Augmented Reality Assisted Sinus Surgery

D. M. Shi¹, W. S. Ng², K. V. Ling³, W. Shao³, H. J. Chen², C. K. Kwoh¹
¹School of Computer Engineering, Nanyang Technological University, Singapore
²School of Mechanical and Production Engineering, Nanyang Technological University, Singapore
³School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

Abstract—In this paper, we propose an augmented reality (AR) assisted sinus surgery system, in which the surgical outcome is improved by providing a new way of visualization which embeds virtual environments into the real scene. Such an AR assisted system, with a preoperative CT scan registered and displayed throughout the surgery, would provide the surgeon visibility of the nearby structures surrounding the endoscopes and the tip of the surgical tool. It would help a surgeon to navigate the instrument to the targeted site while avoiding any possible damage to other delicate tissues. It is especially valuable in revised surgery when the normal landmarks may not be present. For these reasons, the proposed system can make a much safer surgery for the patient.

Keywords—Augmented Reality, Sinus surgery

I. INTRODUCTION

In current sinus surgery, an endoscope is inserted into the nostril of the patient to help the surgeon identify the tissues to be cut and navigate through the path. It is very hard for the surgeon to know if there is critical structure nearby or just under the surface of the opening. Within the constraint of the surgical opening, the surgeon cannot see beyond the exposed surfaces. These limitations are accentuated by the even greater restrictions of minimally invasive surgery. Limited visibility through "keyholes" during endoscopic procedures and through small incisions with ever-diminishing sizes increases the need for intraoperative image guidance.

Existing image guided surgery systems present the surgeon with data that was gathered prior to surgery, track surgical instruments within the operating field, and render the tracked devices along with the data [7]. While current image guided surgery is helpful in surgical planning, navigation and guidance, it is not perfect. Firstly the surgeon has to look away from the patient to a display screen, thus disrupted his hand-eye coordination. Secondly the surgeon cannot see the patient on the display, thus lost the awareness of what is really happening. To overcome this, he has to shift from the display screen and the patient frequently. Thirdly, there is no visual feedback of the registration accuracy, deteriorated the safety of the surgery. Most importantly, image guided surgery cannot create the see-through effect for the surgeon to see through the patient directly.

In this paper, we will propose an approach to enhance image guided surgery by providing a see through visualization with virtual information, in which the images of the patient and the synthetic images are merged together to overcome the problems existed in current image guided surgery. The see through visualization can reduce the cost of the surgery by minimizing the time the surgeon spends in the operating room. Such a capability can also enable the surgeon to remove tumour tissue more comfortably while reducing the chances of harming healthy tissue that surrounds the area under treatment. This suggests an overall improvement in patient outcomes.

The rest of this paper is organised as follows: In section, the core methodologies will be described, including registration, deformable modeling, and accurate tracking. In section 3, the methodologies will be integrated to make up an AR assisted sinus surgery system. Section 4 provides our conclusions.

II. METHODOLOGIES

In this section, An advanced iterative closes algorithm is proposed to improve the registration, kernel based representation is introduced for accurate tracking, and active shape models will be applied to deformable modeling based on Fourier descriptors.

2.1 Advanced iterative closest algorithm

The synthetic objects are registered with the corresponding real objects using both fiducial-based and surface-based technology, with emphasis on the later. Non-rigid object registration is also considered in case the application requires. In this research, an advanced ICP algorithm will be investigated in which tentative rounds of ICP registration (the trial rounds) are conducted, and the initial hypothesis before each round of ICP registration is automatically provided by the perturbation derived from the principal axes analysis [1].

The iterative closest point (ICP) algorithm implements an intrinsically natural and practical idea of image registration [3]. In the ICP algorithm, a data shape $P$ is registered to be in best alignment with a model shape $X$. Let $N_P$, $N_X$ be the number of points in the shapes. Then $P$ and $X$ are, respectively, the $N_P$-tuple $P = (\tilde{p}_1, \tilde{p}_2, ..., \tilde{p}_{N_P})$ and the $N_X$-tuple $X = (\bar{x}_1, \bar{x}_2, ..., \bar{x}_{N_X})$, where $\tilde{p}_i$, $\bar{x}_i$ are 3-D vectors (voxel coordinates $\bar{x}_i = [x_i, y_i, z_i]^T$) and $N_X \geq N_P$. The ICP algorithm estimates the optimal registration vector, i.e., the set of translation and rotation parameters that lead to the optimal registration of the two sets. This process is applied repeatedly until the change of a certain dissimilarity...
converges to a predefined threshold. The optimal registration vector is the one that minimizes the mean square error objective function:

\[ f(\hat{\mathbf{q}}) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\| \mathbf{T}_i - R(\hat{\mathbf{q}}_{\mathbf{p}}) \mathbf{p}_i - \mathbf{q}_i \right\|^2, \]

(1)

where \( \mathbf{T}_i \) and \( R(\hat{\mathbf{q}}_{\mathbf{p}}) \) are the translation and rotation matrices to be optimized, respectively. Due to the monotonical property of ICP convergence, the tentative ICP rounds will continue until no better local minima can be found. In our advanced ICP algorithm, only a certain number of iteration will be executed within each tentative ICP round. Stopping criterion for recursion is defined as the same objective function as that in the classical ICP algorithm, which is represented by the distance MSE. So if the final MSE of the converged registration ascends in the new round, the recursion will stop and retrace to the previous round until it reaches a local convergence solution. Figure 1 shows an example of registration using our advanced ICP algorithm.

