(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 December 2024 (12.12.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/254487 A2

(51) International Patent Classification:
Not classified

(21) International Application Number:
PCT/US2024/033061

(22) International Filing Date:
07 June 2024 (07.06.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/472,117 09 June 2023 (09.06.2023) US

(71) Applicants: **NORTHWESTERN UNIVERSITY** [US/US]; 633 Clark Street, Evanston, Illinois 60208 (US). **ANN AND ROBERT H. LURIE CHILDREN'S HOSPITAL OF CHICAGO** [US/US]; 225 E. Chicago Avenue, No. 205, Chicago, Illinois 60611-2605 (US). **CLEARVOYA LLC** [US/US]; 1211 S. Prairie Ave., Apt. 1201, Chicago, Illinois 60605 (US).

(72) Inventors: **ANSARI, Sameer Ahmad**; c/o Northwestern University, 633 Clark Street, Evanston, Illinois 60208 (US). **CANTRELL, Donald Robinson**; c/o Northwestern University, 633 Clark Street, Evanston, Illinois 60208 (US). **CHO, Leon**; c/o Northwestern University, 633 Clark Street, Evanston, Illinois 60208 (US). **ZHOU, Chaochao**; c/o Northwestern University, 633 Clark Street, Evanston, Illinois 60208 (US).

(74) Agent: **GALFANO, Christopher J.** et al.; BANNER & WITCOFF, LTD., 71 South Wacker Drive, Suite 3600, Chicago, Illinois 60606-7437 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ,

(54) Title: ROBUST SINGLE-VIEW CONE-BEAM X-RAY POSE ESTIMATION

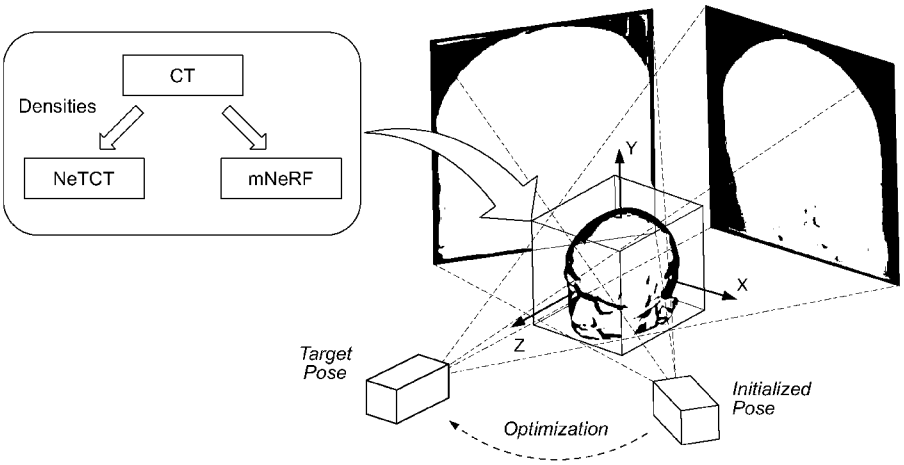


FIG. 1

(57) Abstract: A system and method of differentiable X-ray projection (DiffXP) for generating digitally reconstructed radiographs (DRRs) and optimization of single-view X-ray pose estimations.

[Continued on next page]

WO 2024/254487 A2

RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

ROBUST SINGLE-VIEW CONE-BEAM X-RAY POSE ESTIMATION

CROSS REFERENCE TO RELATED APPLICATION

[0001] This application claims priority to U.S. Provisional Patent Application No. 63/472,117 filed on June 9, 2023, which is hereby incorporated by reference herein.

TECHNICAL FIELD

[0002] This disclosure relates to the field of generating digitally reconstructed radiographs (DRRs). More specifically, systems and methods are disclosed used to improve the fidelity of DRRs and optimization of single-view X-ray pose estimations.

BACKGROUND

[0003] Through the use of widely available 3D medical images, such as computerized tomography (CT) and magnetic resonance imaging (MRI), surgical planning has been enabled on 3D anatomical models reconstructed from the 3D images. Models are limited, however, because intraoperative positions and orientations (i.e., poses) of a patient are unknown. Pose estimation by image registration is a prerequisite for accurate image guidance in diagnosis or surgery. As such, estimation of a patient's poses is essential for accurate image guidance in widespread minimally invasive surgical procedures and essential for many other clinical applications. As such, an improved system for more accurate and detailed X-ray pose estimation through image registration (matching) of digitally reconstructed radiographs (DRRs) to X-ray images would be advantageous.

SUMMARY

[0004] This Summary introduces a selection of concepts relating to this technology in a simplified form as a prelude to the Detailed Description that follows. This Summary is not intended to identify key or essential features.

[0005] In certain aspects, a method for generating digitally reconstructed radiographs disclosed herein may include the steps of selecting a predicted anatomical target pose of an individual, receiving gradient-based registration images comprising known anatomical

poses of a plurality of individuals, estimating the predicted target anatomical target pose by rendering the gradient-based registration images by differentiable X-ray projection, optimizing the rendered gradient-based registration images in which the received gradient-based registration images may be optimized via neural tuning of computed tomography, or wherein the received gradient-based registration images may be optimized via masked neural radiance fields, and generating an optimized digitally reconstructed radiograph of the predicted anatomical target pose. In other aspects, machine learning algorithms, models, or frameworks may be executed to train the disclosed systems and methods disclosed herein.

[0006] In other examples, the received gradient-based registration images may include X-ray images and digitally reconstructed radiographs, and during optimizing, estimation of the predicted anatomical target pose of an individual may be iteratively performed by minimizing differences between the X-ray images and the digitally reconstructed radiographs. In another example, the differences between the X-ray images and the digitally reconstructed radiographs may be minimized by adjusting a pose of a source of the digitally reconstructed radiographs. In another example, the X-ray images may include a contrast enhancement, an intensity invert, and a scaling and image resizing. In yet another example, the predicted anatomical target pose of an individual may be selected from representative X-ray images of the individual. In still other examples, even sampling of volume densities may be utilized to optimize the generated digitally reconstructed radiograph of the predicted anatomical target pose. In another example, random sampling of volume densities may be utilized to optimize the generated digitally reconstructed radiograph of the predicted anatomical target pose.

[0007] In some aspects, a method for generating digitally reconstructed radiographs disclosed herein may include the steps of selecting a predicted anatomical target pose of an individual, obtaining gradient-based registration images that include known anatomical poses from a group of individuals, estimating the predicted target anatomical target pose by rendering the gradient-based registration images by differentiable X-ray projection, optimizing the rendered gradient-based registration images, and generating an optimized digitally reconstructed radiograph of the predicted target pose. In some examples, the rendered gradient-based registration images may be optimized via neural tuning of computed tomography. In other examples, the rendered gradient-based registration images may be optimized via masked neural radiance fields.

[0008] In some examples, the received gradient-based registration images may be optimized via neural tuning of computed tomography. In another example, the received gradient-based registration images may be optimized via masked neural radiance fields. In other examples, the predicted anatomical target pose may be selected from representative X-ray images of the individual at poses near -90°, -45°, 0°, 45°, and 90°. In still other examples, the methods disclosed herein may include a machine learning model configured to optimize the rendered gradient-based registration images. In another example, the machine learning model may be configured to utilize even sampling of volume densities to optimize the generated digitally reconstructed radiograph of the predicted target pose. In yet other examples, the machine learning model may be configured to utilize random sampling of volume densities to optimize the generated digitally reconstructed radiograph of the predicted target pose.

[0009] In other aspects, a system for generating digitally reconstructed radiographs disclosed herein may include an imaging device configured to an generate anatomical image from an individual, a machine learning model configured to optimize the generated anatomical image from the individual, and a processing device configured to generate a digitally reconstructed radiograph. In some examples, the digitally reconstructed radiograph may include a predicted target anatomical pose of the individual, and the predicted target anatomical pose may be generated from the optimized anatomical image of the individual.

[0010] In other examples, the digitally reconstructed radiograph may be generated using gradient-based registration images optimized via neural tuning of computed tomography. In some examples, the digitally reconstructed radiograph may be generated using gradient-based registration images optimized via masked neural radiance fields. In still other examples, the digitally reconstructed radiograph may be generated using gradient-based registration images optimized via neural tuning of computed tomography and via masked neural radiance fields. In yet another example, the machine learning model may be configured to utilize even sampling of volume densities to optimize the generated anatomical image. In another example, the machine learning model may be configured to utilize random sampling of volume densities to optimize the generated anatomical image.

[0011] These and other features, advantages, and objects of the present disclosure will be further understood and appreciated by those skilled in the art by reference to the following

specification, claims, and appended drawings, where various embodiments of the design illustrate how concepts of this disclosure may be used.

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] A more complete understanding of features described herein and advantages thereof may be acquired by referring to the following description in consideration of the accompanying drawings, in which like reference numbers indicate like features. The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0013] FIG. 1 illustrates pose estimation by gradient-based image registration optimization in Differentiable X-ray Projection (DiffXP) as disclosed herein.

[0014] FIG. 2 illustrates DiffXP consisting of an X-ray source and an image intensifier as disclosed herein.

[0015] FIG. 3 depicts a typical setup of a movable X-ray. The imaged geometry on the intensifier is determined by two parameters, source-isocenter distance (SOD) and the source-intensifier distance (SID; also called the focal for a camera). The SOD and SID can be set according to the volume and image sizes, respectively.

[0016] FIG. 4 depicts the configuration of a movable X-ray, where a source moves relative to a fixed object. The positional/angular CS is set at the isocenter. The blue arrows indicate the positive directions of 6DoFs.

[0017] FIG. 5 depicts the configuration of a fixed X-ray, where an object moves relative to a fixed source. The positional CS is set at the source, while the angular CS is set at the isocenter. The blue arrows indicate the positive directions of 6DoFs.

[0018] FIGS. 6A-6C illustrates cone-beam X-ray imaging, processed cone-beam X-ray sequence, and reconstructed cone-beam CT as disclosed herein.

[0019] FIG. 7 illustrates neural tuning of computed tomography (NeTCT) in DiffXP as disclosed herein.

[0020] FIG. 8 illustrates the reconstruction of a neural radiance field (NeRF) in DiffXP as disclosed herein.

[0021] FIG. 9 illustrates skull projections of a NeRF and a masked neural radiance field (mNeRF) after training in a zoom-out and left-posterior view as disclosed herein.

[0022] FIG. 10 illustrates single-view X-ray pose estimation in DiffXP.

[0023] FIG. 11 illustrates successful registrations of computerized tomography (CT), NeTCT, and mNeRF projections (green) to an X-ray image (red) at a pose capturing the lateral skull (the overlapping region between DRR and X-ray is rendered as yellow) as disclosed herein.

[0024] FIG. 12 depicts the comparison of an X-ray image with DRRs by projecting CT, NeTCT, and mNeRF using DiffXP, respectively, at the same pose of the X-ray image as disclosed herein.

[0025] FIG. 13 depicts X-ray images at five target poses that are used to evaluate the performance of pose estimation using gradient-based image registration optimization as disclosed herein.

[0026] FIG. 14 is a graphical comparison of the 3D angle errors and success rates in the registration of CT projections to X-ray images of a single patient at five different poses using the combined loss (left) and the MI loss (right), respectively as disclosed herein.

[0027] FIG. 15 graphically depicts correlation between optimal 3D angle errors and initial 3D angle errors in registration of CT projections to X-ray images at different poses using the combined loss ($\mathcal{L}_C$) and the MI loss ($\mathcal{L}_{MI}$), respectively, as disclosed herein.

[0028] FIG. 16 graphically depicts correlation between optimal 3D angle errors and optimal image losses (i.e., $\mathcal{L}_C$ and $\mathcal{L}_{MI}$, respectively) in registration of CT projections to X-ray images at different poses as disclosed herein.

[0029] FIG. 17 graphically depicts comparison of the 3D angle errors and success rates using the projections of CT, NeTCT, and mNeRF, respectively, to register skull X-ray images of five patients at five different poses as disclosed herein.

[0030] FIGS. 18A and 18B graphically depict registration results of individual patients using CT, NeTCT, and mNeRF projections, respectively. Results are from patients #1 ~ #5, upper to lower, respectively.

