrightarrow (x_k, y_k)$, $0 \le k \le n$. $(u_k, v_k)$ are tessellated to create triangulation in the camera coordinates. The correspondences of $(u_k, v_k)$ in the projector image $(x_k, y_k)$ are connected with triangulation in the same way as $(u_k, v_k)$ to generate the corresponding triangle regions in the projector image. This texture mapping method uses the triangulation in the camera image to find the corresponding pixels and assign the pixel values to generate the pre-warped projector image. In order to find the pre-warped triangulation in the projector image, the blobs' coordinates in the camera image are normalized first within an user-defined rectangle region. In this region, the expected projection display is planned and tessellated by triangulation of the blobs' coordinates. With triangulation and texture mapping of the pixel values of the camera image, the corresponding triangulation in the projector image can be found.

#### 4.1.2. *Registration of AR display with patient body*

The main challenge of registration in a surgical AR environment construction lies in appropriately aligning AR display of the anatomical and pathological models with their respective position on the patient body. As shown in Fig. 5, registration in this study involves two data spaces: surgical model with patient-specific medical data and patient surface data. The patient-specific medical data can be acquired by CT scan, whereas the actual patient (skin) surface data is acquired by the projector-camera system.

The medical data including pathological and anatomical information is acquired through patient CT scanning. The critical anatomical areas on the acquired CT images such as artery, hepatic vessels and the tumor areas can be identified and segmented with hybrid segmentation methods [23]. With the segmented profile data, the anatomical and pathological models can then be constructed. In this study, a noninvasive marker frame is used and attached on the patient's skin during CT scanning. The marker frame is used to construct a coordinate system to relate the surgical model space with the world space (actual patient body). In addition, slight patient movement (e.g., breathing) could also be tracked by observing the markers with the stereo vision device.

In order to register the surgical model data with the patient body, registration follows the two steps. Firstly, a surface matching algorithm [20] is used to find the optimally matched points between two data spaces, surgical model space and patient surface. Eq. (5),

$$p_i^m = \mathbf{R}p_j^s + \mathbf{T} + \zeta_i, \tag{5}$$

describes the relationship between the two data spaces which can be expressed by spatial transformation in terms of the rotation matrix $\mathbf{R}$ and translation matrix $\mathbf{T}$. Assume the coordinates of the patient surface points that are derived from CT and ProCam scanning are $p_i^m$, $\left\{p_i^m = (x_i^m, y_i^m, z_i^m)^T | p_i^m \in \mathbf{M}_a\right\}$, and $p_j^s$, $\left\{p_j^s = (x_j^s, y_j^s, z_j^s)^T | p_j^s \in \mathbf{P}_t\right\}$, $i = 1, 2, \ldots, M$, $j = 1, 2, \ldots, N$, respectively. Singular value decomposition [7] can be used to solve the optimal of $\mathbf{R}$, $\hat{\mathbf{R}}$, and then the optimal of $\mathbf{T}$, $\hat{\mathbf{T}}$, can be derived by $\hat{\mathbf{T}} = p_i^m - \hat{\mathbf{R}}p_j^s$.

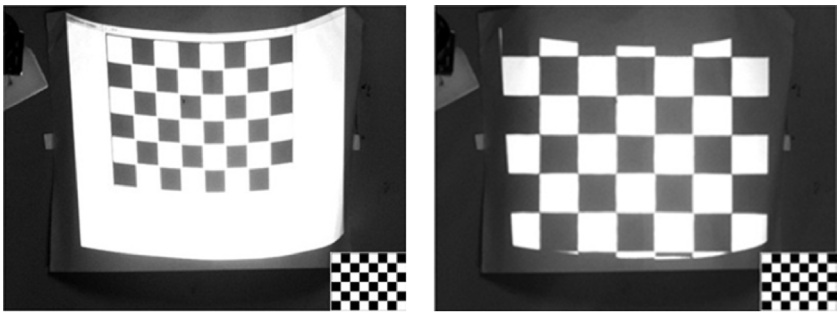



**Fig. 4 – Projection correction on a curved surface. (a) Distorted projection on a curved surface; (b) corrected projection display.**

Next, a point-based registration method is developed to properly align the model projection with its corresponding organs on the patient (skin) surface. The marker frame attached on the patient body provides a spatial relationship among medical data acquired from CT scanning, actual anatomical structures (surgical models) and patient body (surface). Eq. (6) describes the AR display in the world space (patient surface). The data of inner anatomy structure and data of patient surface can be related by the marker frame which can be detected in both of the data spaces.

$$R_O^A = R_{m_i}^A \times R_o^{m_i}, \tag{6}$$

where $m_i$, $o$ and $A$ represents actual position of marker, anatomy structure and AR display, respectively.

### 4.2. *Hand gesture based augmented interaction*

In this cooperative surgical system, hand gestures are used as a natural and direct interaction method which enables surgeons to focus on surgical situation examination and execution guidance rather than control through a console interface. Two types of hand-gesture commands, palm recognition based context selection and dynamic hand gestures recognition based AR and robot control, are developed for surgeon–system cooperation.

