Research

A Kinect-based Medical Augmented Reality System for Craniofacial Applications Using Image-to-Patient Registration

Chung-Hung Hsieh[1], Jiann-Der Lee[1-3,†], Chieh-Tsai Wu[2,4]

Abstract

In this study, we proposed a kinect-based medical augmented reality (AR) system for cranio-facial applications.By using a Kinect sensor to acquire the surface structure of the patient, image-to-patient registration is accomplished by an Enhanced Iterative Closest Point (EICP) algorithm automatically. Moreover, a pattern-free AR scheme is designed by integrating the Kanade-Lucas-Tomasi (KLT) feature tracking and RANdom Sample Consensus (RANSAC) correction, which is better than traditional pattern-based and sensor-based AR environment. The demonstrated system was evaluated with a plastic dummy head and a human subject. Result shows that the image-to-patient registration error is around 3~4 mm, and the pattern-free AR scheme can provide smooth and accurate AR camera localization as the commercial tracking device does.

Keywords:

Medical augmented reality, Image-to-patient registration, Tracking device,Image-Guided Navigation Systems, Computed tomography

Introduction

Recently, image-guided surgery plays an important role in different fields of surgery significantly. Surgeons make surgical planning through various types of preoperative medical imaging, such as computed tomography (CT) or magnetic resonance imaging (MRI), and surgeries are performed according to these preoperative medical images. In general, Image-Guided Navigation Systems (IGNS) help the surgeons to find the spatial relationship between preoperative medical images and physical space. Moreover, they can also provide visualized information for the surgeons to recognize the position of nidus or to verify the surgical path during the surgery [1]. Recently, applying augmented reality (AR) technology to the IGNS is a trend for image-guided surgery. The medical AR system merges medical images or anatomical graphic models into the scene of real world [2-4]. Therefore, the ways to accurately align pre-operative medical images with physical anatomy and to effectively provide anatomical information to the surgeons are two important keys for amodern medical AR IGNS.

In past, in order to perform intraoperative image-to-patient registration, conventionally, a stereotactic frame is fixed to the skull of a patient before scanning the medical images [5,6]. With known landmarks of the stereotactic frame, the

[1]Department of Electrical Engineering, Chang Gung University, Taoyuan, Taiwan 33302

[2]Department of Neurosurgery, Chang Gung Memorial Hospital at LinKou, Taoyuan, Taiwan 333

[3]Department of Electrical Engineering, Ming Chi University of Technology, New Taipei City, Taiwan 24301

[4]Department of Neurosurgery and Medical Augmented Reality Research Center, Chang Gung Memorial Hospital, Tao-Yuan, Taiwan 333

†Author for correspondence: Jiann-Der Lee, Department of Electrical Engineering, Chang Gung University, Taoyuan, Taiwan 333; Department of Neurosurgery, Chang Gung Memorial Hospital at LinKou, Taoyuan, Taiwan 333; Department of Electrical Engineering, Ming Chi University of Technology, New Taipei City, Taiwan 24301. Phone: 03-2118800, Ext: 5316; email: jdlee@mail.cgu.edu.tw

stereotactic frame acts as a bridge to connect the coordinates of medical imaging and the coordinate system of physical space. Although the stereotactic frame is quite reliable for spatial localization, there are several drawbacks while applying in the surgery. First of all, mounting a stereotactic frame to the patient's head is invasive and possibly extends the rehabilitation period. Second, the stereotactic frame is needed to fix on the skull not only during medical image acquisition but also in the whole surgical operation. A method to avoid the shortcomings is using skin-attached fiducial markers [7,8]. Comparing to the stereotactic frame, the skin markers are friendlier and less invasive to patients, but it is less accurate due to the elasticity of the skin.

Later, using natural surface for the image-to-patient registration seems becoming a more proper solution [9-11]. The surface data in 3-D point-cloud form is usually acquired by utilizing non-invasive devices such as laser range scanner (LRS) or time-of-flight camera [12]. By applying a surface registration algorithm, such as the well-known Iterative Closet Point (ICP) [13-16], the surface data can be used to register with the same surface extracted from the patient's medical images. However, the ICP algorithm has two drawbacks: the registration result is sensitive to the initial position of these two data sets; in other words, the algorithm might get trapped into a local minimum. Moreover, the outliers, *i.e.* the data with noise, may also affect the registration accuracy [17].

In additions, image-guided medical systems display the medical information such as image slices or anatomical structures in a virtual reality (VR) coordinate system on a regular screen. Therefore, surgeons need good spatial concept to interpret this virtual-to-real world transformation. In contrast, AR is an alternative but more attractive way for visualization. An AR system in general uses a movable camera to capture images from physical space, and draws graphical objects at their correct position on the images. As a result, the way to estimate the spatial relationship between the camera and the physical space is the most important issue in an AR system [18,19]. The spatial relationship is also known as the extrinsic parameters of AR camera. A precise estimation of the extrinsic parameters ensures that the medical information can be accurately rendered on the scene captured from the physical space. A conventional approach to estimate extrinsic parameters is placing a black-white rectangle pattern in the scene. This pattern forms a reference coordinate system and can easily be detected in the camera field of view (FOV) [20,21] by using computer vision technique. Once the pattern is detected in AR camera frame, the extrinsic parameters can be determined through perspective projection camera model. Alternatively, some studies attach couple of retro-reflective markers, *e.g.* infrared reflective markers, on the AR camera and track these markers by using an optical tracking sensor [22,23]. Therefore, position and orientation of the camera can be estimated by using the position of these markers.

In this study, we proposed a Kinect-based AR system for craniofacial applications. In order to automatically perform image-to-patient registration in the manner of totally non-contacting, Microsoft Kinect is adopted herein. The depth data obtained by Kinect is utilized to reconstruct a point cloud of the patient's craniofacial surface in physical space. An enhanced ICP (EICP) algorithm is then applied to align this surface data to another surface data, which is extracted from CT images of the same patient. The EICP algorithm has two characteristics: First, a random perturbation technique is applied to provide the algorithm the ability to escape from the local minimum. Second, a weighting strategy is added to the cost function of ICP in order to decrease the influence of outlier points.

Furthermore, a Head-Mounted Display (HMD) device mounted with a camera is utilized to achieve pattern-free AR visualization. For the AR camera localization, Scale-Invariant Feature Transform (SIFT) [24] image matching is first applied to find the correspondence between the color images of Kinect and the AR camera. When the AR camera moves, these SIFT feature points are then tracked by using KLT tracking algorithm [25] frame by frame, and thus the extrinsic parameters can keep updating. Moreover, RANdom Sample CONsensus (RANSAC) [26] is applied to make the KLT tracking result smoother in each frame of the AR visualization. Moreover, considering the target for AR rendering might be out of the FOV of the AR camera, a re-initialization step is necessary and we accomplish it by using Speeded-Up Robust Features (SURF) [27] feature matching for target face detection.

In order to evaluate the performance of the proposed system, we have tested the system

with two subjects: a plastic dummy head and a human subject. The extrinsic parameters of the AR camera estimated by the proposed system are evaluated with the parameters estimated by using a commercial optical tracking device.