DETAILED DESCRIPTION

[0031] In the following description of the various embodiments, reference is made to the accompanying drawings identified above and which form a part hereof, and in which is shown by way of illustration various embodiments in which features described herein may be practiced. It is to be understood that other embodiments may be utilized and structural and functional modifications may be made without departing from the scope described herein. Various features are capable of other embodiments and of being practiced or being carried out in various different ways.

[0032] Pose estimation by image registration is a prerequisite for accurate image guidance in diagnosis or surgery. Recent progress in differentiable rendering of optically reflective materials has garnered attention, due to the state-of-the-art performance of view synthesis and pose estimation, which were built upon and applied to reconstruction and registration of medical images that are attenuations of radiolucent materials. As disclosed herein, differentiable X-ray projection (DiffXP) was developed for specifically and efficiently rendering DRRs. Based on DiffXP, two novel methods were introduced: 1) neural tuning of computed tomography (NeTCT) also referred to as shortening neural tuned CT, and 2) masked neural radiance fields (mNeRF), to improve the fidelity of DRRs, and improve implemented gradient-based optimization for single-view X-ray pose estimation.

[0033] Techniques were systematically explored to improve the robustness of pose estimation to achieve a 3D angle error less than 3°. It was discovered, surprisingly, that the success rates of pose optimization were distinctly improved when the optimization objective was defined as an image loss based on the mutual information. Furthermore, the robustness of pose estimation can be further enhanced by registration of more realistic DRRs to target X-ray images. Synthesizing DRRs, using both NeTCT and mNeRF, can achieve similar performance in pose estimation, with overall success rates of 93.2% and 95.4%, respectively. Considering the much shorter training time of NeTCT (~18 min) compared to mNeRF (~6 hours), NeTCT is an attractive option for improving the robustness of pose estimation. The novel concepts disclosed herein include: 1) introduction of NeTCT, a new method to generate realistic synthetic X-ray images from a cone beam CT reconstruction; and 2) a new system and method of modifying and optimizing NeTCT and NeRF for X-ray image pose estimation.

[0034] Accurate estimation of an intraoperative patient pose is essential for many clinical applications, such as augmented reality (e.g., presenting planned placement of implants on X-ray images captured in vivo) and robotic navigation (i.e., determining a robot's position relative to a patient in an intraoperative reference frame). As disclosed herein, a novel method, Neural Tuned CT (NeTCT) may be used for pose estimations. Compared to NeRF, NeTCT can achieve comparable high accuracy (NeTCT: 93.2% success rate vs. NeRF: 95.4% success rate for 3 deg registration accuracy), and can be trained much faster (NeTCT: 18 min vs. NeRF: 6 hours) that is beneficial for clinical deployment.

[0035] Surgical planning has been enabled on 3D anatomic models reconstructed from widely available 3D medical images, such as CT and MRI. However, the intraoperative position and orientation of a patient are unknown, so the correlation between the landmarks created on the 3D anatomic model and those inside the patient (invisible) *in vivo* cannot be conveniently established. Therefore, estimation of patient's poses is essential for accurate image guidance in widespread minimally invasive surgical procedures. The patient's pose *in vivo* can be reproduced through a 3D/2D image registration technique, in which a 3D volume (e.g., CT) representing the patient's anatomy is translated and rotated, until the synthetic 2D radiograph by projection of the 3D volume matches the actual 2D radiograph (e.g., X-ray) [1]. In the literature, the synthetic 2D radiograph is also referred as to the digitally reconstructed radiograph (DRR). Due to the complexity of image registration and a high-level requirement for accuracy in clinical settings, pose estimation still relies heavily on manual registration by human experts.

[0036] As discussed above, machine learning algorithms, models, or frameworks may be executed to train the disclosed systems and methods disclosed herein. The terms machine learning frameworks, algorithms, and models may be used interchangeably, and may generally refer to automating and improving the learning process of computers based on their experiences or historical datasets without being actually programmed, i.e., without any human assistance. The process may start with inputting good quality data (e.g., X-ray images and DRR's) and then training the machines or algorithms by building machine learning models using the data and different algorithms.

[0037] Machine learning implementations as used herein may be classified into three major categories, depending on the nature of the learning signal or response available to a learning system. The first is supervised learning. This machine learning algorithm consists of a target or outcome or dependent variable which is predicted from a given set of predictor or

independent variables. Using these datasets of variables, a function is generated that maps input variables to desired output variables. The training process continues until the model achieves a desired level of accuracy on the training data. Examples of supervised learning include: Regression, Decision Tree, Random Forest, KNN, Logistic Regression, etc. The second is unsupervised learning. In this machine learning algorithm, there is no target or outcome or dependent variable to predict or estimate. It is used for clustering a given data set into different groups. The third is semi-supervised or reinforcement learning. Using this algorithm, the machine is trained to make specific decisions. Here, the algorithm trains itself continually by using trial and error methods and feedback methods. This machine learns from past experiences and tries to capture the best possible knowledge to make accurate decisions. Markov Decision Process is an example of semi-supervised machine learning. Machine learning can also be used to recognize patterns in data.

[0038] Supervised learning has been widely attempted for automatic estimation of poses captured by single-view or multi-view X-ray images. For example, Miao et al. [2] proposed a regression approach to single-view X-ray pose estimation, in which a convolutional neural network (CNN) was trained to directly estimate six-degree-of-freedom (6DoF) transformation parameters of an implant from the residual in the appearance between DRRs and X-ray images. Liao et al. [3] tackled the problem of multi-view 3D/2D rigid registration by introducing a $POINT^2$ network consisting of a tracking module and a triangulation module. The tracking module featuring a Siamese-like architecture extracted features on DRRs similar to those on X-rays, and these features were further fed to the triangulation module for point-based registration. Zhou et al. [4] presented a transfer learning strategy for landmark-based 3D/2D image registration to estimate *in vivo* skull poses captured by dual-plane fluoroscopic images. They utilized the CNN trained on an artificial DRR-landmark dataset to detect landmarks on X-ray images experiencing DRR-style translation by a cycle-consistent generative adversarial network (GAN).

[0039] In general, there are several limitations in the implementation of supervised learning for pose estimation. First, deterministic models are used in supervised learning. No matter how accurate a model is, there are inevitably poor predictions, whereas these predictions cannot be changed once the model is trained. Moreover, supervised learning is greedy for data to generalize. For a generic application, it is necessary to collect a large dataset of medical images and ground-truth labels involving a population for model development. However, except the expense, medical images are commonly less available due to the risk

of high radiation exposure. Although DRRs are an effective way for data augmentation, sufficient ground-truth images are needed to eliminate the style difference.

[0040] Reinforcement learning (RL) is a way of solving the problem of deterministic models, as it mimics how an intelligent agent takes actions in manual image registration. In RL, the value function estimating the expected cumulative reward during a Markov decision process (MDP) is maximized, by optimizing the policy function that determines actions of the agent according to the current state of the environment [5]. In particular, deep CNNs have been introduced to represent the policy function. Recently, Shao et al. [6] have developed a deep RL framework for 6DoF pose estimation of real-world objects captured by photos with 2D annotation. Impressive performance of the deep RL framework was achieved by a delicate reward definition and a composite reinforced optimization method for efficient and effective policy training.

[0041] As one of recent advances, Mildenhall et al. [7] proposed neural radiance fields (NeRF) that offered a new avenue to represent a real-world 3D scene using a multi-layer perceptron (MLP) and synthesize high-fidelity RGB images. Particularly, they realized differentiable rendering of a 3D field with densities and view-dependent colors to generate 2D photorealistic images with pixel colors, by fully vectorizing the rendering process in a machine learning framework (TensorFlow). Based on the differentiability of NeRF, Lin et al. [8] proposed "inverting" NeRF, a gradient-based optimization framework that performs mesh-free 6DOF pose estimation given a RGB image. Compared to RL which estimates the value (i.e., the expected cumulative reward) during pose refinement assumed as an MDP, the gradient information in terms of an image with respect to a pose can much more accurately guide image registration.

[0042] Furthermore, it was observed that NeRF has potential to directly generate realistic DRRs simulating X-ray images, without the additional processing to eliminate the difference in their styles via domain adaptation, like GAN [9]. Previous work has demonstrated that a large dataset of DRRs and X-ray images is warranted to train a GAN for the style translation, but the performance still appeared to be less robust due to mode collapse [4].

[0043] Many previous works based on NeRF have achieved state-of-the-art performance on the rendering of optically reflective materials [10], [11]. However, whether and how differentiable rendering can be leveraged to synthesize DRRs simulating the attenuations

of radiolucent materials and to improve medical image reconstruction/registration have not been thoroughly explored. As disclosed herein, an in-depth investigation was conducted with a focus on single-view cone-beam X-ray pose estimation. First, a differentiable X-ray projection (DiffXP) was developed for specifically and efficiently rendering DRRs. Based on DiffXP, neural tuning of CT was then introduced and mNeRF to improve the fidelity of the DRRs. Furthermore, DRR/X-ray image registration using gradient-based optimization was also implemented in DiffXP for pose estimation. X-ray images from a cone-beam CT scanner were procured and single-view X-ray pose estimation was performed. To improve the robustness of pose estimation, 1) the choice of image losses as the optimization objective, and 2) the fidelity of DRRs rendered by CT, NeTCT, or mNeRF were examined.

[0044] Differentiable X-ray Projection (DiffXP)

[0045] A differentiable rendering algorithm has been implemented in NeRF, which is used to reproduce a real-world 3D scene and synthesize photorealistic views [7]. Through vectorization of the rendering computation, automatic differentiation was realized in machine learning frameworks based on backpropagation. However, in medical image analysis, X-ray images are often simulated by DRRs, which are generated by projection of widely available 3D medical image volumes (e.g., CT and MRI). Different from a NeRF that is a continuous field, theses 3D volumes are discretized, so they cannot be directly used in the rendering algorithm of NeRF to synthesize DRRs. Furthermore, as typical 2D and 3D medical images only have density/intensity values without RGB colors, a simplification of the rendering algorithm in NeRF [7] by ignoring RGB colors (thus, eliminating the dependence of the rendering on ray directions that affects image colors) can significantly enhance computational efficiency. Therefore, DiffXP for specifically rendering DRRs was developed. Discretized volumes or continuous fields, representative of 3D anatomic structures, can be imported into DiffXP. As shown in FIG. 1, in optimization, estimation of the target pose is iteratively performed by minimizing the difference between the target X-ray image and the projection (i.e., DRR) of a 3D volume/field (e.g., CT, NeTCT, or mNeRF) representing a 3D anatomic structure. FIG. 1 also depicts the relationships of NeTCT and mNeRF with respect to CT. NeTCT and mNeRF are conditional on CT densities and a 3D mask of the anatomic region segmented from CT, respectively.

[0046] As shown in FIG. 2, a movable cone-beam X-ray (such as C-arm fluoroscopy) was adopted in which a 3D volume/field representing an anatomic structure was fixed at the X-ray isocenter (i.e., the rotation center of the movable X-ray), while the X-ray source

translates and rotates relative to the X-ray isocenter. Like NeRF, a normalized device coordinate (NDC) system, "XYZ" is set at the X-ray isocenter (O), whereby the original viewing frustum is mapped to the cube $[-1,1]^3$ [7]. However, 3D medical images such as CT are typically created based on the actual space coordinate, with a source-isocenter distance (SOD) and a source- intensifier distance (SID) preset by the CT scanner. Hence, to input a CT volume into DiffXP, dimensions of the CT volume are expanded to its maximum dimension, and then the cubic volume is normalized to $[-1, 1]^3$ (note that the unit of the NDC is a half of the actual volume size). Correspondingly, the SOD and SID in the actual coordinate need to be adjusted to those in the NDC, such that the resulting projections are identical regardless of adopted coordinates (setting X-ray parameters is described below and shown in FIGS. 3-5. Additionally, DiffXP also normalizes the densities of 3D volumes/fields and the intensities of rendered DRRs to [0, 1]. Both the normalizations for volume dimensions and densities/intensities are in order for efficient training in later machine learning implementations.