#### 4.2.1. *Palm recognition for context-selection*

In this proposed system, palm-based biometrics are used to identify users and load the user specific work-context. The conditions of the operating room offer a unique challenge to biometric recognition, as the textural palm print features described in conventional biometric literature cannot be used. The dimensions of the palm and fingers provides a physical invariant that can be applied for biometric recognition tasks, but it is generally not possible to reconstruct the physical dimensions of the palm and fingers from only a 2D projection such as a color image. However, the additional depth information from the Kinect depth sensor allows for the local scale to be estimated; a screen pixel with a measured depth $d$ can be projected onto a 2D patch with area proportional to $d^2$. This projected area is sufficient to compute palm biometric descriptors that are scale-invariant. For each palm, a 90-dimensional feature descriptor is computed. The

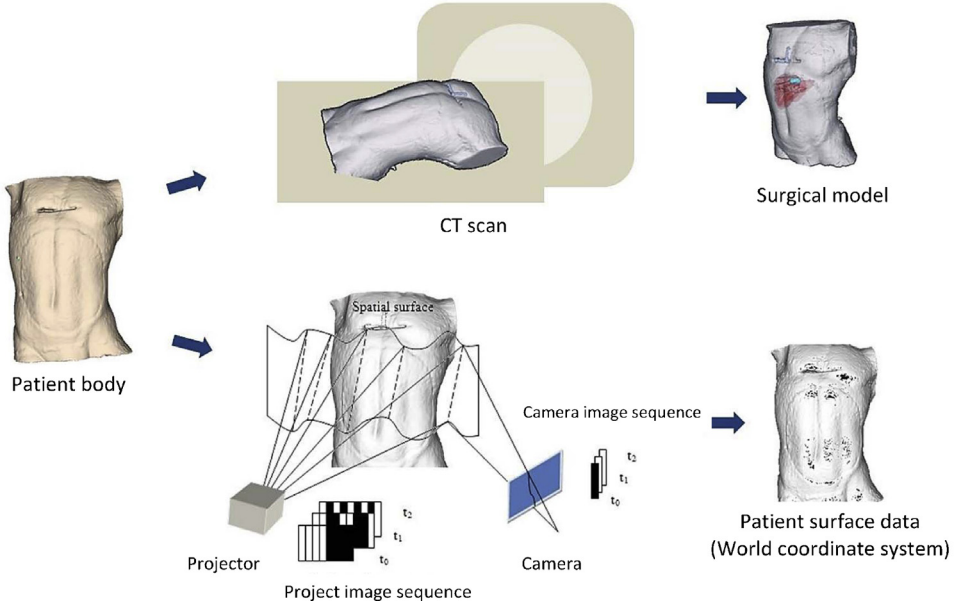



**Fig. 5 – Data acquisition for registration.**

90-dimensional feature consists of five 18-dimensional features, one per finger, measuring the length and widths of each finger segment in each finger. For improved robustness to pose variations, each unique palm is registered using a video registration system that records a few (≤10) seconds of the user's palm under minor pose changes. The biometric recognition system then computes the palm descriptors every few frames and stores the resulting descriptors into a palm database.

A large margin nearest neighbors classifier is used to identify a presented user palm from the computed palm descriptors. Large margin nearest neighbors is a modification of classic nearest neighbors that computes a new projection of the data to improve the classification accuracy under the $k$-nearest neighbor rule. We use the large margin method described by Weinberger and Saul [18] to compute a projection matrix $M$ on the training data. The distance between points $x_i$ and $x_j$ under large margins reprojection is thus

$$d(x_i, x_j) = (x_i - x_j)^T M(x_i - x_j). \quad (7)$$

This new distance is used to determine the similarity between instances for the nearest neighbors classifier.

The user palm recognition is only activated when the system detects a new palm in the active field of view. The large margin nearest neighbors classifier matches the descriptors of the unidentified palm against the registered palm database to determine the user identity and load the contexts relevant to the identified user. This user identity is then attached to the palm object for as long as it remains in the active field of view, removing the need for constant user palm recognition.

#### 4.2.2. *Robot-assisted operation guided by hand gestures and AR display*

Hand gesture based surgeon–robot cooperation consists of two modes: manual and semi-automatic. In the manual mode, surgeons play a dominant role in conducting an in situ surgical planning based on AR interface. With visual guidance from AR planning information, the surgical robot is then guided by hand gestures to operate, command by command. Surgeons' decision and hand gesture control have priority over robotic implementation. Surgeons can intraoperatively interact with the AR display and direct the robot to realize their intentions. The other mode is semi-automatic robot execution which is intraoperatively monitored and intervened by surgeons. In this mode, the surgical robot plays a dominant role in systematical implementation following the ablation model planning. Eight dynamic hand gestures are defined as shown in Fig. 6.

In order to guarantee accuracy and robustness of dynamic gesture guidance in these two modes, hand point cloud (hand blob) is used and detected by its depth variation, which is more distinguishable compared to the other parts of the hand in the surgical field. The Kinect that is fixed on the RRF can retrieve the spatial data of the hand blob in real time. The geometric features can be derived by hand blob segmentation and feature points extraction (finger tips) with convex hull. Gradient values of the hand shape are used as invariant geometric features which are robust to scale and rotation transformations between the hand blob and Kinect. In this way, each hand shape contour can be normalized as gradient vectors by principal component analysis.

In definition of dynamic features, hand motion trajectory is considered as a spatio-temporal pattern. It can be represented with sequences of the centroid points of the hand blobs by the equation,

$$\theta_i = \arctan\left(\frac{y_{i+1} - y_i}{x_{i+1} - x_i}\right), \quad i = 1, 2, \ldots, T, \quad (8)$$

where $\theta_i$ is direction of hand blob movement in image $i$. $x_i$, $y_i$ are the centroid points of the hand blob. Hidden Markov Model (HMM) is used for training, classifying the hand blob data and recognizing the defined gestures.

