computers

MDPI

Article

An Automated Marker-Less Registration Approach Using Neural Radiance Fields for Potential Use in Mixed Reality-Based Computer-Aided Surgical Navigation of Paranasal Sinus

Suhyeon Kim, Hyeonji Kim and Younhyun Jung *

School of Computing, Gachon University, Seongnam 13120, Republic of Korea; kih629@gachon.ac.kr (S.K.); hjkim1770@gachon.ac.kr (H.K.)
* Correspondence: younhyun.jung@gachon.ac.kr

Abstract: Paranasal sinus surgery, a common treatment for chronic rhinosinusitis, requires exceptional precision due to the proximity of critical anatomical structures. To ensure accurate instrument control and clear visualization of the surgical site, surgeons utilize computer-aided surgical navigation (CSN). A key component of CSN is the registration process, which is traditionally reliant on manual or marker-based techniques. However, there is a growing shift toward marker-less registration methods. In previous work, we investigated a mesh-based registration approach using a Mixed Reality Head-Mounted Display (MR-HMD), specifically the Microsoft HoloLens 2. However, this method faced limitations, including depth holes and invalid values. These issues stemmed from the device's low-resolution camera specifications and the 3D projection steps required to upscale RGB camera spaces. In this study, we propose a novel automated marker-less registration method leveraging Neural Radiance Field (NeRF) technology with an MR-HMD. To address insufficient depth information in the previous approach, we utilize rendered-depth images generated by the trained NeRF model. We evaluated our method against two other techniques, including prior mesh-based registration, using a facial phantom and three participants. The results demonstrate our proposed method achieves at least a 0.873 mm (12%) improvement in registration accuracy compared to others.

Keywords: marker-less registration; neural radiance fields; mixed reality

Academic Editor: Paolo Bellavista

Received: 16 November 2024
Revised: 16 December 2024
Accepted: 23 December 2024
Published: 27 December 2024

Citation: Kim, S.; Kim, H.; Jung, Y. An Automated Marker-Less Registration Approach Using Neural Radiance Fields for Potential Use in Mixed Reality-Based Computer-Aided Surgical Navigation of Paranasal Sinus. *Computers* **2025**, *14*, 5. https://doi.org/10.3390/computers14010005

1. Introduction

Chronic rhinosinusitis ranks as one of the most prevalent chronic conditions worldwide, with an estimated prevalence of 12.3% in the United States [1]. This clinical condition adversely affects the quality of daily life, frequently leading to severe fatigue, major depressive symptoms, and, in severe cases, cognitive decline [2]. This condition develops as a result of ineffective management of acute sinusitis or recurring inflammation. When pharmacological treatments fail to relieve symptoms, paranasal sinus surgery (PSS) is generally considered an alternative therapeutic option. This procedure utilizes a minimally invasive approach, relying on a nasal endoscope for sinus access and inflammation reduction [3]. In PSS, it is crucial to ensure accurate navigation and handling of surgical instruments, as the paranasal sinuses are situated close to vital anatomical features within the skull, including the pituitary gland, carotid arteries, orbit, and optic nerve [4]. To achieve this precision, computer-aided surgical navigation (CSN) provides surgeons with visual information, such as the real-time location of surgical instruments, by utilizing preoperative medical imaging such as computed tomography (CT) or magnetic resonance imaging (MRI) scan [5–8]. In the

case of PSS, CSN utilizes a CT scan to accurately and promptly localize surgical instruments within the nasal cavity [9].

The essential step of CSN is the registration of the patient to preoperative medical images, ensuring accurate alignment between the surgical site and the preoperative medical data. Traditionally, this process is performed either manually by the operator or through automated methods using fiducial markers, such as implantable or adhesive. The manual registration approach typically involves the user carefully defining anatomical landmarks on the patient's surface using a pointer tip, which is then aligned with the predefined corresponding landmarks on the preoperative medical images using point-to-point alignment algorithms. This approach has been extensively studied [10–12]; however, it is inherently time-consuming and requires a high degree of human input while being prone to both intra- and inter-observer errors [13]. Additionally, physical contact with the patient during the registration process introduces a potential risk of infection. Fiducial markers that physically predefine anatomical landmarks allow the registration process to be automated while addressing some of the limitations of manual registration methods [14–17]. However, inherent limitations remain in the registration process. The invasive nature of implantable markers, involving insertion into the human body, may cause potential complications. In contrast, adhesive markers are non-invasive but carry the issue of marker point displacement [18]. Alternatives aim to eliminate the reliance on fiducial markers and thus address their limitations. Such marker-less registration approaches utilize RGB-D camera feeds and computer vision algorithms to automatically reconstruct 3D geometry to represent the patient surface and align them with preoperative medical images [19–21].

The exploration of CSN in Mixed Reality (MR) environments has gained increasing attention [22–24]. There are now affordable commercial MR head-mounted devices (MR-HMD) capable of providing glass-based high-fidelity displays. With MR-HMDs, CSN interfaces can be seamlessly integrated into the surgical field of view through real-time visual feeds. Furthermore, MR-HMD offers built-in RGB-D cameras, originally designed to implement spatial awareness, and it can be used to reconstruct the 3D geometry of patient surfaces directly. As such, all stages of CSN can be developed within the MR-HMD environment by utilizing all those capabilities. Our previous conference work [25] has proposed an automated MR-HMD-based marker-less registration approach with acceptable registration accuracy when compared to the manual counterpart. Our previous approach, however, inherently suffered from the low quality of the built-in RGB-D camera of MR-HMD, thus heavily compromising the registration accuracy. There were depth holes or invalid values due to the low-spec camera and the required 3D projection step to greater resolution of RGB camera spaces.