[0047] Consider a single ray ($\boldsymbol{r} = \boldsymbol{o} + t\boldsymbol{d}$) which emits from the X-ray source ($\boldsymbol{o}$) along an arbitrary direction ($\boldsymbol{d}$), as shown in FIG. 1. The resulting image intensity (I) at the pixel on the image intensifier where the ray arrives can be formulated as a numerical integration of all intensities throughout this ray in the 3D space:

$$I(\boldsymbol{r}) = \sum_{i=1}^{N_s} w_i \sigma_i \tag{1}$$

where w_i are weights on each sampled volume density, σ_i. N_s is the sample size per ray. To save computational cost, the intensities (σ_i, $i = 1, 2, \ldots, N_s$) can be sampled only within the 3D volume by setting a near and far bounds (which were set at z = 1 and z = -1, respectively, in this work), as shown in FIG. 1. In particular, to sample densities from a discretized volume like CT, a differentiable sampling module was developed, in which coordinate transformation from the NDC to the image coordinate and linear interpolation of volume densities were fully vectorized in the machine learning framework (TensorFlow). Moreover, DiffXP offers two options for sampling of volume densities [7], including even sampling (for efficiency in later gradient-based pose optimization and random sampling (for jittering in later training of NeTCT and mNeRF). To do that, the ray

between the near and far bounds is evenly sliced into N_s segments along the z-axis. For even sampling, the volume densities at the starting points of these segments are sampled; for random sampling, the volume densities within these segments are randomly selected.

[0048] Furthermore, the weight on each sampled volume density in Eq. 1 above can be written as $w_i = C_i T_i$ comprising a coefficient C_i and a nested numerical integration, T_i:

$$C_i = 1 - \exp(-\sigma_i \delta_i) \tag{2}$$

$$T_i = \exp\left(- \sum_{j=1}^{i-1} \sigma_j \delta_j \right) \tag{3}$$

where $\delta_i = t_{i+1} - t_i$ is the spacing between adjacent sampled densities (t_i is the distance of the sampled density, σ_i to the source along the ray). It is worth noting that δ_i is a constant for even sampling but varies for random sampling. Both the equations indicate that the weight, w_i is related to $\sigma_i \delta_i$ in terms of a sublinear and bounded function which penalizes high $\sigma_i \delta_i$ values (Eq. 2), meanwhile w_i decays as the distance of the sampled density to the source increases (Eq. 3).

[0049] To generate all pixel intensities of a DRR, the numerical integration (Eq. 1) needs to be performed for a total of L × W rays (L and W are the numbers of pixels along the length and width of the DRR, respectively). In this work, DiffXP was used for projection and reconstruction of 3D volumes/fields, including CT, NeTCT, and mNeRF, respectively, as well as for image registration via gradient-based pose optimization. As noted in NeRF, the gradient computation of the differentiable rendering through backpropagation is prohibitively expensive [8]. For all the implementations of DiffXP without excessively sacrificing computational efficiency, the DRR size was set to 128 × 128 and a sample size of N_s = 128 was chosen.

[0050] **Setting of X-ray Parameters and Configuration**

[0051] As shown in FIG. 3, a typical movable X-ray (e.g., C-arm fluoroscopy), the X-ray source moves relative to a coordinate system (CS) established at the isocenter (i.e., the rotation center of the X-ray). The geometry imaged on the intensifier depends on two parameters, the source-isocenter distance (SOD) and the source-intensifier distance (SID;

also called the focal for a camera). The SOD and SID can be set according to the volume and image sizes, respectively.

[0052] For a cubic volume that is scaled to different sizes, l_1 and l_2 (e.g., those in the actual coordinate and NDC, respectively), the corresponding SODs, z_1 and z_2, should be set, such that the same geometry is imaged:

$$\frac{z_1}{z_2} = \frac{l_1}{l_2} \tag{4}$$

[0053] Furthermore, a volume could be projected onto a rectangular intensifier using different image size, s_1 and s_2. If the imaged geometry remains invariant when one image is resized to match another image, the corresponding SIDs, f_1 and f_2, should satisfy:

$$\frac{f_1}{f_2} = \frac{s_1}{s_2} \tag{5}$$

[0054] As shown in FIG. 4, in the movable X-ray, the X-ray source can be moved by first translations and then rotations with respect to a CS established at the isocenter, while the object to image remains static. In particular, the angular DoFs are represented by three intrinsic Euler angles with a sequence of "YXZ". However, sometimes, it is also desirable to position an object relative to the X-ray source (e.g., when aligning patient's CT to patient's pose in actual 3D space). In this case, a "fixed" X-ray is assumed, as shown in FIG. 5.

[0055] The 6DoF motion (first translations and then rotations) in the movable X-ray is convertible to that (first rotations and then translations) in the fixed X-ray. To ensure the same magnitude of 6DoFs in both the X-ray configurations, in the fixed X-ray, as shown in FIG. 5, a positional CS and an angular CS need to be set at the X-ray source and the isocenter, respectively; both the CSs have the coordinate directions opposite to those in the movable X-ray (i.e., the two CSs are left- handed). The angular DoFs of the object are defined as three *extrinsic* Euler angles with the same sequence of "yxz" with respect to the angular CS, while the positional DoFs of the object are measured with respect to the positional CS.

[0056] Cone-beam X-ray Imaging and CT Reconstruction

[0057] X-ray sequences were acquired of the skulls of five de-identified stroke patients from a cone-beam CT scanner, as shown in FIG. 6A. Each X-ray sequence includes 133 X-ray images with a size of 960 × 960 covering a range of principal (axial) angles from -100° to 100° in an increment of 1.5°. The acquired X-ray images sequentially underwent contrast enhancement (tone mapping and histogram equalization), intensity (grayscale) invert and scaling to [0, 255], and image resizing to 256 × 256, resulting in a processed X-ray sequence as shown in FIG. 6B. To reconstruct traditional cone-beam CT, the processed X-ray sequence was imported into a GPU-based CT reconstruction toolbox, TIGRE [12]. Within TIGRE, the SOD and SID were set to 1000 mm and 1536 mm, respectively, and the OSSART reconstruction algorithm was adopted. A reconstructed cone-beam CT with a size of 256 × 256 × 256 is shown in FIG. 6C.

[0058] Neural Tuning of CT (NeTCT)

[0059] There could be discrepancies in the style and texture between X-rays acquired from the CT scanner and DRRs by projection of the cone-beam CT using DiffXP, because the formulation and parameters of DiffXP (Eq. 1) may not be exactly coincident with those of actual X-ray imaging. It is hypothesized that tuning of CT densities can facilitate the domain alignment of DRRs to X-rays. In this study, a MLP consisting of full-connected layers was introduced to perform NeTCT in DiffXP, as illustrated in FIG. 7. Like the positional encoding in NeRF [7], an encoding of the CT densities (σ) was used as the inputs of the MLP:

$$\boldsymbol{E}(\sigma) = [\sigma \quad \sin 2^i\sigma \quad \cos 2^i\sigma \quad \cdots] \tag{6}$$

where a six-layer encoding is adopted in this study, i.e., $0 \leq i \leq 5$. The MLP outputs new densities ($\hat{\sigma}$), which are used in the numerical integration for each ray in DiffXP (Eq. 1).

[0060] To train the MLP, 133 processed X-ray images were collected and the corresponding ground-truth source poses were acquired from the CT scanner as shown in FIG. 6B. The X-ray images were downsized to 128 × 128 to match the size of DRRs generated by DiffXP. Correspondingly, the actual SOD and SID were adjusted to 7.8 and 768.0 in the NDC of DiffXP, respectively. During MLP training, an X-ray image (I_{Xray}) and the corresponding ground-truth source pose ($\boldsymbol{p}_t$) are randomly sampled from the

processed X-ray sequence. The optimization of the MLP weights ($\boldsymbol{W}$) can be formulated as:

$$\boldsymbol{W}_* = \underset{\boldsymbol{W}}{\operatorname{argmin}} \mathcal{L}\left(I_{DRR}(\boldsymbol{p}_t|\boldsymbol{W}),\ I_{Xray}(\boldsymbol{p}_t)\right) \tag{7}$$

where $I_{DRR}(\boldsymbol{p}_t|\boldsymbol{W})$ is the DRR generated by projection of NeTCT depending on the MLP weights ($\boldsymbol{W}$), given a ground-truth source pose ($\boldsymbol{p}_t$) in DiffXP. Corresponding to the X-rays captured in the CT scanner, the pose ($\boldsymbol{p}_t$) is one of the principal angles in an interval of 1.5° from -100° to 100° as shown in FIG. 6B. $\mathcal{L}$ is a combined image loss function, quantifying the difference between a pair of I_{DRR} and I_{Xray}; it includes a mean square error (MSE) loss, a focal frequency loss [13], and a structural similarity (SSIM) loss [14]. It is worth noting that Eq. 7 can be efficiently solved in gradient-based optimization, because the computation of the gradient, $\frac{\partial I_{DRR}}{\partial \boldsymbol{W}}$ is enabled by the differentiability of DiffXP.

[0061] Masked Neural Radiance Field (mNeRF)

[0062] Different from a discretized CT volume, an NeRF is an unconstrained and continuous radiance field. As shown in FIG. 8, the NeRF is parameterized by an MLP mapping from NeRF coordinates (input) to NeRF densities (output) [7], and the reconstruction of an NeRF starts from an MLP with randomized weights. As the representation of a NeRF is more complex than the tuning task in NeTCT, a more sophisticated MLP was introduced with more hidden layers and neurons as shown in FIG. 8. To predict the density at a position, the corresponding coordinates are fed to the MLP by positional encoding [7]:

$$\begin{aligned}&\boldsymbol{E}(x,y,z)\\&=[x \quad y \quad z \quad \sin 2^i x \quad \sin 2^i y \quad \sin 2^i z \quad \cos 2^i x \quad \cos 2^i y \quad \cos 2^i z \quad \cdots]\end{aligned} \tag{8}$$

where a six-layer encoding is adopted in this study, i.e., $0 \le i \le 5$. Like that in NeTCT, the densities ($\hat{\sigma}$) predicted by the MLP are used in the numerical integration for each ray in DiffXP (see Eq. 1). For training the MLP of the NeRF, the protocol, loss function, and optimization problem are completely same as those in NeTCT (see Eq. 7). In particular,

compared to the MSE loss that was solely used to train NeRF [7], incorporating the SSIM loss can help converge in the training of the medical NeRF (without RGB values).

[0063] A main problem of the unconstrained NeRF is that artifacts would occur surrounding the reconstructed anatomic structure, as shown in FIG. 9. These artifacts can compensate for the difference between DRRs and X-rays, causing overfitting to the X-rays in training but incorrect rendering at new poses. The potential reason is that the CT scanner can only capture a sequence of 133 X-rays covering a range of the principal angles from -100° to 100° (i.e., the pose trajectory was a semi-circle). In contrast, in NeRF, the input natural images and ground-truth poses covered a semi-sphere of the 3D scene [7]. To eliminate artifacts due to the limited range of X-rays and poses in training, mNeRF was introduced, in which the 3D region of the skull manually segmented from cone-beam CT (processed using 3D Slicer [15] with a slight dilation of 3 mm applied) as shown in FIG. 6C, is used as a 3D mask to constrain the NeRF, such that only the densities within the anatomic structure are generated.

[0064] For training of mNeRF, a differentiable masking module was developed in the machine learning framework (TensorFlow) to ensure that all operations can be differentiated. In the initial implementation of this module, the products of the densities predicted by the MLP and the corresponding mask values (0: anatomy, 1: background) at all sampled positions were fed to DiffXP to render a DRR by numerical integrations. Because the 3D CT mask is discretized, the mask values at the sampled positions need to be also interpolated using the differentiable sampling module. However, due to these additionally introduced operations, the backpropagation during training was too expensive, causing a machine learning framework crash. In fact, the densities outside the CT mask are useless for DiffXP, without a need to be predicted by the MLP. Therefore, in the viable differentiable masking module, the sampled positions of the background were firstly filtered out, such that only the sampled positions of the anatomic region were fed to the MLP to predict densities that were numerically integrated in DiffXP.

[0065] **Pose Estimation by DRR/X-ray Image Registration**

[0066] Intuitively, pose estimation of an X-ray can be iteratively achieved by an DRR/X-ray image registration process, in which the image difference between the rendered DRR and the X-ray image is minimized by adjusting the pose of the DRR source. FIG. 10 illustrates the single-view X-ray pose estimation in DiffXP, in which a 3D volume/field is

used to represent the 3D anatomic structure. The source pose consists of six-degree-of-freedom (6DoF) components, including three angular and three positional DoFs with respect to the NDC system established at the isocenter of the X-ray. It is worth noting that the relative motion between the source and the object can be estimated based on an assumption of either a movable or a "fixed" X-ray. In this study, the source poses were estimated of a movable X-ray, assuming that the imaged object was fixed, since the motion of the source can be more concisely described. However, this approach of source pose estimation is equivalently applicable to estimating object poses while fixing the X-ray source (i.e., a fixed X-ray). See the above description of FIGS. 4 and 5 describing interpretation of the equivalence of source vs. object movements, as well as the coordinate transformation of 6DoFs between these two X-ray setups.