A HMM can be described as $\lambda = (A, B, \pi)$, $A$, $B$ and $\pi$ are state transition probability matrix, symbol output probability matrix, and initial state probability matrix respectively. Given the observation sequence $O^T = O_1O_2O_3 \ldots O_T$ and model $\lambda = (A, B, \pi)$, the probability of the observation sequence given the model $P(O|\lambda)$ can be computed.

In order to improve the accuracy of the hand movement recognition, the calculation of the symbol output probability matrix is developed with a full-covariance Gaussian distribution [11] in the training procedure, instead of the traditional Gaussian distribution:

$$b_j(O) = \sum_{i=1}^{n} c_{jm} G(O, \mu_{jm}, \Sigma_{jm}), \quad 1 \leqq j \leqq N, \quad (9)$$

where $b_j(O)$ is the observation symbol probability distribution, $B = b_{jk}$. G is the Gaussian distribution,

$$G(x, \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}\Sigma^{1/2}} \exp\left[-\frac{1}{2}(x - \mu)^K \Sigma^{(-1)}(x - \mu)\right], \quad (10)$$

$$\sum_{m=1}^{M} c_{jm} = 1, \quad c_{jm} \geqq 0; \quad 1 \leqq m \leqq M. \quad (11)$$

The unknown parameters of the HMM are obtained using the Baum–Welch algorithm. The best hand motion path can then be computed using the Viterbi algorithm from the trained parameters and the discrete vector derived from a hand motion trajectory. To classify a sequence into one of $K$ classes, we train up parameters of $K$ HMMs. A sequence of input gestures is classified by computing output probabilities in each state.

### 4.3. *Safety strategies for surgical human–robot interaction*

Safety is an important design consideration robot-assisted surgery. The surgical robot states and their work flow controlled by hand gesture are illustrated as shown in Figs. 7 and 8. For both the manual and semi-automatic control modes, ten states are defined.

In the system calibration procedure, the robot moves with its mobile platform to adjust its position with respect to the patient body and surgeon. Before the operation is started by surgeons, it stays in "stop" state. In the manual control mode, each state transition is controlled by hand gesture command including translation ($x$- and $y$-axis), rotation and needle

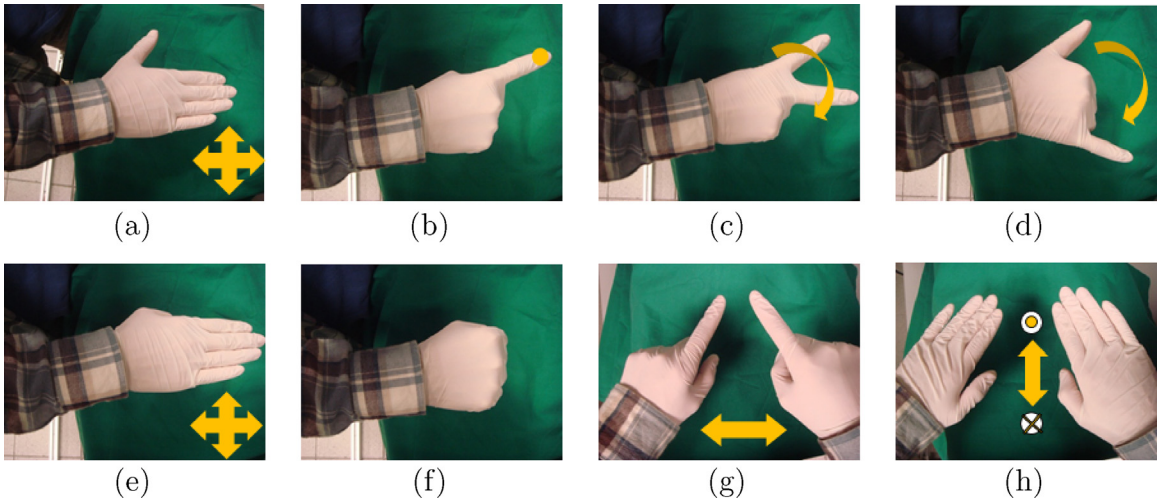



**Fig. 6 – Dynamic hand gestures for robot and AR control. (a) AR – move; (b) AR – pointing; (c), (d) robot – arm rotation; (e) robot – x/y/z axial motion; (f) robot – stop; (g) AR – zoom in/out; (b) robot – suspend.**

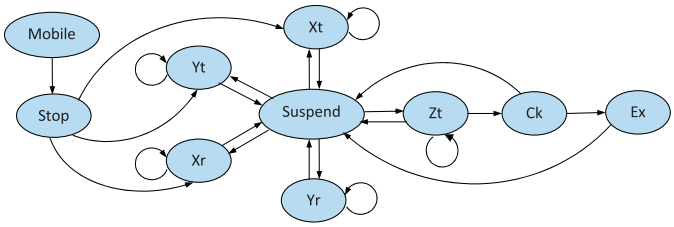



**Fig. 7 – State machine design for manual control. Xt – translation along x-axis, Yt – translation along y-axis, Zt – translation along z-axis, Xr – rotation around x-axis, Yr – rotation around y-axis, Stop – motion and execution suspend, Ck – current position check before execution, Ex – execution (RFA).**

insertion (z-axis). Surgeons can continuously adjust the needle's spatial position through state "Xt", "Yt", "Xr", "Yr" and "Zr". State "Suspend" is necessary to prevent the robot from working outside of designated safety areas. Before each RF ablation starts, the state "Ck" will make safety checks to ensure that the distance between the current RF needle tip's position and the planned ablation point is within a satisfactory range. After each RF ablation, state "Ex" completes and enters the state "Suspend" await for the surgeon's command.