Materials and Methods

■ System Overview

The system has two main devices. The first one is Microsoft Kinect sensor. The imaging components of the Kinect comprise a color camera, an infrared transmitter, and an infrared CMOS camera. The color camera acquires images with 640x480 pixels in 30 frames per second (fps). The infrared transmitter projects a pattern of IR dots into the physical space, and the depth information of this scene is then reconstructed by the infrared camera. According to Khoshelham, *et al.* [28], the error of Kinect depth estimation is only a few millimeters when the object is located less than one-meter distance to the cameras. Another device utilized for AR visualization is a head mounted display (HMD) device. This HMD device is mounted with a CCD camera with resolution 640x480 on its front to capture the scene of real world. By fusing the anatomical information of the preoperative images on the images captured by the attached camera, the AR visualization can thus be displayed on the HMD screens. The HMD and camera module used in the study is the iWear-VR920 made by VUZIX [29]. **Figure 1** illustrates a flowchart of the proposed system. In general, the proposed system is divided into two stages: registration stage and real-time AR stage.

In the registration stage, marker-free image-to-patient registration is performed. Facial data of the patient is reconstructed from the preoperative CT images, and we utilized an EICP to align the preoperative CT model with the facial data of the patient in physical space, which is extracted by the Kinect depth camera. After the registration process, the transformation between CT coordinate system and the real world coordinate system will be applied to the pre-produced Virtual-Reality Modeling Language (VRML) model of CT for AR visualization of this model.

In the real-time AR display stage, firstly an initialization step is executed. The initialization step loads the VRML model which is generated by previous registration step, and performs SIFT

feature matching to find the correspondence between HMD camera and Kinect color image. Next, in real-time AR stage, for each frame, KLT feature tracking is performed and RANSAC is used for tracking error correction. Finally, the position and orientation of HMD camera can be estimated, and thus the VMRL model of preoperative CT can be rendered on the camera image for AR visualization.

■ Automatic Image-to-Patient Registration

The relationship between all of the coordinate systems for image-to-patient registration is shown in **Figure 2**. There are two coordinate systems used here; one is the coordinate system of preoperative CT image C_{MI} and the other is the coordinate system of the Kinect, which is considered as the world coordinate system C_W, *i.e.* the physical space herein. The image-to-patient registration aligns the preoperative CT images to the position of the patient, and the transformation for alignment is denoted as $^W_{MI}T$.

In this study, Microsoft Kinect is utilized for extracting the patient's facial data in the physical space. Since the color sensor of the Kinect and its depth sensor were calibrated in advanced, each pixel *(x, y)* in the color image is assigned a depth value *D(x, y)* obtained from the depth image [30]. It is assumed that the color camera of Kinect follows the pin-hole camera model [31]. Hence, the relationship of a 3-D point in the physical space $P = (X, Y, Z, 1)^T$ and its projection point $p = (x, y, 1)^T$ on the image are expressed by Eq. (1):

$$sp = KEP, K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, E = \begin{bmatrix} R & T \end{bmatrix} \quad (1)$$

Where *s* is a scale factor. K is the matrix representing the intrinsic parameters of camera, which comprises the focal lengths (f_x, f_y) and the principle point (c_x, c_y). E is called extrinsic parameters including a rotation matrix R and a translation vector T. Here the 3-D coordinate system of Kinect color camera is considered as the world coordinate system C_w, and thus Eq. (1) can be written as

$$sp = KP \quad (2)$$

Giving a pixel in the color image *(x, y)* and its corresponding depth *D (x, y)*, the 3-D coordinate *(X, Y, Z)* of this point can be calculated by using Eq. (3):

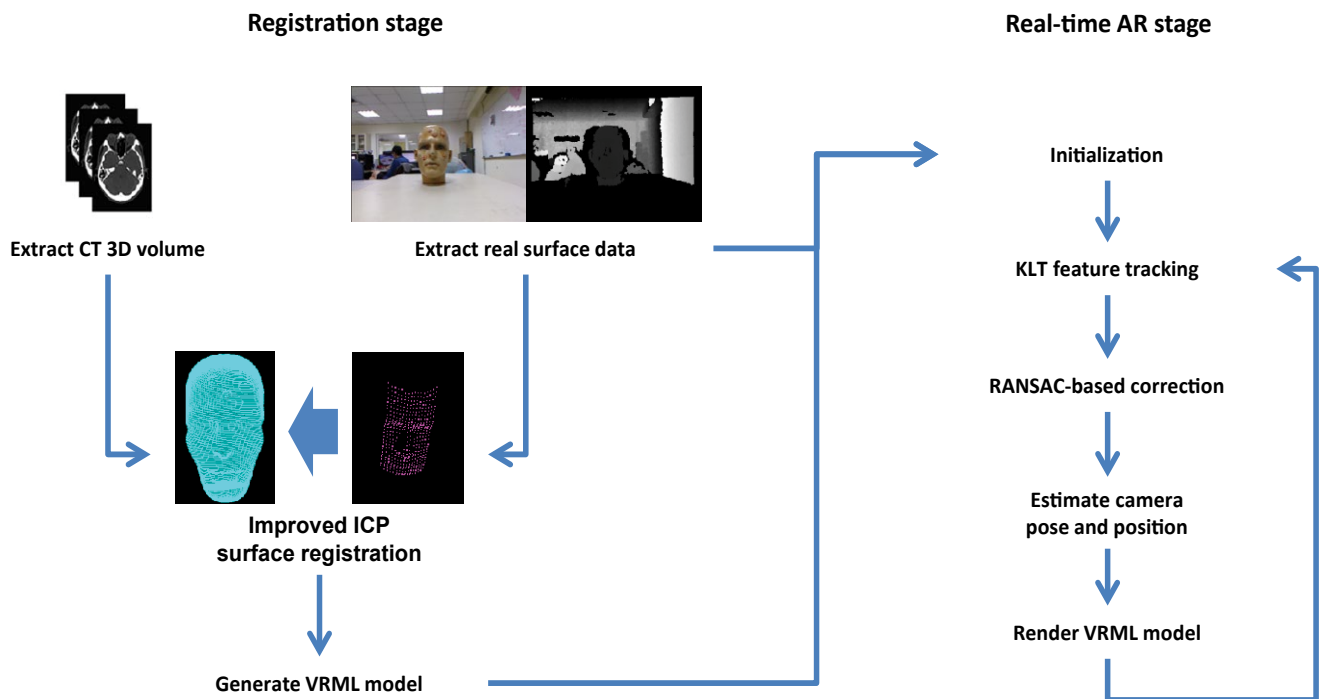**Registration stage**   **Real-time AR stage**
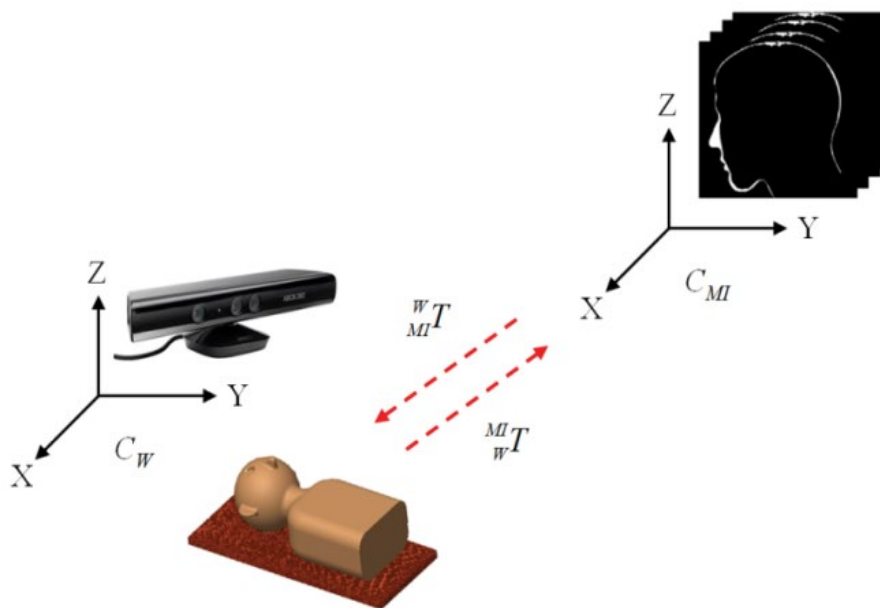