[0067] Given an X-ray image (I_{Xray}) with an unknown pose, the estimation of a 6DoF pose ($\boldsymbol{p}$) can be described as an optimization problem:

$$\boldsymbol{p}_* = \underset{\boldsymbol{p}}{\operatorname{argmin}} \mathcal{L}\left(I_{DRR}(\boldsymbol{p}),\ I_{Xray}\right) \tag{9}$$

where $I_{DRR}(\boldsymbol{p})$ is the DRR generated by projection of NeTCT at a source pose ($\boldsymbol{p}$) in DiffXP. $\boldsymbol{p} = [\boldsymbol{\theta}; \boldsymbol{\chi}]$ consists of three angular DoFs, $\boldsymbol{\theta} = [\beta\ \alpha\ \gamma]$ and three positional DoFs, $\boldsymbol{\chi} = [x\ y\ z]$. In particular, $\boldsymbol{\theta}$ is defined as three intrinsic Euler angles with a sequence of "YXZ" [16]. $\mathcal{L}$ is the image loss function used to quantify the difference between a pair of I_{DRR} and I_{Xray}; here, a mutual information (MI) loss was adopted [17], [18] for the pose optimization. Due to the differentiability of DiffXP, the gradient, $\frac{\partial I_{DRR}}{\partial \boldsymbol{p}}$ can be computed by backpropagation in machine learning frameworks, so Eq. 9 can be efficiently solved in gradient-based optimization.

[0068] Because X-ray imaging is radiolucent and the skull is symmetric, the two X-ray images with a difference of 180° around the principal (axial) axis are almost identical (e.g., it is often hard to distinguish an anterior radiograph from a posterior radiograph). Therefore, for estimating X-ray poses, it is necessary to provide an initial guess for the pose optimization (Eq. 9). Here, the initial pose ($\boldsymbol{p}_0$) is assigned randomly jittered angular and positional DoFs away from the target pose ($\boldsymbol{p}_t$):

$$\boldsymbol{p}_0 = \boldsymbol{p}_t + \Delta\boldsymbol{p} \tag{10}$$

where $\boldsymbol{p}_0 = [\boldsymbol{\theta}_0; \boldsymbol{\chi}_0]$ and $\boldsymbol{p}_t = [\boldsymbol{\theta}_t; \boldsymbol{\chi}_t]$. $\Delta\boldsymbol{p}$ is a vector consisting of six uniformly random numbers between $\min(\Delta\boldsymbol{p})$ = [-30, -30, -30; -0.2, -0.2, -0.2] and $\max(\Delta\boldsymbol{p})$ = [30, 30, 30; 0.2, 0.2, 0.2]. Note that the units of angular and positional DoFs are degrees and a half of the volume size (in the NDC space), respectively. Eq. 10 simulates a rough registration for a target X-ray image by humans making a little effort, meaning that the initial pose is distinctly different from the target pose, as shown in FIG. 11. In addition, it can be noted that Eq. 10 results in a highly skewed variable space in terms of the optimization variables, $\boldsymbol{p}$, which is bounded by $\boldsymbol{lb} = \boldsymbol{p}_t + \min(\Delta\boldsymbol{p})$ and $\boldsymbol{ub} = \boldsymbol{p}_t + \max(\Delta\boldsymbol{p})$. Therefore, in the implementation, the optimization variables, $\boldsymbol{p}$ have been normalized to $\overline{\boldsymbol{p}}$ in a regular variable space bounded by $\overline{\boldsymbol{lb}} = \mathbf{0}_{1\times6}$ and $\overline{\boldsymbol{ub}} = \mathbf{1}_{1\times6}$ for efficient optimization [19].

[0069] Angular estimation is much more challenging than positional estimation, especially for the two out-of-plane rotations that are harder to determine by single-view image registration [20]. As disclosed herein, the angular and positional DoFs were coupled in a single optimization within a normalized variable space for pose estimation (Eq. 9). Overall, the positional DoFs can converge to their ground-truth DoFs, once the errors in the angular DoFs are minimized, as shown in FIG. 11. Therefore, for simplicity in subsequent analysis and presentation of experimental results, only the errors in angular DoFs are reported, which are measured by introducing a 3D angle error ($\mathcal{E}_{\boldsymbol{\theta}}$). Given the angular DoFs, $\boldsymbol{\theta}_*$ in the optimal pose and those, $\boldsymbol{\theta}_t$ in the ground-truth pose, a 3D angle based on the axis-angle representation [21] can be calculated:

$$\mathcal{E}_{\boldsymbol{\theta}} = \arccos\frac{\mathrm{trace}(\boldsymbol{R}_* \cdot \boldsymbol{R}_t^T) - 1}{2} \tag{11}$$

where $\boldsymbol{R}_*$ and $\boldsymbol{R}_t \in \mathbb{R}_{3\times3}$ are the rotation matrices calculated from the Euler angles $\boldsymbol{\theta}_*$ and $\boldsymbol{\theta}_t$, respectively. The superscript T represents transpose of a matrix. As disclosed herein, a successful image registration / pose estimation was defined with $\mathcal{E}_{\boldsymbol{\theta}} < 3°$, that is anticipated by surgeons in image-guided interventional neurosurgery.

EXPERIMENTS AND RESULTS

[0070] Reconstruction of CT, NeTCT, and mNeRF

[0071] X-ray sequences of the skulls of five stroke patients were collected from a cone-beam CT scanner. Each X-ray sequence included 133 X-ray image captured around the principal axis in an interval of 1.5° ranging from -100° to 100°. Using the X-ray sequences, CT, NeTCT, and mNeRF were reconstructed for the skull of each patient, respectively. The Adam optimizer with a learning rate of 5×10^{-4} was used for training both NeTCT and mNeRF. NeTCT and mNeRF were trained for 2,500 and 20,000 iterations, respectively, to ensure that the peak signal-to-noise ratio of DRRs with respect to X-rays reached a plateau. During each iteration, only an X-ray image with the ground-true pose was randomly sampled from the X-ray sequence for training. The training of mNeRF is more difficult than NeTCT, as shown by the training time that is approximately twice in each iteration and 20 times in total as much as that of NeTCT as shown in below Table 1. In general, the resulting DRRs by projection of NeTCT and mNeRF are similar, both with smaller differences in the style and texture from the X-ray image, compared to the CT-projected DRR where most intensities are too dark as shown in FIG. 12.

[0072] **Table 1**: Summary of the computational cost in reconstruction of CT, NeTCT, and mNeRF for a single patient, as well as use for pose estimation of a single X-ray image. All experiments were performed on Google Colab using a Tesla T4 GPU.

	Reconstruction	Pose Estimation (50 ~ 300 iters)
CT	4.8 min	0.68 sec/iter
NeTCT	18.3 min (2,500 iters; 0.44 sec/iter)	0.77 sec/iter
mNeRF	366.7 min (20,000 iters; 1.11 sec/iter)	1.037 sec/iter

[0073] Pose Estimation Using CT, NeTCT, and mNeRF

[0074] Five representative X-ray images at the poses near -90°, -45°, 0°, 45°, and 90° as shown in FIG. 13 were chosen from each X-ray sequence of the five patients, to estimate the target poses. For each target pose, pose estimation was performed 20 times by random initialization (Eq. 10) using CT, NeTCT, and mNeRF, respectively. During pose estimation for a single X-ray image, the best (minimal) image loss till the current iteration was recorded. The optimization terminated, when no better image loss occurred for 50 iterations (that is also the possible minimum iteration number), or a maximum iteration

number of 300 was reached. After optimization, the pose corresponding to the best image loss in the optimization is considered as the optimal solution. The Adam optimizer was used in pose optimization, with a learning rate of 0.03 and an exponential decay rate of 0.5 for the first moment estimates. Here, a large learning rate is set, as the optimization is encouraged to jump out of a local optimum. During each optimization iteration, the gradient computation took more time for mNeRF projection, compared to that for NeTCT and CT projections, as shown in Table 1.

[0075] Effects of Image Losses on Pose Estimation

[0076] In the optimization for pose estimation, the ultimate goal is to achieve the ground-truth pose, whereas only an image loss used as the optimization objective function (Eq. 9) to indirectly quantify the pose error. In general, it is desirable to adopt an image loss that can be well correlated with the pose error, as well as ensure that the optimization is less sensitive to initial guesses. Here, the performance of the MI loss ($\mathcal{L}_{MI}$) is compared with a combined loss ($\mathcal{L}_{C}$) in the pose estimation. The combined loss (different from that used in NeTCT and mNeRF in Eq. 7) is defined as a linear combination of a SSIM loss [14], a soft dice loss [22], and a L1 loss (meaning the average of all the absolute differences in pixel intensities between the DRR and the X-ray). Testing indicates that the combined loss can work well for a toy problem, where the poses of DRRs (instead of X-ray images) were estimated by DRR/DRR registration (that said, the style difference between DRRs and X-ray images was not considered).

[0077] The performance in the registration of CT projections to X-ray images of a single patient at five different poses using either the combined loss or the MI loss is shown in FIG. 14. Using the combined loss, the overall success rate was only 11%, and the registrations at the poses of -90°, - 45°, and 45° completely failed. When the MI loss was used, the overall success rate was remarkably boosted to 73%, and successful registrations occurred at all poses; in particular, all the 20 registrations at 0° were successful.

[0078] As shown in FIG. 15, there were no strong correlations between the optimal 3D angle errors and the initial 3D angle errors in registration of CT projections to X-ray images at different poses using either the combined loss or the MI loss. However, it is observed that using the MI loss, there are much more optimal 3D angle errors concentrated near the global optimum (i.e., 0°) in all registration scenarios with different target poses, regardless of the initial 3D angle errors ranging from 0° to 50°.

[0079] Furthermore, optimal 3D angle errors were more strongly correlated to optimal MI image losses than optimal combined image losses, as shown in FIG. 16. This is important evidence that the MI loss is superior to measure 3D angle errors, as it was found that the minimum in the optimal combined losses unnecessarily corresponds to the minimum in the optimal 3D angle errors (e.g., the registrations using $\mathcal{L}_C$ for poses -90° and 90°). In addition, it can be noted that most failures ($\mathcal{E}_{\boldsymbol{\theta}} \geq 3°$ according to the established definition) in the registrations using the MI loss occurred with large 3D angle errors (e.g., $\mathcal{E}_{\boldsymbol{\theta}} > 30°$). This is also a merit, as humans can more easily perceive a registration failure, such that they can decide to redo the registration.

[0080] **Effects of DRR Fidelity on Pose Estimation**

[0081] It was hypothesized that more realistic DRRs could facilitate pose estimation by DRR/X-ray image registration. This hypothesis was supported by the experimental results in registrations of CT, NeTCT, and mNeRF projections to X-ray images. As shown in FIG. 17, the registration results were pooled for the skulls of five patients at different poses, when the MI loss was adopted as the objective function in the gradient-based pose optimization. Compared to the overall success rate of 71.8% in CT registrations, NeTCT and mNeRF registrations markedly boosted the overall success rates to 93.2% and 95.4%, respectively. Although the performance of mNeRF registration was slightly better than NeTCT registration in terms of the pooled results, for individual patients, NeTCT registration was able to outperform mNeRF registration. For example, the overall success rates of NeTCT and mNeRF registrations for patient #2 were 99% vs. 96% as shown in FIGS. 18A-18B.

