Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

CT-Free Image-to-Patient Registration for Stereotactic Neurosurgery Using Depth Camera and Automatic Facial Landmarks

Ersan Yazan

eyazan@adiyaman.edu.tr

Adıyaman University https://orcid.org/0000-0002-3315-241X

Muhamed Fatih Talu

İnönü University https://orcid.org/0000-0003-1166-8404

Ömür Aydoğmuş

Fırat University https://orcid.org/0000-0001-8142-1146

Research Article

Keywords: Stereotactic neurosurgery, CT-free navigation, Robotic surgical guidance, Depth camera, Facial landmarks

Posted Date: June 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9939835/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

CT-Free Image-to-Patient Registration for Stereotactic Neurosurgery Using Depth Camera and Automatic Facial Landmarks

Ersan Yazan[1]* , Muhamed Fatih Talu[2] , Ömür Aydoğmuş[3]

[1] Department of Computer Engineering, Faculty of Engineering, Adiyaman University, Adıyaman 02040, Turkey

[2] Department of Computer Engineering, Faculty of Engineering, Inonu University, Malatya 44000, Turkey

[3] Department of Mechatronics Engineering, Technology Faculty, Firat University, Elazığ 23119, Turkey

Abstract

Purpose Stereotactic brain surgery requires high-precision targeting of brain regions, traditionally relying on two separate radiological imaging procedures—MRI and CT—and manual landmark-based registration. This process increases radiation exposure and depends heavily on operator expertise.

Methods This study proposes an automatic head model registration method that uses MRI and depth camera data to replace the conventional MRI-CT image fusion workflow. 3D head models were generated from MRI (MHead) and depth camera input (DHead). Facial landmarks (n=468) were automatically detected using the MediaPipe framework and used for point-based registration. Rigid, affine, Random Forest (RF), and Singular Value Decomposition (SVD) transformation techniques were evaluated.

Results Registration achieved sub-2 mm accuracy, with RMSE and FRE values within clinically acceptable thresholds. Real-world targeting tests with a robotic arm confirmed successful localization of target points, validating the method's practical applicability in surgical settings.

Conclusions The proposed approach eliminates the need for CT imaging and manual landmark labeling in stereotactic procedures, offering a reliable, efficient, and fully automatic alternative with strong potential for clinical translation.

INTRODUCTION

Stereotactic surgery is a minimally invasive surgical technique applied to precisely identify and target specific regions within the brain or other organs. It is particularly effective in neurosurgery for the diagnosis and treatment of various neurological disorders, including tumor resection, localization of epileptic foci, and treatment of Parkinson’s disease. The stereotactic targeting process enables the surgeon to reach the target tissue with millimetric accuracy through three-dimensional coordinate systems constructed using advanced imaging technologies such as magnetic resonance imaging (MRI) and computed tomography (CT) [1]. This method is based on the Cartesian coordinate system, allowing the mathematical definition of target areas in space by using the intersection points of three mutually perpendicular planes (anterior-posterior, lateral, and vertical) [2].

Stereotactic surgery is a highly meticulous and demanding procedure that requires detailed planning and precision. It is typically performed under anesthesia and, as illustrated in Figure 1, is carried out in two separate sessions. In the first session, imaging and surgical planning are conducted, while the second session—usually scheduled for the following day or a predetermined later date—involves image reacquisition, co-registration with previously recorded data, validation processes, and the surgical intervention itself.

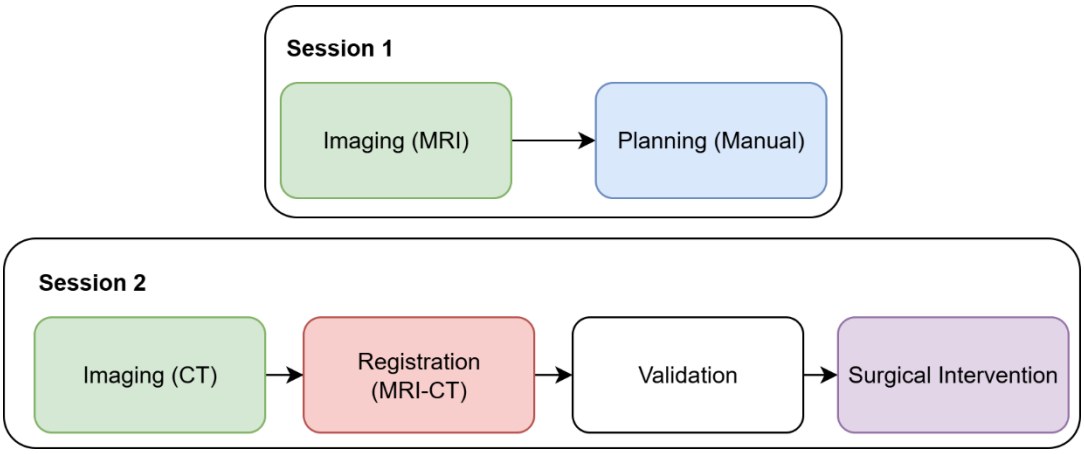


Figure 1. Framed and frameless stereotactic surgical process

Stereotactic surgery can be performed using two different approaches: frame-based and frameless. In the frame-based method, the procedure is typically conducted in two stages. In the first stage, a specialized metal frame is securely attached to the patient's head. Once the frame is fixed, detailed imaging of the brain structures is performed. At this stage, advanced imaging modalities such as magnetic resonance imaging (MRI) or computed tomography (CT) are used to scan the brain. The reference points on the stereotactic frame are clearly visible in these scans, allowing surgeons to accurately determine the precise coordinates of the targeted region [3].

At this stage, the surgeon conducts a detailed planning process on the acquired images, which takes approximately 221 ± 39 minutes on average [4]. During this process, key anatomical structures such as the anterior commissure (AC), posterior commissure (PC), the TAL line, as well as target and entry points, are identified [5]. In the second stage, which is typically performed the following day, the stereotactic frame is reattached to the patient, and a computed tomography (CT) scan is acquired. Subsequently, the MRI and CT images are co-registered, allowing the coordinates marked during the first session to be transferred onto the CT images. Based on this information, the surgical intervention is performed manually.

Over time, various frame models have been developed. Several examples of these frames are shown in Figure 2. Despite various structural and technological refinements over the years, the fundamental design of frame-based systems can still compromise patient and surgeon comfort during the procedure [6].



Figure 2. Stereotactic frames

In frameless stereotactic surgery, two key differences stand out. First, no physical frame is affixed to the patient’s head. Instead, targeting is performed using specific surface markers placed on the skull, in conjunction with an auxiliary navigation system. This allows for computer-assisted guidance to the target regions. In recent years, significant advancements have been made in frameless stereotactic systems, particularly with the integration of

robotic technologies into surgical practice [7]. As a result of these developments, robot-assisted frameless systems have increasingly been adopted in various clinical applications, including stereotactic biopsy, deep brain stimulation, treatment of Parkinson's disease, and the intervention of brain lesions.

In navigation-guided frameless surgery, reference markers enable the system to accurately recognize the patient's cranial and brain anatomy in three dimensions. During preoperative planning, the surgeon determines the marker placement locations on the patient's head; the markers are then secured using adhesive pads or small screws under local anesthesia. Following marker placement, high-resolution MRI or CT imaging is performed, allowing the surgeon to visualize both the reference markers and brain anatomy simultaneously. The navigation system subsequently uses these markers to track head position and calibrate instrument movements in real time. Representative examples of reference markers used in current clinical practice are shown in Figure 3.

