Research Article

Markerless Augmented Reality via Stereo Video See-Through Head-Mounted Display Device

Chung-Hung Hsieh and Jiann-Der Lee

Department of Electrical Engineering, Chang Gung University, Taoyuan 333, Taiwan

Correspondence should be addressed to Jiann-Der Lee; jdlee@mail.cgu.edu.tw

Conventionally, the camera localization for augmented reality (AR) relies on detecting a known pattern within the captured images. In this study, a markerless AR scheme has been designed based on a Stereo Video See-Through Head-Mounted Display (HMD) device. The proposed markerless AR scheme can be utilized for medical applications such as training, telementoring, or preoperative explanation. Firstly, a virtual model for AR visualization is aligned to the target in physical space by an improved Iterative Closest Point (ICP) based surface registration algorithm, with the target surface structure reconstructed by a stereo camera pair; then, a markerless AR camera localization method is designed based on the Kanade-Lucas-Tomasi (KLT) feature tracking algorithm and the Random Sample Consensus (RANSAC) correction algorithm. Our AR camera localization method is shown to be better than the traditional marker-based and sensor-based AR environment. The demonstration system was evaluated with a plastic dummy head and the display result is satisfactory for a multiple-view observation.

1. Introduction

In the past few decades, augmented reality (AR) has become an attractive research topic with potentials in the fields of machine vision and visualization techniques. As a part of mixed reality and an alternative to virtual reality (VR), AR allows users to see virtual objects overlay the physical environment at the same time. According to the definition by Azuma [1, 2], AR system has three important properties: first, it combines real and virtual objects in a real environment; second, it runs interactively, and in real-time; and third, it registers (aligns) real and virtual objects with each other. AR creates a "next generation, reality-based interface of human-computer interaction" [3], provides information beyond what the user can normally see, and augments our real world experiences [4, 5].

Over the years, the applications of AR visualization had been expanded to include more and more areas, such as engineering, industrial manufacturing, navigation, entertainment, and especially medical applications. The aim of medical AR-based applications starts from a simple concept, to see through the virtual object and see the patient's medical information along with the patient. AR brings benefits to medical applications since it can help visualize medical information along with the patient at the same time and within the same physical space. In medical fields, AR creates a way for advanced medical display [6], which can be applied to telementoring [7], medical procedure training [8], or medical data visualization [9]. Recently, applying augmented reality (AR) technology to the image-guided navigation system (IGNS) has become a new trend in medical technologies. A medical AR system merges medical images or anatomical graphic models into a scene of the real world [6, 10, 11]. From the literatures, we find that a successful medical AR system has to deal with two major problems: first, how to accurately align preoperative medical information with the physical anatomy and, second, how to effectively and clearly provide virtual anatomical information.

Conventional medical systems and applications display medical information such as image slices or anatomical structures using the virtual reality (VR) coordinate system on a regular screen. Therefore, surgeons using such systems need to have a good spatial concept in order to interpret the virtual-to-real world transformation. In order to train a surgeon

to have a god spatial concept, this usually requires a lot of clinical experience. In contrast, AR provides an attractive alternative for medical information visualization because the displays use the real world coordinate system. In general, a video see-through AR system uses a movable camera to capture images from the physical space and draws virtual objects onto the correct positions on the images. As a result, the way to estimate the spatial relationship between the camera and the physical space, that is, camera localization, is the most important problem to be solved in an AR system [12, 13]. The spatial relationship is also known as the extrinsic parameters of the AR camera, which are three translation parameters and three rotation parameters, and can be referred to as the position and orientation of the 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 these extrinsic parameters is to place a black-white rectangle pattern within the scene which can be used as a reference to be detected in the camera field of view (FOV) [14, 15] using computer vision techniques. Once the reference pattern is detected in the AR camera frame, the extrinsic parameters can be determined using the perspective projection camera model. Alternatively, some studies attach a couple of retroreflective markers, for example, infrared reflective markers, on the AR camera and track these markers by using an optical tracking sensor [13, 16, 17] in order to estimate the position and orientation of the camera by using the positions of these markers. However, in the pattern-based or sensor-based AR display, the FOV of camera is limited because these artificial markers should be observed with the FOV. In addition, the accuracy of AR rendering depends on the size and likelihood of correct identification of the pattern or markers within the FOV.

In this paper, we present a markerless AR scheme for medical applications. The markerless AR scheme is mainly based on a Video See-Through Head-Mounted Display (HMD) device with a stereo camera on it. We use the stereo camera to reconstruct feature point cloud of the target for AR visualization. An improved ICP algorithm is then applied to align the surface data to the virtual object created from medical information. The improved ICP algorithm has two additional characteristics. First, a random perturbation technique is applied to provide the algorithm opportunities to escape the local minima. Second, a weighting strategy is added to the cost function of the ICP in order to weaken the influence of outlier points.

When the AR camera moves, feature points are then tracked by using the Kanade-Lucaz-Tomasi tracking algorithm [18] on a frame by frame basis while updating the extrinsic parameters. Moreover, the Random Sample Consensus (RANSAC) [19] algorithm is applied to keep the KLT tracking results stable in each frame in order to make the AR visualization smoother and more accurate. Furthermore, considering the target for AR rendering might be out of the FOV of the AR camera, a reinitialization step is necessary and can be accomplished by applying the Speeded-Up Robust Features (SURF) [20] feature matching algorithm for target detection.

The remainder of this paper is organized as follows: in Section 2, we report on some related works about AR visualization. Section 3 includes all material and methods of the proposed system. Experimental results are presented and discussed in Section 4. Finally, Section 5 gives the conclusion of this work.

2. Related Works

Camera localization is a key component in dense 3D mapping, Simultaneous Localization and Mapping (SLAM) and full-range 3D reconstruction. It is also the keypoint problem of video-based AR visualization since we need to know the pose and position of the camera in order to render the virtual object onto the camera scene. There are different types of methods for camera localization, and the most well-known category of camera localization methods belongs to the planar-based methods. In the past 10 years, the planar-based methods have become one of the major types in researches for AR camera localization methods; one of the examples is the widely used AR environment development library ARToolkit [21]. In this type of method, a predesigned rectangle marker would be placed inside the camera's FOV, and a tracking algorithm is then applied to the captured frames in order to track the marker [22, 23].

Another category is the landmark-based approaches [24]. In this type of method, several landmark feature points are extracted from the camera image by using feature extraction algorithms. Descriptors of these landmark feature points are also extracted for tracking and comparison with the 3D locations of these landmark feature points. In each frame, these landmark feature points are picked by tracking or comparing their corresponding feature descriptors' similarities, and the pose of the camera can then be determined [25, 26]. Besides, Kutter et al. [27] proposed a marker-based approach to render the patient's volume data on a HMD device. Their scheme provides an efficient volume rendering in an AR workspace and solves the problem of occlusion by the physician's hands. Later, Wieczorek et al. [28] extended this scheme by improving the occlusion due to the medical instrument and added new functions such as virtual mirror to the AR system. Suenaga et al. [29] also proposed a fiducial marker-based approach for on-patient visualization of maxillofacial regions by using an optical tracking system to track the patient and using a mobile device to visualize the internal structures of the patient's body. Nicolau et al. [30] used 15 markers to perform the registration in order to use AR to overlay a virtual liver over the patient. Debarba et al. [31] proposed a method to visualize anatomic liver resections in an AR environment with the use of a fiducial marker, which made locating the position and tracking of the medical data in the scene possible.