A significant difference between semi-automatic control and manual control is that most of the state transition in the former is conducted by the robotic system itself. As shown in Fig. 8, the state transition within translation states (Xt and Yt), rotation states (Yr and Xr), and from translation states to rotation states and then to needle insertion states, are automatically conducted by robot itself following the preplanned data. Surgeon's hand gestures are only used to change current states to state "Suspend" and make compensation in manual mode. In this way, the robotic implementation is able to keep an efficient and systematic work compared to guidance in a purely manual mode.

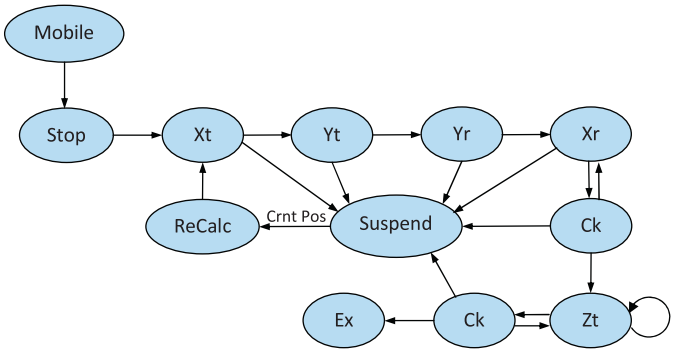



**Fig. 8 – State machine design for semi-auto control. Xt – translation along x-axis, Yt – translation along y-axis, Zt – translation along z-axis, Xr – rotation around x-axis, Yr – rotation around y-axis, Stop – motion and execution suspend, Ck – current position check before execution, Ex – execution (RFA), Crnt Pos – current position, ReCalc – Recalculation.**

In ablation model planning, a safety margin of 10 mm was assumed in the ablation model to avoid healthy tissue lesion by RFA. The margin could be adjusted by selecting different RFA volumes on RF needle to modify the number of needle insertions. With the generated ablation points and their relative position to tumor's peripheral anatomical structures, the insertion trajectories can be generated from the insertion points located on the patient skin to the ablation points. To avoid harming the patient's organs and bones with the RF needle, the needle's position and orientation could be optimized by a gradient-based method [3].

In order to adapt to hand gestures' guidance, surgeons have the priority to immediately suspend the robotic implementation and are allowed to interactively revise the corresponding preplanned data on the AR interface. Robot design is also an important issue for safety consideration. The developed robotic system is equipped with necessary dexterity for maneuvering the needle constrained at the incision port. This is known as Remote Center of Motion (RCM) and is a particularly important requirement in the design of robotic system for minimally invasive surgery [22]. The end effector allows the needle insertion to be manually driven, without having to detach it from the system. The workspace of the surgical robot with respect to the preplanned needle placements is taken into consideration. It covers the entire tumor region and all the insertion trajectories while avoiding the fiducial markers attached on the patient. Another clinical requirement unique to surgical robots is the issue of sterilizability. Surgical tool and equipment in contact with patient should be sterilizable. The needle guiding unit and the needle grip at the end effector can be easily detached from the robotic system in order to be placed into sterilizing systems like an autoclave.

## 5. Experimental results and discussion

The system and its main components are shown in Fig. 9. It consists of a surgical robot, an LCD projector with resolution of 1024 × 768 (XGA), a stereo vision device with two cameras with resolution of 640 × 580, a Microsoft Kinect and a computer workstation with quad-core processors, 12 GB memory and a graphics card supporting OpenGL 3D.

### 5.1. *AR construction*

Fig. 10 displays AR calibration result on the targeted region of the mannequin body. The projection, which was distorted by the mannequin's irregular surface, was corrected. Quantitative accuracy evaluation on this correction was examined by comparing the source image of the chessboard pattern with its corresponding corrected projection. Geometric errors were calculated on each corresponding square unit along the $x$ and the $y$ axis. Average distortion error was 0.68 mm and 0.73 mm respectively. The registration between specific patient anatomic models and patient body is shown as Fig. 11. Evaluation of this registration result was obtained by projecting the registered virtual geometric models implanted inside the mannequin body to examine their projection with respect to the location of the real models.

### 5.2. *Augmented robotic operation guided by hand gestures*

Fig. 12 shows hand gesture guided AR interaction and robot implementation on the mannequin body. With AR interface guidance, the augmented surgical field displays the target ablation areas and critical anatomy structures. Natural hand gesture based augmented interaction enables users to interactively arrange the ablation models, plan insertion ports and optimize insertion paths (Fig. 13). The direct AR display also provides users navigation information for surgical robot control. The surgical robot can adjust its position with respect to surgeons with the mobile platform. With robot cooperative assistance, surgeons' operative ability could be enhanced and extended without spending too much of their attention on the manipulation of the robot.

Fig. 14 illustrates the robotic needle insertion which was guided by the augmented surgical field. The preplanned path was labelled on the patient (mannequin) body. The anatomical structure which was displayed and registered on the mannequin body was used to enable surgeons to visually guide the robot to avoid critical tissues or organs. In order to validate the accuracy of the hand gesture guided augmented needle insertion, four tumor models (Fig. 15 (b)) with geometrical contours that were similar to a real tumor were constructed from plasticine. They were implanted inside the mannequin body and scanned by a CT scanner, to be registered and used to test the accuracy of the augmented needle insertion.