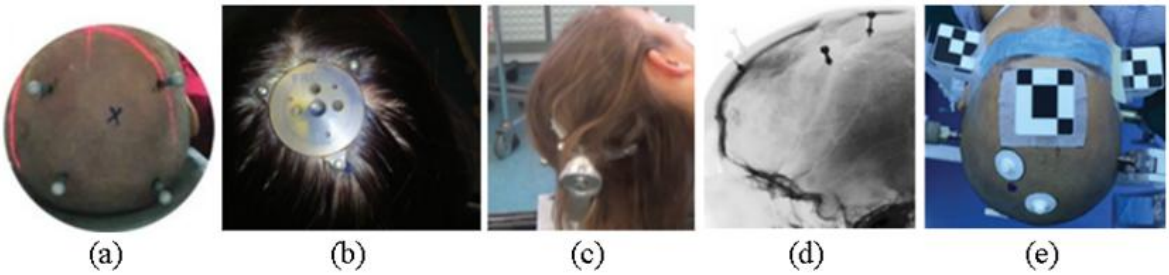


Figure 3. Reference markers used today

Frameless systems allow for more precise targeting and significantly reduce surgical time, thanks to the absence of mechanical constraints associated with frame use, advanced image fusion algorithms, and three-dimensional imaging capabilities [8-10].

In both stereotactic approaches, successful outcomes depend on the surgeon's ability to accurately localize—in physical space—the anatomical points previously marked on MRI scans with millimetric precision. For example, if a surgeon marks the center of a tumor on an MRI, the surgical goal is to reach that exact point. In conventional workflows, MRI and CT images are fused to enable this localization: a second CT scan is acquired, reference points are manually marked, and MRI-CT registration is performed. The accuracy of the alignment is then verified manually by superimposing the two image sets at different opacity levels, after which the surgical procedure begins.

Both frame-based and frameless systems carry inherent limitations, which can be summarized as follows:

- The requirement for two radiological imaging sessions subjects the patient to radiation exposure on two separate occasions.
- The image fusion process requires the surgeon to manually identify and mark reference points on both image sets.
- In frame-based systems, the mechanical compression of the frame on the patient's skull can potentially damage the underlying bone structure.
- In frameless systems, screw-based marker fixation to the skull may adversely affect patient comfort.
- Both hardware and software components of current systems entail substantial costs.

This study is motivated by the need to address the aforementioned limitations of current stereotactic systems through the application of recent computer vision-based modeling and detection techniques. Specifically, a novel approach employing a depth camera and automatic facial landmark registration is proposed as an alternative to existing workflows. The key advantages over conventional methods are as follows:

- Unlike conventional two-stage radiological imaging, the proposed system requires only a single MRI session; in the second stage, a depth camera is used in place of CT to acquire 3D surface data.
- Facial reference points are detected automatically, eliminating the need for manual surgeon intervention.
- The entire registration process is carried out automatically.

The proposed method comprises three main stages. In the first stage, head models are extracted from MRI data (MHead) and depth camera data (DHead), respectively. In the second stage, facial landmarks are identified on both models. In the third stage, these landmarks are used to perform point-based registration between MHead and DHead.

MATERIAL AND METHODS

Current stereotactic targeting workflows present significant challenges: patients must undergo two separate radiological imaging sessions, reference points must be marked manually or with dedicated hardware, and overall system costs are high. To overcome these limitations, this study presents a novel targeting method in which a depth camera replaces the second radiological imaging step and all landmark marking is performed automatically, without manual intervention or additional hardware.

The proposed system integrates MRI and depth camera data within a unified processing pipeline. As shown in Figure 4, two-dimensional and three-dimensional facial landmarks are extracted from skull models derived from each modality. Registration between these landmark sets yields the patient's spatial position in real-world coordinates, and the resulting transformation matrix is transmitted to the robotic arm to enable precise targeting of the surgeon-defined region.

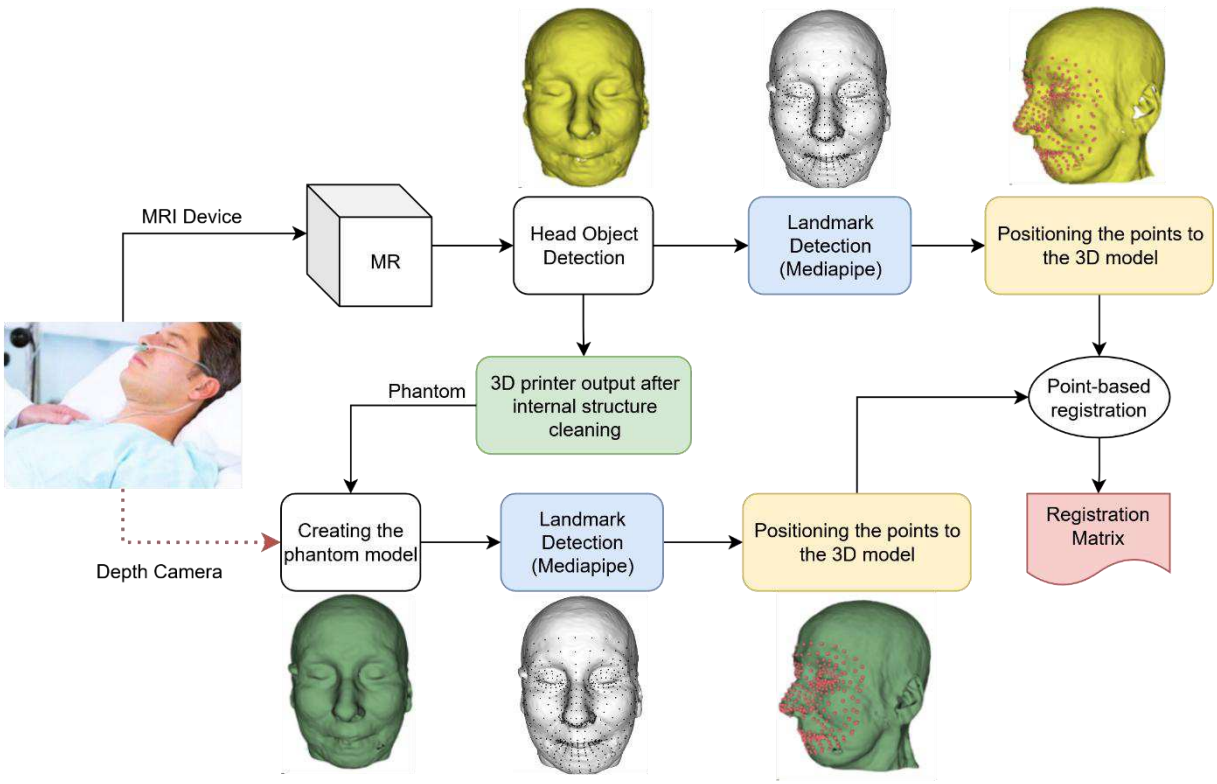


Figure 4. General structure of the proposed system

For the practical implementation of the proposed method, MRI and depth camera data must be collected from the same patient. However, since the existing MRI datasets used in this study did not include depth camera data, an

additional procedure became necessary. Within this scope, an "External Head Surface Model," which will be detailed in subsequent sections, was created, and a physical replica of this model (phantom) was produced using a 3D printer. The phantom was used in place of the real patient, enabling the acquisition of depth camera data (indicated by the red dashed line) from this model.

Generation of the MRI Head Model

The patient's cranial structure was digitally modeled using MRI images. In this modeling process, the cranial anatomy was first segmented, followed by the extraction of the outer surface. The surface data was then transferred to a 3D printer to produce a physical model (phantom). Additionally, specific anatomical facial landmarks were defined on the head model. The following sections provide technical details of the sub-steps carried out during this process.