[0082] X-ray images of the patients at other arbitrary poses (which need to be tracked as the ground truths) were not available in this study. Therefore, the cone-beam X-ray images at five poses were selected, representative of common intraoperative poses, to evaluate the performance of pose estimation by DRR/X-ray image registration. However, the generalizability of the gradient-based pose optimization should be less affected, as optimization is performed based on the image loss during each image registration. In contrast, in supervised learning, a deterministic model is trained only once (in an optimization) using a training data with a limited number of images and labels, so there is a potential problem in overfitting to the training data and poor generalization. In addition, it appears that there are lower success rates in the registrations of the left/right lateral X-ray images of the skull (i.e., at poses -90° and 90°), using different volumes/fields for

projection as shown in FIG. 17. It may be related to the limited imaging range of the cone-beam CT scanner that can only cover the principal angles within a range of 200°, so the lateral anatomy (at the end-range poses) may be less accurately reconstructed. Future work should investigate whether it can be improved by increasing the range of X-ray imaging to cover the circumference of the skull. It is anticipated to construct a customized X-ray imaging system with the capability of tracking the ground-truth poses of an imaged subject according to the methods and systems disclosed herein.

[0083] As disclosed herein, DiffXP may be used in methods and systems for rendering of DRRs, based on which novel methods including NeTCT and mNeRF were introduced to improve the fidelity of DRRs, as well as gradient-based optimization for single-view X-ray pose estimation. A threshold of 3° for the 3D angle error was set as the testing success rate. It was found that the MI image loss is a superior objective function for pose optimization, as it can be well correlated with the pose error, as well as ensure that the optimization is less sensitive to initial guesses. Furthermore, the robustness of pose estimation was remarkably enhanced by more realistic DRRs, which can be generated by projecting either NeTCT or mNeRF. Considering that much more time is required for the training of NeTCT than mNeRF, there is potential to implement NeTCT in clinical registration scenarios.

[0084] The foregoing has been presented for purposes of example. The foregoing is not intended to be exhaustive or to limit features to the precise form disclosed. The examples discussed herein were chosen and described in order to explain principles and the nature of various examples and their practical application to enable one skilled in the art to use these and other implementations with various modifications as are suited to the particular use contemplated. The scope of this disclosure encompasses, but is not limited to, any and all combinations, subcombinations, and permutations of structure, operations, and/or other features described herein and in the accompanying drawing figures.

[0085] Although examples are described above, features and/or steps of those examples may be combined, divided, omitted, rearranged, revised, and/or augmented in any desired manner. Various alterations, modifications, and improvements will, in view of the foregoing disclosure, readily occur to those skilled in the art. Such alterations, modifications, and improvements are intended to be part of this description, though not expressly stated herein, and are intended to be within the spirit and scope of the disclosure. Accordingly, the foregoing description is by way of example only, and is not limiting.

REFERENCES

Each of the below references are incorporated herein by reference in their entireties for all purposes.

[1] M. Unberath et al., "The Impact of Machine Learning on 2D/3D Registration for Image-Guided Interventions: A Systematic Review and Perspective," Front Robot AI, vol. 8, Aug. 2021, doi: 10.3389/frobt.2021.716007.

[2] S. Miao, Z. J. Wang, Y. Zheng, and R. Liao, "Real-time 2D/3D registration via CNN regression," in 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Apr. 2016, pp. 1430–1434. doi: 10.1109/ISBI.2016.7493536.

[3] H. Liao, W.-A. Lin, J. Zhang, J. Zhang, J. Luo, and S. K. Zhou, "Multiview 2D/3D Rigid Registration via a Point-Of-Interest Network for Tracking and Triangulation," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2019, pp. 12630–12639. doi: 10.1109/CVPR.2019.01292.

[4] C. Zhou, T. Cha, Y. Peng, and G. Li, "Transfer learning from an artificial radiograph-landmark dataset for registration of the anatomic skull model to dual fluoroscopic X-ray images," Comput Biol Med, vol. 138, p. 104923, Nov. 2021, doi: 10.1016/j.compbiomed.2021.104923.

[5] R. S. Sutton and A. G. Barto, Reinforcement learning: An introduction, 2nd ed. MIT Press, 2018.

[6] J. Shao, Y. Jiang, G. Wang, Z. Li, and X. Ji, "PFRL: Pose-Free Reinforcement Learning for 6D Pose Estimation," Feb. 2021, [Online]. Available: http://arxiv.org/abs/2102.12096

[7] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "NeRF: representing scenes as neural radiance fields for view synthesis," Commun ACM, vol. 65, no. 1, pp. 99–106, Jan. 2022, doi: 10.1145/3503250.

[8] L. Yen-Chen, P. Florence, J. T. Barron, A. Rodriguez, P. Isola, and T.-Y. Lin, "iNeRF: Inverting Neural Radiance Fields for Pose Estimation," in 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Sep. 2021, pp. 1323–1330. doi: 10.1109/IROS51168.2021.9636708.

[9] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in 2017 IEEE International Conference on

Computer Vision (ICCV), Oct. 2017, vol. 2017-Octob, pp. 2242–2251. doi: 10.1109/ICCV.2017.244.

[10] R. Martin-Brualla, N. Radwan, M. S. M. Sajjadi, J. T. Barron, A. Dosovitskiy, and D. Duckworth, "NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7210–7219, Aug. 2021, [Online]. Available: http://arxiv.org/abs/2008.02268

[11] A. Yu, V. Ye, M. Tancik, and A. Kanazawa, "pixelNeRF: Neural Radiance Fields from One or Few Images," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4578–4587, Dec. 2021, [Online]. Available: http://arxiv.org/abs/2012.02190

[12] A. Biguri et al., "Arbitrarily large tomography with iterative algorithms on multiple GPUs using the TIGRE toolbox," J Parallel Distrib Comput, vol. 146, pp. 52–63, Dec. 2020, doi: 10.1016/j.jpdc.2020.07.004.

[13] Z. Tu et al., "MAXIM: Multi-Axis MLP for Image Processing," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5769–5780, 2022, doi: https://doi.org/10.48550/arXiv.2201.02973.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.819861.

[15] A. Fedorov et al., "3D Slicer as an image computing platform for the Quantitative Imaging Network.," Magn Reson Imaging, vol. 30, no. 9, pp. 1323–41, Nov. 2012, doi: 10.1016/j.mri.2012.05.001.

[16] C. Zhou, T. Cha, W. Wang, R. Guo, and G. Li, "Investigation of Alterations in the Lumbar Disc Biomechanics at the Adjacent Segments After Spinal Fusion Using a Combined In Vivo and In Silico Approach," Ann Biomed Eng, vol. 49, no. 2, pp. 601–616, Feb. 2021, doi: 10.1007/s10439-020-02588-9.

[17] A. v Dalca, J. Guttag, and M. R. Sabuncu, "Anatomical Priors in Convolutional Networks for Unsupervised Biomedical Segmentation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9290–9299, 2018, doi: https://doi.org/10.48550/arXiv.1903.03148.

[18] P. Viola and W. M. Wells III, "Alignment by Maximization of Mutual Information," Int J Comput Vis, vol. 24, no. 2, pp. 137–154, 1997, doi: 10.1023/A:1007958904918.

[19] C. Zhou and R. Willing, "Multiobjective Design Optimization of a Biconcave Mobile-Bearing Lumbar Total Artificial Disk Considering Spinal Kinematics, Facet Joint Loading, and Metal-on-Polyethylene Contact Mechanics," J Biomech Eng, vol. 142, no. 4, p. 041006, Apr. 2020, doi: 10.1115/1.4045048.

[20] J. Hanley, G. S. Mageras, J. Sun, and G. J. Kutcher, "The effects of out-of-plane rotations on two dimensional portal image registration in conformal radiotherapy of the prostate.," Int J Radiat Oncol Biol Phys, vol. 33, no. 5, pp. 1331–43, Dec. 1995, doi: 10.1016/0360-3016(95)02062-4.

[21] S. Mahendran, H. Ali, and R. Vidal, "3D Pose Regression using Convolutional Neural Networks," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2174–2182, 2017, [Online]. Available: https://shapenet.cs.stanford.edu/media/syn

[22] L. R. Dice, "Measures of the Amount of Ecologic Association Between Species," Ecology, vol. 26, no. 3, pp. 297–302, Jul. 1945, doi: 10.2307/1932409.

CLAIMS

What is claimed is:

1. A method for generating digitally reconstructed radiographs, wherein each of the foregoing steps are performed by a computing device comprising at least one processor, a communication interface, and memory, comprising:

selecting a predicted anatomical target pose of an individual;

receiving gradient-based registration images comprising known anatomical poses of a plurality of individuals;

estimating the predicted target anatomical target pose by rendering the gradient-based registration images by differentiable X-ray projection;

optimizing the rendered gradient-based registration images wherein the received gradient-based registration images are optimized via neural tuning of computed tomography, or wherein the received gradient-based registration images are optimized via masked neural radiance fields; and

generating an optimized digitally reconstructed radiograph of the predicted anatomical target pose.

2. The method of claim 1, wherein the received gradient-based registration images comprise X-ray images and digitally reconstructed radiographs, and wherein during optimizing, estimation of the predicted anatomical target pose of an individual is iteratively performed by minimizing differences between the X-ray images and the digitally reconstructed radiographs.

3. The method of claim 2, wherein the differences between the X-ray images and the digitally reconstructed radiographs are minimized by adjusting a pose of a source of the digitally reconstructed radiographs.

4. The method of claim 3, wherein the X-ray images include a contrast enhancement, an intensity invert, and a scaling and image resizing.

5. The method of claim 1, wherein the predicted anatomical target pose of an individual is selected from representative X-ray images of the individual.

6. The method of claim 1, wherein even sampling of volume densities is utilized to optimize the generated digitally reconstructed radiograph of the predicted anatomical target pose.

7. The method of claim 1, wherein random sampling of volume densities is utilized to optimize the generated digitally reconstructed radiograph of the predicted anatomical target pose.

8. A method for generating digitally reconstructed radiographs comprising:

selecting a predicted anatomical target pose of an individual;

receiving gradient-based registration images comprising known anatomical poses of a plurality of individuals;

estimating the predicted target anatomical target pose by rendering the gradient-based registration images by differentiable X-ray projection;

optimizing the rendered gradient-based registration images; and

generating an optimized digitally reconstructed radiograph of the predicted anatomical target pose of an individual.

9. The method of claim 8, wherein the received gradient-based registration images are optimized via neural tuning of computed tomography.

10. The method of claim 8, wherein the received gradient-based registration images are optimized via masked neural radiance fields.

11. The method of claim 10, wherein the predicted anatomical target pose is selected from representative X-ray images of the individual at poses near -90°, -45°, 0°, 45°, and 90°.

12. The method of claim 8, further comprising a machine learning model configured to optimize the rendered gradient-based registration images.

13. The method of claim 12, wherein the machine learning model is configured to utilize even sampling of volume densities to optimize the generated digitally reconstructed radiograph of the predicted target pose.

14. The method of claim 12, wherein the machine learning model is configured to utilize random sampling of volume densities to optimize the generated digitally reconstructed radiograph of the predicted target pose.

15. A system for generating digitally reconstructed radiographs comprising:

an imaging device configured to generate an anatomical image from an individual;

a machine learning model configured to optimize the generated anatomical image from the individual; and

a processing device configured to generate a digitally reconstructed radiograph,

wherein the digitally reconstructed radiograph comprises a predicted target anatomical pose of the individual, and wherein the predicted target anatomical pose of the individual is generated from the optimized anatomical image of the individual.

16. The system of claim 15, wherein the digitally reconstructed radiograph is generated using gradient-based registration images optimized via neural tuning of computed tomography.

17. The system of claim 15, wherein the digitally reconstructed radiograph is generated using gradient-based registration images optimized via masked neural radiance fields.

18. The system of claim 15, wherein the digitally reconstructed radiograph is generated using gradient-based registration images optimized via neural tuning of computed tomography and via masked neural radiance fields.

19. The system of claim 15, wherein the machine learning model is configured to utilize even sampling of volume densities to optimize the generated anatomical image.

20. The system of claim 15, wherein the machine learning model is configured to utilize random sampling of volume densities to optimize the generated anatomical image.