As displayed in Table 1, four groups of twenty-two insertion tests were administered on the tumor models. In group 1, nine tests with four insertions in each test were conducted. In group 2, three tests with ten insertions in each test were conducted. The tests in group 1 were designed to examine the stability of the entire procedure of the gesture-guided robot surgery – from the robot's initial position to needle insertion. The tests in group 2 were designed to examine the robustness of consecutive robotic needle insertion under gesture guidance. This procedure is commonly used for large tumor RF ablation which requires multiple needle insertion. Four and six tests were conducted in group 3 and 4 with seven and five needle insertion in their each test, respectively. Groups 3 and 4 were used to further examine the accuracy and robustness of the whole system. As shown in Table 1 and Fig. 15, mean error was used to evaluate the average distance between the targets and the needle insertion ports in each test, which were 1.85 mm, 1.87 mm, 1.92 mm and 1.89 mm, respectively. Mean deviation in each group test was 1.34 mm, 1.90 mm, 2.07 mm and 1.84 mm, respectively. The experimental results demonstrate that this hand gesture guided robotic implementation was robust in consecutive needle insertion. The mean errors were below 2 mm and were clinically acceptable. Needle practical paths in the experiments were also compared with the preplanned paths and most of them were matched with errors of 1.5 mm along $x$ axis, 2.4 mm along $y$ axis and 0.34 along $z$ axis. However, this accuracy evaluation was conducted on a polystyrene-made phantom model which lacks patient breathing and soft tissue deformation.

In order to guarantee safety in our state machine design, the robot's response time to hand gesture commands can be increased. This is because the system was designed to

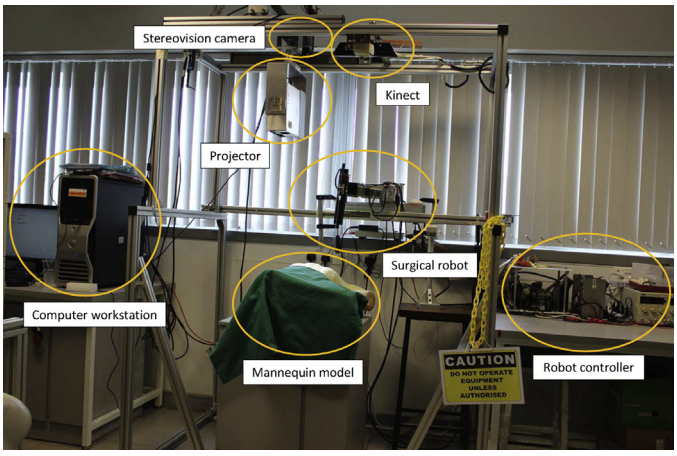
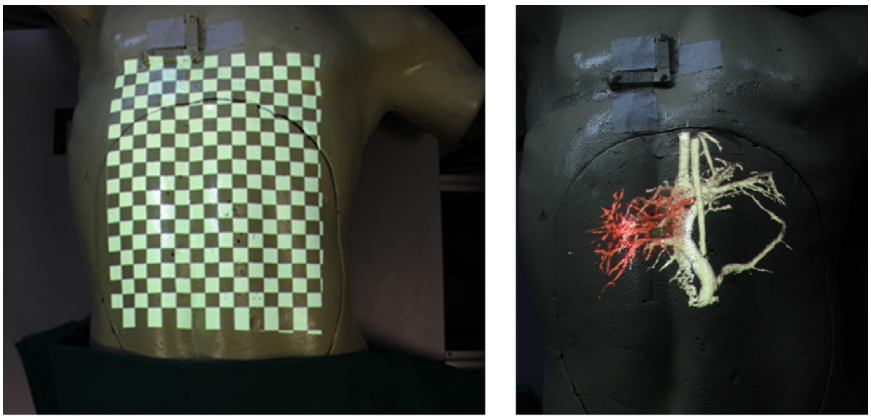



**Fig. 9 – System setup.**

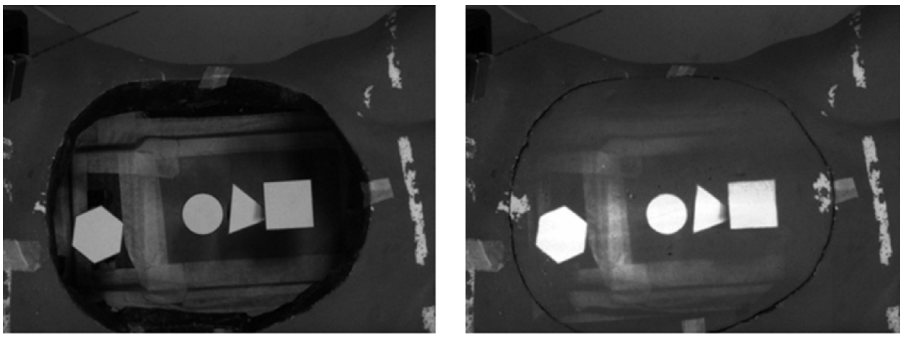


(a) (b)

**Fig. 10 – AR construction. (a) AR calibration on the mannequin body with irregular surface; (b) AR display of the interested anatomic structures on the patient body.**

generate a lag period between hand gesture commands and their corresponding robotic implementation. The lag period gives surgeons a safety buffer in which to suspend the previous command, avoiding an uncontrollable or incorrect operation on the patient. This system was able to realize real-time gesture recognition by Kinect sensors. The latency between gesture guidance and robot response could be evaluated by a series of gesture-guided robot movements, which was about 2–3 s in each robot response.