Detection of the MRI Head Object

MRI images were primarily treated as two main voxel sets, representing regions belonging to the cranial structure and those that do not. Thresholding techniques were utilized to differentiate between these two regions. In this study, two widely used thresholding methods were applied: Otsu and MuStd. As an alternative approach, a method aimed at extracting the surface structures of objects from MRI images was also adopted. For this purpose, the Marching Cubes algorithm, commonly employed in computer graphics, was selected, and outputs obtained with different parameter settings were compared. To evaluate the effectiveness of the chosen methods on different datasets, one MRI image was sampled from each of the ADNI, IBSR, and IXI databases. The main characteristics of these datasets are summarized in Table 1.

Table 1. Datasets used in the head detection method

Dataset	Image Dimension(pixel)	Voxel Size(mm)	Type
ADNI	256 x 256 x 166	0.94 x 0.94 x 1.20	T1
IBSR	256 x 256 x 128	0.93 x 0.93 x 1.50	T1
IXI	256 x 256 x 150	0.93 x 0.93 x 1.20	T1

Figure 5 illustrates representative outputs from each dataset. Thresholding-based methods produced solid, complete head objects, whereas Marching Cubes yielded a hollow surface model. Although this structural difference did not constitute a decisive criterion for method selection, thresholding was substantially faster. Among the evaluated methods, MuStd thresholding achieved the most consistent results in terms of accuracy and was therefore selected for head object extraction. The method demonstrated robust performance across varying MRI resolutions and imaging sequences without significant degradation.

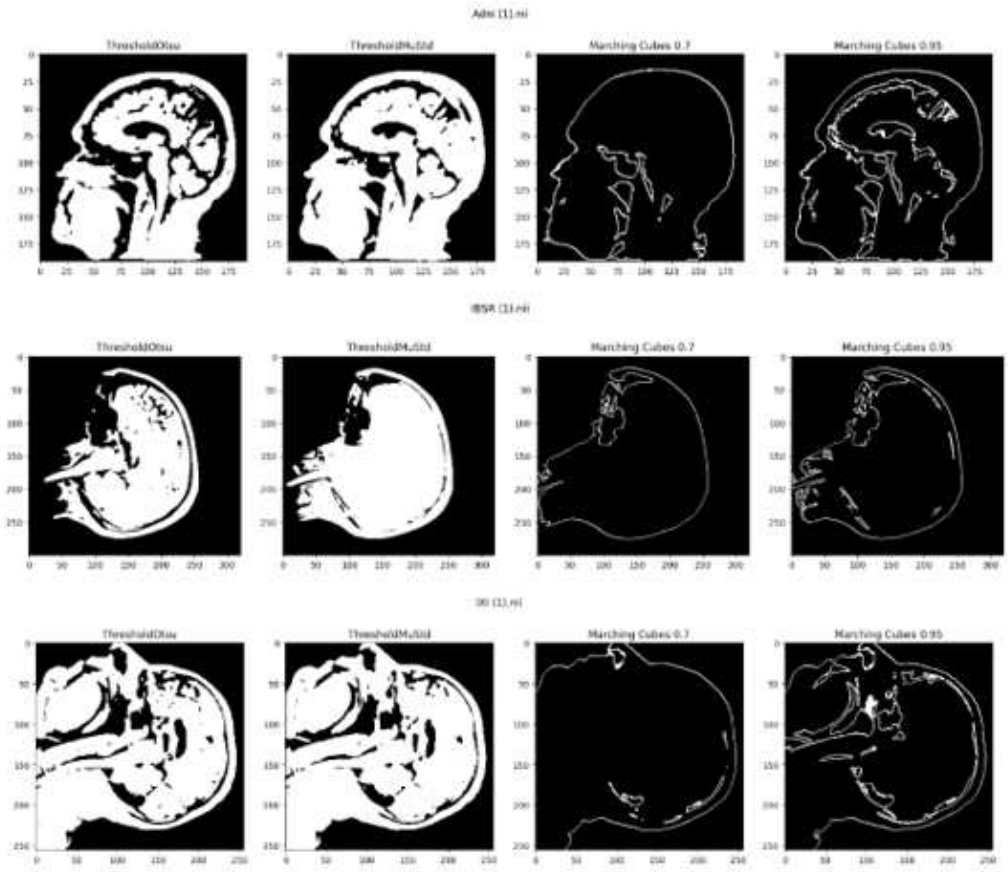


Figure 5. Results of the head detection method (top: ADNI, middle: IBSR, bottom: IXI dataset sample)

MR Head Model Generation

At this stage, binary LabelMap volumes are converted into three-dimensional surface models, hereafter referred to as MHead. The conversion pipeline used built-in functions from the open-source Visualization Toolkit (VTK) library and proceeded as follows: surface extraction was first performed using the *vtkDiscreteMarchingCubes* function; isolated or spurious surface components were then removed using *vtkCleanPolyData*. surface holes were subsequently filled using *vtkFillHolesFilter*, and the model was smoothed using *vtkWindowedSincPolyDataFilter*. The resulting MHead models, generated from five patients, are shown in Figure 6.



Figure 6. Generating MRI head models (M-Head) of 5 different patients

Prior to 3D printing, the internal structures of the head models were removed to prepare hollow shells suitable for fabrication. This process is described in the following section.

Internal Structure Cleaning of MR Head Model

Removing unnecessary internal structures from head models reduces 3D printing time and improves print quality. A dedicated method was developed for this purpose, as described in Algorithm 1. The algorithm scans the head volume along the sagittal, axial, and coronal planes and retains only the first and last head voxels encountered in each slice, effectively producing a hollow shell.

Algorithm 1 Internal Structure Cleaning of MR Head Model

```
# sagittal axis (x)
for x in range(shape[0]):
    for y in range(shape[1]):
        slice_ = labelmap[x, y, :]
        first, last = process_slice(slice_)
        if first != -1:
            output_map[x, y, first] = 1
        if last != -1 and last != first:
            output_map[x, y, last] = 1

# axial axis (z)
for z in range(shape[2]):
    for x in range(shape[0]):
        slice_ = labelmap[x, :, z]
        first, last = process_slice(slice_)
        if first != -1:
            output_map[x, first, z] = 1
        if last != -1 and last != first:
            output_map[x, last, z] = 1

# coronal axis (y)
for y in range(shape[1]):
    for z in range(shape[2]):
        slice_ = labelmap[:, y, z]
        first, last = process_slice(slice_)
        if first != -1:
            output_map[first, y, z] = 1
        if last != -1 and last != first:
            output_map[last, y, z] = 1
```

The algorithm was successfully applied to five patient head models, producing hollow shells ready for 3D printing. A representative result is shown in Figure 7, and outputs across different MRI datasets are presented in Figure 8.

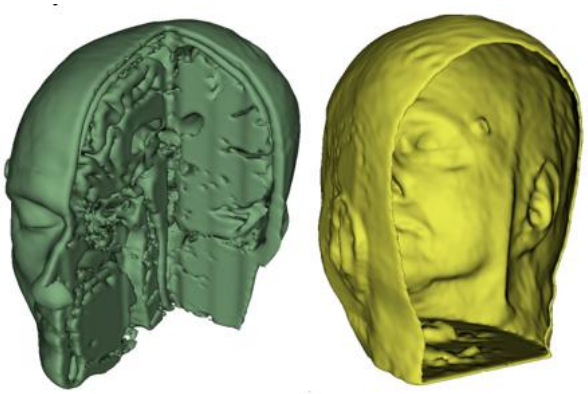


Figure 7. Internal structure cleaning result (left): Before cleaning, (right): After cleaning