Figure 1: An overall flowchart of the proposed system.

Figure 2: Spatial relationships among the involved coordinate systems for the image-to-patient registration.

$$X = \frac{(x - c_x)}{f_x} Z$$

$$Y = \frac{(y - c_y)}{f_y} Z$$

$$Z = D(x, y) \tag{3}$$

To generate facial data of the patient, first a region of interest (ROI) representing the patient's face in the Kinect color image is selected manually. For all pixels within this ROI, their 3-D coordinates are assigned by using Eq. (3) and form a set of point cloud, denoted as P_{face}. The set of point cloud represents the facial

data of the patient in physical space. Meanwhile, the color information of the same ROI, called facial template I_{face}, is also stored for the next step, *i.e.* the AR camera localization. The other surface data for registration is extracted from the preoperative CT images of the same patient by using commercial software named AMIRA [32], and a VRML model of this surface data is constructed for AR visualization. The point-cloud form of the VRML model, denoted as the reference point dataset P_{CT}, is used for the image-to-patient alignment.

In order to align these two surface data, we have designed an EICPalgorithm to accomplish the surface registration task. The EICP algorithm is basically based on the original ICP registration algorithm, but some steps are added to overcome the disadvantages of traditional ICP. Assuming the facial data P_1 is the floating data $F = \{f_1, f_2, \ldots, f_m\}$ and the CT surface P_2 is the reference data $Q = \{q_1, q_2, \ldots, q_n\}$, original ICP uses rigid transformation to align these two 3-D data point sets in an iterative manner. In each iteration of ICP, firstly every point f_i in F finds their closest point q_j in Q, and a cost function C is evaluated based on the distance of each corresponding pair (f_i, q_j). In the EICP algorithm, we modified the cost function C of ICP by adding a weighting function to the distance of all closest corresponding pair (f_i, q_j) to deal with the problem of outliers, as shown in Eq. (4):

$$C = \frac{1}{m}\sum_{i=1}^{m} w_i \min_{j \in \{1,2,\ldots,n\}} \left\| f_i - q_j \right\| \quad (4)$$

Where w_i is a weighting function, determined according to the median of distances of all the corresponding pairs, as defined by Eq. (5):

$$w_i = \begin{cases} 1 & f \quad \left\| f_i - q_j \right\| < median \\ \dfrac{median}{\left\| f_i - q_j \right\|} & otherwise \end{cases} \quad (5)$$

Furthermore, a random perturbation strategy is added to provide the algorithm the ability to escape from the local minimum. The way that ICP reaches a local minimum acts as a gradient descent approach. In each iteration of ICP, the cost function is evaluated at the current solution and then move along the direction of gradient to the local minimum. When the registration goes to convergence, we can get a transformation solution T which projects a set of points onto another set of points, and the total distance between these two point sets is the smallest.

When the facial data of patient and the preoperative CT surface data have been registered, the transformation between the CT coordinate system and the real world coordinate system is thus estimated. The transformation is then applied to a pre-constructed VRML model of preoperative CT, transforming the VRML model to the position of patient for AR visualization.

■ Pattern-free AR visualization

In this study, a HMD-based AR system featuring on its pattern-free camera localization is proposed. The flowchart is shown in **Figure 3**. First, we check whether the patient's face is in the FOV of the AR camera by using a SURF face detection approach. If the patient's face is detected, a SIFT-based face-matching algorithm is applied to find the correspondence between these two images. Once the corresponding relationship is established, the 3-D coordinate of the SIFT features can thus be assigned to the corresponding features on the image of the AR camera. Therefore, the extrinsic parameters of the AR camera can thus be estimated by using the 3-D coordinate of these features. Since the AR camera is movable, after estimating the extrinsic parameters on the first frame where the face was detected, a feature-point tracking algorithm, *i.e.* KLT tracker, is embedded into a RANSAC framework to correct tracking-failed points. As a result, the extrinsic parameters of the AR camera can be estimated continuously, and the anatomical information of the preoperative CT can thus be rendered on the images of AR camera in a stable and smooth manner.

Experimental Results

■ Hardware devices

The key image acquisition device in the proposed system is Microsoft Kinect. The imaging components of the Kinect sensor comprise a color camera, an infrared transmitter, and an infrared CMOS camera. The color camera acquires images with 640×480 pixels in 30 frames per second (fps). The infrared transmitter projects a pattern of IR dots into the physical space, and the depth information of this scene is then reconstructed by the infrared camera. After a calibration procedure on the color camera and the infrared camera, depth information obtained by the infrared camera can thus be mapped to each pixel on the color image. According to Khoshelham [28], the error of Kinect depth estimation is only a few millimetres when the object is located less than one-meter distance to the cameras.
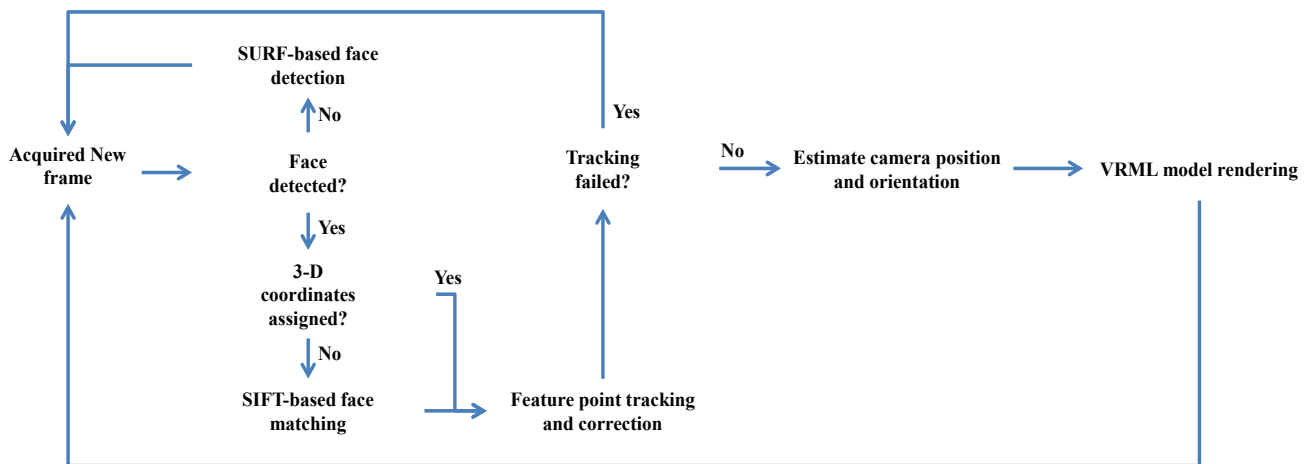
Figure 3: Flowchart of the proposed pattern-free AR visualization.