Neural radiance fields (NeRF) [26] have attracted considerable attention in computer graphics due to their revolutionary capability to reconstruct 3D geometry. NeRF represents objects as continuous radiance fields modeled by multi-layer perceptron (MLP). These MLPs learn to approximate the object's structure from a collection of captured images and their associated camera poses. NeRF has proven effective in preserving high-quality geometry and appearance details. Additionally, the learned radiance field allows for 3D geometry reconstruction at arbitrary resolutions, offering potential applications in super-resolution imaging.

In this study, we propose a new MR-HMD-based automated approach for marker-less registration in PSS-CSN by leveraging NeRF, with the aim to obtain the high quality of depth images and enhanced registration accuracy. We note the capability of NeRF in the resolution-arbitrary and noise-robust geometry (i.e., depth) reconstruction. We demonstrate our approach using Microsoft HoloLens 2 MR-HMD [27], taking advantage of its capability to efficiently capture RGB-D images and rapidly develop in-situ visualization prototypes,

illustrating how virtual preoperative images are aligned with the target physical areas. The accuracy of our registration approach is assessed using a face phantom, with comparisons made against two counterparts [25]. The contributions of our work are as follows:

- Incorporate NeRF into marker-less registration approaches to refine 3D geometry (depth) quality and consequently improve registration accuracy.
- Evaluate the effectiveness of our registration approach through a comparison of two types of previous marker-less registration approaches.
- Develop an in-suit visualization prototype using the commercial MR-HMD.

2. Related Work

2.1. Marker-Less Registration in Mixed Reality

Marker-less registration in MR has been actively explored in recent years to improve the precision and practicality of surgical navigation systems. Various approaches have been proposed, leveraging point clouds and mesh-based techniques to achieve accurate registration without the need for fiducial markers.

Gsaxner et al. [28] demonstrate marker-less registration for head and neck regions, relying on point cloud data for alignment. The face of the patient is first detected and point cloud data is generated only for the detected region to facilitate registration. Similarly, groenenberg et al. [29] employ depth data to perform real-time registration, showcasing the potential of depth-based marker-less techniques for surgical applications. Pepe et al. [30] propose a marker-less registration method for PET-CT related to head and neck cancer treatment. In this approach, feature landmarks are identified in RGB images, and a mesh is constructed from depth data to locate 3D landmarks using ray-marching. Similarly, our prior work [25] also employs a comparable method to extract 3D facial landmarks.

2.2. Background: Neural Radiance Fields

NeRF [26] aims to reconstruct the 3D geometry and appearance of an object by utilizing a collection of captured images and associated camera poses. NeRF is designed to learn an implicit mapping function F, between a sample along a viewing ray to color and density as below:

$$F(x,\ \theta,\ \varnothing) \rightarrow (c,\ \sigma) \tag{1}$$

where x is the 3D coordinate in an object, θ and $\varnothing$ represents a viewing ray in the azimuthal and polar coordinates, c is RGB color, and σ is the density. The implicit mapping function approximates the color and density of samples, which are then aggregated into a single pixel of the output image following the principles of the volume ray-casting integral:

$$C(r) = \int_{t1}^{t2} T(t) \cdot \sigma(r(t)) \cdot c(r(t),\ d)\ dt, \tag{2}$$

$$\text{where } T(t) = \exp\left(-\int_{t1}^{t} \sigma(r(s))\ ds \right)$$

where $r(t) = o + td$, o is the origin of the ray, d is its directional vector, $t1$ is its near bound, and $t2$ is its outer bound. The implicit mapping function is approximated by training a MLP, where the difference between the accumulated pixel colors $\hat{C}(r)$ and the ground truth colors $C(r)$ are minimized. The squared error between the color pairs is used as the loss function:

$$L = \sum \left\| \hat{C}(r) - C(r) \right\|_2^2 \tag{3}$$

NeRF iteratively optimizes the implicit mapping function using a set of training images until the loss function converges to a certain threshold value.

3. Method

3.1. Overview

Figure 1 illustrates the overview of our proposed marker-less registration approach that leverages NeRF-rendered depth using a facial phantom. Our proposed approach is composed of five components: (a) multi-view image data acquisition; (b) a 2D facial landmark extraction at patient sites; (c) reprojected depth images and NeRF-rendered depth images acquisition; (d) a 3D facial landmark localization using depth images and (e) CT facial landmark extraction and 3D to 3D registration.