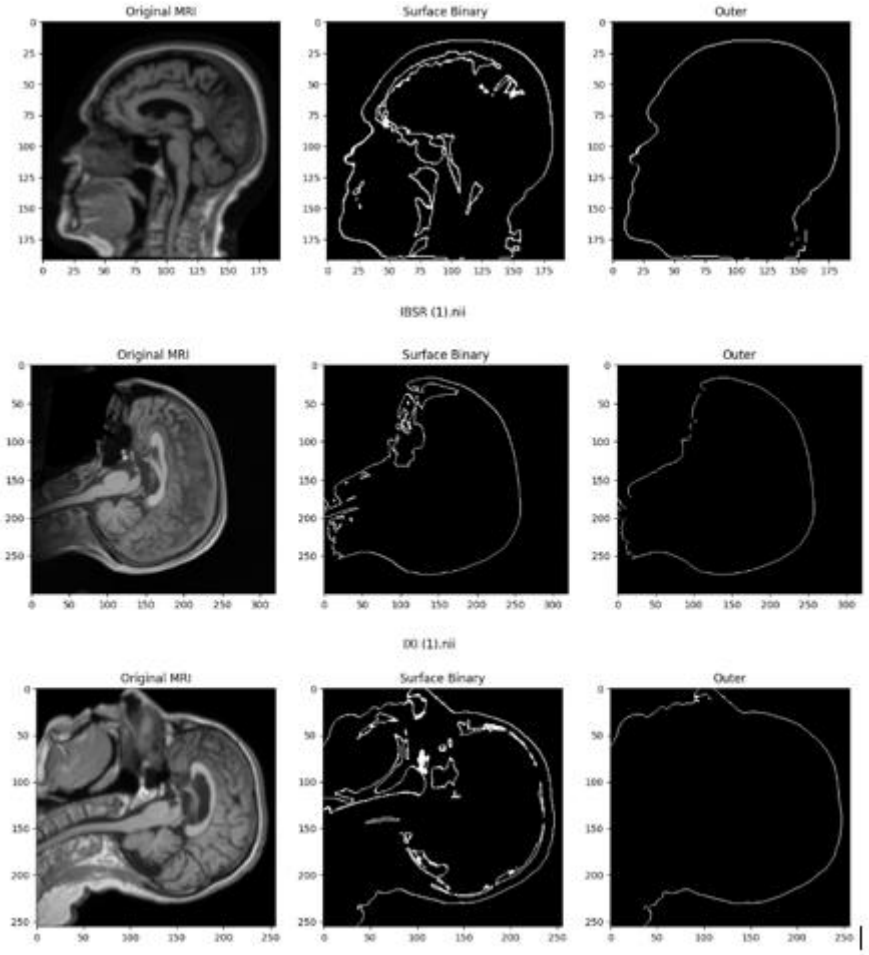


Figure 8. Internal structure cleaning results (left): MR, (middle) Marching cubes, (right) Internal structure cleaning output

The resulting hollow models were then fabricated as physical phantoms using a 3D printer and used as surrogates for real patients in all subsequent experiments.

Head Model Generation Using Depth Camera

Three-dimensional DHead models were generated from phantom head objects using depth camera data. For this purpose, two different methods were employed. In the first method, the depth camera mounted at the end of a

robotic arm was positioned to directly face the phantom head, and a single RGBD frame was captured. The color (RGB) image from this RGBD data was processed using the *MediaPipe* library, and 468 two-dimensional (2D) facial landmarks were detected through a 2D facial mesh analysis. Subsequently, using the corresponding depth information, the 3D spatial coordinates of each of these 2D points were computed. This transformation process was carried out using the method presented in Algorithm 2**.**

Algorithm 2 Conversion of 2D Facial Landmarks to 3D Coordinates

Input: *Point2d, Depth, T_Cam, width, height, cx, cy, fx, fy*
Output: *Point3d*

```
for Point in range(Point2d) do
    u = Point.x * width
    v = Point.y * height
    z = Depth[u, v]
    x = (u – cx)* z/fx
    y = (v – cy) * z/fy
    Point = TCam * [x, y, z, 1]
    Point3d.append(Point)
end for
```

The variable *Point2d* refers to the two-dimensional (2D) facial landmarks extracted using the Mediapipe library. The *width* and *height* parameters represent the pixel dimensions of the color (RGB) image acquired from the depth camera, while *cx* and *cy* denote the coordinates of the image center. The horizontal and vertical focal lengths of the camera are stored in the variables *fx* and *fy*, respectively.

The *TCam* matrix encodes the position and orientation of the camera mounted on the robotic arm end-effector in world coordinates. Multiplying the camera-frame landmark coordinates by T_Cam transforms them into real-world coordinates, enabling accurate spatial localization of each facial landmark.

In the second method, multiple RGBD frames were merged (stitched) to construct the facial mesh model. During this process, numerous RGBD images of the phantom head captured from different angles were imported into a software called "Kiri Engine", which then generated a 3D facial model of the phantom. Figure 9 presents both the MHead and DHead models corresponding to four patients (IDs 08, 10, 25, and 38) from the ADNI dataset.

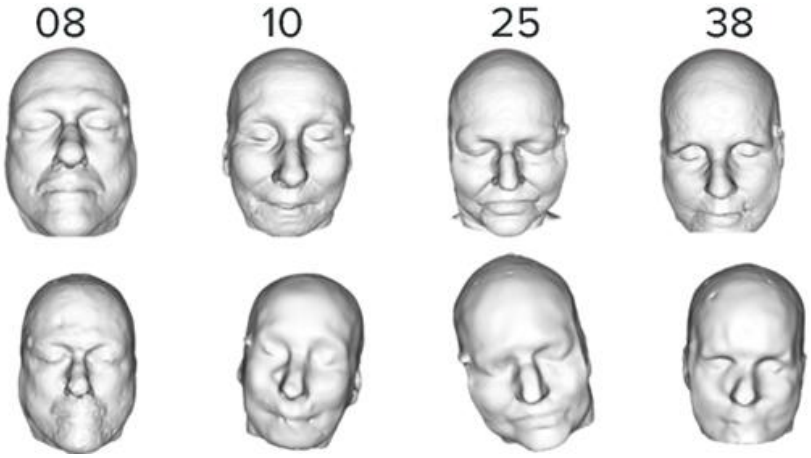


Figure 9. Head Models (top): MHead, (bottom): DHead

The multi-view models produced by the second method were visually smoother and geometrically more accurate. Consequently, all subsequent registration experiments used models and phantoms generated by this approach.

Detection of Facial Landmarks on Head Models

In point-based registration, both the number of landmarks and their spatial accuracy are critical to alignment quality. Crucially, corresponding points on the fixed and moving models must map to anatomically identical locations. The commercial BrainLab system [11] addresses this requirement through a dedicated hardware solution, whereas the proposed approach achieves the same objective automatically and without any additional hardware, offering clear advantages in cost and implementation simplicity.

A further key advantage of the proposed method is rapid, fully automatic facial landmark detection via the MediaPipe library [12]. MediaPipe performs detailed facial anatomy analysis and models the entire facial surface using 468 two-dimensional landmarks encompassing the eyes, eyebrows, nose, lips, and surrounding regions. A representative example is shown in Figure 10.