Recently, there are some markerless AR systems proposed in the literatures using various camera configurations. For example, Maier-Hein et al. [32] implemented a markerless mobile AR system by using a time-of-flight camera mounted on a portable device, that is, a tablet PC, to see the patient's anatomical information. The registration method of this approach is an anisotropic variant of the ICP algorithm and

the speed performance is about 10 FPS. By using a RGB-D senor, such as the Microsoft Kinect, Blum et al. [33] utilize a display device to augment the volume information of a CT dataset onto the user for anatomy education. They employed a skeleton tracking algorithm to estimate the pose of the user and scaled the generic CT volume according to the size of the user before performing the AR overlay. Since the depth data provided by Kinect are not sufficiently accurate for pose estimation, Meng et al. [34] proposed an improved method by using landmarks to increase the system performance. Based on a similar concept, Macedo et al. [35] developed an AR system to provide on-patient medical data visualization using Microsoft Kinect. They used the KinectFusion algorithm provided by Microsoft to reconstruct the patient's head data and a variant of the ICP algorithm was used in conjunction with a face tracking strategy for use during the medical AR. An extension to this approach has been proposed recently in [36], in which a multiframe, nonrigid registration scheme was presented to solve the problem of displacement of natural markers on the patient's face. In order to speed up the huge computations required for a real-time nonrigid registration, a lot of GPUs are needed and the implementation of this system would be very complex compared with the previous methods. Moreover, because the patient undergoing an operation would be under anesthesia, the appearance of his face would be unchanged, so there is no need to perform nonrigid registration to align the CT data to patient's face. In summary, although the Microsoft Kinect is popular for AR applications, it is inconvenient for use in the clinical environment because its volume is too big to mount on the physician's head.

In light of these previous works, we propose a simple but efficient framework for markerless augmented reality visualization, which is based on the Stereo Video See-Through AR. The markerless AR procedure mainly follows the principle of landmark-based camera localization approach. An improved ICP algorithm is applied for alignment of the virtual object and the feature point cloud in the physical space. This framework improves the previous studies in markerless AR visualization by using a light-weight stereo HMD instead of a heavier device such as the Microsoft Kinect, and it can achieve a more accurate registration result by using an improved form of the ICP. This system can be applied to medical applications, such as training, telementoring, or preoperative explanation.

3. Materials and Methods

The entire workflow of the proposed stereo AR scheme is shown in Figure 1. Feature point cloud of the target object is extracted and reconstructed by the stereo camera. The improved ICP-based surface registration algorithm had been designed and applied in order to align the virtual model to the feature point cloud of the target in the physical space.

3.1. 3D Feature Point Cloud Reconstruction. Firstly, a stereo camera setup is used to acquire 3D information of the target in the world coordinate system (WCS). The SURF algorithm is then utilized to extract feature points from the target object

region in left camera image. The corresponding points in right camera image are then obtained by comparing the SURF feature points in the right image with the aid of stereo epipolar constrain. The object region is recorded for later use in the tracking recovery stage.

According to pin-hole camera model, the perspective projection between a 3D object point (X, Y, Z) and its projection point (x, y) can be described as

$$
\begin{bmatrix} x \\ y \\ w \end{bmatrix} = \begin{bmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} [\mathbf{R} \mid \mathbf{t}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = KE \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (1)
$$

where K denotes the camera's intrinsic parameters. The parameter f_x denotes the focal length divided by pixel size s_x in x direction, and f_y denotes the focal length divided by pixel size s_y in y direction. The point (x_0, y_0) is the principle point of the projected 2D image plane. E denotes the camera's extrinsic parameters, which contains a 3×3 rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$ between the WCS and the camera coordinate system (CCS). We denote the projection matrix as

$$
P = KE = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix}. \quad (2)
$$

So if a stereo camera pair is well-calibrated, we can obtain intrinsic parameters and extrinsic parameters of the both camera. For a feature point $p_L(x_L, y_L)$ in left camera image, we can calculate a projection line in the WCS from the perspective projection equation and the projection matrix $\mathbf{P}_L$ of the left camera:

$$
\begin{aligned}
&(P_{11} - P_{31} X_L) X + (P_{12} - P_{32} X_L) Y + (P_{13} - P_{33} X_L) Z \\
&\quad = P_{34} X_L - P_{14}, \\
&(P_{21} - P_{31} X_L) X + (P_{22} - P_{32} X_L) Y + (P_{23} - P_{33} X_L) Z \\
&\quad = P_{34} X_L - P_{24}.
\end{aligned} \quad (3)
$$

Also, assuming the corresponding point of p_L in right camera image is $p_R(x_R, y_R)$, we can then calculate another projection ray equation by the projection matrix $\mathbf{P}_R$ of the right camera:

$$
\begin{aligned}
&(P_{11} - P_{31} X_R) X + (P_{12} - P_{32} X_R) Y + (P_{13} - P_{33} X_R) Z \\
&\quad = P_{34} X_R - P_{14}, \\
&(P_{21} - P_{31} X_R) X + (P_{22} - P_{32} X_R) Y + (P_{23} - P_{33} X_R) Z \\
&\quad = P_{34} X_R - P_{24}.
\end{aligned} \quad (4)
$$

The middle point of the common perpendicular between the two projection lines in the WCS is then selected as the 3D object point of the corresponding pair (p_L, p_R). By calculating the 3D object point of each feature-corresponding pair, we can obtain a 3D feature point cloud of the target object.
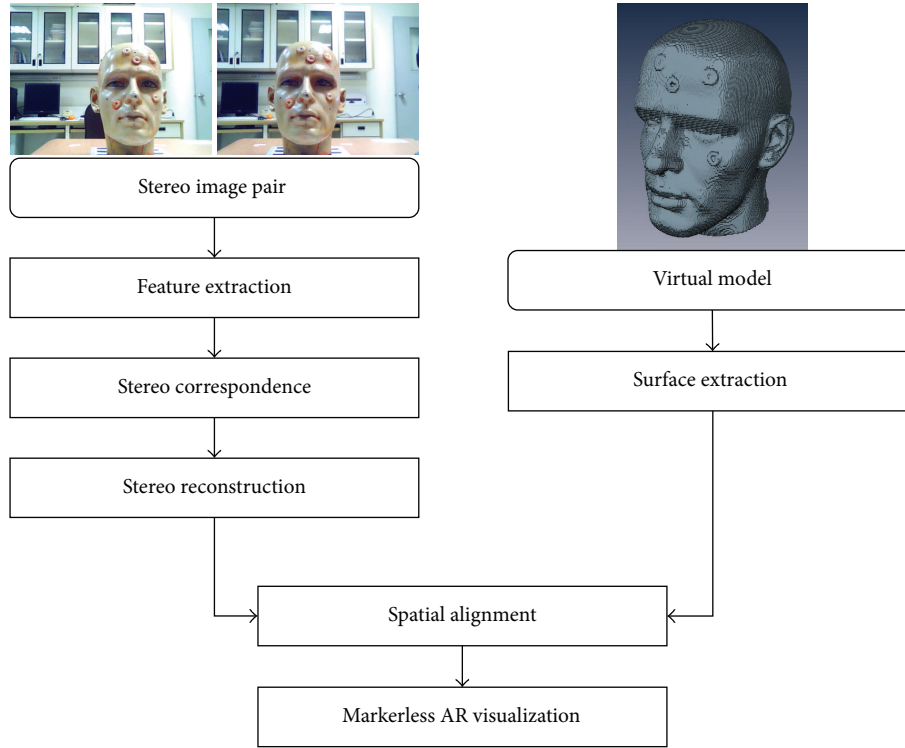
FIGURE 1: Workflow of the proposed AR system.