Another device utilized for AR visualization is a head mounted display (HMD) device. This HMD device is mounted with a CCD camera with resolution 640x480 on its front to capture the scene of the physical space. By fusing the anatomical information of the preoperative images on the images captured by the attached camera, the AR visualization can thus be displayed on the HMD screens. The HMD and camera module used in the study is the Iwear-VR920 made by VUZIX Inc.

■ **Accuracy evaluation of image-to-patient registration**

In order to evaluate the accuracy of the image-to-patient registration of the proposed system, a plastic dummy head was utilized as a phantom. Before scanning CT images of the phantom, five skin markers were attached on the face of the phantom, as shown in **Figure 4(a)**. Since the locations of these skin markers could easily be identified in the CT images, these markers were considered as the reference to evaluate the accuracy of registration. Meanwhile, as shown in **Figure 4(b)**, a commercial 3-D digitizer - G2X produced by MicroScribe [33], was utilized to establish a reference coordinate system and estimate the location of the markers in the physical space. According to its specification, the accuracy of G2X is 0.23 mm, which is suitable for locating the coordinates of skin markers as the ground truth for evaluation.

Figure 5 shows the setup of the Kinect, the digitizer, and the preoperative medical images in the experiment for accuracy evaluation. Before evaluation, a calibration step was performed to find the transformation $_{W}^{Digi}T$ between Kinect

C_W, i.e. the world coordinate system, and the digitizer's coordinate system C_{Digi}. To perform the calibration, a triangular prism attached with chessboard patterns, as shown in **Figure 4(c)**, was placed in the FOV of Kinect sensor. Corner points of the chessboard were selected by the digitizer and on the color image obtained by the Kinect. The two sets of 3-D points, represented by the coordinate systems of the digitizer and the Kinect respectively, were used to estimate a transformation $_{W}^{Digi}T$ by using least mean square (LMS) method. Therefore, the 3-D points reconstructed by Kinect can be transformed to the digitizer's coordinate system C_{Digi}.

Firstly, as shown in **Figure 5(a)**, a plastic dummy head was placed in front of the Kinect at a 60-cm distance. The G2X digitizer was used to obtain the 3-D coordinates of these markers, and the Kinect was utilized to reconstruct the head's surface. Its depth information was sampled into a point cloud with approximately 220 points. The point cloud was transformed to the C_{Digi} by applying $_{W}^{Digi}T$, and then registered to the preoperative CT images by using the EICP algorithm. Image-to-patient registration is evaluated by calculating the target registration error (TRE) [34-37] of the five skin markers. The TRE for evaluation is defined as

$$TRE = M_{CT,i} - MI_W T(M_{face,i}) \qquad (9)$$

where $M_{CT,i}$ denotes the coordinate of the i-th marker in the CT coordinate system, and $M_{face,i}$ is the coordinate of the i-th marker in C_{Digi}. Transformation $_{W}^{MI}T$ represents the rigid transformation obtained from the EICP algorithm. **Figure 5(b)** shows the initial spatial position of the reconstructed facial data
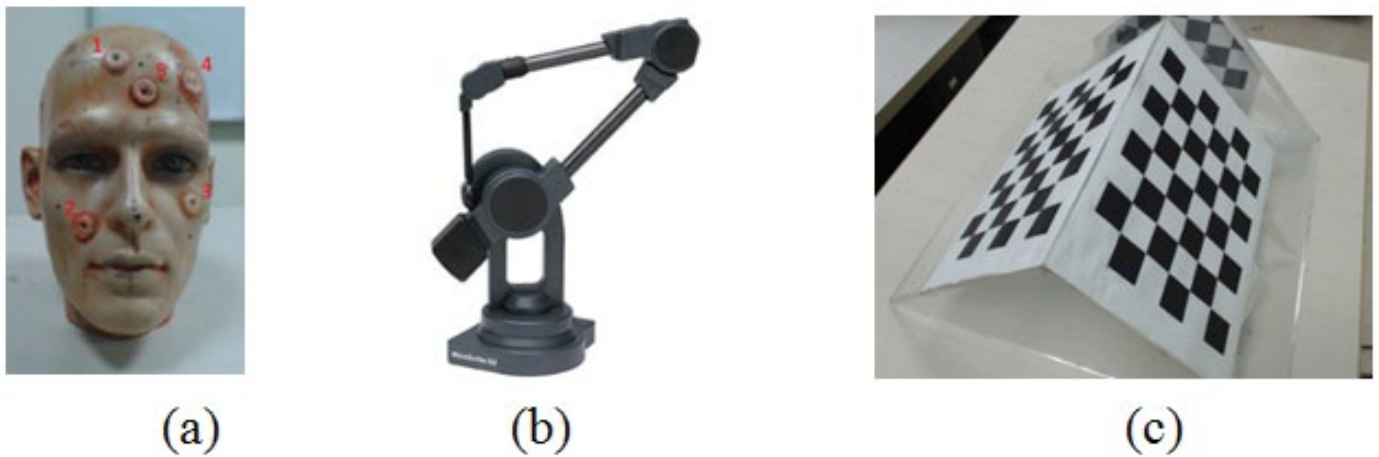
Figure 4: Tools to evaluate the accuracy of the image-to-patient registration (a) plastic dummy head (b) MicroScribe G2X digitizer (c) traingular prism for coordinate system calibration.

(magenta) and CT surface data (cyan), while **Figure 5(c)** shows the registration result. The image-to-patient registration procedure was performed repeatedlyin 100 times, and each time we slightly shift the location and orientation of the phantom.The TREs at each registration procedure were recorded and the means and the standard deviation are shown in **Table 1**. The mean TREs of the skin markers are within a range of 2~4 mm.