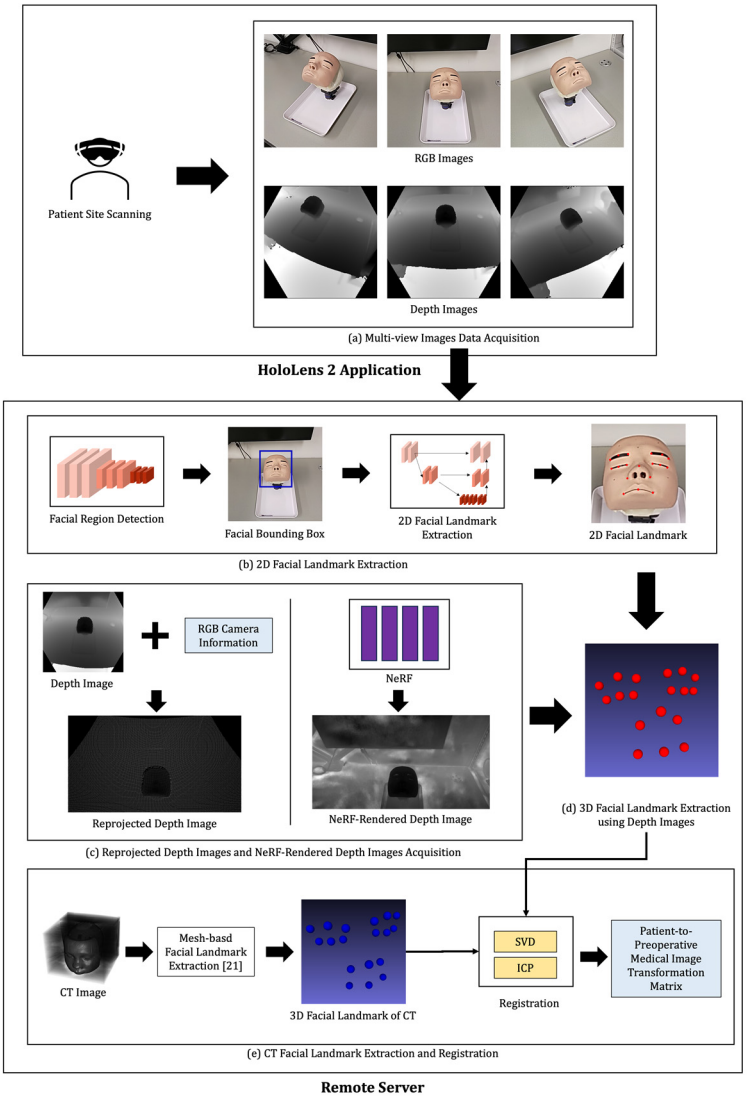

Figure 1. Comprehensive overview: (**a**) Multi-view Image Data Acquisition, (**b**) 2D Facial Landmark Extraction, (**c**) Reprojected Depth Images and NeRF-Rendered Depth Image Acquisition, (**d**) 3D Facial Landmark Extraction using Depth Images, (**e**) CT Facial Landmark Extraction and Registration.

3.2. Multi-View Image Data Acquisition

Our marker-less registration is required to capture 500 images around the patient's face to ensure sufficient coverage of both frontal and partial lateral regions. To ensure consistency, the image data acquisition process was designed with a fixed trajectory forming a quarter sphere around the patient's face, capturing images at regular intervals of one frame every 0.3 s as the user moved slowly and steadily. We expect that an image set covers various views of the patient's face, including frontal and lateral, to facilitate the NeRF

reconstruction process. Every capture stores an image, a camera position, and intrinsic camera parameters for each RGB and depth.

3.3. 2D Facial Landmark Extraction

We used RGB images to extract 2D facial landmarks for registration. We implemented a two-stage process: (i) detecting the region of the face with MTCNN [31] and (ii) extracting facial landmarks using HRN [32,33]. This fully automated approach to facial landmark extraction required no user input.

MTCNN [31] is a reliable deep-learning model chosen for facial detection in RGB images due to its adaptability to various resolutions, shapes, and viewpoints. It operates in two cascaded networks: a proposal network that identifies a set of potential face regions (facial bounding boxes) and a refinement network that optimizes these candidates by removing duplicates and refining accuracy through regression. This stage efficiently limits the region of interest for the next facial landmark extraction step, minimizing the chances of inaccuracies. We then utilized the HRN model [32,33] to precisely extract the 2D face landmarks. It has been shown to perform well in capturing fine-grained features due to its high-resolution convolutional neural network architecture. Unlike traditional models [34,35], HRN maintains multiple parallel streams at different resolutions, using multi-resolution fusion to combine localized and global features.

Due to time constraints, we did not apply the above two-stage process to extract 2D facial landmarks on all 500 images. Instead, we opted to extract landmarks from every 10th image, resulting in a subset of 50 images for this step. This sampling approach was selected to balance computational efficiency with the need for accurate registration data. By using 50 evenly spaced images, we were able to capture sufficient facial landmark information across multiple perspectives while significantly reducing processing time.

3.4. Reprojected Depth Images and NeRF-Rendered Depth Images

We use the depth image to locate the facial landmarks of the RGB image. Note that the original depth image does not have a pixel-to-pixel alignment with the corresponding RGB image, as the camera positions and resolutions for the RGB and depth images differ. This requires a transformation process that projects the pixels of the depth image into the 3D perspective of the RGB camera (i.e., a reprojected depth image), as shown in Figure 2a.

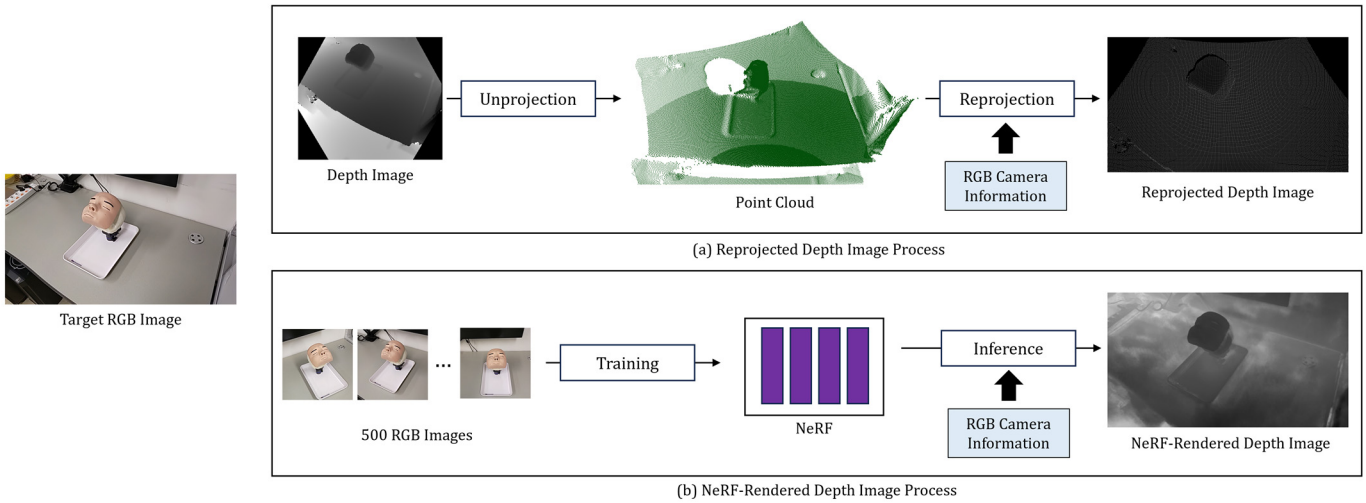

Figure 2. Reprojected depth image and NeRF-rendered depth image: (**a**) Reprojected Depth Image Process, (**b**) NeRF-Rendered Depth Image Process.