3.2. Improved-ICP Algorithm for Virtual Model Alignment.

In this study, we designed an improved ICP-based surface registration algorithm for spatial alignment of the virtual object and the feature point cloud. The improved ICP algorithm is based on the original ICP registration algorithm, which is the most widely used method to solve the 3D surface registration problem. However, ICP suffers from two important weaknesses in dealing with local minimum and outliers. Therefore, we added two strategies to improve the ICP algorithm in order to overcome these drawbacks. A weighting function is added to decrease the influence of outliers, and a random perturbation scheme is utilized to help ICP escape from the local minimum.

3.2.1. Distance-Based Weighting Function.

Assume that the 3D feature point cloud of the target P_{target} is the floating data A with m points $\{a_1, a_2, \ldots, a_m\}$ and the surface point cloud of the virtual object is the reference data B with n points $\{b_1, b_2, \ldots, b_n\}$. The original ICP used rigid transformation to align these two 3D data point sets in an iterative manner. In each iteration of ICP, every point a_i in A first finds its closest point b_j in B, and a cost function F is then evaluated based on the distance between each corresponding pair (a_i, b_j). In our improved ICP algorithm, we modified the cost function F of the ICP by adding a weighting function to the distances of all closest corresponding pairs (a_i, b_j) in order to deal with the problem of outliers, as shown in

$$F = \frac{1}{m} \sum_{i=1}^{m} w_i \min_{j \in \{1,2,\ldots,n\}} \left\| a_i - b_j \right\|, \tag{5}$$

where w_i is a distance-based weighting function, determined according to the median of the distances of all the corresponding pairs, as defined by

$$w_i = \begin{cases} 1 & \text{if } \left\| \mathbf{a}_i - \mathbf{b}_j \right\| < \text{median} \\ \dfrac{\text{median}}{\left\| \mathbf{a}_i - \mathbf{b}_j \right\|} & \text{otherwise.} \end{cases} \tag{6}$$

3.2.2. Random Perturbation Scheme.

The way that ICP reaches a local minimum is by using a gradient descent approach. In each iteration of the ICP, the cost function is evaluated at the current solution and then moves along the direction of gradient to the local minimum. When the registration reaches convergence, we can get a transformation solution T which projects a set of points onto another set of points, where the total sum of the distances between these two point sets is the smallest. Although the actual solution space of the cost function F is multidimensional since transformation T comprises three rotation operations (R_x, R_y, R_z) and three translation operations (T_x, T_y, T_z), for the sake of convenience we will explain the concept of perturbation strategy using a one-dimensional solution space example as illustrated in Figure 2. Suppose the initial position in solution space before using ICP is T_{Init} and the converged solution is T_{temp}; thus, the ICP registration would reach T_{temp} from T_{Init} by exploring the range of $|T_{\text{Init}} - T_{\text{temp}}|$. Let r denote rotation element R from the initial solution to the converged solution; that is, $r = |T_{\text{Init}} - T_{\text{temp}}|$. The rotation in each direction for perturbation is determined by using a parabolic
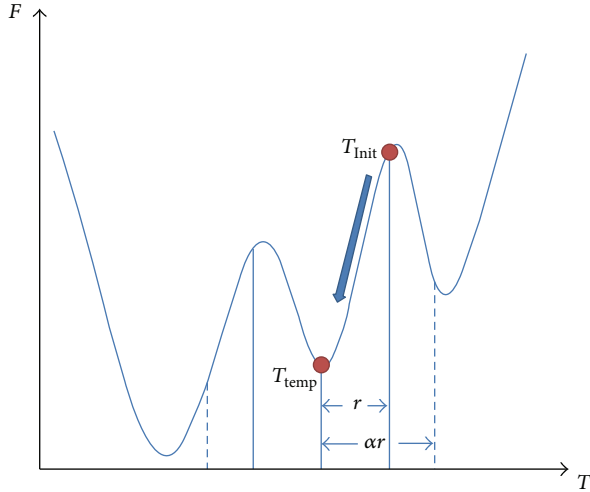
FIGURE 2: Solution space of ICP cost function F.

probability density function, as denoted in (7). Larger values have relative higher probabilities, which provide a greater chance to escape from the local minimum:

$$p(y) = \begin{cases} \dfrac{3y^2}{2\alpha^3 r^3} & \text{if } -\alpha r < y < \alpha r \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

3.2.3. Improved-ICP Algorithm. Figure 3 shows the flowchart of the improved-ICP registration scheme. The detailed steps of the improved ICP are described as follows.

Step 1. Perform the standard ICP registration to align floating data A to reference data B. The initial position in the solution space is denoted as T_{Init} and the converged solution is denoted as T_{temp}. The final value of the cost function F is recorded as the current cost C'.

Step 2. Check whether C' is better than the current best cost of ICP C_{best} or not. If it is, accept the transformation T_{temp} as the temporal best solution T_{best}, update the best cost C by C', and go on to Step 3. Otherwise, move to Step 4.

Step 3. Perturb the aligned data A with a transform T_{perturb}. A transformation T_{perturb} is selected according to (7). The transformation T_{perturb} is then applied to the floating data A and the algorithm moves back to Step 1 for performing ICP registration.

Step 4. Check whether the result meets the stopping criteria or not. If the best cost C is below a threshold C_{thres} or the count of repetition reaches a certain value k, the algorithm stops and outputs the final transformation T_{best}. Otherwise, go on to Step 5.

Step 5. Check if the perturbation range needs to be expanded or not. If the cost function is not improved after n times of perturbations, then we scale α in (7) to extend the searching range. Otherwise, the searching range does not need to

be extended. After the decision of whether to extend the perturbation range or not is made, the algorithm goes back to Step 3.

3.3. Markerless AR Visualization

3.3.1. Tracking and Camera Pose Estimation. The flowchart for the procedure for markerless AR visualization is shown in Figure 4. The KLT tracker takes turns to track the extracted feature points in the AR image. Assuming the tracking result in each frame is denoted as a point set q_t, initially we randomly select a number of N_R points from q_t. The first estimation of extrinsic parameters is thus calculated by using the EPNP camera pose estimation algorithm [40]. Then, we use these extrinsic parameters to project the 3D points D_q of these features onto the AR frame, obtaining a set of 2D projective points q_{proj}, which has the number of N_q points. Ideally, if all points are being tracked correctly, the projective points q_{proj} and the tracking points q_t in camera image should be overlapping or very close to each other. The L_2-norm distance, d_i, for each pair of projective point $q_{\text{proj},i}$ and tracked point $q_{t,i}$, is as shown in

$$d_i = \left\| q_{t,i} - q_{\text{proj},i} \right\|_2. \quad (8)$$

If d_i is greater than a predefined threshold κ, then the point q_i is considered as an outlier, that is, a tracking-failed point. Otherwise, the point q_i is an inlier. Here, we choose three pixels for the threshold κ, which makes the AR display more stable. By determining whether every point of q_t is inlier or not, the "inlier rate" of this tracking result q_t is measured, which implies the rate of how many points are being tracked correctly in this frame. The inlier rate r_t of a frame in time t is defined as

$$r_t = \frac{N_{\text{Inliers}}}{N_{\text{Inliers}} + N_{\text{Outliers}}}, \quad (9)$$

where N_{Inliers} stands for the number of feature points which are tracked correctly and N_{Outliers} represents the number of outliers. If the inlier rate is higher than a predefined threshold ζ, it indicates that this estimation of extrinsic parameters in the current frame is highly reliable and therefore we can use this estimation result of extrinsic parameters to correct the outliers. If a point is considered as an outlier, its projective point is used to replace this outlier point. The threshold ζ is set to 0.8 in this study.