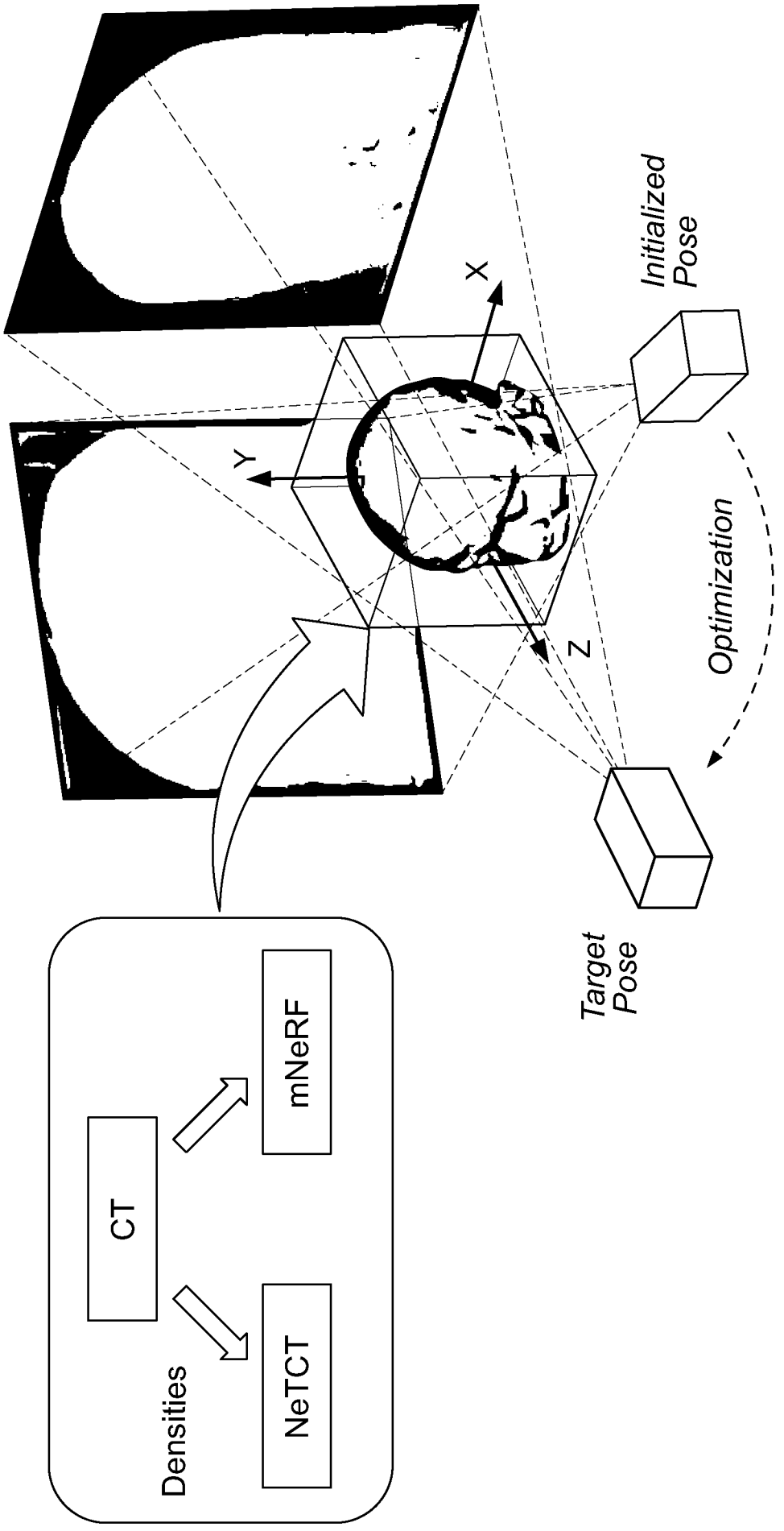

Initialized Pose
X
Y
Z
Optimization
Target Pose
CT
mNeRF
Densities
NeTCT

FIG. 1

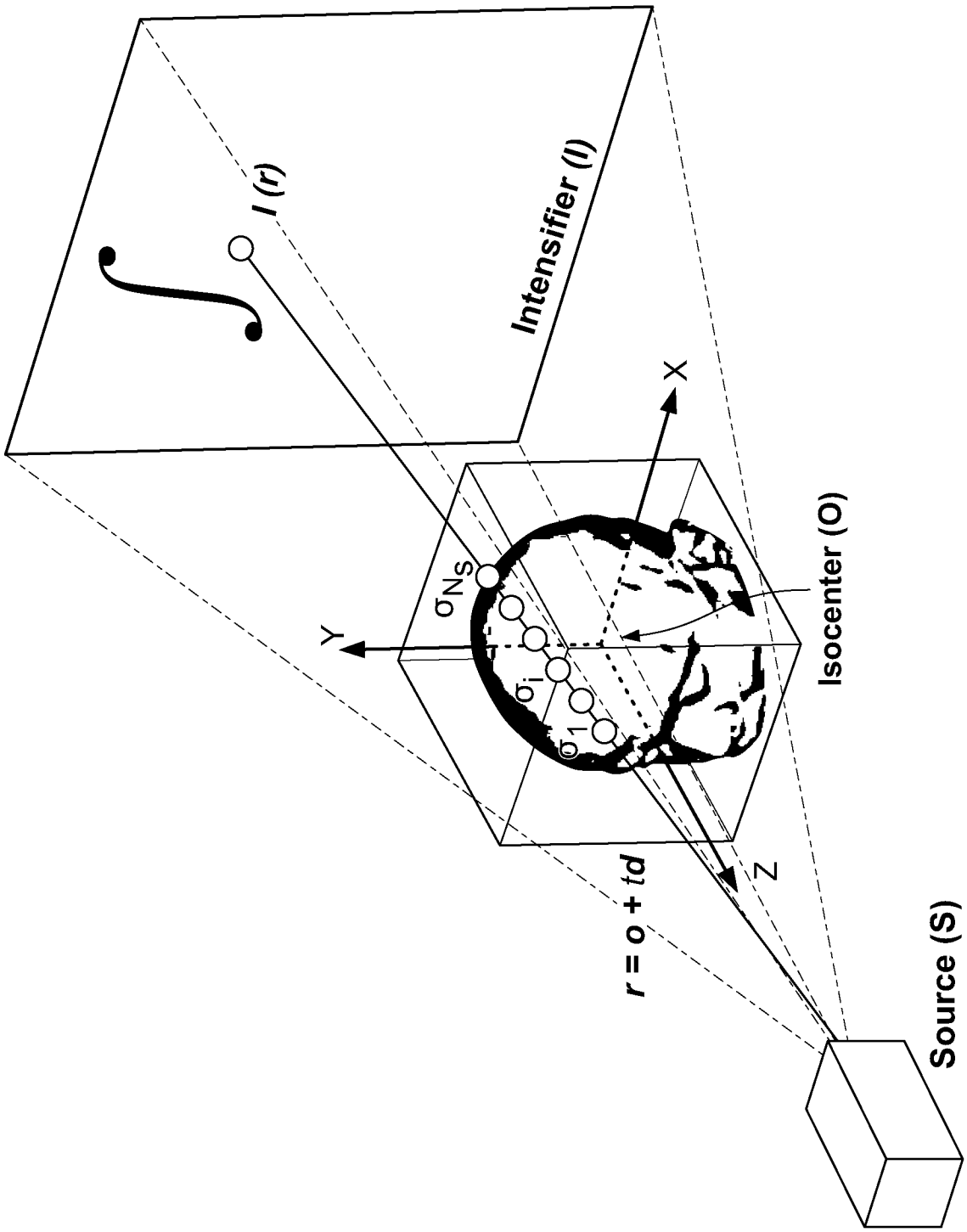

I (r)
Intensifier (I)
X
Y
Z
σ_{N_S}
σ_i
σ_1
Isocenter (O)
r = o + td
Source (S)