Next, we tested the proposed EICP algorithm with other ICP-based algorithm on the Stanford bunny [34]. The point numbers of Stanford bunny is 40256, and one hundred points were randomly selected from the original data set. Then we applied a geometric transform to generate the floating data by rotating 50° along x-axis, y-axis and z-axis, respectively. The original Stanford bunny was regarded as the reference data, the initial positions of reference data and floating data are shown in **Figure 6(a)**. The EICP, Adaptive-ICP [35], and the Random-ICP [36] were applied to register the floating data to the reference data, and the registration result is shown in **Figure 6(b)-6(d)**. In additions, the root-mean-square (RMS) vs. the number of iterations for EICP algorithm is shown in **Figure 6(e)**. We also calculated the RMS of the distance between registered floating points and their reference points to compare the accuracy of these three ICP-based algorithms. The results are listed in **Table 2**. Obviously, both of the Adaptive-ICP and Random-ICP were trapped into a local optimum, while the EICP shows its ability to find a better solution which is much closer to the global optimum.
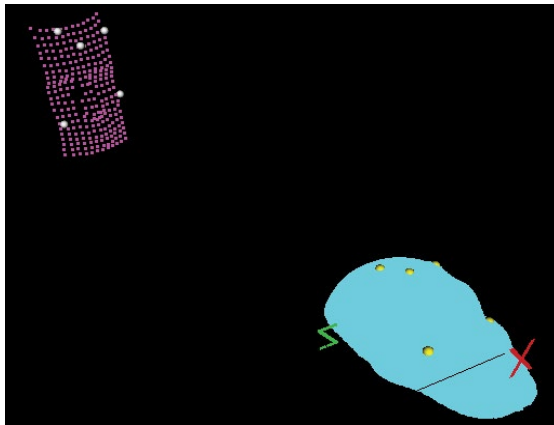
■ Working range of the image-to-patient registration

In order to determine the working range of the image-to-patient registration, we evaluated the registration accuracies,*i.e.*TREs,in various distances from the Kinect sensor to the dummy head. The distancesfrom the Kinect sensor to the dummy headwere set from 55 cm to 95 cm with a 5-cm interval. At each position, the image-to-patient registration procedure was repeated 10 times to calculate a mean TRE of the skin markers. The graph of the mean TRE at each position versus distance is shown in **Figure 7(a)**.
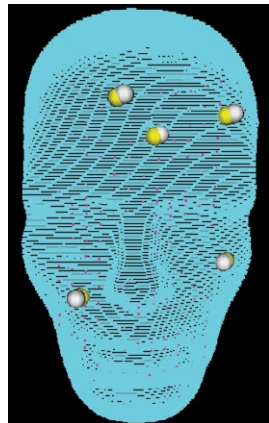
From **Figure 7(a)**, the mean TRE is around 3.5 mm and increases slightly within the range between 55 cm to 75 cm. When the distance exceeds 75 cm, the mean TRE continuously increases to around 5 mm as the distance increases. **Figure 7(b)** reveals the relationship between the numbers of the surface data points captured by Kinect and the distance between the Kinect and the dummy head. When the distance increases, the number of points P_{face}in ROI relatively reduces. The reduction in the number of surface data points causes the increment of mean TRE. **Figure 7(c)** shows the computing time cost for registration at different distances. The time cost of image-to-patient registration is reduced with the increment of distance because of the reduction in the number of surface data points. Therefore, considering the trade-off between the registration accuracy andcomputing cost, the distance at 70 cm to 75 cm seems to be an appropriate choice for a suitable working distance for the proposed system.
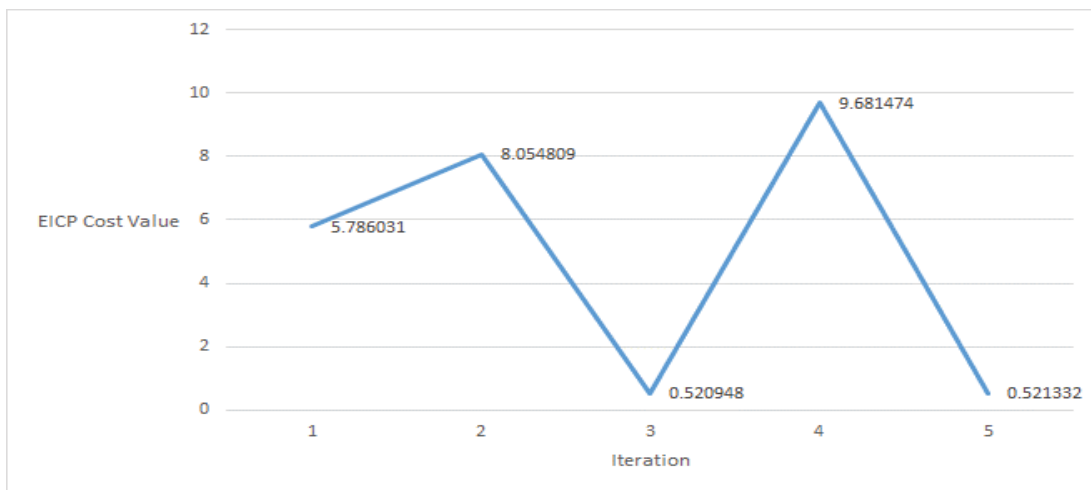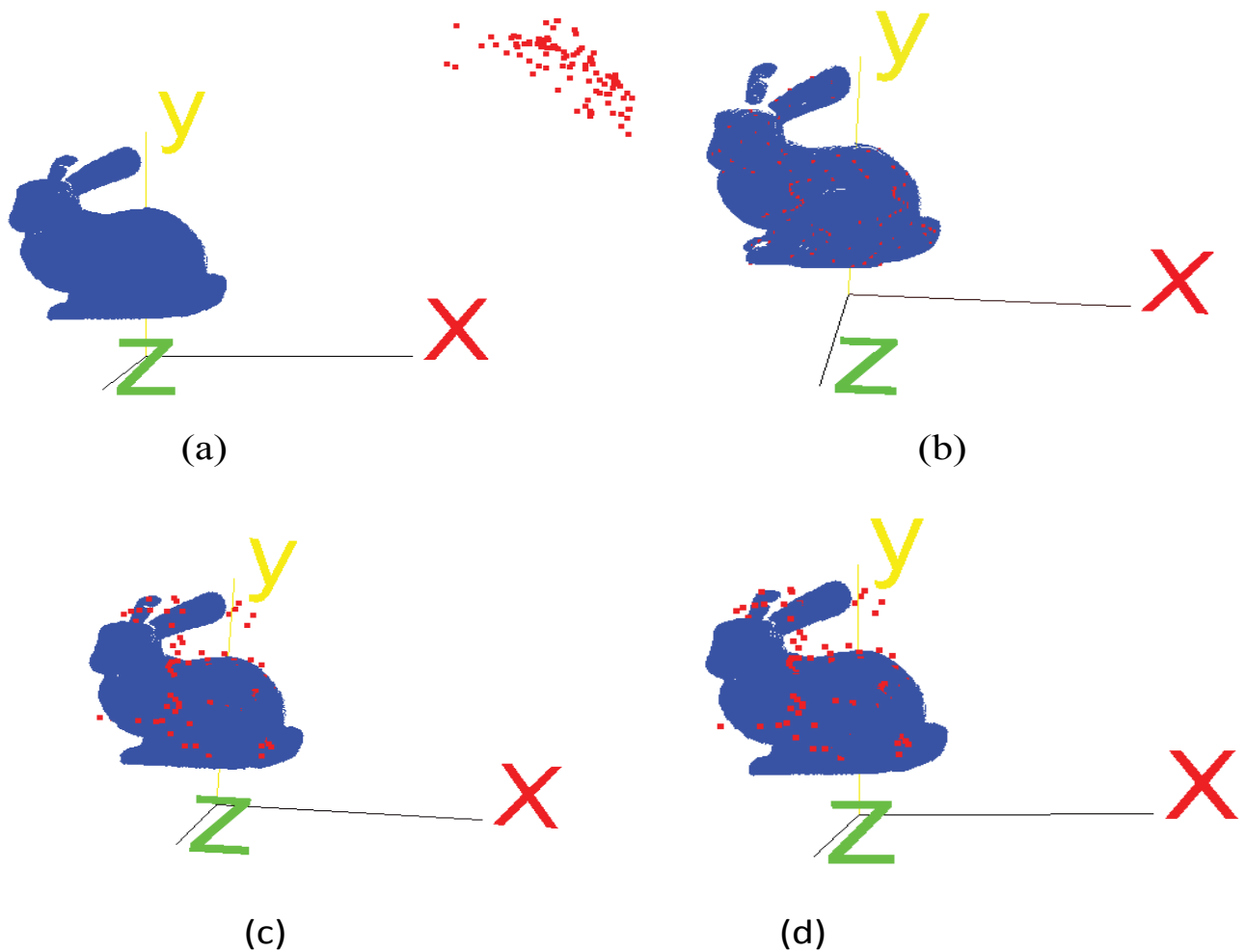
(a)