On the other hand, if the inlier rate is less than ζ, the system randomly selects a different group in q_t to estimate another set of extrinsic parameters. The inlier rate r_t is then calculated again after finding projecting points q_{proj} of 3D points, D_q. If this process is performed over a times and all inliers rates r_t are less than ζ, the system is determined as a tracking failure, and a tracking recovery is required.

3.3.2. Tracking Recovery. When the tracking fails, we use the SURF feature comparison method as a reference to help the system recover to the original tracking status. As mentioned above, in the step of feature point cloud reconstruction,
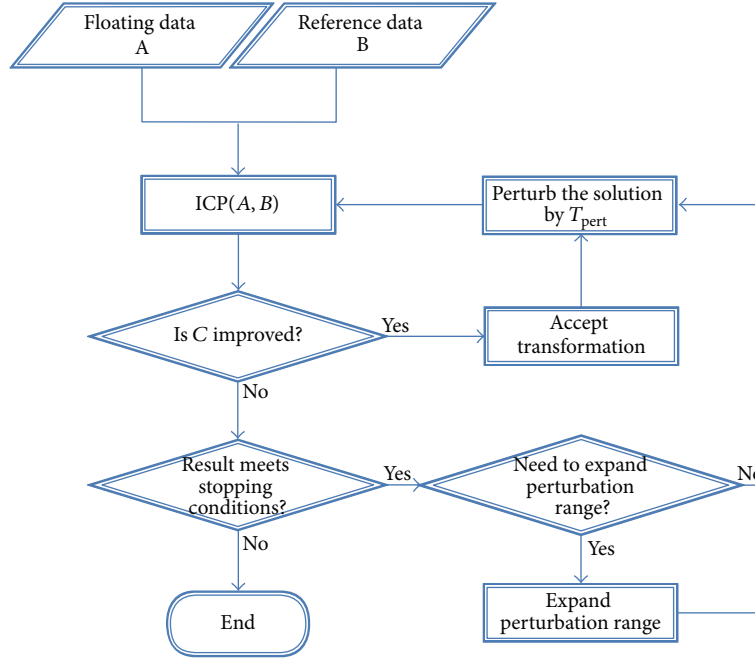
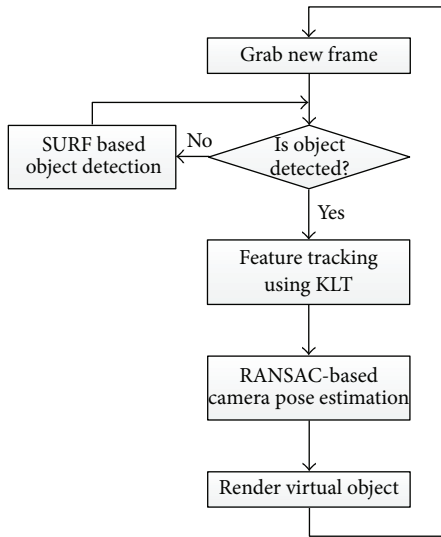Figure 3: Flowchart of the improved ICP Algorithm.

Figure 4: Markerless AR Visualization Procedure.

a set of SURF keypoints is extracted from the target region I_{target} and denoted as $\mathbf{p}_{\text{temp}}$, while their corresponding SURF descriptors D_{temp} are then estimated, as shown in

$$\mathbf{p}_{\text{temp}} = \left\{ p_{\text{temp},i}\left(x,y\right), 1 \leq i \leq N_{p_{\text{temp}}} \right\},$$
$$D_{\text{temp}} = \left\{ \mathbf{d}_{\text{temp},i}\left(x,y\right), 1 \leq i \leq N_{p_{\text{temp}}} \right\}, \tag{10}$$

where $p_{\text{temp},i}$ stands for the ith SURF keypoint in I_{target} and $\mathbf{d}_i$ is the SURF descriptor of ith SURF keypoint. As the AR camera is turned on and being prepared for AR visualization, from each frame I_{cam}, which is the image captured by

the AR camera, another set of SURF keypoints $\mathbf{p}_{\text{cam}}$ with $N_{\mathbf{p}\text{cam}}$ points are extracted and denoted as

$$\mathbf{p}_{\text{cam}} = \left\{ p_{\text{cam},j}\left(x,y\right), 1 \leq j \leq N_{p_{\text{cam}}} \right\},$$
$$D_{\text{cam}} = \left\{ \mathbf{d}_{\text{cam},j}\left(x,y\right), 1 \leq j \leq N_{p_{\text{cam}}} \right\}. \tag{11}$$

For each SURF keypoint $p_{\text{cam},i}$ in I_{cam}, it is compared to every SURF keypoint p_{temp} in I_{target} by calculating L_2-norm distance of their descriptors $\mathbf{d}_{\text{temp},i}$ and $\mathbf{d}_{\text{cam},j}$:

$$E_{i,j} = \left\| \mathbf{d}_{\text{temp},i} - \mathbf{d}_{\text{cam},j} \right\|_2. \tag{12}$$

A matching pair is selected if the L_2-norm distance between the descriptors is smaller than 0.7 times of the distance to the second-nearest keypoint. Assuming the number of successful corresponding pairs is N_c, then if N_c is greater than a predefined threshold τ, it implies that the target object is probability in the FOV of the AR camera and the system then moves to the next image-matching step. Otherwise, the previous process continuously repeats until the criteria is met; that is, $N_c > \tau$.

4. Experimental Results

The proposed markerless AR scheme is based on a Stereo Video See-Through HMD device, a Vuzix Wrap 1200DXAR, as show in Figure 5(a). Since this work is aimed for an AR visualization of medical application, a plastic dummy head is chosen as target object for AR visualization. The computed tomography image of the dummy head is obtained to construct the virtual object for visualization, as show in Figure 5(b).
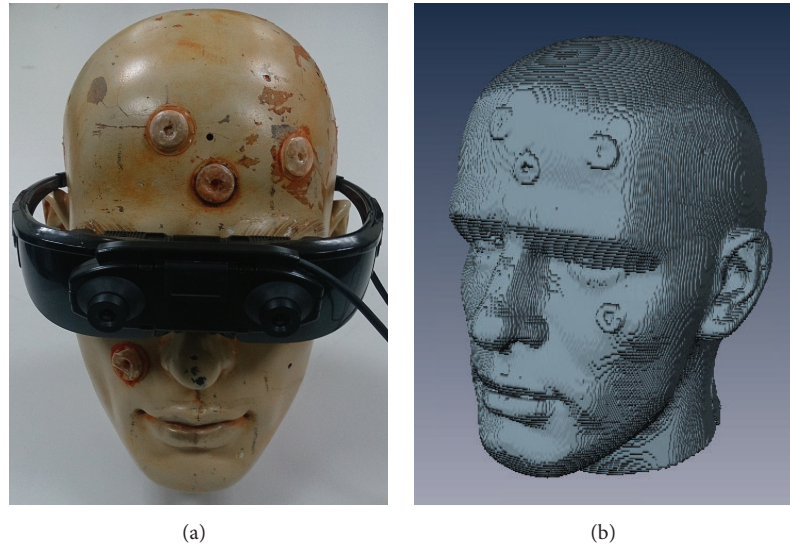
(a)   (b)

FIGURE 5: (a) Vuzix Wrap 1200DXAR Stereo HMD and (b) reconstructed CT model of the dummy head.