The biometric recognition-palm recognition was exploited in several novel ways in the experiment. First, different gesture-profiles can be loaded for each user, allowing authorized users to control the robot implementation during surgery. This increases safety by restricting robot functions



(a) (b)

**Fig. 11 – AR registration between (a) the real geometric objects that implanted inside the mannequin body and (b) their projection of virtual geometric objects on the body surface.**

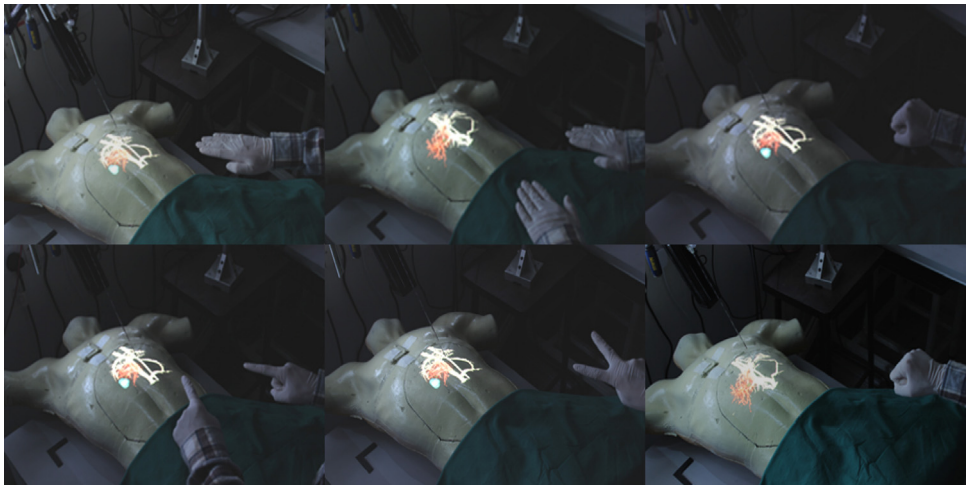
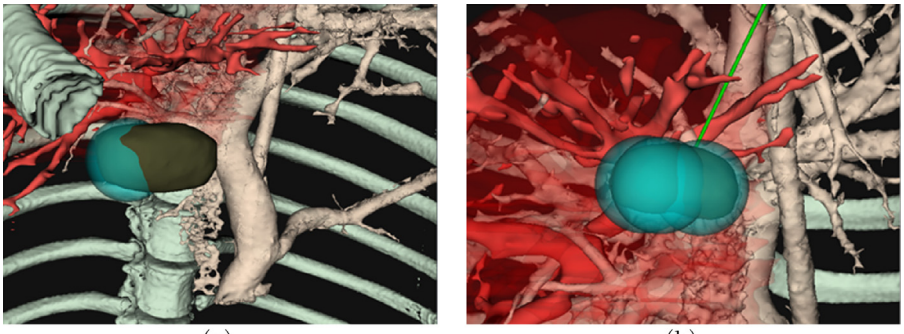


**Fig. 12 – Hand gesture's interaction with both AR display and surgical robot implementation.**

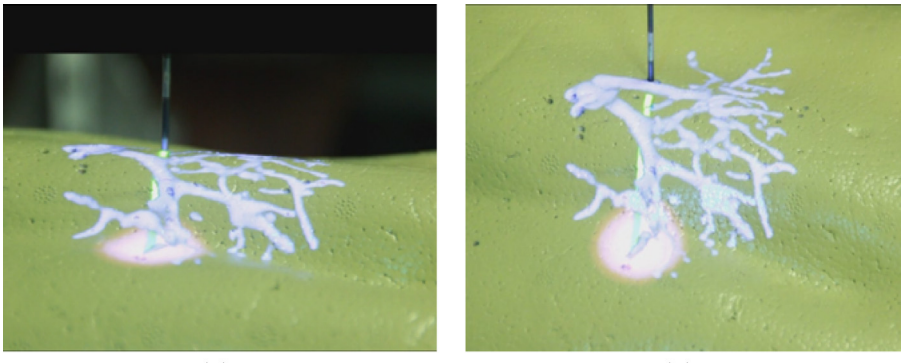


(a) (b)

**Fig. 13 – Surgical in situ planning: (a) manual mode, (b) semi-automatic mode.**

and control to authorized surgeons with appropriate training. Another way of employing user-specific context selection is to project different interfaces or data for each user. For example, surgical assistants may be assigned different roles for the operation, such as manipulation of the ablation system or maintenance of patient homeostasis, and the projected AR system can intelligently switch interfaces depending on the current user. In this way, context-selection can greatly improve the efficiency of human–robot interaction in the surgical setting.

In this study, we assume that the RF needle is a rigid body, and it maintains a state of stiffness without deforming or deflecting during needle insertion. With this assumption, the actual depth of needle insertion is estimated by tracking the RF needle position (tracking of needle markers). Needle insertion depth and angles were monitored and adjusted in real time to adapt to preplanned insertion paths. Another limitation is that accuracy evaluation was performed on a static mannequin body. Possible errors due to patient breathing was not considered during the intraoperative needle insertion procedures.