(b)

(c)

Figure 5: Accuracy evaluation for image-to-patient registration. (a) Scene of experimental setup. (b) Initial position of the two surface data sets for registration, displayed in VR. The points in magenta are facial surface reconstructed by Kinect, and the points in cyan are CT surface points. The white spheres are the skin-attached markers in physical space and the yellow spheres stand for the coordinates of these markers in CT coordinate system (c) Registration result after performing the EICP algorithm.

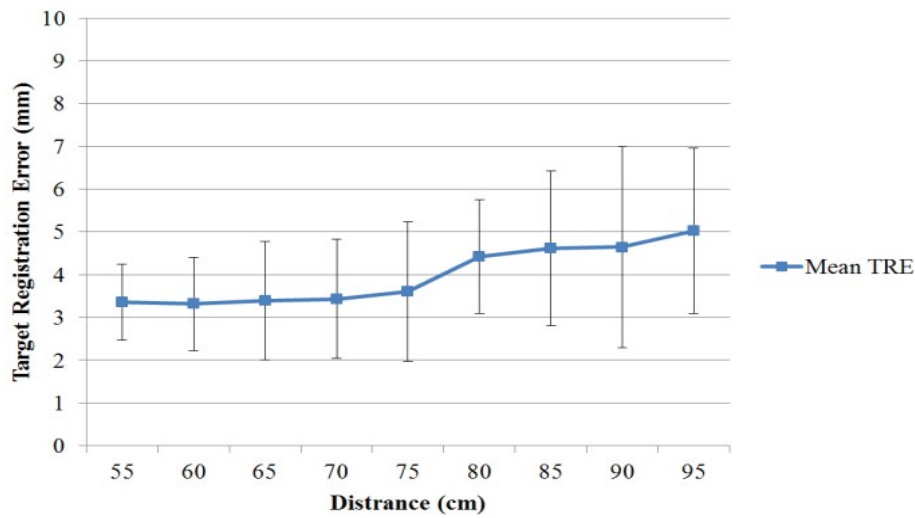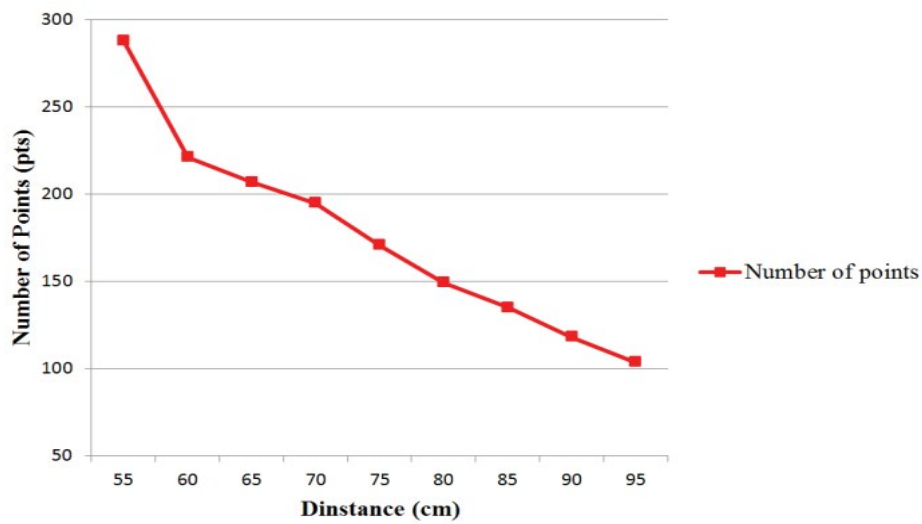| Table 1: Target registration error of each skin-attached marker after image-to-patient registration.					
	Target 1	Target 2	Target 3	Target 4	Target 5
Mean TRE (mm)	3.47	2.22	3.06	3.55	3.77
Standard deviation (mm)	1.49	0.61	0.84	1.59	1.44

Figure 6: Registration results of Stanford bunny (a) Initial position (b) Result of the proposed EICP algorithm (c) Result of Adaptive-ICP algorithm (d) Result of Random-ICP algorithm (e) Rms vs. iteration iterations curve for EICP the algorithms.

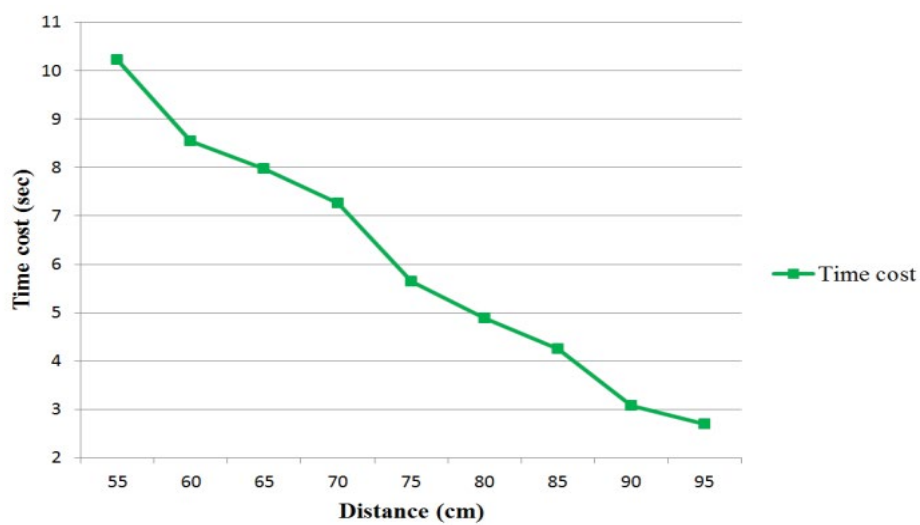| Table 2: RMS results of each registration algorithm.			
	EICP	**Adaptive-ICP**	**Random-ICP**
RMS(unit)	0.52	28.77	27.25

(a)

(b)

(c)

Figure 7: Experimental results for analyzing the accuracy at various distances. (a) Mean TRE when the dummy head placed at different distance from the Kinect. The error bars stand for the standard deviations of TRE. (b) The relationship between distances and number of points reconstructed from ROI. (c) Time cost for registration at different distance.