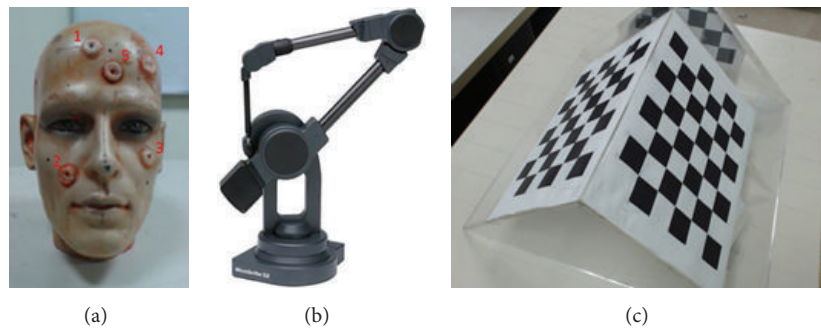(a)   (b)   (c)

FIGURE 6: Tools used to evaluate the accuracy of spatial alignment of medical information: (a) plastic dummy head, (b) MicroScribe G2X digitizer, and (c) triangular prism for coordinate system calibration.

The proposed medical AR system has two unique features: one is a marker-free image-to-patient registration and the other is a pattern-less AR visualization. In this section, experiments were carried out to evaluate the performance with respect to these features. At first, the accuracy of the medical information alignment is evaluated in Section 4.1. In Section 4.2, visualization result of the proposed AR scheme is shown.

4.1. Accuracy Evaluation of Alignment. In order to evaluate the accuracy of the image-to-patient registration of the proposed system, a plastic dummy head was utilized as the phantom target object. Before scanning CT images of the phantom, five skin markers were attached on the face of the phantom, as shown in Figure 6(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. A commercial 3D digitizer, G2X produced by Microscribe [41], as shown in Figure 6(b), was utilized to establish the 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.

Before the evaluation, a calibration step was performed to find the transformation $^{\text{Digi}}_{\text{W}}T$ between the stereo 3D coordinate system, that is, the world coordinate system, and the digitizer's coordinate system C_{Digi}. In order to perform the calibration, a triangular prism with chessboard patterns attached is used, as shown in Figure 6(c). This prism was placed in the FOV of the stereo HMD. Corner points of the chessboard were selected by the digitizer to be reconstructed by the stereo camera. The two sets of 3D points, represented by the coordinate systems of the digitizer and the stereo HMD, respectively, were used to estimate a transformation $^{\text{Digi}}_{\text{W}}T$ by using least mean square (LMS) method, so that the 3D points reconstructed by stereo camera can be transformed to the digitizer's coordinate system C_{Digi}.

The plastic dummy head was placed in front of the FOV of the stereo camera at a 60 cm distance. First, we used the G2X digitizer to obtain the 3D coordinates of these markers. Next, the stereo camera was utilized to reconstruct
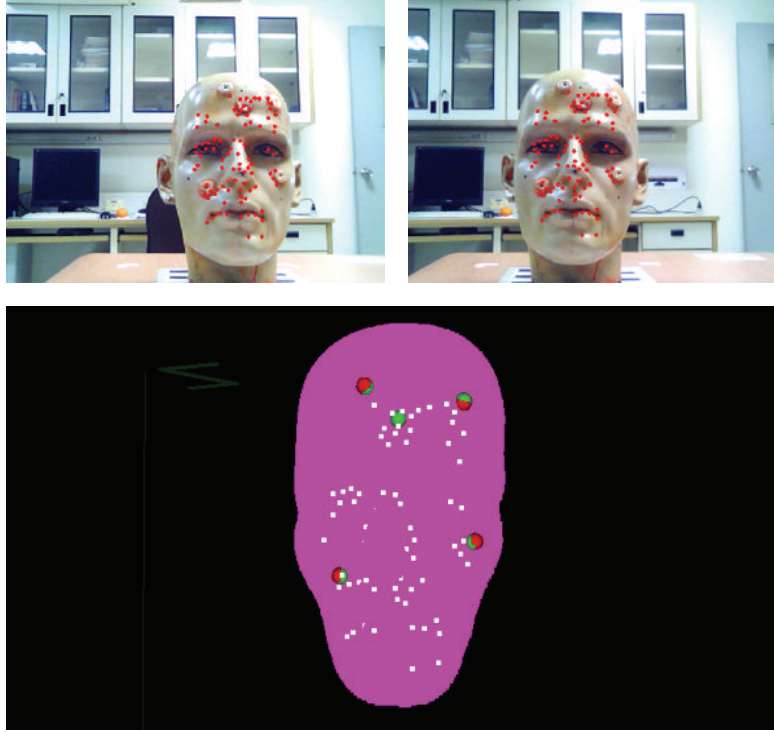
FIGURE 7: Result of spatial alignment between feature point cloud and virtual object surface from CT.

TABLE 1: Target registration error of each skin-attached marker after using various alignment methods.

	Target 1	Target 2	Target 3	Target 4	Target 5	Average
MeanTRE (mm) using Adaptive-ICP [37]	9.35	5.73	4.79	8.38	7.78	7.20
MeanTRE (mm) using Random-ICP [38]	9.87	5.82	5.15	8.88	8.24	7.59
MeanTRE (mm) using Fast-MICP [39]	4.51	5.13	3.56	1.57	1.66	3.28
MeanTRE (mm) using improved ICP	3.47	2.22	3.06	3.55	3.77	3.21

the head's feature point cloud. Next, another surface is extracted from the CT image of the dummy head. The point cloud was transformed to the C_{Digi} by applying $_{\text{W}}^{\text{Digi}}T$ and then registered to the preoperative CT images by using the improved ICP algorithm. Image-to-patient registration is evaluated by calculating the target registration error (TRE) [42, 43] of the five skin markers. The TRE for evaluation is defined as

$$\text{TRE} = M_{\text{CT},i} - {}_{\text{W}}^{\text{MI}}T\left(M_{\text{face},i}\right), \tag{13}$$

where $M_{\text{CT},i}$ denotes the coordinate of the ith marker in the CT coordinate system and $M_{\text{face},i}$ is the coordinate of the ith marker in C_{Digi}. The transformation $_{\text{W}}^{\text{MI}}T$ represents the rigid transformation obtained from the improved ICP algorithm. Figure 7 shows the result of spatial alignment between feature point cloud and the virtual object surface from CT; the initial spatial position of the reconstructed facial surface data (white) and CT surface data (magenta). The alignment

procedure was performed repeatedly 100 times, and each time we slightly shifted the location and orientation of the phantom. The TREs at each registration procedure were recorded and the means and the average errors are shown in Table 1. To demonstrate the performance of the improved ICP algorithm, three variants of ICP methods, for example, Adaptive-ICP [37], Random-ICP [38], and Fast-MICP [39], are used for accuracy comparison. From the experimental results, it is noted that the TREs using Adaptive-ICP and Random-ICP are large because no good initial values are given. For Fast-MICP, good initial condition is needed such that the registration error can be reduced. On the other hand, a good initial condition is not required in our case because the proposed improved ICP algorithm can still obtain good results by efficient error-minimization strategy. As shown in Table 1, the mean TREs of the skin markers using the proposed method are within the range of 2 to 4 mm. On a personal computer with an Intel Core 2 Duo

FIGURE 8: Augmented reality visualization results of the proposed medical AR system using the plastic phantom head as testing subject. (a) First row: original stereo images view, second row: AR visualization result. (b) AR visualization results in another view.