![Fig.1. Visualization of the registered two surfaces after registration using the advanced ICP algorithm.](image)

2.2 Kernel-based representation

The kernel-based tracking technique uses the basin of attraction of the similarity function [4]. The differentiable similarity function and efficient gradient-based optimizations procedures can be used to find the maxima. The similarity function defines a distance amongst the model and candidates. The metric is based on Bhattacharyya coefficient that is a divergence-type that is already used in computer vision. The target localization can be found by the following way. The distance should be minimized as a function of the target candidate location.

The localization procedure starts from the position of the target in the previous frame and searches in the neighborhood. Since the distance function is smooth, the procedure uses gradient information that is provided by the mean shift vector. The complete target localization is represented by Bhattacharyya coefficient Maximization equation. The coefficient is computed only after the algorithm completion to evaluate the similarity between the target model and the chosen candidate. The maximization of the Bhattacharyya coefficient can be also interpreted as a matched filtering procedure. The mean shift procedure is used to find the local maximum of the scalar filed of correlation coefficients. The operational basin of attraction is the region in the current frame in which the new location of the target can be found by the proposed algorithm. Due to the use of kernels, this basin is at least equal to the size of the target model. That is, if in the current frame the center of the target remains in the image area covered by the target model in the previous frame, the local maximum of the Bhattacharyya coefficient presents a unique maximum in the local neighborhood.

To eliminate the influence of different target dimensions, all targets are first normalized to a unit circle. This is achieved by independently rescaling the row and column dimensions. Let \( \{x^*_i\}_{i=1}^n \) be the normalized pixel locations in the region defined as the target model. The region is centered at 0. An isotropic kernel, with a convex and monotonic decreasing kernel profile \( k(x) \), assigns smaller weights to pixels farther from the center. Using these weights increases the robustness of the density estimation since the peripheral pixels are the least reliable, being often affected by occlusions or interference from the background. The function \( b : R^2 \rightarrow \{1, \ldots, m\} \) associates to the pixel at location \( x^*_i \) the index \( b(x^*_i) \) of its bin in the quantized feature space. The probability of the feature \( u=1, \ldots, m \) in the target model is then computed as [4]:

\[ \hat{q}_u = C \sum_{i=1}^{n} k\left(\|x^*_i\|^2\right) \delta\left[b(x^*_i)-u\right] \]

(2)

where \( \delta \) is the Kronecker delta function. The normalization constant \( C \) is given by \( C = 1/\sum_{i=1}^{n} k\left(\|x^*_i\|^2\right) \).

Kalman filter assumes that the noise sequence is Gaussian and solves the state estimation problem in two steps: prediction and update [2]. The kernel-based target localization method will be integrated with the Kalman filtering framework. For a faster implementation, two independent trackers were defined for horizontal and vertical movement. A constant-velocity dynamic model with
acceleration affected by white noise has been assumed. The uncertainty of the measurements has been estimated. The idea is to normalize the similarity surface and represent it as a probability density function. Since the similarity surface is smooth, for each filter only three measurements are taken into account, one at the convergence point and the other two at a distance equal to half of the target dimension, measured from the peak. We fit a scaled Gaussian to the three points and compute the measurement uncertainty as the standard deviation of the fitted Gaussian.

2.3 Active shape models on Fourier descriptors

In this research, the deformable soft tissue modelling play an important role, as it can predicate shape locations for registration and tracking procedure. The original active shape models (ASMs) can be used to model faithfully objects whose shape variations are linear with respect to the shapes point description [5]. The motivation of ASMs is to describe variability but to remain specific to the class of structures they represent. The ASM technique relies upon each object or image structure being represented by a set of points. The model gives the average positions of the points, and has a number of parameters which control the main modes of variation found in the training set.

Active shape modelling extracts the eigenvectors, $U$, of the training examples by principal component analysis (PCA) [6]. PCA is readily performed by solving an eigenvalue problem, or by using iterative algorithms which estimate the principal components. Thereafter, any model $\Gamma$ can be generated by adjusting the shape parameters, $b$, corresponding to the principal modes:

$$\Gamma = \Psi + Ub,$$

where $\Psi$ is the mean vector of the training examples. Given such a model and an image containing an example of the object modelled, image interpretation involves choosing values for each of the parameters so as to find the best fit of the model to the image.