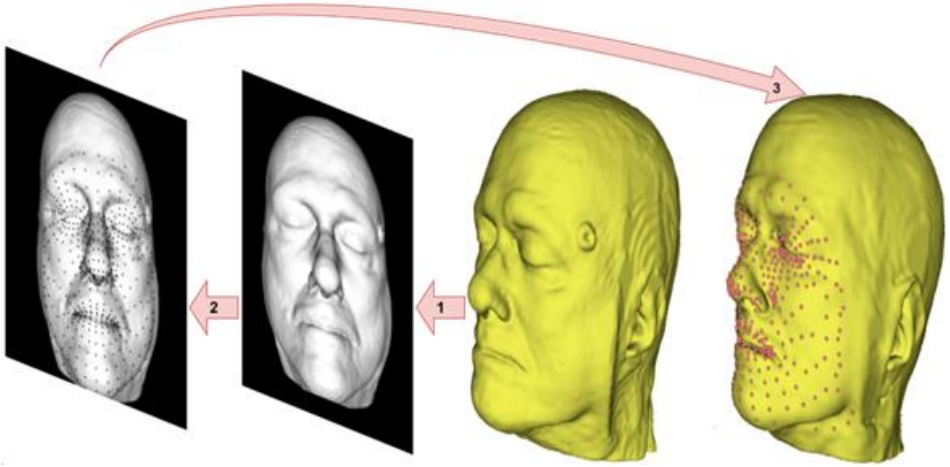


Figure 10. Detection of facial points on head models

Because MediaPipe operates on two-dimensional images, a 2D projection of the facial region was first rendered from the 3D head model and passed to MediaPipe, which returned 468 (x, y) landmark coordinates. To recover the corresponding 3D positions on the model surface, Algorithm 3 was applied; this procedure uses the VTK vtkCellPicker class to project each 2D point back onto the 3D surface and retrieve its world coordinates.

Algorithm 3 Conversion of 2D Facial Landmarks to 3D Coordinates with VTK

Input: *Point2d, vtkrenderer*
Output: *Point3d*

```
picker = vtk.picker = vtk.vtkCellPicker()
Point3d = []
for each, i ∈ len(Point2d) do
    X = int(Point2d [i].x ∗ MR.width)
    Y = int(Point2d [i].y ∗ MR.height
    picker.Pick(Landmark2d[i][0], Landmark2d[i][1], 0, vtkrenderer)
   Point = picker.GetPickPosition()
   Point3d.append([Point[0], Point[1], Point[2]])
```

Algorithm 3 was applied to both MHead and DHead models, yielding two corresponding 3D point sets—MHeadPoint and DHeadPoint. The procedure was applied to patients 10, 25, and 38; representative outputs are shown in Figure 11.

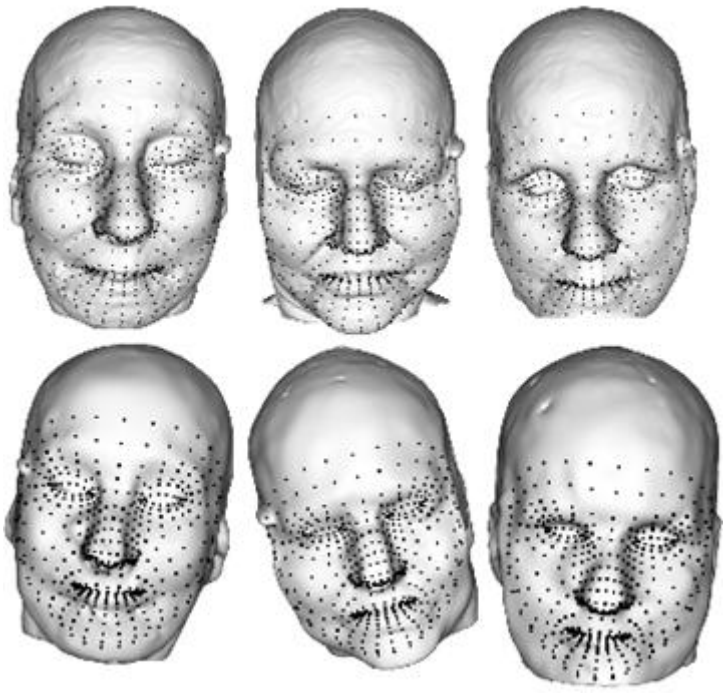


Figure 11. Detection of facial landmarks (top): MHeadPoint, (bottom): DHeadPoint

Registration of Head Models

Image registration is the process of spatially aligning a source (moving) image to a target (fixed) image, i.e., finding a transformation *T* that maps corresponding structures onto one another. Based on the degrees of freedom allowed, registration methods are classified as rigid, similarity, affine, or non-rigid [13]. In this study, rigid and affine transformations were evaluated alongside two landmark-based methods: Singular Value Decomposition (SVD) and Random Forest (RF).

In the SVD-based method, both point sets are first mean-centered by translating their centroids to the origin. A cross-covariance matrix M is then computed by projecting the moving point set onto the fixed set, yielding a MMM that represents the influence of each moving point on the fixed set. This 468x468 matrix undergoes singular value decomposition to obtain the eigenvectors of MM^T and M^TM (denoted as U and V, respectively). Using these orthonormal matrices, the rotation matrix RRR is computed as V^TUV. Scale and translation parameters are then incorporated to complete the full transformation. The SVD-based approach is computationally efficient; its pseudocode is given in Algorithm 4.

Algorithm 4 SVD-based registration

Input: *fixedN, movingN*
Output: *alignedN*

movingN = movingN – mean(movingN)
fixedN = fixedN – mean(fixedN)
$M = movingN^T * fixedN$
U, S, V = SVD(M)
$R = V^T. U. V$
$alignedN = movingN * R^T$
scale = norm(fixedN) / norm(movingN)
*alignedN = alignedN * scale + mean(fixedN)*

The Random Forest (RF) method [14] is an ensemble learning approach that aggregates predictions from multiple independently trained decision trees; during inference, all trees produce position estimates that are averaged to yield the final output. RF offers high accuracy, strong generalizability, and fast inference, though it requires sufficient training data. The 468 landmark points available in this study provided an adequate training set, making RF well suited for modeling the spatial correspondence between MHeadPoint and DHeadPoint. RF has also been applied in several related registration studies [15–18].

Registration experiments were conducted on data from three ADNI patients (IDs 10, 25, and 38) using the MHead and DHead models. Two experimental conditions were evaluated: registration using manually identified landmarks and registration using automatically detected landmarks.

Registration accuracy was quantified using three standard metrics: Root Mean Square Error (RMSE), Fiducial Registration Error (FRE), and Target Registration Error (TRE). RMSE and FRE measure the residual alignment error at the fiducial points used during registration, whereas TRE is evaluated at independent target points not used in the registration. For TRE computation, the centroid of the 468-point set was used as the target reference in both models.

RESULTS

This section presents the findings related to the registration processes carried out within the scope of the proposed method. Various registration experiments were performed on MHead and DHead models using different approaches. The registration was conducted based on both manually selected facial landmarks and automatically detected facial points, and the impact of different transformation methods (rigid, affine, SVD-based, and Random Forest) on registration accuracy was thoroughly evaluated. To quantitatively analyze the results, commonly used error metrics such as RMSE, FRE, and TRE were employed, enabling a comparative assessment of each method's performance. The following subsections detail the registration procedures applied and the corresponding outcomes obtained from each.

Registration Using 6 Manually Selected Points

Registration was first performed using six manually selected facial landmarks, applying the SVD and RF methods described in Section 2 as well as the *vtkLandmarkTransform* class from the *VTK* library. The resulting transformation matrix was applied to all 468 landmarks and accuracy was assessed using the three error metrics. Results for Patient 10 are presented in Table 2.

Table 2. Registration results for Patient 10 using 6 manually selected points

	Initial	SVD	RF	Rigid	Affine
RMSE	95.885	4.270	10.693	**3.362**	4.754
FRE	166.077	7.396	18.521	**5.824**	8.235
TRE	165.802	2.605	3.863	**2.273**	0.984

Upon examining Table 2, it is observed that the Random Forest (RF) method yielded the highest error values among the registration techniques. As previously discussed, this method typically performs better with larger datasets, making this outcome an expected result. The rigid registration performed using the VTK library produced the lowest error values. This result is also visually supported by Figure 12, which illustrates the alignment of the head models following the registration process.