CPU 2.93 GHz CPU with 2 GB RAM, the processing frame rate reached 30 frames/s.

4.2. Markerless AR Visualization Results. The proposed AR system was tested on a plastic dummy head. In the case of the dummy head, a 3D structure of the dummy head's CT image is reconstructed. The outer surface of the 3D structure is extracted to build the virtual object for AR rendering, as show in Figure 5(b). The AR visualization results from different viewpoints are shown in Figure 8. The CT model is well aligned to the position of dummy head. When the camera starts moving, markerless AR scheme is applied to both stereo image, and the extrinsic parameters of the two cameras are estimated frame by frame. The CT model of the dummy head is rendered in both stereo camera views.

4.3. Accuracy Evaluation of Camera Extrinsic Parameters Estimation. For an AR system, the accuracy of the extrinsic parameter estimation of the AR camera is the most important thing. In order to evaluate the AR visualization part of the proposed system, an optical tracking device, the Polaris Vicra, produced by Northen Digital Inc., was utilized to evaluate the extrinsic parameter estimation results of the proposed system. The Polaris Vicra is an optical spatial localization apparatus, which can detect infrared reflective balls by using a pair of infrared cameras. Since the reflective balls are fixed on a cross-type device, called dynamic reference frame (DRF), the posture and position of the DRF can thus be obtained. The DRF was attached to the HMD in order to track the AR camera by using the Polaris Vicra sensor. According to the specification of this product, the localization error of the Polaris Vicra is smaller than 1 mm. Therefore, the posture and position of the AR camera estimated by the Polaris Vicra are considered as the comparative reference to evaluate the accuracy of the proposed patternless AR system. In this experiment, we have evaluated the proposed patternless AR

TABLE 2: Accuracy in each degree of freedom.

Degrees of freedom	Mean error	Standard deviation
Rotation in x direction	3.33 degrees	1.38
Rotation in y direction	2.31 degrees	1.52
Rotation in z direction	0.72 degrees	0.46
Translation in x direction	6.11 mm	4.27
Translation in y direction	5.96 mm	3.59
Translation in z direction	4.78 mm	3.55

system by comparing results against the estimation results obtained by the Polaris Vicra.

In this experiment, the extrinsic parameters of the HMD camera estimated by the proposed system were compared to the results estimated by the Polaris Vicra in 450 frames. The differences for each of the six degrees of freedom were measured. The tracking results of the Polaris Vicra are considered as the ground truth for comparison. Figure 9 shows the evaluation results for rotation and Figure 10 shows the evaluation results for translation. The blue curves represent the estimation results of the proposed system, and the red curves are the results estimated by the Polaris Vicra. The mean errors of each degree of freedom are shown in Table 2.

5. Conclusion

In traditional AR camera localization methods, a known pattern must be placed within the FOV of the AR camera for the purpose of estimating extrinsic parameters of the camera. In this study, the shortcomings of the traditional methods are improved. A markerless AR visualization scheme is proposed by utilizing a stereo camera pair to construct the surface data of the target, and an improved ICP based surface registration technique is performed to align the preoperative medical
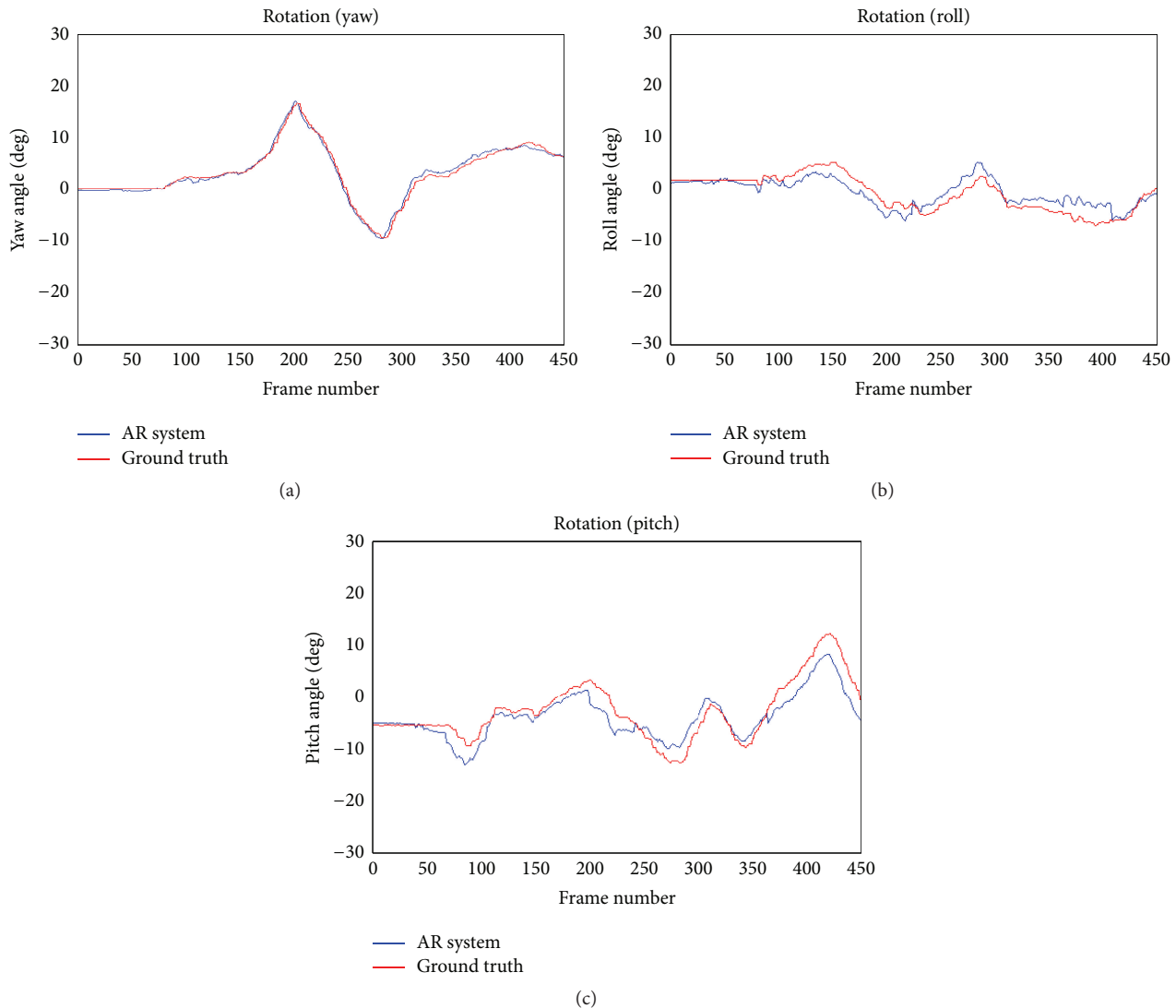
(a)

(b)

(c)

Figure 9: Accuracy evaluation results of rotation estimation compared to the result of tracking device Polaris Vicra (rotation): (a) roll, (b) pitch, and (c) yaw.

image model to the real position of the patient. A RANSAC-based correction is integrated to solve the problem of the AR camera location without using any pattern, and the experimental results demonstrate that the proposed approach provides an accurate, stable, and smooth AR visualization.