However, the original ASMs face the problem of shape alignment. To this end, Fourier descriptors (FDs) will be introduced in this project to transform a Cartisian representation to frequency space. First, Fourier transform is applied to convert the shape information from spatial space to frequency space, then a shape will be represented by a vector of frequencies. The high frequency components capture the finer detail whereas the low frequency components of the FDs capture the general shape properties of the object without regard to spatial positions. In this research PCA will be applied to FDs to capture tissue variations, and then active models will be generated to predicate or locate the target shape.

III. IMPLEMENTATION

The integration of the AR assisted sinus surgery system will be described with hardware and software respectively.

**Hardware.** The system includes four main components: (1) a Tracking/Detection/Imaging (TDI) unit, (2) A stereo display device, (3) a graphic workstation, and (4) a movable cart with one to two passive arms. A stereo camera will be integrated in the TDI unit for 3D tracking purpose. The stereo camera is for tracking fiducials. In theory, it can be used to detect a surface. However, for better accuracy and reliability, we will add a structure light projector in the TDI unit which projects a structure light on the target object. The surface of the object will be calculated from one or both images from the stereo camera with structure light. A separate set stereo imaging camera will be integrated in the TDI unit the same way as in current ART system. It is to meet the multiple requirements in 3D tracking and stereo visualization. We chose stereo because it is especially important for medical applications where the physician needs depth information to make judgments and decisions. Moreover, stereo images, while properly displayed, can greatly enhance the understanding of the augmented display. Suitable stereo display device and graphic workstation will be purchased. The cart and passive arm will be developed from current design of the ART system at CIMI Lab. A 3D scanner will be used before the TDI unit is built to allow surface detection and surface-based registration study to be conducted in parallel.

**Software.** The software will be made up from three modules, i.e., registration, tracking and tissue modelling. The current algorithm of fiducial-based registration for rigid object is to be further improved and tested. New algorithms for surface-based registration will be developed to align the surface detected by the system and the surface reconstructed from MRI or CT image. Surface-based registration for non-rigid object such as the various tissues in the sinus is also to be investigated. The surface of the sinus tissues will be updated in a time-based sequence and the affine transformations links to the series of surfaced are to be calculated. The aim is to simulate/predict the deformation of the tissues under lesion procedure. When the geometric relation between the surface or fiducial points and the internal anatomy changes, the above mentioned registration is not applicable. In this case intraoperative images such as real-time ultrasound or CT is needed and we have to re-register preoperative and intraoperative images with intensity-based method.

The frontal sinus is located between eyes and the cranial cavity (brain). It is bounded by the orbit of the eyeball, the optic nerve and the anterior skull base. In this surgery, the surgeon has to slice through many layers of tissues to the target sinus with limited field of view. Currently, he uses an endoscope which is inserted into the nostril of the patient to help him identify the tissues to be cut and navigate through the path. In current procedure, it is very hard for the surgeon
to know if there is critical structure nearby or just under the surface of the opening. The AR system, with a preoperative CT scan registered and displayed throughout the surgery, would provide the surgeon visibility of the nearby structures surrounding the endoscopes and the tip of the surgical tool. It would help him to navigate the instrument to the targeted site while avoiding the causing damage to other delicate tissues, in the accuracy of sub-millimetre. This is especially helpful for surgeons operating on patients with loose/eroded anatomical landmark due to infection or revised surgery as navigation becomes more difficult. It is also valuable in revised surgery when the normal landmarks may not be present. In severe sinus disease, regular anatomy may be distorted. Nasal polyps may thin the bones, infectious swelling may move a structure to a different place; even in the normal sinus the anatomy can vary. For these reasons, the AR assisted system can make a much safer and more efficient surgery for the patient.

Our proposed methodology will be tried and tested on phantoms, mock-up models, or even cadavers to evaluate the safety, efficacy and accuracy, and then this cutting-edge technology will be introduced into live surgery of real patients. Experimental trials and clinical tests will be conducted at Changyi General Hospital. In addition, after our proposed AR system is developed, it could also be applied to other surgical speciality, such as skull base surgery.

IV. CONCLUSION

In this paper, we propose an AR assisted sinus surgery system with “see through” capability. With the additional virtual information, the system can produce enhanced images to provide a surgeon clearer views. The core competencies for this project consist of four folds: registration, accurate tracking, deformation modeling and AR-based sinus surgery.

Our proposed system is unique in its integration of 3D tracking, surface detection and stereo imaging into a single unit. Such cohesion is required for seamless integration of pre-operative and intra-operative data. With advanced surface-based registration, 3D accurate tracking and statistical tissue deformation modeling, the proposed AR assisted sinus surgery system is capable to mix the virtual view with the real view to visualize the “invisible” parts to the surgeons, so that a safer surgery can be conducted.

REFERENCES