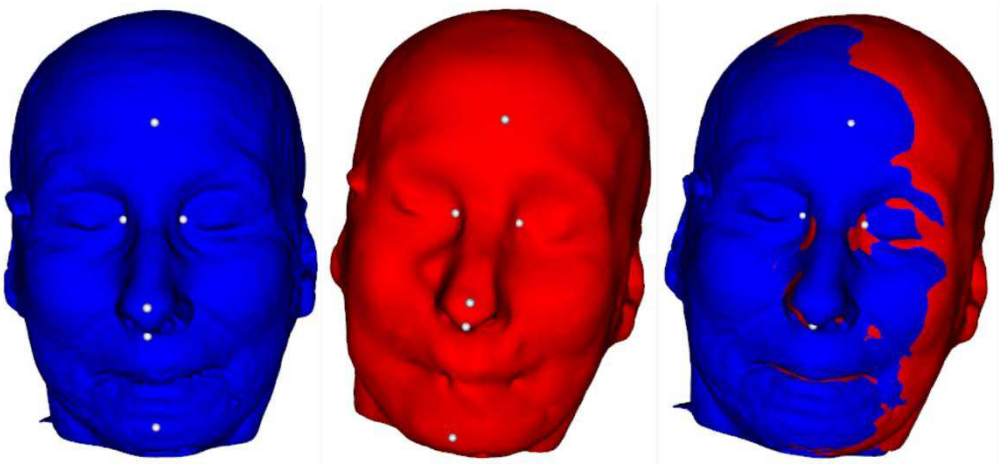


Figure 12. Registration Result Using Manually Selected Points. (Left): MHead, (Middle): DHead, (Right): Aligned Models

Figure 12 confirms that the models are well aligned overall. The result also suggests that more precise point correspondence would further improve registration accuracy, motivating the use of automatically detected landmarks as described in the following section.

Using 6 Automatically Selected Points

At this stage, the registration process was carried out using only six points selected from specific anatomical regions of the face. Since both the location and identity of each point obtained with the Mediapipe library are predefined, the selection could be performed automatically without the need for manual placement. The selected points were chosen as follows: one point from the center of the forehead, the tip of the nose, the root of the nose, the tip of the chin, and one point from each of the left and right cheeks. The error metrics obtained after performing the registration using these six reference points are presented in Table 3.

Table 3. Registration results for Patient 10 using 6 automatically selected points

	Initial	SVD	RF	Rigid	Affine
RMSE	95.885	3.023	12.080	3.377	**2.669**
FRE	166.077	5.235	20.924	5.850	**4.622**
TRE	165.802	2.147	3.503	2.711	**1.718**

Affine registration achieved the lowest errors in this configuration; the aligned models are shown in Figure 13.

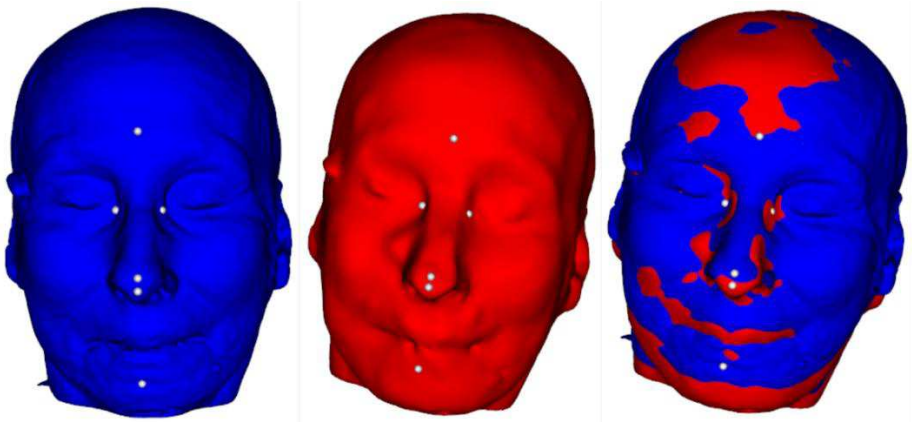


Figure 13. Registration result using automatically selected points. (Left): MHead, (Middle): DHead, (Right): Aligned models

Using a Large Number of Automatically Selected Points

Following landmark detection on both MHead and DHead models, registration was performed using SVD, RF, and the *vtkLandmarkTransform* class from VTK, with both rigid and affine transformation options. To assess the influence of landmark set composition on registration accuracy, four point sets were defined and evaluated separately (Table 4).

Table 4. Facial point sets used in the registration process

	Face Area	Landmark Count
Set1	All	468
Set2	Silhouette of the face	36
Set3	Eye, eyebrow, mouth, nose, cheek	181
Set4	Set3 + Set2	217

For each of the three ADNI patients (IDs 10, 25, and 38), registration was performed for all combinations of the four point sets and four methods. The evaluation strategy was as follows:

- Set 1: All 468 landmarks were used directly for registration.
- Sets 2, 3, and 4: Registration matrices were estimated from the respective reduced point sets and then applied to all 468 landmarks to evaluate global alignment accuracy.

For TRE calculations, the centroid of the full 468-point set served as the target reference point in both models. Tables 5, 6, and 7 report RMSE, FRE, and TRE values (in mm) for patients 10, 25, and 38, respectively, across all method-set combinations.

Table 5. Registration results for Patient 10

		Initial	SVD	RF	Rigid	Affine
Set1 (468 Points)	**RMSE**	95.885	2.682	**1.051**	2.945	2.371
	FRE	166.077	4.645	**1.820**	5.101	4.107
	TRE	165.802	0.000	0.100	0.000	0.000
Set2 (36 Points)	**RMSE**	95.885	3.215	2.823	3.372	2.728
	FRE	166.077	5.569	4.882	5.840	4.725
	TRE	165.802	2.700	1.436	2.635	1.089
Set3 (181 Points)	**RMSE**	95.885	2.806	3.077	3.088	2.582
	FRE	166.077	4.860	5.329	5.348	4.473
	TRE	165.802	0.301	0.462	0.829	0.267
Set4 (217 Points)	**RMSE**	95.885	2.722	2.485	2.966	2.406
	FRE	166.077	4.714	4.305	5.137	4.167
	TRE	165.802	0.408	0.417	0.379	0.270

Table 6. Registration results for Patient 25

		Initial	SVD	RF	Rigid	Affine
Set1 (468 Points)	**RMSE**	151.010	3.905	**0.894**	3.898	3.026
	FRE	261.556	6.763	**1.549**	6.752	5.242
	TRE	261.229	0.000	0.055	0.000	0.000
Set2 (36 Points)	**RMSE**	151.010	7.934	6.394	8.109	5.557
	FRE	261.556	13.742	11.075	14.045	9.626
	TRE	261.229	10.903	7.755	11.360	6.315
Set3 (181 Points)	**RMSE**	151.010	4.279	3.317	3.991	3.594
	FRE	261.556	7.411	5.745	6.913	6.225
	TRE	261.229	1.252	1.251	0.940	1.306
Set4 (217 Points)	**RMSE**	151.010	4.068	3.600	4.068	3.445
	FRE	261.556	7.046	6.235	7.046	5.968
	TRE	165.802	1.635	0.417	1.633	0.568

Table 7. Registration results for Patient 38

		Initial	SVD	RF	Rigid	Affine
Set1 (468 Points)	**RMSE**	87.445	3.948	**0.947**	4.074	3.206
	FRE	151.460	6.838	**1.641**	7.057	5.553
	TRE	151.288	0.000	0.095	0.000	0.000
Set2 (36 Points)	**RMSE**	87.445	6.634	5.566	6.828	5.183
	FRE	151.460	11.491	9.641	11.827	8.978
	TRE	151.288	8.473	6.315	8.701	5.508
Set3 (181 Points)	**RMSE**	87.445	4.130	3.372	4.145	3.397
	FRE	151.460	7.154	5.841	7.179	5.884
	TRE	151.288	0.904	0.691	0.921	0.714
Set4 (217 Points)	**RMSE**	87.445	4.046	3.710	4.148	3.505
	FRE	151.460	7.007	6.425	7.185	6.071
	TRE	151.288	1.071	0.713	0.868	0.551

Across all three patients, Set 1 (all 468 landmarks) consistently yielded the lowest error values. Within Set 1, the RF method achieved the best performance, recording the lowest RMSE and FRE values.