Compared 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 because the FOV of the AR camera is not limited by the requirement of the visibility of the AR pattern. A RANSAC-based correction technique is used 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 since been tested on a human subject and showed promising AR visualization results. In the future, extensive clinical trials are expected for further investigation.

Furthermore, the medical AR environment is expected to be integrated to an image-guided navigation system for surgical applications.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.
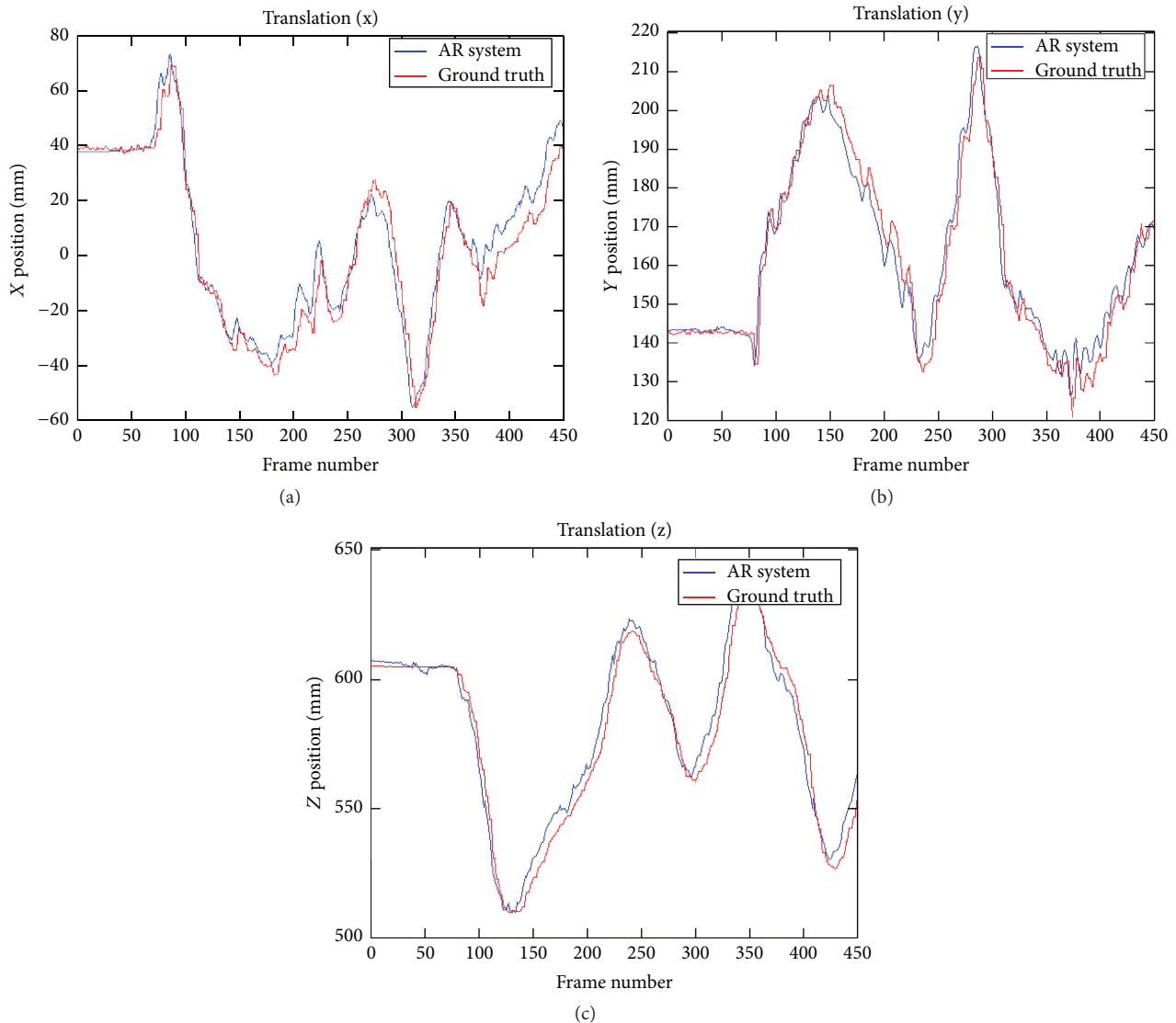
Acknowledgments

FIGURE 10: Accuracy evaluation results of translation estimation compared to the result of tracking device Polaris Vicra (translation) in (a) x, (b) y, and (c) z direction.

References

[1] R. T. Azuma, "A survey of augmented reality," *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.

[2] R. T. Azuma, Y. Baillot, R. Behringer, S. K. Feiner, S. Julier, and B. MacIntyre, "Recent advances in augmented reality," *IEEE Computer Graphics and Applications*, vol. 21, no. 6, pp. 34–47, 2001.

[3] T. Jebara, C. Eyster, J. Weaver, T. Starner, and A. Pentland, "Stochasticks: augmenting the billiards experience with probabilistic vision and wearable computers," in *Proceedings of the International Symposium on Wearable Computers (ISWC '97)*, pp. 138–145, Cambridge, Mass, USA, October 1997.

[4] S. K. Feiner, "Augmented reality: a new way of seeing," *Scientific American*, vol. 286, no. 4, pp. 48–55, 2002.

[5] D. W. F. van Krevelen and R. Poelman, "A survey of augmented reality technologies technologies, applications and limitations," *The International Journal of Virtual Reality*, vol. 9, no. 2, pp. 1–20, 2010.

[6] T. Sielhorst, M. Feuerstein, and N. Navab, "Advanced medical displays: a literature review of augmented reality," *Journal of Display Technology*, vol. 4, no. 4, pp. 451–467, 2008.

[7] T. Loescher, S. Y. Lee, and J. P. Wachs, "An augmented reality approach to surgical telementoring," in *Proceedings of the IEEE*

International Conference on Systems, Man and Cybernetics (SMC '14), pp. 2341–2346, IEEE, San Diego, Calif, USA, October 2014.

[8] F. Pretto, I. H. Manssour, M. H. I. Lopes, and M. S. Pinho, "Experiences using augmented reality environment for training and evaluating medical students," in *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW '13)*, pp. 1–4, San Jose, Calif, USA, July 2013.

[9] T. Blum, R. Stauder, E. Euler, and N. Navab, "Superman-like X-ray vision: towards brain-computer interfaces for medical augmented reality," in *Proceedings of the 11th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '12)*, pp. 271–272, IEEE, Atlanta, Ga, USA, November 2012.

[10] V. Ferrari, G. Megali, E. Troia, A. Pietrabissa, and F. Mosca, "A 3-D mixed-reality system for stereoscopic visualization of medical dataset," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 11, pp. 2627–2633, 2009.