(a) (b)

**Fig. 14 – Augmented needle insertion on the mannequin body: (a) side view, (b) front view.**

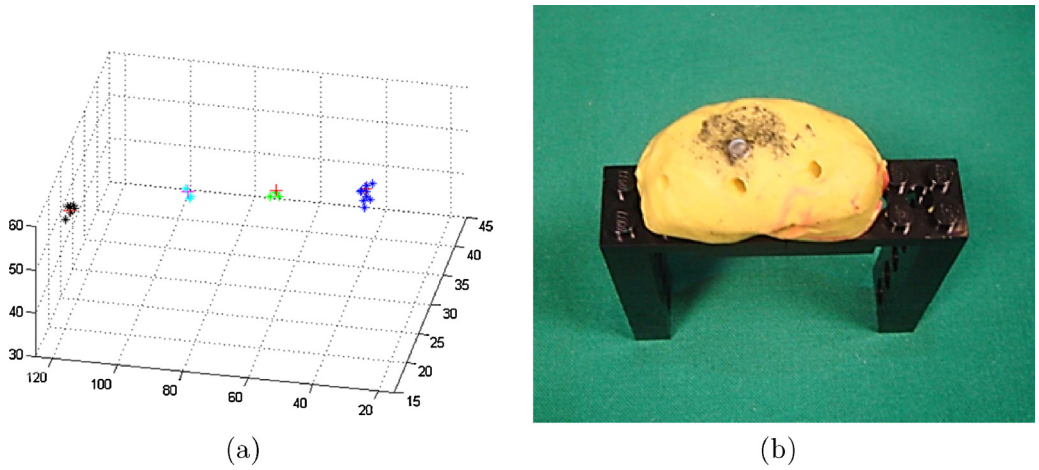



**Fig. 15 – Accuracy evaluation of the cooperated needle insertion. (a) Diagram of insertion error evaluation; (b) plasticine model for insertion test.**

**Table 1 – Experimental data for robot-assisted needle insertion guided by hand gestures.**

| Measure | Group 1 | | | | | | | | | Group 2 | | | Group 3 | | | | Group 4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | T1 | T2 | T3 | T4 | T5 | T6 | T7 | T8 | T9 | T1 | T2 | T3 | T1 | T2 | T3 | T4 | T1 | T2 | T3 | T5 | T4 | T6 |
| ME[a] | 1.77 | 1.89 | 1.95 | 1.86 | 1.80 1.85 | 1.92 | 1.73 | 1.77 | 1.96 | 1.84 | 1.86 1.87 | 1.91 | 2.55 | 2.05 1.92 | 1.36 | 1.70 | 1.91 | 1.90 | 2.11 | 1.62 1.89 | 1.78 | 2.04 |
| MD[b] | 1.38 | 1.30 | 1.52 | 1.31 | 1.33 1.34 | 1.26 | 1.29 | 1.37 | 1.26 | 1.88 | 1.87 1.90 | 1.94 | 3.17 | 1.29 2.07 | 2.38 | 1.43 | 1.26 | 2.12 | 2.64 | 1.29 1.84 | 1.74 | 2.01 |

[a] Mean error (mm).
[b] Mean deviation (mm).

Hostettler et al. [8] proposed the method of using preoperative 3D CT or MRI images based simulation to predict the tumor position during the intraoperative intervention. According to their experiment results, the accuracy could be within 2–3 mm at 50 Hz for patient free respiration (breathing) which meets the clinic requirement. Thus, with the initial system's calibration parameters and the real-time simulation-based organ motion parameters, the virtual position of the organ and needle can be predicted and controlled for.

## 6. Conclusions and future work

This paper presented a hand gesture guided surgical AR cooperative system. Our method combined the advantages of AR visual guidance information with the accuracy and consistency of surgical robots in order to assist surgeons. Natural hand gesture guidance enabled users to intuitively interact with the AR surgical field and surgical robot. Safety strategies including state machine design, special design of robotic system and biometric palm recognition guarantees that the system is able to work safely under control. The experimental results demonstrated its feasibility and potential to be developed as a practical cooperative surgical system.

Based on the current work, several issues have to be addressed for a successful bare-hand gesture guided robot-assisted surgery. For augmented interaction, we are faced with the dilemma of dynamic shadow elimination which may be caused by users and surgical tools. The discrepancy between surgical model simulation and uncertainty of interaction between the surgical tools and anatomy is another problem. In addition, laminar air flow control over the insertion ports is also an important factor to be considered to avoid potential surgical site infection. Investigating these topics would allow intraoperative surgical information to be more accurately delivered to surgeons, and thus improving surgical outcomes.

## Conflict of interest

None to declare.

## Acknowledgment

This work is supported in parts by National University of Singapore FRC Tier 1 Grant (WBS: R265-000-446-112).

## REFERENCES

[1] R. Aggarwal, P. Crochet, A. Dias, A. Misra, P. Ziprin, A. Darzi, Development of a virtual reality training curriculum for laparoscopic cholecystectomy, British Journal of Surgery 96 (2009) 1086–1093.
[2] Y.Y. Cai, C.K. Chui, X.Z. Ye, Z. Fan, J.H. Anderson, Tactile VR for handeye coordination in simulated PTCA, Computers in Biology and Medicine 36 (2006) 167–180.