Another factor which might affect the registration accuracy is the capturing angle between the Kinect and the subject. Suppose the central line perpendicular to the Kinect is zero degree. Therefore, under the situation that the dummy head is placed at the 70-cm distance from the Kinect, we rotated the Kinect from -20 to 20 degrees and estimated the TRE every 5 degrees. The mean TREs at different angles were estimated and revealed in **Figure 8**. The experimental results show that in angles between -15°to 15°, the mean TREs are slightly increased when the rotating angle is away from the central line. As a result, making the Kinect facing the subject as straight as possible ensures relative good performance.

Augmented reality visualization results

The proposed AR system was tested with two subjects -aplastic dummy head and a real human. In the dummy head case, its VRML model constructed from CT imagesor CT image slices were rendered in the AR images. The AR visualization results from different viewpoints are shown in **Figure 9**. The preoperative CT model is well aligned to the position of dummy head. When the camera starts moving, SIFT features are tracked by KLT algorithm, and the extrinsic parameters of the camera are estimated frame by frame.

The proposed system had also been tested on a real human case. CT images of the subject are obtained beforehand. Head surface data and skull surface data were extracted respectively to construct their VRML models for AR visualization. **Figure 10(a)** shows the AR visualization results with the reconstructed facial surface, while **Figure 10(b)** shows the rendered skull surface.

Conclusion

For a medical AR system, image-to-patient registration and AR camera localization are two primary key issues. Traditionally, the image-to-patient registration is accomplished by the aids of stereotactic frame or skin markers. These approaches usually cause discomfort or inconvenience to the patient. On the other hand, in the issue of the AR camera localization, a known pattern is required to be placed within the FOV of the AR camera for the purpose on estimating extrinsic parameters of the camera.In this study, these shortcomings are improved. In the Kinect based medical AR system we demonstrated, a marker-free image-to-patient registration is accomplished by utilizing a Microsoft Kinect sensor to construct the surface data of a patient, and an EICP based surface registration technique is performed to align the preoperative medical image model to the real position of the patient. Moreover, a HMD mounted with a CCD camera is cooperated with the color camera of the Kinect to accomplish a pattern-
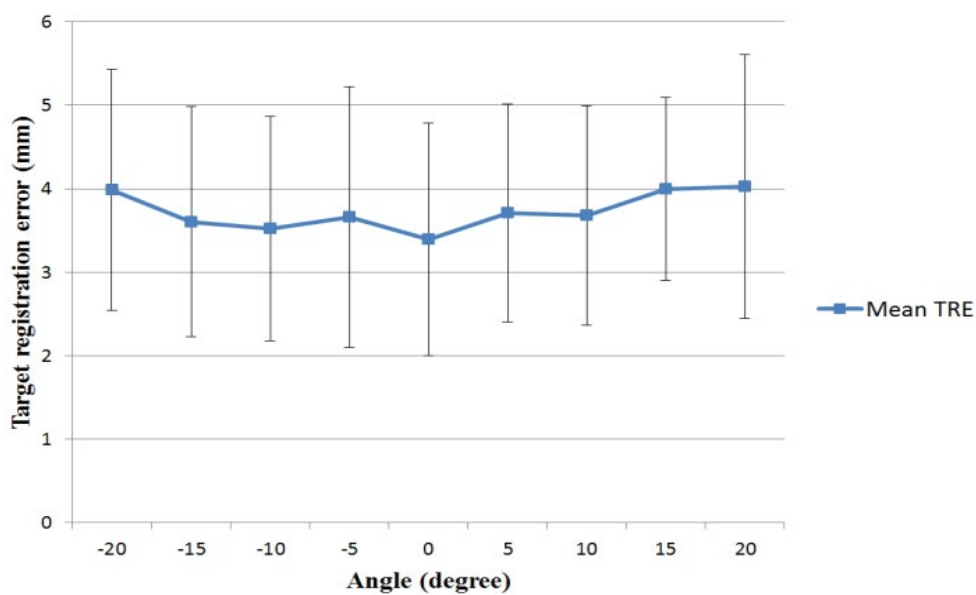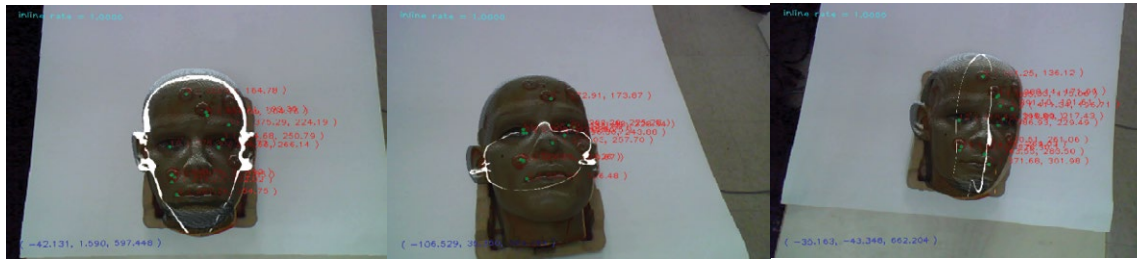

Figure 8: The mean TREs in different camera angles. Error bars stand for standard deviation of TRE.

Figure 9: Augmented reality visualization results of the proposed medical AR system using the plastic phantom head as testing subject. (a) Rendering VRML model of preoperative CT. (b) Plotting CT slice in HMD camera frame.

Figure 10: Augmented reality visualization result of real human case. (a) Rendering VRML model of preoperative CT. (b) Rendering VRML model of skull extracted from CT image.

free AR visualization. A SIFT-based feature point matching and a RANSAC-based correction are integrated to solve the problem of the AR camera location without using any pattern, and our experimental results demonstrate that the proposed approach provides an accurate, stable, and smooth AR visualization.

Comparing to conventional pattern-based AR systems, the proposed system uses only nature features to estimate the extrinsic parameters of the AR camera. As a result, it is more convenient and practical to be utilized because the FOV of the AR camera is not limited on the visibility of the AR pattern. A RANSAC-based correction technique is proposed to improve the robustness of the extrinsic parameter estimation of the AR camera. The proposed system has been evaluated on both

image-to-patient registration and AR camera localization with a plastic dummy head. The system has also been tested on a human subject and promising AR visualization is demonstrated. This study is just a preliminary implementation of a marker-free and pattern-free medical AR system for craniofacial surgeries. In the future, extensive clinical trials are expected for further investigation.