The original depth image is unprojected into a 3D point cloud using the depth camera position and lookup table provided by the HoloLens 2. This point cloud is then transformed into world space using the depth camera-to-world transformation, followed by a world-to-RGB camera transformation to place the point cloud in RGB space. Finally, the point cloud is projected onto the RGB image plane using the RGB camera parameters.

We note that the reprojected depth image contains numerous depth holes, as the point cloud generated from the depth image cannot fully cover the RGB image due to differences in resolution and field of view. We address these depth holes by leveraging NeRF to complement the missing depth holes (See Figure 2b). Using 500 RGB image and camera position pairs, we trained the NeRF model and then inferred a depth image along with the same viewpoint of the reprojected depth image. We now have two depth images and use the NeRF-rendered depth image when the 3D location of facial landmarks is not available from the paired reprojected depth image.

The formula for obtaining the NeRF-rendered depth image is as follows.

$$D(r) = \int_{t1}^{t2} T(t) \cdot \sigma(r(t)) \cdot t_i \, dt, \tag{4}$$

$$\text{where } T(t) = \exp\left(- \int_{t1}^{t} \sigma(r(s)) \, ds \right)$$

where $r(t) = o + td$, o is the origin of the ray, d is its directional vector, $t1$ is its near bound, and $t2$ is its outer bound.

3.5. 3D Facial Landmark Extraction Using Depth Images

We obtained 3D facial landmarks by unprojecting the 2D corresponding landmarks on an RGB image with the paired depth images and intrinsic camera parameters. The algorithm for corresponding depth values (pixels) to 2D facial landmarks is illustrated in Figure 3. It confirms whether the coordinates of the 2D facial landmarks fall within the image resolution boundaries. Since the 2D facial landmarks include all 19 points predefined by the model, if only the mouth is visible in the image and the chin is not, the model will assign points to likely positions outside the image boundaries. It then fetches a depth value at the corresponding pixel location in the reprojected depth image. If the depth value is available, the 3D location of 2D facial landmarks is retrieved; otherwise, we used the NeRF-rendered depth image. This process is repeated for every 2D facial landmark.

We then unprojected 2D facial landmarks using depth values, focal length, and principal points. The unprojection matrix used to perform this operation is as follows.

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} 1/f_x & 0 & -C_x/f_x & 0 \\ 0 & -1/f_y & C_y/f_y & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} i \cdot d \\ j \cdot d \\ d \\ 1 \end{bmatrix} \tag{5}$$

Here, f_x and f_y are the focal length of the RGB camera, while C_x and C_y are the principal points of the RGB camera. With these parameters, the 2D facial landmark is unprojected and can be obtained in 3D RGB space, which then applies the RGB-to-World transformation to acquire the 3D facial landmarks in world space. We compute the N images from which 2D facial landmarks were extracted according to different views, and the resulting N pairs of 3D facial landmarks were averaged, reducing the impact of errors associated with single-view depth images.

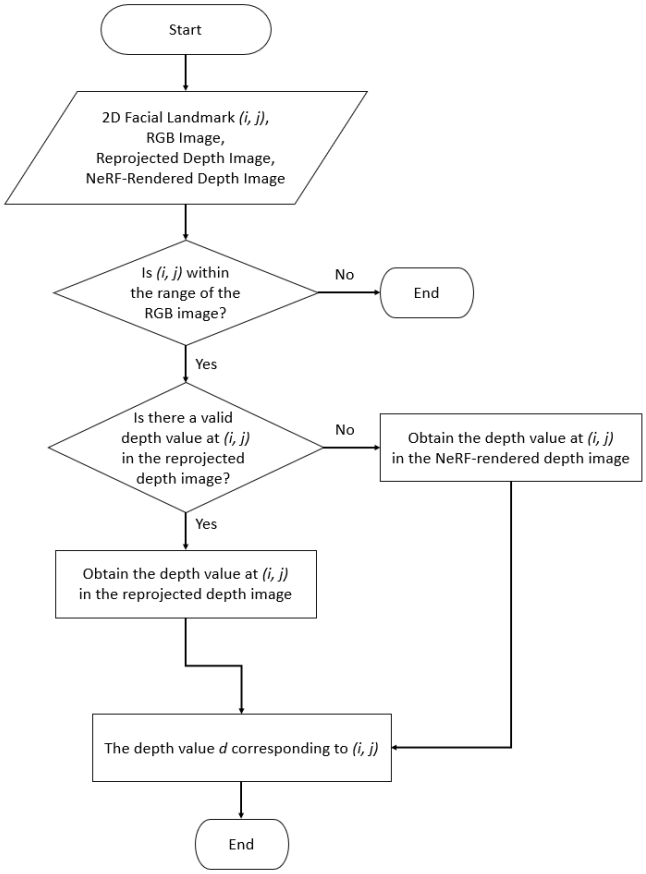

Figure 3. Flowchart for obtaining depth values corresponding to 2D facial landmarks.