[11] H. Liao, T. Inomata, I. Sakuma, and T. Dohi, "3-D augmented reality for MRI-guided surgery using integral videography autostereoscopic image overlay," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 6, pp. 1476–1486, 2010.

[12] J. Fischer, M. Neff, D. Freudenstein, and D. Bartz, "Medical augmented reality based on commercial image guided surgery," in *Proceedings of the 10th Eurographics Conference on Virtual Environments (EGVE '04)*, pp. 83–86, Aire-la-Ville, Switzerland, June 2004.

[13] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab, "Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality," in *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '07)*, pp. 129–138, Nara, Japan, November 2007.

[14] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," in *Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99)*, pp. 85–94, IEEE, San Francisco, Calif, USA, 1999.

[15] H. T. Regenbrecht, M. T. Wagner, and H. T. Regenbrecht, "Interaction in a collaborative augmented reality environment," in *Proceedings of the Extended Abstracts on Human Factors in Computing Systems (CHI '02)*, pp. 504–505, ACM, 2002.

[16] Y. Sato, M. Nakamoto, Y. Tamaki et al., "Image guidance of breast cancer surgery using 3-D ultrasound images and augmented reality visualization," *IEEE Transactions on Medical Imaging*, vol. 17, no. 5, pp. 681–693, 1998.

[17] D. Pustka, M. Huber, C. Waechter et al., "Automatic configuration of pervasive sensor networks for augmented reality," *IEEE Pervasive Computing*, vol. 10, no. 3, pp. 68–79, 2011.

[18] C. Tomasi and T. Kanade, "Detection and tracking of point features," Tech. Rep., Carnegie Mellon University, CMU-CS-91-132, 1991.

[19] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[20] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision & Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[21] ARtoolkit, http://www.hitl.washington.edu/artoolkit/.

[22] H. Kato and M. Billinghurst, "Marker tracking and HMD calibration for a video-based augmented reality conferencing system," in *Second International Workshop on Augmented Reality*, pp. 85–94, San Francisco, Calif, USA, 1999.

[23] M. Hirzer, "Marker detection for augmented reality applications," Research Paper, Granz University of Technology, Graz, Austria.

[24] B. Glocker, S. Izadi, J. Shotton, and A. Criminisi, "Real-time RGB-D camera relocalization," in *Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '13)*, pp. 173–179, IEEE, Adelaide, Australia, October 2013.

[25] E. Eade and T. Drummond, "Unified loop closing and recovery for real time monocular SLAM," in *Proceedings of the 19th British Machine Vision Conference (BMVC '08)*, pp. 6.1–6.10, September 2008.

[26] B. Williams, G. Klein, and I. Reid, "Automatic relocalization and loop closing for real-time monocular SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1699–1712, 2011.

[27] O. Kutter, A. Aichert, C. Bichlmeier et al., "Real-time volume rendering for high quality visualization in augmented reality," in *Proceedings of the International Workshop on Augmented Environments for Medical Imaging Including Augmented Reality in Computer-Aided Surgery (AMI-ARCS '08)*, MICCAI Society, New York, NY, USA, September 2008.

[28] M. Wieczorek, A. Aichert, O. Kutter et al., "GPU-accelerated rendering for medical augmented reality in minimally-invasive procedures," in *Bildverarbeitung fuer die Medizin (BVM '10)*, Springer, Aachen, Germany, March 2010.

[29] H. Suenaga, H. Hoang Tran, H. Liao et al., "Real-time in situ three-dimensional integral videography and surgical navigation using augmented reality: a pilot study," *International Journal of Oral Science*, vol. 5, no. 2, pp. 98–102, 2013.

[30] S. A. Nicolau, X. Pennec, L. Soler et al., "An augmented reality system for liver thermal ablation: design and evaluation on clinical cases," *Medical Image Analysis*, vol. 13, no. 3, pp. 494–506, 2009.

[31] H. G. Debarba, J. Grandi, A. Maciel, and D. Zanchet, *Anatomic Hepatectomy Planning Through Mobile Display Visualization and Interaction*, vol. 173 of *Studies in Health Technology and Informatics*, MMVR, 2012.

[32] L. Maier-Hein, A. M. Franz, M. Fangerau et al., "Towards mobile augmented reality for on-patient visualization of medical images," in *Bildverarbeitung für die Medizin*, H. Handels, J. Ehrhardt, T. M. Deserno, H.-P. Meinzer, and T. Tolxdorff, Eds., Informatik aktuell, pp. 389–393, Springer, Berlin, Germany, 2011.

[33] T. Blum, V. Kleeberger, C. Bichlmeier, and N. Navab, "Mirracle: augmented reality in-situ visualization of human anatomy using a magic mirror," in *Proceedings of the 19th IEEE Virtual Reality Conference (VRW '12)*, pp. 169–170, IEEE, Costa Mesa, Calif, USA, March 2012.

[34] M. Meng, P. Fallavollita, T. Blum et al., "Kinect for interactive AR anatomy learning," in *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR '13)*, pp. 277–278, IEEE, Adelaide, Australia, October 2013.

[35] M. Macedo, A. Apolinario, A. C. Souza, and G. A. Giraldi, "High-quality on-patient medical data visualization in a markerless augmented reality environment," *SBC Journal on Interactive Systems*, vol. 5, no. 3, pp. 41–52, 2014.

[36] A. C. Souza, M. C. F. Macedo, and A. L. Apolinário Jr., "Multi-frame adaptive non-rigid registration for markerless

augmented reality," in *Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry (VRCAI '14)*, pp. 7–16, ACM, Shenzhen, China, November-December 2014.

[37] J.-D. Lee, C.-H. Huang, L.-C. Liu, S.-T. Lee, S.-S. Hsieh, and S.-P. Wang, "A modified soft-shape-context ICP registration system of 3-d point data," *IEICE Transactions on Information and Systems*, vol. E90-D, no. 12, pp. 2087–2095, 2007.

[38] G. P. Penney, P. J. Edwards, A. P. King, J. M. Blackall, P. G. Batchelor, and D. J. Hawkes, "A stochastic iterative closest point algorithm (stochastICP)," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2001*, vol. 2208 of *Lecture Notes in Computer Science*, pp. 762–769, Springer, Berlin, Germany, 2001.

[39] J.-D. Lee, C.-H. Huang, S.-T. Wang, C.-W. Lin, and S.-T. Lee, "Fast-MICP for frameless image-guided surgery," *Medical Physics*, vol. 37, no. 9, pp. 4551–4559, 2010.

[40] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: an accurate O(n) solution to the PnP problem," *International Journal of Computer Vision*, vol. 81, no. 2, pp. 155–166, 2009.

[41] Microscribe, http://www.emicroscribe.com/products/microscribe-g2.htm.

[42] N. M. Hamming, M. J. Daly, J. C. Irish, and J. H. Siewerdsen, "Effect of fiducial configuration on target registration error in intraoperative cone-beam CT guidance of head and neck surgery," in *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS '08)*, pp. 3643–3648, IEEE, Vancouver, Canada, August 2008.

[43] J. M. Fitzpatrick, J. B. West, and C. R. Maurer Jr., "Predicting error in rigid-body point-based registration," *IEEE Transactions on Medical Imaging*, vol. 17, no. 5, pp. 694–702, 1998.