References

1. Hong J, Matsumoto N, Ouchida R, *et al.* Medical navigation system for otologic surgery based on hybrid registration and virtual intraoperative computed tomography. *IEEE. Trans. Biomed. Eng* 56(2), 426-432 (2009).

2. Sielhorst T, Feuerstein M, Navab N. Advanced Medical Displays: A Literature Review of Augmented Reality. *J. Disp. Technol* 4(4), 451 - 467 (2008).

3. Ferrari V, Megali G, Troia E, *et al.* A 3-D Mixed-Reality System for Stereoscopic Visualization of Medical Dataset. *IEEE. Trans. Biomed. Eng* 56(11), 2627-2633 (2009).

4. Liao H, Inomata T, Sakuma I, *et al.* 3-D Augmented Reality for MRI-Guided Surgery Using Integral Videography Autostereoscopic Image Overlay. *IEEE. Trans. Biomed. Eng* 57(6), 1476 - 1486 (2010).

5. Eggers G, Muhling J, Marmulla R. Image-to-patient registration techniques in head surgery. *Int. J. Oral. Max. Sur* 35(12), 1081-1095 (2006).

6. Lee JD, Huang CH, Lee ST. Improving stereotactic surgery using 3-D reconstruction. *IEEE. Eng. Med. Biol* 21(6), 109-116 (2002).

7. Hamming NM, Daly MJ, Irish JC, *et al.* Effect of fiducial configuration on target registration error in intraoperative cone-beam CT guidance of head and neck surgery. *Conf. Proc. IEEE. Eng. Med. Biol. Soc* 21(6), 109-116 (2008).

8. Shamir RR, Joskowicz L, Shoshan Y. Optimal landmarks selection and fiducial marker placement for minimal target registration error in image-guided neurosurgery. *Proc. Med. Imaging* (2009).

9. Cash DM, Sinha TK. Incorporation of a laser range scanner into image-guided liver surgery: surface acquisition, registration and tracking. *Med. Phys* 30(7), 1671-1682 (2003).

10. Miga MI, Sinha TK, Cash DM, *et al.* Cortical surface registration for image-guided neurosurgery using laser-range scanning. *IEEE. Trans. Med. Imaging* 22(8), 973-985 (2003).

11. Clements LW, Chapman WC, Dawant BM, *et al.* Robust surface registration using salient anatomical features for image-guided liver

surgery: Algorithm and validation. *Med. Phys* 35(6), 2528-2540 (2008).

12. Placht S, Stancanello J, Schaller C, *et al.* Fast time-of-flight camera based surface registration for radiotherapy patient positioning. *Med. Phys* 29(1), 4-17 (2012).

13. Fieten L, Schmieder K, Engelhardt M, *et al.* Fast and accurate registration of cranial CT image with A-mode ultrasound. *Int. J. Comput. Ass. Rad* 4(3), 225-237 (2009).

14. Ma B, Ellis RE. Robust registration for computer-integrated orthopedic surgery: Laboratory validation and clinical experience. *Med. Image. Anal* 7(3), 237-250 (2003).

15. Almhdie A, Leger C, Deriche M, *et al.* 3D registration using a new implementation of the ICP algorithm based on a comprehensive lookup matrix: Application to medical imaging. *Pattern. Recogn. Lett* 28(12), 1523-1533 (2007).

16. Shi L, Wang D, Hung VW, Yeung BHY, *et al.* Fast and accurate 3-D registration of HR-pQCT images. *IEEE. Trans. Inf. Technol. Biomed* 14(5), 1291-1297 (2010).

17. Besl PJ, McKay HD. A method for registration of 3-D shapes. *IEEE. Trans. Pattern Anal. Mach. Intell* 14(2), 239-256 (1992).

18. Fischer J, Neff M, Freudenstein D, *et al.* Medical augmented reality based on commercial image guided surgery. *Proc. Eurographics. Symp. Virt. Environ* 83-86 (2004).

19. Bichlmeier C, Wimme F, Heining SM, *et al.* Contextual Anatomic Mimesis Hybrid In-situ Visualization Method for Improving Multi-sensory Depth Perception in Medical Augmented Reality. *Proc. 6th IEEE. ACM. Int. Symp. Mixed. Augmented. Reality* 129-138 (2007).

20. Kato H, Billinghurst M. Marker tracking and HMD calibration for a video-based augmented reality conferencing system. *Proc. 2nd IEEE. ACM Int. Workshop. Augmented. Reality* 85-94, (1999).

21. Regenbrecht HT, Wagner MT. Interaction in a collaborative augmented reality environment. *Proc. CHI. Extend. Abstract. Human. Factors. Comput. Systems* 504-505 (2002).

22. Sato Y, Nakamoto M, Tamaki Y, *et al.* Image guidance of breast cancer surgery using 3-D ultrasound images and augmented reality

visualization. *IEEE. Trans. Med. Imag* 17(5), 681-693 (1998).

23. Pustka D, Huber M, Waechter C, *et al.* Automatic configuration of pervasive sensor networks for augmented reality. *IEEE. Pervas. Comput* 10(3), 68-79 (2011).

24. Lowe DG. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis* 60(2), 91-110 (2004).

25. Tomasi C, Kanade T. Detection and tracking of point features. Technical Report CMU-CS-91-132, Carnegie Mellon University (1991).

26. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* 24(6), 381-395 (1981).

27. Bay H, Ess A, Tuytelaars T, *et al.* Speeded-up robust features. *Comput. Vis. Image. Underst* 110(3), 346-359 (2008).

28. Khoshelham K, Elberink SO. Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications. *Sensors* 12(2), 1437-1454 (2012).

29. http://www.vuzix.com/consumer/products_vr920.html

30. Smisek J, Jancosek M, Pajdla T. 3D with Kinect. *IEEE. Int. Conf. Computer. Vision. Workshops* 1154-1160 (2011).

31. Zhang Z. A flexible new technique for camera calibration. *IEEE. Trans. Pattern. Anal. Mach. Intell* 22(11), 1330-1334 (2000).

32. http://www.amira.com/

33. http://www.emicroscribe.com/

34. http://graphics.stanford.edu/data/3Dscanrep/

35. Lee JD, Huang CH, Liu LC, et al. A modified soft-shape-context ICP registration system of 3-D point data. *IEICT. Trans. Inf. Syst* E90-D(12), 2087-2095 (2007).

36. Penney GP, Edwards PJ, King AP, *et al.* A Stochastic Iterative Closed Point Algorithm. *Lect. Note. Comput. Sci* 2208(1), 762-769 (2001).

37. Fitzpatrick JM, West JB, Maurer Jr CR. Predicting error in rigid-body point-based registration. *IEEE. Trans. Med. Imaging* 17(5), 697-702 (1998).