For the reduced point sets (Sets 2–4), affine transformation generally outperformed the other methods, likely due to its greater flexibility in accommodating spatial deformations when fewer landmarks are available.

Notably, while RF produced the lowest RMSE and FRE values under Set 1, the TRE values for the other methods were occasionally zero. This indicates that rigid and SVD methods aligned the centroid precisely but could not maintain that accuracy across the full surface, confirming the absence of a direct linear relationship between FRE and TRE.

Figure 14 visualizes the registration results for all three patients using RF with Set 1. The blue model represents MHead and the red model represents DHead; the close overlap confirms the high alignment quality achieved by this combination.

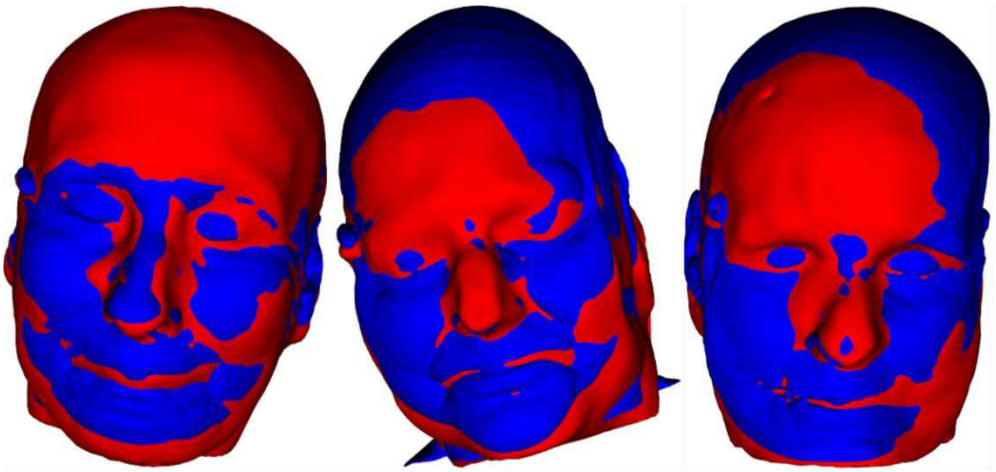


Figure 14. Visualization of skull models after registration using the RF method (left): Patient 10, (middle): Patient 25, (right): Patient 38

Compared with related work, the proposed method achieves competitive accuracy. A study on point-based 3D head model registration [19] reported RMSE values below 3 mm as clinically acceptable, and a study comparing skin markers with natural anatomical landmarks using real patient data [20] reported RMSE values of 3.2 mm and 2.9 mm, respectively.

A study examining the effect of reference point count on navigation accuracy [21] reported TRE values of 1.63–2.06 mm and noted the absence of a linear FRE–TRE relationship, consistent with the patterns observed in the present study.

An evaluation of a commercially deployed MRI registration platform [22] reported a TRE of 1.6 mm, providing a useful benchmark for the accuracy limits of current clinical systems.

In aggregate, the proposed method—using 468 landmarks with RF registration—achieves RMSE and FRE values below 2 mm and TRE values below the clinical benchmarks reported in the literature, demonstrating its viability as an accurate and reliable targeting tool.

Guiding the Robotic Arm to the Target Point on the Head

To evaluate the practical targeting accuracy of the proposed system, a robotic arm guidance test was performed. The surgeon defined a target point on the MRI volume, and the robotic arm was commanded to localize that point on the physical phantom.

Figure 15 shows the test performed on the Patient 10 phantom. The left panel depicts the robotic arm axes, the depth camera landmark coordinates, and the camera coordinate frame; the red dot indicates the surgeon-defined target on the MRI.

The right panel confirms that the needle mounted on the robotic arm end-effector was accurately guided to the target point on the forehead, demonstrating the feasibility of the proposed system for robot-assisted surgical targeting.

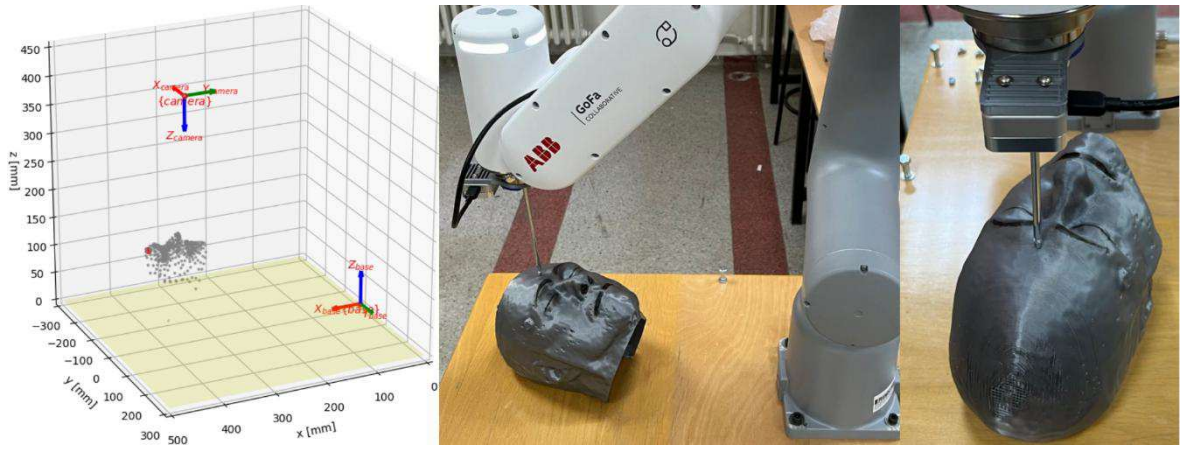


Figure 15. Targeted positioning of the robotic arm to a selected point on the forehead

Figure 16 shows a second targeting test in which a nasal landmark was selected; the robotic arm again reached the target accurately.

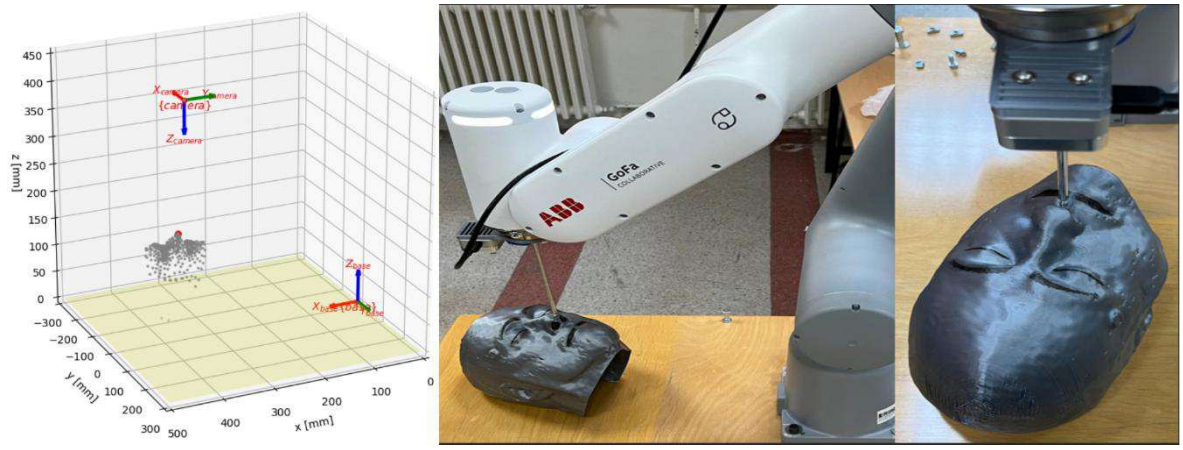


Figure 16. Targeted positioning of the robotic arm to a selected point on the nose

CONCLUSION

Conventional deep brain targeting systems rely on two radiological imaging modalities (MRI and CT) and require manual reference point marking by the surgeon. Image quality limitations, inter-operator variability, and marking errors can all degrade MRI-CT registration accuracy, compromising the precision of targeting and ultimately putting treatment outcomes at risk.