FIG. 2

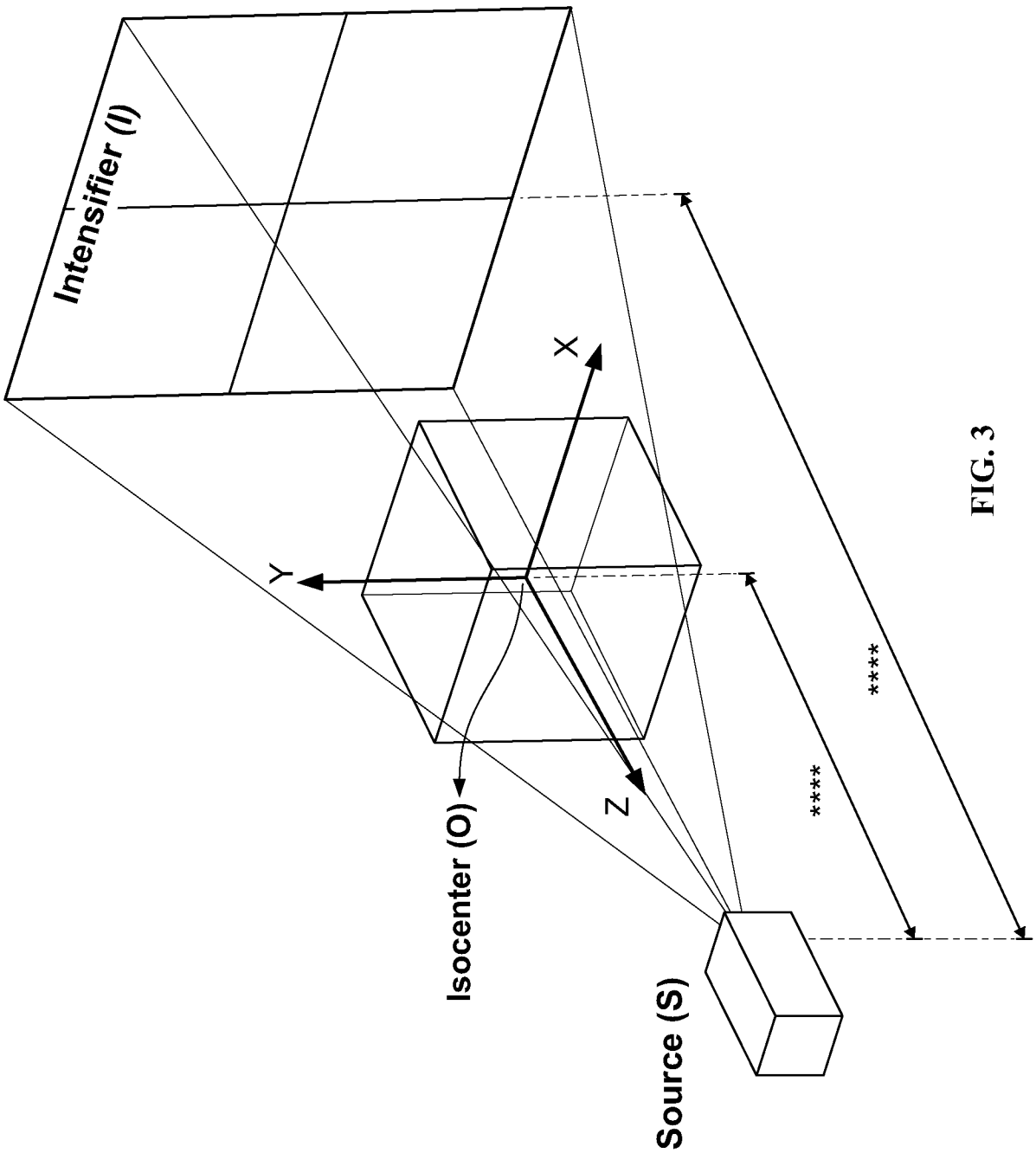


FIG. 3

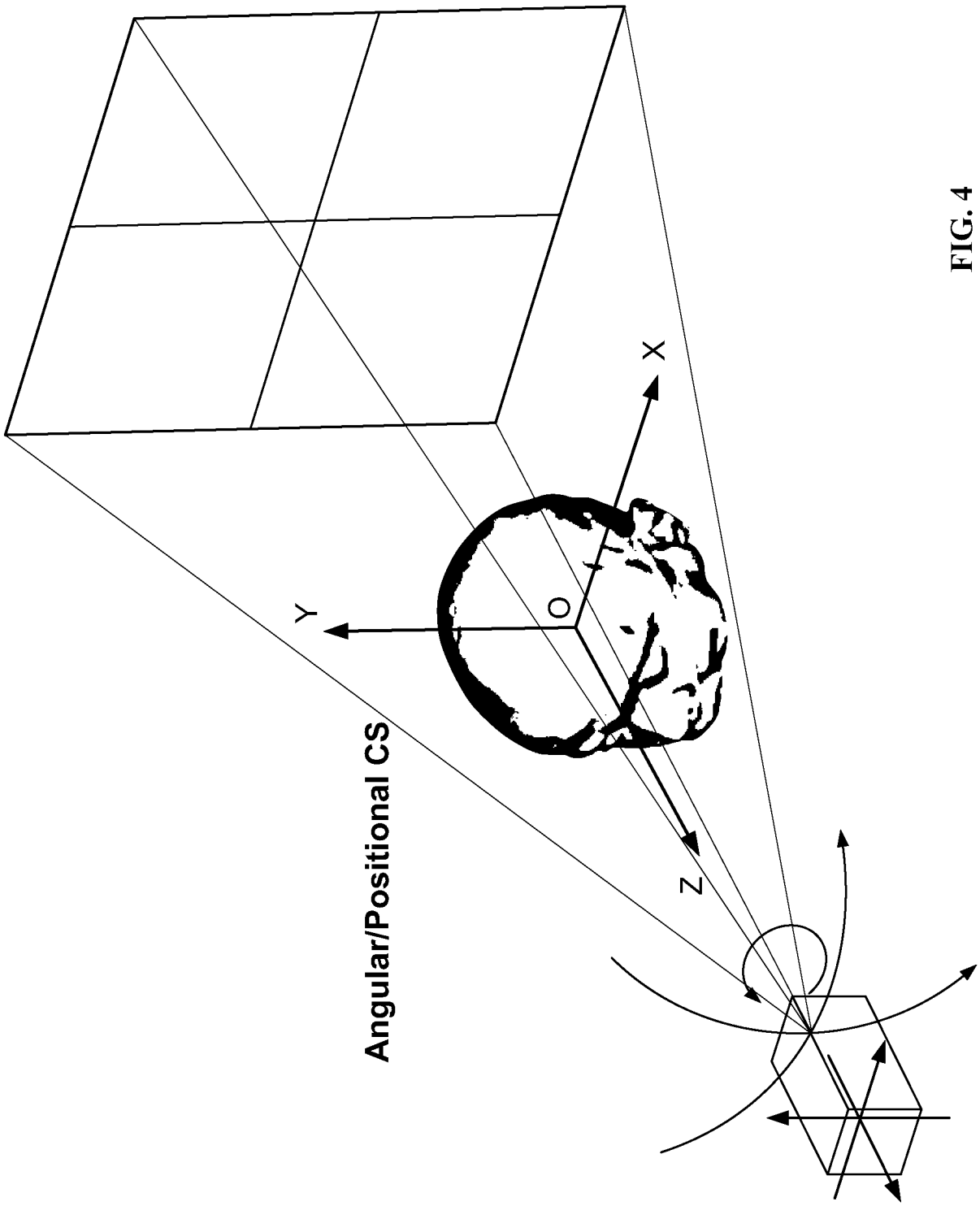


FIG. 4

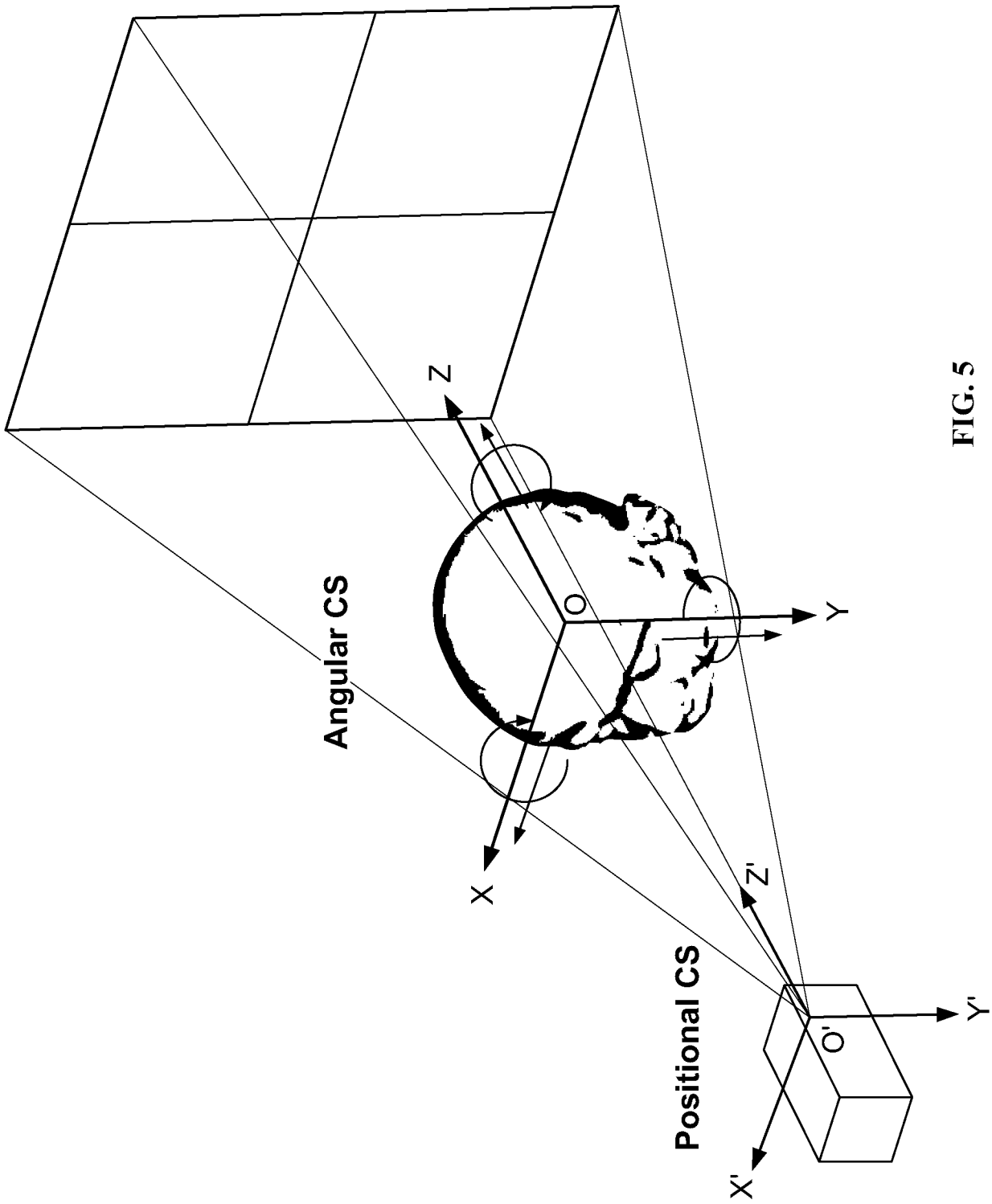


FIG. 5

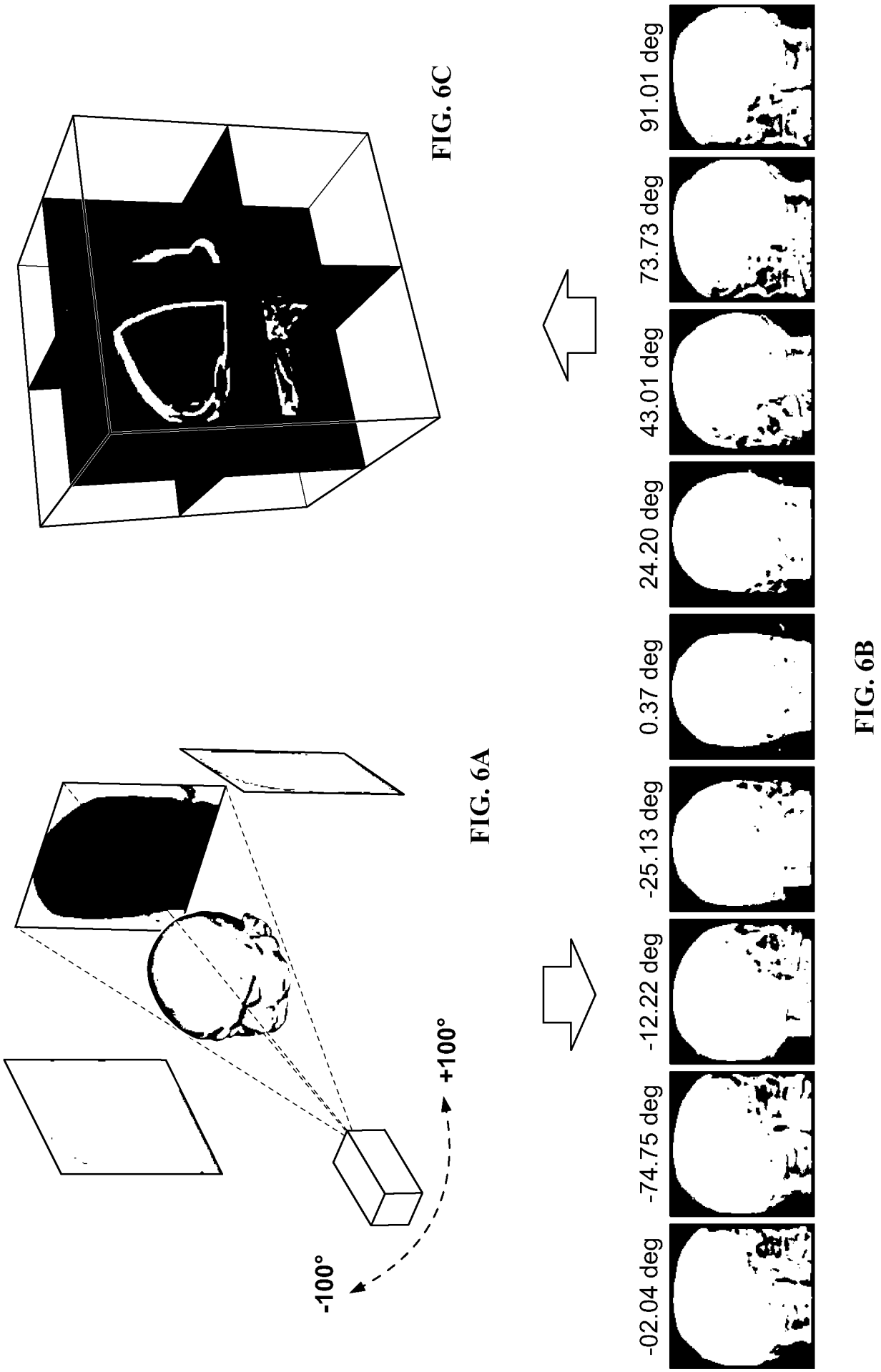


FIG. 6B

FIG. 6A