3.6. CT Facial Landmark Extraction and Registration

The 3D facial landmarks in CT scans were obtained using the same method as prior work [25]. First, a facial skin mesh was generated based on the marching-cube algorithm, and 2D facial landmarks were extracted from an image captured from the front view. Then, using ray-marching, the corresponding 3D points for each 2D facial landmark were identified and used as the 3D facial landmarks.

The next step is to align the 3D facial landmarks from the patient's face with the CT-derived 3D facial landmarks through a registration process. We first applied the singular value decomposition (SVD) algorithm [36] to compute an initial transformation matrix between the pairs of 3D facial landmarks, followed by fine-tuning using the iterative closest point (ICP) algorithm [37]. This two-stage approach enhanced the robustness of our registration process against variations in the initial alignment of 3D facial landmarks and enabled the derivation of an accurate final transformation matrix.

The SVD algorithm constructed a covariance matrix from the correspondences between the patient's facial landmarks and CT landmark pairs, capturing their statistical relationships. By applying SVD decomposition, singular vectors and values were obtained. The singular vectors determined the optimal rotation matrix for aligning the landmarks, while the singular values indicated the scaling factors for each axis.

The initial transformation matrix was then iteratively refined using the ICP algorithm to enhance the registration accuracy. During each iteration, the ICP algorithm calculated an intermediate transformation matrix by optimizing the alignment and minimizing the distance between corresponding points (landmarks). This distance measured the discrepancy between each landmark on the patient's face and its nearest landmark in the CT scan. The iterative process continued until convergence, yielding a highly precise transformation matrix.

4. Experiment

4.1. Experiment Settings and Implementation

For the experiment, we utilized a facial phantom secured to a support structure, accompanied by its CT scan. For the experiment, we utilized a facial phantom secured to a support structure, accompanied by its CT scan. To compensate for the phantom's lack of distinct features compared to a human face, masking tape was applied to the eyebrow, eye, and mouth regions to enhance its similarity to a real face. The approach presented in Section 3 was implemented using a HoloLens 2 application and a remote server. Using HoloLens 2 (OS release version: 22621.1376), we captured RGB images and depth images (RGB-D) with associated camera parameters, which were then transmitted to the server via a TCP connection (See Section 3.2). On the server, 3D facial landmarks were extracted from the RGB-D images (see Sections 3.3–3.5) and registered with the CT facial landmarks (see Section 3.6). It produces a patient-to-preoperative medical image transformation matrix for aligning the CT mesh to the patient site. We ensure a sufficient processing speed by using the instant-ngp [38] as the NeRF model, which provides both speed and accuracy due to its hash encoding and CUDA-based implementation. The instant-ngp model primarily performed 35,000 iterations, and the training took approximately 3 min on an NVIDIA TITAN RTX with 500 RGB images. The HoloLens 2 application was implemented using Unity and the HoloLens2ForCV library [39].

We evaluate the proposed approach by comparing it with two other methods. One is our prior work [25], where instead of the depth image, the derived 3D geometry mesh was used to locate 3D facial landmarks. Another approach relied on the reprojected depth image but addressed the inherent depth holes by filling them through interpolation, i.e., obtaining 3D facial landmarks without the use of NeRF. For interpolation, we used the inpaint function from the OpenCV library [40].

Target Registration Error (TRE) was chosen as the evaluation metric, as it quantifies the accuracy of alignment between corresponding points after registration. The formula is shown below:

$$TRE = \frac{1}{N}\sum_{k=1}^{n} \| {}^{patient}T_{CT} point_k - point_k^{GT} \| \quad (6)$$

where N represents the number of points, ${}^{patient}T_{CT}$ is the transformation matrix that converts coordinates from the patient space to the CT space. It maps points in the patient space to the CT space, allowing their distance from the ground truth (GT) points to be calculated and compared.

4.2. Experiment Procedure

The experiment consists of two main stages. First, participants perform data acquisition for the marker-less approaches using the HoloLens 2 application. To capture 500 images, each participant maintains a distance of approximately 30–80 cm from the facial phantom and slowly moves while observing the phantom for around three minutes. Participants are then asked to select the five predefined GT points on the phantom using a pointer tip. These GT points are identical to those designated in our previous work [25] and are located at the edges of the forehead, between the eyebrows, and on each cheek. The five points are transformed into the CT space via the calculated matrix and matched with the corresponding GT points within the CT space to evaluate the TRE. We calculated the TRE for three cases: the proposed NeRF-based approach and two comparison approaches. Three participants experienced with the registration process were invited, and each experiment was conducted three times per participant to obtain reliable results on the registration accuracy.

5. Results and Discussions

In Table 1, we compare the registration accuracy of our marker-less approach with prior works [25] and the interpolated depth approach in terms of the mean and standard deviation of TRE from the three participants. The results show that in every case, except for the marginally lower one (S2 of P3), our proposed approach outperforms the other prior approaches. In terms of the mean TRE, our method demonstrates at least a 0.873 mm (12 percent) improvement over the others, with its standard deviation being relatively low, thus highlighting the accurate and consistent registration capability of our approach. When comparing the mesh- and reprojected depth-based approaches, the mean error is slightly lower for the mesh-based approach. The deterministic conclusion on the performance gain, however, cannot be easily made, as the mesh-based approach results in only 4 cases that are more accurate. This suggests that the reprojected depth-based approach tends to provide more accurate 3D geometry than the mesh-based approach, but considerable distortions can occur when interpolation errors (i.e., missing depths) arise. Our approach uses NeRF-rendered depth to fill in missing depths. The use of NeRF allows our approach to leverage geometrical features obtained from RGB images that are relatively invariant to the physical environment.

Table 1. The comparison result of our marker-less registration approach with prior works [25] and the interpolated depth approach for the registration accuracy using TRE metric from three participants (P1 to P3) with three sessions (S1 to S3).