[3] Q. Du, X. Zhang, Research on robotic technologies for ultrasound guided radio frequency ablation surgery, in: IEEE Proceedings of the International Conference on Industrial Technology (ICIT 2008), 2008, pp. 1–6.
[4] V. Ferrari, G. Megali, E. Troia, A. Pietrabissa, F. Mosca, A 3-D mixed-reality system for stereoscopic visualization of medical dataset, IEEE Transactions on Biomedical Engineering 56 (2009) 2627–2633.
[5] G. Fichtinger, A. Deguet, K. Masamune, E. Balogh, G.S. Fischer, H. Mathieu, R.H. Taylor, S.J. Zinreich, L.M. Fayad, Image overlay guidance for needle insertion in CT scanner, IEEE Transactions on Biomedical Engineering 52 (2005) 1415–1424.
[6] L. Gallo, A.P. Placitelli, M. Ciampi, Controller-free exploration of medical image data: experiencing the Kinect, in: IEEE Proceedings of the 24th International Symposium on Computer-Based Medical Systems (CBMS), 2011, pp. 1–6.
[7] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, second ed., Cambridge University Press, Cambridge, UK, 2004.
[8] A. Hostettler, S. Nicolau, Y. Rémond, J. Marescaux, L. Soler, A real-time predictive simulation of abdominal viscera positions during quiet free breathing, Progress in Biophysics and Molecular Biology 103 (2010) 169–184.
[9] M. Jacob, Y.T. Li, G. Akingba, J.P. Wachs, Gestonurse: a robotic surgical nurse for handling surgical instruments in the operating room, Journal of Robotic Surgery 6 (2012) 53–63.
[10] T. Kipshagen, M. Graw, V. Tronnier, M. Bonsanto, U. Hofmann, Touch- and marker-free interaction with medical software, in: O. Dössel, W.C. Schlegel (Eds.), World Congress on Medical Physics and Biomedical Engineering, volume 25/6 of IFMBE Proceedings, September 7–12, 2009, Munich, Germany, Springer, 2007, pp. 75–78.
[11] N. Nguyen-Duc-Thanh, S. Lee, D. Kim, Two-stage hidden markov model in gesture recognition for human robot interaction, International Journal of Advanced Robotic Systems 9 (2012) 1–10.
[12] N.I. Sainani, P.B. Shyn, S. Tatli, P.R. Morrison, K. Tuncali, S.G. Silverman, PET/CT-guided radiofrequency and cryoablation: is tumor fluorine-18 fluorodeoxyglucose activity dissipated by thermal ablation? Journal of Vascular and Interventional Radiology 22 (2011) 354–360.
[13] W.L. Tay, C.K. Chui, S.H. Ong, Single camera-based remote pointing and recognition for monitor-based augmented reality surgical systems, in: ACM Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry, 2012, pp. 35–38.
[14] The Johns Hopkins University, Gesture Based Surgical Manipulation of a da Vinci Robot Using a Kinect, 2013, http://www.youtube.com/watch?v=YsY_A0kLh-g (accessed 25.07.13).
[15] V. Vitiello, S.L. Lee, T. Cundy, G.Z. Yang, Emerging robotic platforms for minimally invasive surgery, IEEE Reviews in Biomedical Engineering 6 (2013) 111–126.
[16] J. Wachs, H. Stern, Y. Edan, M. Gillam, C. Feied, M. Smith, J. Handler, Gestix: a doctor–computer sterile gesture interface for dynamic environments, in: A. Saad, K. Dahal, M. Sarfraz, R. Roy (Eds.), Soft Computing in Industrial Applications, volume 39 of Advances in Soft Computing, Springer, 2007, pp. 30–39.
[17] J.P. Wachs, M. Jacob, Y.T. Li, G. Akingba, Does a robotic scrub nurse improve economy of movements? in: D.R. Holmes III, K.H. Wong (Eds.), Medical Imaging 2012: Image-Guided Procedures, Robotic Interventions, and Modeling, volume 8316 of Proceedings of SPIE, February 04, 2012, San Diego, CA, USA, 2012, 83160E-1–83160E-7.
[18] K.Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest neighbor classification The Journal of Machine Learning Research 10 (2009) 207–244.
[19] R. Wen, C.B. Chng, C.K. Chui, K.B. Lim, S.H. Ong, S.K.Y. Chang, Robot-assisted RF ablation with interactive planning and mixed reality guidance, in: Proceedings of the IEEE/SICE International Symposium on System Integration (SII 2012), 2012, pp. 31–36.
[20] R. Wen, C.K. Chui, S.H. Ong, K.B. Lim, S.K.Y. Chang, Projection-based visual guidance for robot-aided RF needle insertion
, International Journal Computer Assisted Radiology Surgery 8 (2013) 1015–1025.
[21] J.A. Wong, E.D. Matsumoto, Primer: cognitive motor learning for teaching surgical skill-how are surgical skills taught and assessed? Nature Reviews Urology 5 (2008) 47–54.
[22] L. Yang, R. Wen, J. Qin, C.K. Chui, K.B. Lim, S.K.Y. Chang, A robotic system for overlapping radiofrequency ablation in large tumor treatment, IEEE/ASME Transactions on Mechatronics 15 (2010) 887–897.
[23] X. Yang, H. Yu, Y. Choi, W. Lee, B. Wang, J. Yang, H. Hwang, J. Kim, J. Song, B. Cho, H. You, A hybrid semi-automatic method for liver segmentation based on level-set methods using multiple seed points, Computer Methods and Programs in Biomedicine 13 (2014) 69–76.
[24] X. Zheng, S. Vaezy, A targeting method based on acoustic backscatter for treatment planning in tissue ablation using focused ultrasound, IEEE Transactions on Biomedical Engineering 57 (2010) 71–79.