In this study, an alternative targeting approach was proposed to overcome the limitations of classical targeting methods mentioned above. The developed system utilizes not only the patient's MRI images but also depth camera data, thereby preventing the patient from being exposed to two separate radiological imaging sessions. Moreover, instead of manual marking, a facial mask consisting of 468 facial landmarks was automatically placed on both the MRI and depth data (MHead and DHead). In the subsequent stage, these masks were aligned using point-based registration methods, and a transformation matrix was calculated. According to the results obtained, the registration performed with the Random Forest (RF) algorithm yielded the lowest error values. The analyses revealed RMSE and FRE error values below 2 mm, which fall within the acceptable limits reported in the literature.

Robotic arm targeting tests on phantoms of three ADNI patients (IDs 10, 25, and 38) confirmed that the system could accurately reach surgeon-defined target points, validating the potential of the proposed method for integration into robot-assisted surgical workflows.

Future work should focus on testing the system with paired MRI and depth camera data acquired from real patients to evaluate performance under clinical conditions. The robustness of the approach to variations in facial morphology and expression should also be examined.

In summary, the proposed method constitutes a viable alternative to classical stereotactic targeting systems by reducing radiological imaging to a single session, eliminating manual landmark marking, and enabling fully automatic registration. With further development and clinical validation, this approach has the potential to become a practical and broadly adoptable tool in neurosurgical practice.

References

[1] Lozano, A. M., Lipsman, N., Bergman, H., Brown, P., Chabardes, S., Chang, J. W., ... & Krauss, J. K. (2019). Deep brain stimulation: current challenges and future directions. Nature Reviews Neurology, 15(3), 148–160. https://doi.org/10.1038/s41582-018-0128-2

[2] Faria, C., Erlhagen, W., Rito, M., De Momi, E., Ferrigno, G., & Bicho, E. (2015). Review of robotic technology for stereotactic neurosurgery. *IEEE Reviews in Biomedical Engineering*, ***8***, 125–137. https://doi.org/10.1109/RBME.2015.2428305

[3] Lunsford, L. D., & Kondziolka, D. (2000). *Frame-based stereotactic neurosurgery.* Neurosurgery Clinics of North America, 11(2), 157-172.

[4] Zanello, M., Carron, R., Peeters, S., Gori, P., Roux, A., Bloch, I., Oppenheim, C., & Pallud, J. (2021). Automated neurosurgical stereotactic planning for intraoperative use: a comprehensive review of the literature and perspectives. *Neurosurgical Review*, ***44(2)***, 867–888. https://doi.org/10.1007/s10143-020-01315-1

[5] Rainer G**.** *https://www.brainvoyager.com,* Accessed 12 April 2024.

[6] Ulutabanca H. [In Turkish]. (2021). Derin Beyin Stimülasyonu: Yeni Teknikler ve Teknolojiler, *Türk Nöroşir Dergisi., **31(3)***, 343-347.

[7] Piano, C., Bove, F., Mulas, D., Bentivoglio, A. R., Cioni, B., & Tufo, T. (2021). Frameless stereotaxy in subthalamic deep brain stimulation: 3-year clinical outcome. Neurological Sciences, 42(1), 259–266. https://doi.org/10.1007/s10072-020-04561-9

[8] Hiremath, G. (2020). Robotic Deep Brain Stimulation (R-DBS)-"Awake" Deep Brain Stimulation Using the Neuromate Robot and O-Arm. Neurology India, 68(8), 328. https://doi.org/10.4103/0028-3886.302450

[9] Neudorfer, C., Hunsche, S., Hellmich, M., El Majdoub, F., & Maarouf, M. (2018). Comparative Study of Robot-Assisted versus Conventional Frame-Based Deep Brain Stimulation Stereotactic Neurosurgery. Stereotactic and Functional Neurosurgery, 96(5), 327–334. https://doi.org/10.1159/000494736

[10] Paff, M., Wang, A. S., Phielipp, N., Vadera, S., Morenkova, A., Hermanowicz, N., & Hsu, F. P. K. (2020). Two-year clinical outcomes associated with robotic-assisted subthalamic lead implantation in patients with Parkinson's disease. Journal of Robotic Surgery, 14(4), 559–565. https://doi.org/10.1007/s11701-019-01025-x

[11] BrainLab. Available at: https://www.brainlab.com. Accessed 16 April 2024

[12] MediaPipe. Available at: https://ai.google.dev/edge/mediapipe/solutions/guide. Accessed 16 December 2023

[13] Aksoy, T. (2015). Template-based 3D-2D rigid registration of vascular structures in frequency domain from a single view. Doctoral Dissertation, Sabanci University, Istanbul, Turkey

[14] Breiman, L. (2001). Random Forest. *Machine Learning*, ***45(1)***, 5–32. https://doi.org/10.1023/A:1010933404324

[15] Sokooti, H., Saygili, G., Glocker, B., Lelieveldt, B. P. F., & Staring, M. (2019). Quantitative error prediction of medical image registration using regression forests. Medical Image Analysis, 56, 110–121. https://doi.org/10.1016/j.media.2019.05.005

[16] Sokooti, H., Saygili, G., Glocker, B., Lelieveldt, B. P. F., & Staring, M. (2016). Accuracy Estimation for Medical Image Registration Using Regression Forests, 107–115. https://doi.org/10.1007/978-3-319-46726-9_13

[17] Huynh, T., Gao, Y., Kang, J., Wang, L., Zhang, P., Lian, J., & Shen, D. (2016). Estimating CT Image From MRI Data Using Structured Random Forest and Auto-Context Model. IEEE Transactions on Medical Imaging, 35(1), 174–183

[18] Wei, L., Cao, X., Wang, Z., Gao, Y., Hu, S., Wang, L., Wu, G., & Shen, D. (2017). Learning-based deformable registration for infant MRI by integrating random forest with auto-context model. Medical Physics, 44(12), 6289–6303. https://doi.org/10.1002/mp.12578

[19] Schlesinger, O., Kundu, R., Isaev, D., Choi, J.Y., Goetz, S.M., Turner, D.A., Sapiro, G., Peterchev, A.V., Di Martino, J.M. (2024). Scalp surface estimation and head registration using sparse sampling and 3D statistical models. Computers in Biology and Medicine, 178: 108689. https://doi.org/10.1016/j.compbiomed.2024.108689

[20] Wolfsberger, S., Rössler, K., Regatschnig, R., Ungersböck, K. (2002). Anatomical landmarks for image registration in frameless stereotactic neuronavigation. Neurosurgical Review 25, 68-72. https://www.doi.org/10.1007/s10143-001-0201-x

[21] Javer, A., AlQunaee, M., Kolahi, A. (2020). Relationship between Target Registration Error (TRE) and Fiducial Registration Error (FRE), a clinical study and statistical analyses.

[22] Lin, J.S., Fuentes, D.T., Chandler, A., Prabhu, S.S., Weinberg, J.S., Baladandayuthapani, V., Hazle, J.D., Schellingerhout, D. (2017). Performance assessment for brain MR imaging registration methods. American Journal of Neuroradiology, 38(5), 973-980. https://doi.org/10.3174/ajnr.A5122