Participants	Session	Our Approach (Reprojected Depth + Nerf-Rendered Depth)	Prior Work [25] (Geometry Mesh)	Reprojected Depth with Interpolation
P1	S1	**5.626**	7.466	8.291
	S2	**5.876**	6.984	6.831
	S3	**7.969**	8.438	9.282
P2	S1	**5.009**	6.453	6.003
	S2	**6.822**	6.984	7.595
	S3	**6.000**	6.175	6.100
P3	S1	**6.797**	6.859	7.897
	S2	6.254	8.008	**6.011**
	S3	**7.131**	7.973	8.264
TRE mean ± SD (mm)		6.387 ± 0.888	7.260 ± 0.7614	7.364 ± 1.185

In Figure 4, we compare the quality of depth (geometry) among different depth images. Here, the original depth image (See Figure 4b) is currently cropped to appear similar to other images, but it does not actually match the geometry of the facial phantom (see Figure 4a) perfectly, so its use is inappropriate for extracting 3D facial landmarks. The approach that uses 3D mesh suffers from regions with crucial depth differences or depth holes (see the nose being noticeably blurred in Figure 4c). In contrast, the use of the reprojected and interpolated depth image (see Figure 4e) demonstrates relatively uniform filling in the nose regions, i.e., alleviated depth loss. However, regions with sharp depth changes or large depth holes, such as along contours, still appear distorted or scattered. In comparison, NeRF-rendered depth images (i.e., our approach) ensure that all structures retain clear and well-defined shapes.

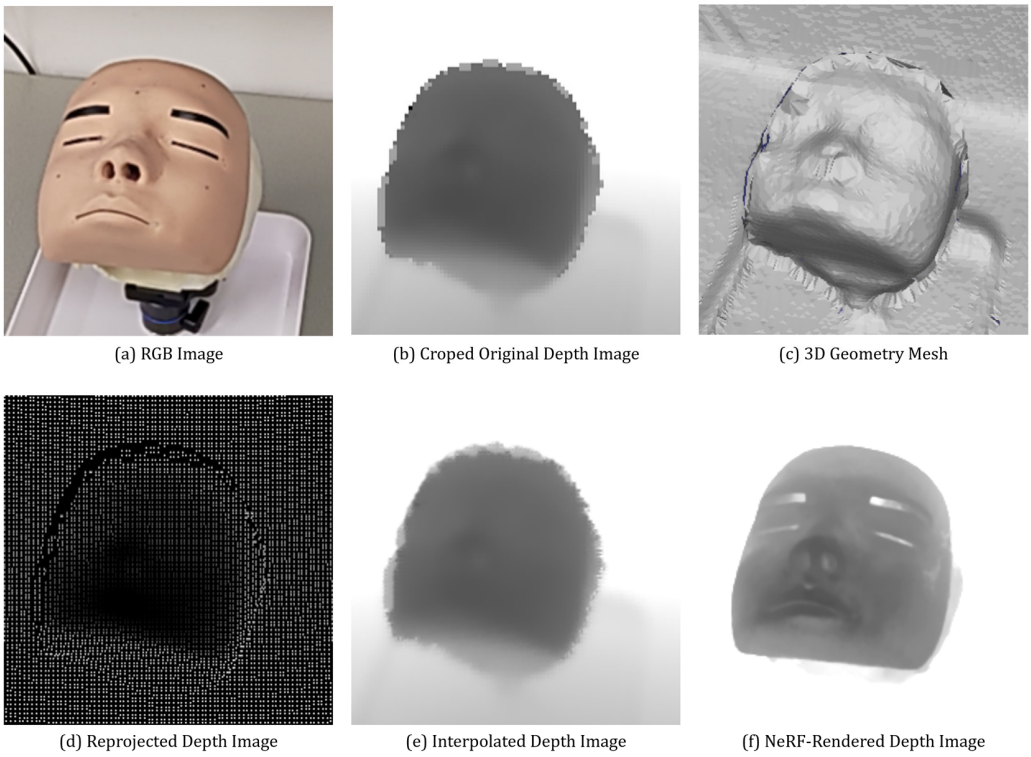

Figure 4. (**a**) the captured RGB image of the facial phantom; (**b**) the captured depth image from the built-in depth camera; (**c**) the 3D geometry mesh from (**b**) in the RGB space, (**d**) the reprojected depth image; (**e**) its interpolated depth image from (**d**); (**f**) the NeRF-rendered depth image.

Figure 5 shows how the physical phantom is aligned with the CT rendering achieved through the transformation matrix from our approach. The in-situ views of the MR-HMD reveal that the sinus structure is accurately positioned within the intracranial region of the nasal bone. This highlights the practical utility of our approach for PSS-CSN as the provision of the in-suit visualization of accurately aligned sinus structures is critical for guiding surgical instrument paths.

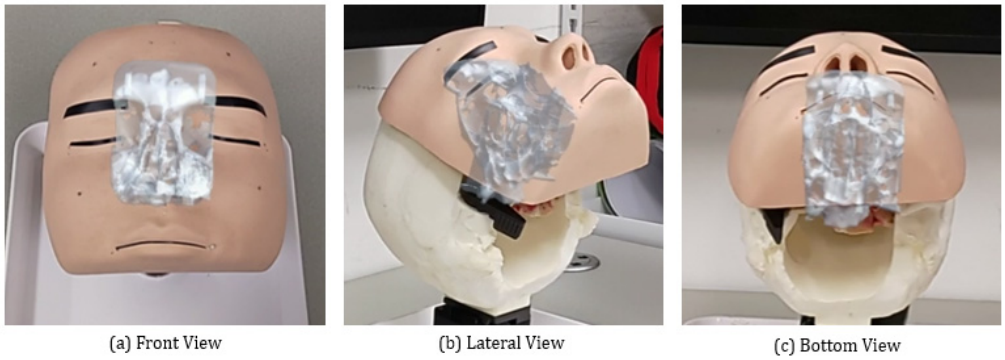

Figure 5. The augmentation results of the CT rendering onto the physical phantom using the transformation matrix from our approach: (**a**) front view, (**b**) lateral view, and (**c**) bottom view.

6. Conclusions and Future Work

This study proposes a new MR-HMD-based marker-less registration approach for PSS-CSN, which automates the identification and alignment of facial anatomical landmarks with CT scans. The results demonstrate that the accuracy of our approach registration outperforms the two existing marker-less registration approaches (see Table 1 and Figure 4). Our approach also eliminates the need for manual tasks like landmark selection or marker

attachment, making it a user-friendly and accessible tool, especially for practitioners who lack experience with the registration process or MR-HMDs. We are currently exploring the incorporation of our marker-less registration approach into the prototype MR-HMD model of the commercial PSS-CSN [24]. This integration could support a user study based on real patient cases, thus allowing us to assess the clinical robustness of our marker-less registration approach and its utility in real-world clinical settings.

Author Contributions: Methodology, validation, and writing—original draft preparation, S.K.; software, H.K.; supervision and writing—review and editing, Y.J. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2023 (Project Name: Cultural Technology Specialist Training and Project for Metaverse Game, Project Number: RS-2023-00227648).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Albu, S. Chronic Rhinosinusitis—An Update on Epidemiology, Pathogenesis and Management. *J. Clin. Med.* **2020**, *9*, 2285. [CrossRef] [PubMed]
2. Keating, M.K.; Phillips, J.C.; Phillips, J. Chronic Rhinosinusitis. *Am. Fam. Physician* **2023**, *108*, 370–377.
3. Weber, R.K.; Hosemann, W. Comprehensive Review on Endonasal Endoscopic Sinus Surgery. *GMS Curr. Top. Otorhinolaryngol. Head Neck Surg.* **2015**, *14*, Doc08. [PubMed]
4. Wise, S.K.; DelGaudio, J.M. Computer-Aided Surgery of the Paranasal Sinuses and Skull Base. *Expert Rev. Med. Devices* **2005**, *2*, 395–408. [CrossRef] [PubMed]
5. Duan, X.; Xie, D.; Zhang, R.; Li, X.; Sun, J.; Qian, C.; Song, X.; Li, C. A Novel Robotic Bronchoscope System for Navigation and Biopsy of Pulmonary Lesions. *Cyborg Bionic Syst.* **2023**, *4*, 0013. [CrossRef] [PubMed]
6. Fan, Y.; Xu, L.; Liu, S.; Li, J.; Xia, J.; Qin, X.; Li, Y.; Gao, T.; Tang, X. The State-of-the-Art and Perspectives of Laser Ablation for Tumor Treatment. *Cyborg Bionic Syst.* **2024**, *5*, 0062. [CrossRef]
7. Han, Z.; Tian, H.; Han, X.; Wu, J.; Zhang, W.; Li, C.; Qiu, L.; Duan, X.; Tian, W. A Respiratory Motion Prediction Method Based on LSTM-AE with Attention Mechanism for Spine Surgery. *Cyborg Bionic Syst.* **2024**, *5*, 63. [CrossRef] [PubMed]
8. Zhang, J.; Liu, L.; Xiang, P.; Fang, Q.; Nie, X.; Ma, H.; Hu, J.; Xiong, R.; Wang, Y.; Lu, H. AI Co-Pilot Bronchoscope Robot. *Nat. Commun.* **2024**, *15*, 241. [CrossRef] [PubMed]
9. Singh, A.; Kumar, R.; Thakar, A.; Sharma, S.; Bhalla, A. Role of Image Guided Navigation in Endoscopic Surgery of Paranasal Sinuses: A Comparative Study. *Indian J. Otolaryngol. Head Neck Surg.* **2020**, *72*, 221–227. [CrossRef] [PubMed]
10. Bae, D.K.; Song, S.J. Computer Assisted Navigation in Knee Arthroplasty. *Clin. Orthop. Surg.* **2011**, *3*, 259–267. [CrossRef] [PubMed]
11. Kosugi, Y.; Watanabe, E.; Goto, J.; Watanabe, T.; Yoshimoto, S.; Takakura, K.; Ikebe, J. An Articulated Neurosurgical Navigation System Using MRI and CT Images. *IEEE Trans. Biomed. Eng.* **1988**, *35*, 147–152. [CrossRef] [PubMed]
12. Grauvogel, T.D.; Engelskirchen, P.; Semper-Hogg, W.; Grauvogel, J.; Laszig, R. Navigation Accuracy after Automatic-and Hybrid-Surface Registration in Sinus and Skull Base Surgery. *PLoS ONE* **2017**, *12*, e0180975. [CrossRef]
13. Krueger, S.; Wolff, S.; Schmitgen, A.; Timinger, H.; Bublat, M.; Schaeffter, T.; Nabavi, A. Fast and Accurate Automatic Registration for MR-Guided Procedures Using Active Microcoils. *IEEE Trans. Med. Imaging* **2007**, *26*, 385–392. [CrossRef]
14. Wang, M.; Song, Z. Automatic Localization of the Center of Fiducial Markers in 3D CT/MRI Images for Image-Guided Neurosurgery. *Pattern Recognit. Lett.* **2009**, *30*, 414–420. [CrossRef]
15. Hong, J.; Hashizume, M. An Effective Point-Based Registration Tool for Surgical Navigation. *Surg. Endosc.* **2010**, *24*, 944–948. [CrossRef]
16. Lin, Q.; Yang, R.; Cai, K.; Si, X.; Chen, X.; Wu, X. Real-Time Automatic Registration in Optical Surgical Navigation. *Infrared Phys. Technol.* **2016**, *76*, 375–385. [CrossRef]
17. Yamamoto, S.; Taniike, N.; Takenobu, T. Application of an Open Position Splint Integrated with a Reference Frame and Registration Markers for Mandibular Navigation Surgery. *Int. J. Oral Maxillofac. Surg.* **2020**, *49*, 686–690. [CrossRef]
18. Wang, M.; Song, Z. Surgical Navigation. In *Computer-Aided Oral and Maxillofacial Surgery*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 161–181.

19. Fan, Y.; Jiang, D.; Wang, M.; Song, Z. A New Markerless Patient-to-Image Registration Method Using a Portable 3D Scanner. *Med. Phys.* **2014**, *41*, 101910. [CrossRef]
20. Jiang, L.; Zhang, S.; Yang, J.; Zhuang, X.; Zhang, L.; Gu, L. A Robust Automated Markerless Registration Framework for Neurosurgery Navigation. *Int. J. Med. Robot. Comput. Assist. Surg.* **2015**, *11*, 436–447. [CrossRef]
21. Kang, S.; Kim, M.; Kim, J.; Park, H.; Park, W. Marker-Free Registration for the Accurate Integration of CT Images and the Subject's Anatomy During Navigation Surgery of the Maxillary Sinus. *Dentomaxillofac. Radiol.* **2012**, *41*, 679–685. [CrossRef]
22. Li, G.; Dong, J.; Wang, J.; Cao, D.; Zhang, X.; Cao, Z.; Lu, G. The Clinical Application Value of Mixed-Reality-Assisted Surgical Navigation for Laparoscopic Nephrectomy. *Cancer Med.* **2020**, *9*, 5480–5489. [CrossRef] [PubMed]
23. Teatini, A.; Kumar, R.P.; Elle, O.J.; Wiig, O. Mixed Reality as a Novel Tool for Diagnostic and Surgical Navigation in Orthopaedics. *Int. J. Comput. Assist. Radiol. Surg.* **2021**, *16*, 407–414. [CrossRef] [PubMed]
24. Lee, S.; Jung, H.; Lee, E.; Jung, Y.; Kim, S.T. A Preliminary Work: Mixed Reality-Integrated Computer-Aided Surgical Navigation System for Paranasal Sinus Surgery Using Microsoft HoloLens 2. In *Advances in Computer Graphics: 38th Computer Graphics International Conference, CGI 2021, Virtual Event, 6–10 September 2021*; Proceedings; Springer: New York, NY, USA, 2021.
25. Kim, S.; An, H.; Song, M.; Lee, S.; Jung, H.; Kim, S.; Jung, Y. Automated Marker-Less Patient-to-Preoperative Medical Image Registration Approach Using RGB-D Images and Facial Landmarks for Potential Use in Computed-Aided Surgical Navigation of the Paranasal Sinus. In *Advances in Computer Graphics: 40th Computer Graphics International Conference, CGI 2023, Shanghai, China, 28 August–1 September 2023*; Proceedings, Part IV; Springer: New York, NY, USA, 2023.
26. Mildenhall, B.; Srinivasan, P.P.; Tancik, M.; Barron, J.T.; Ramamoorthi, R.; Ng, R. Nerf: Representing Scenes as Neural Radiance Fields for View Synthesis. *Commun. ACM* **2021**, *65*, 99–106. [CrossRef]
27. Microsoft. Microsoft Hololens 2 Docs. Available online: https://www.microsoft.com/it-it/hololens (accessed on 15 November 2024).
28. Gsaxner, C.; Pepe, A.; Wallner, J.; Schmalstieg, D.; Egger, J. Markerless Image-to-Face Registration for Untethered Augmented Reality in Head and Neck Surgery. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, 13–17 October 2019.
29. Groenenberg, A.; Brouwers, L.; Bemelman, M.; Maal, T.J.; Heyligers, J.M.; Louwerse, M.M. Feasibility and Accuracy of a Real-Time Depth-Based Markerless Navigation Method for Hologram-Guided Surgery. *BMC Digit. Health* **2024**, *2*, 11. [CrossRef]
30. Pepe, A.; Trotta, G.F.; Mohr-Ziak, P.; Gsaxner, C.; Wallner, J.; Bevilacqua, V.; Egger, J. A Marker-Less Registration Approach for Mixed Reality–Aided Maxillofacial Surgery: A Pilot Evaluation. *J. Digit. Imaging* **2019**, *32*, 1008–1018. [CrossRef]
31. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
32. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep High-Resolution Representation Learning for Human Pose Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.
33. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 3349–3364. [CrossRef] [PubMed]
34. Martinez, B.; Valstar, M.F.; Binefa, X.; Pantic, M. Local Evidence Aggregation for Regression-Based Facial Point Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 1149–1163. [CrossRef] [PubMed]
35. Sun, Y.; Wang, X.; Tang, X. Deep Convolutional Network Cascade for Facial Point Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013.
36. Van Loan, C.F. Generalizing the Singular Value Decomposition. *SIAM J. Numer. Anal.* **1976**, *13*, 76–83. [CrossRef]
37. Chetverikov, D.; Svirko, D.; Stepanov, D.; Krsek, P. The Trimmed Iterative Closest Point Algorithm. In Proceedings of the 2002 International Conference on Pattern Recognition, Quebec City, QC, Canada, 11–15 August 2002.
38. Müller, T.; Evans, A.; Schied, C.; Keller, A. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph. (TOG)* **2022**, *41*, 1–15. [CrossRef]
39. Microsoft. Hololens2forcv. Available online: https://github.com/microsoft/HoloLens2ForCV (accessed on 15 November 2024).
40. Opencv. Available online: https://opencv.org/ (accessed on 15 November 2024).