FIG. 6C

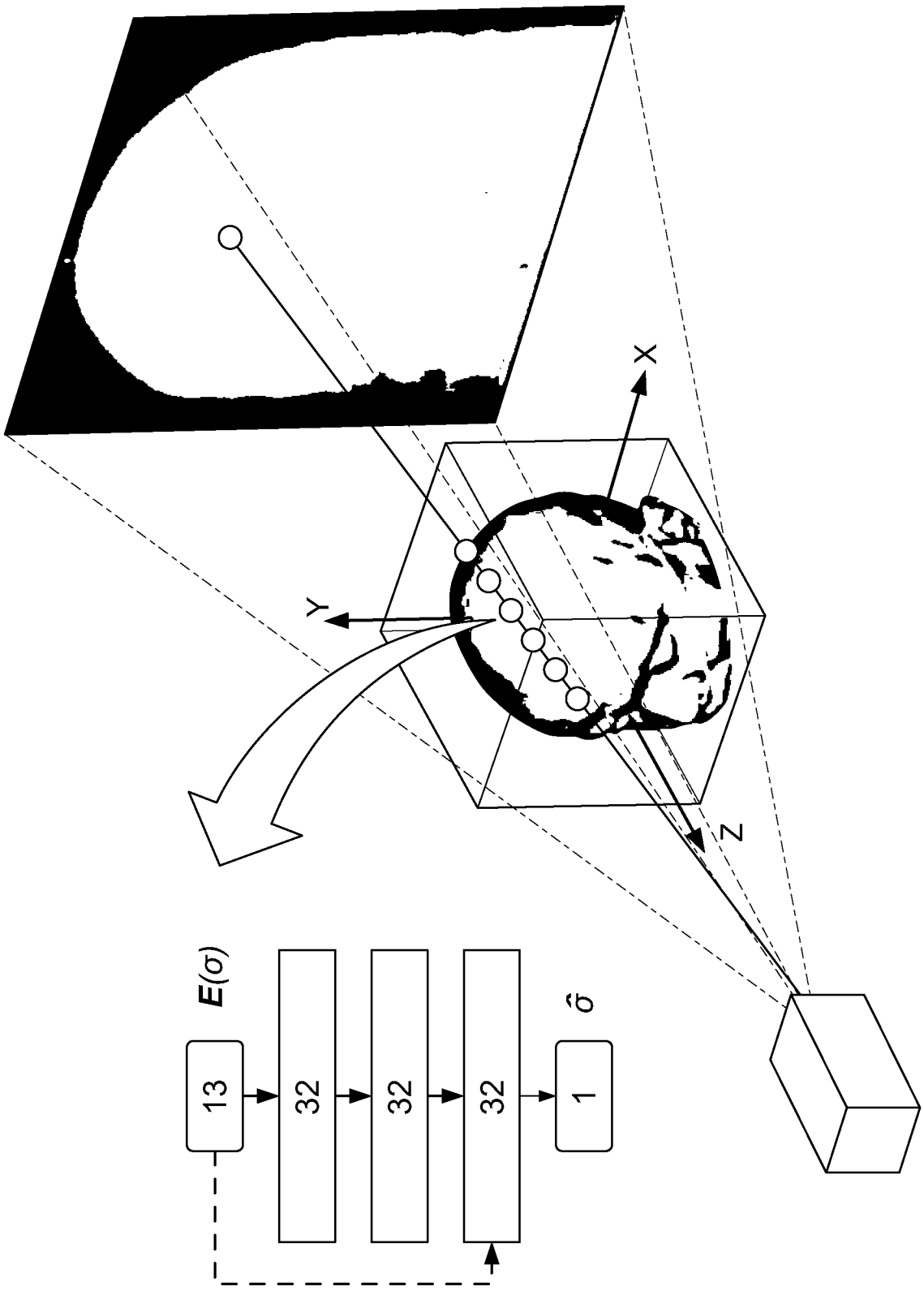


FIG. 7

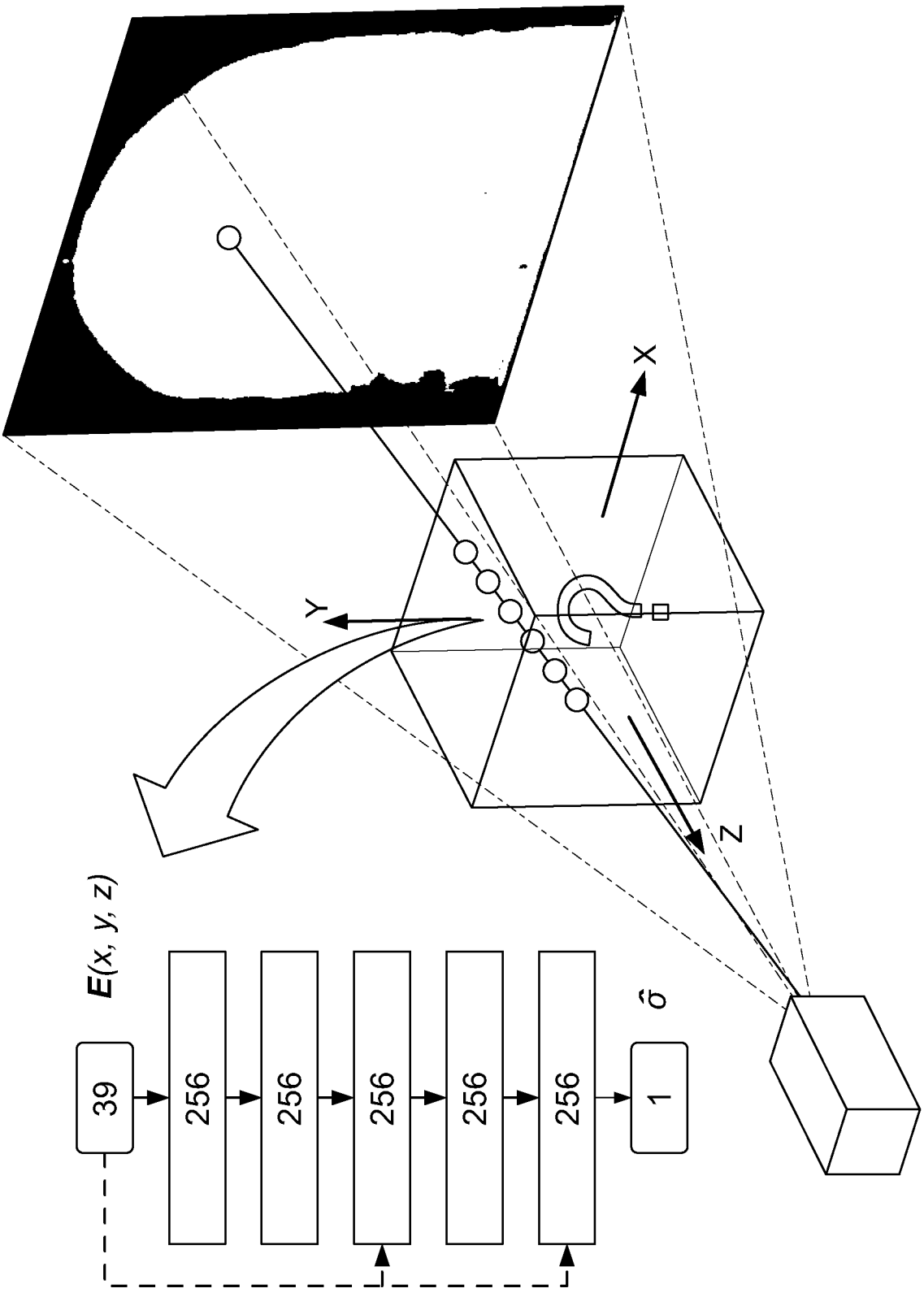


FIG. 8

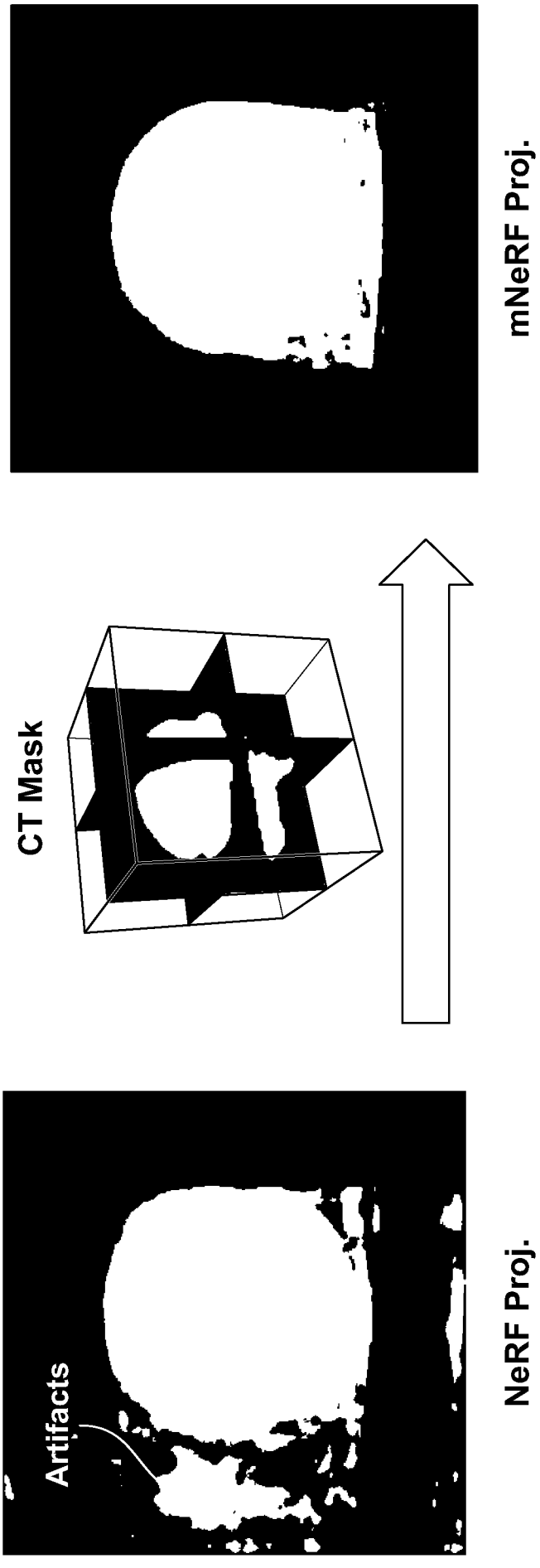


FIG. 9

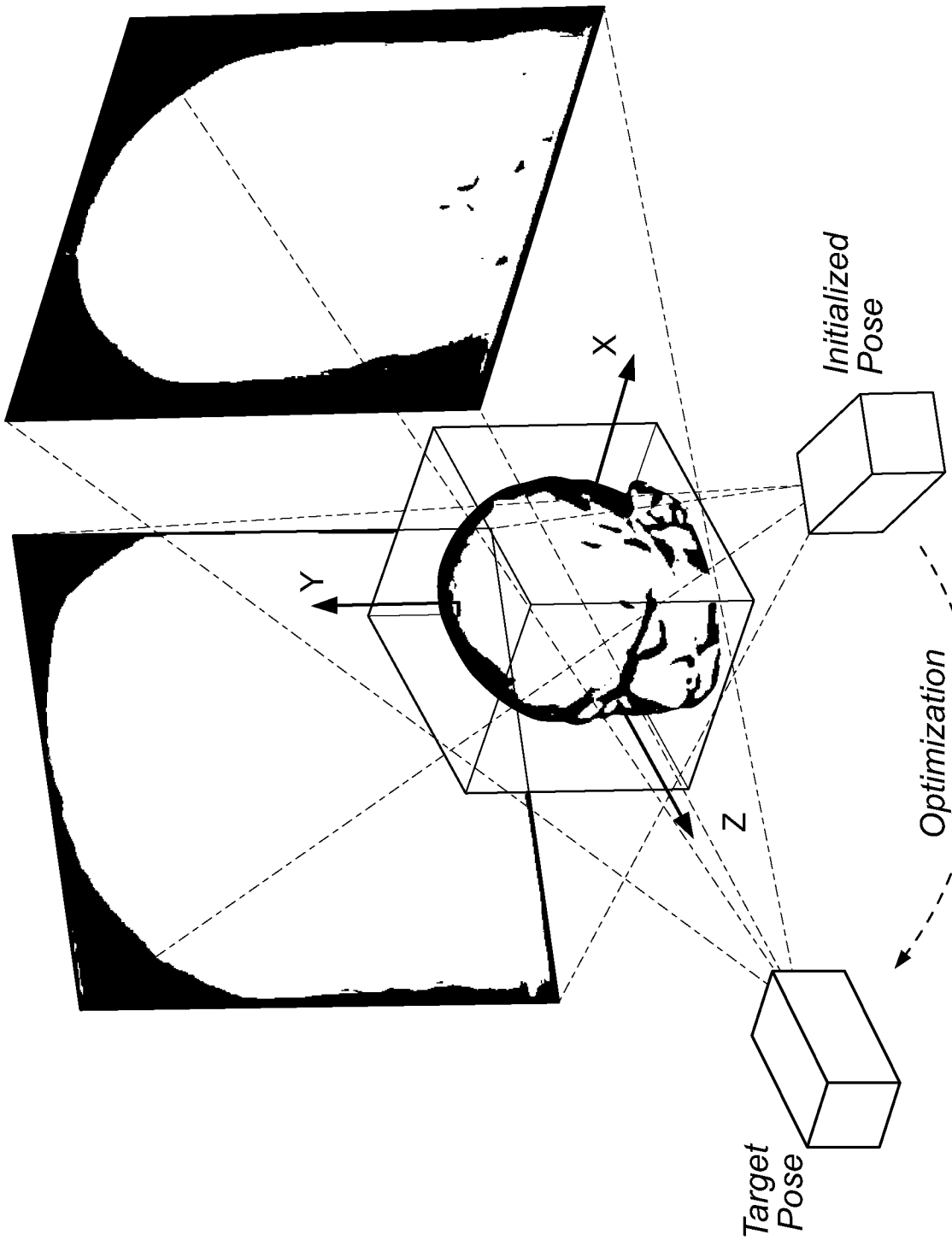


FIG. 10

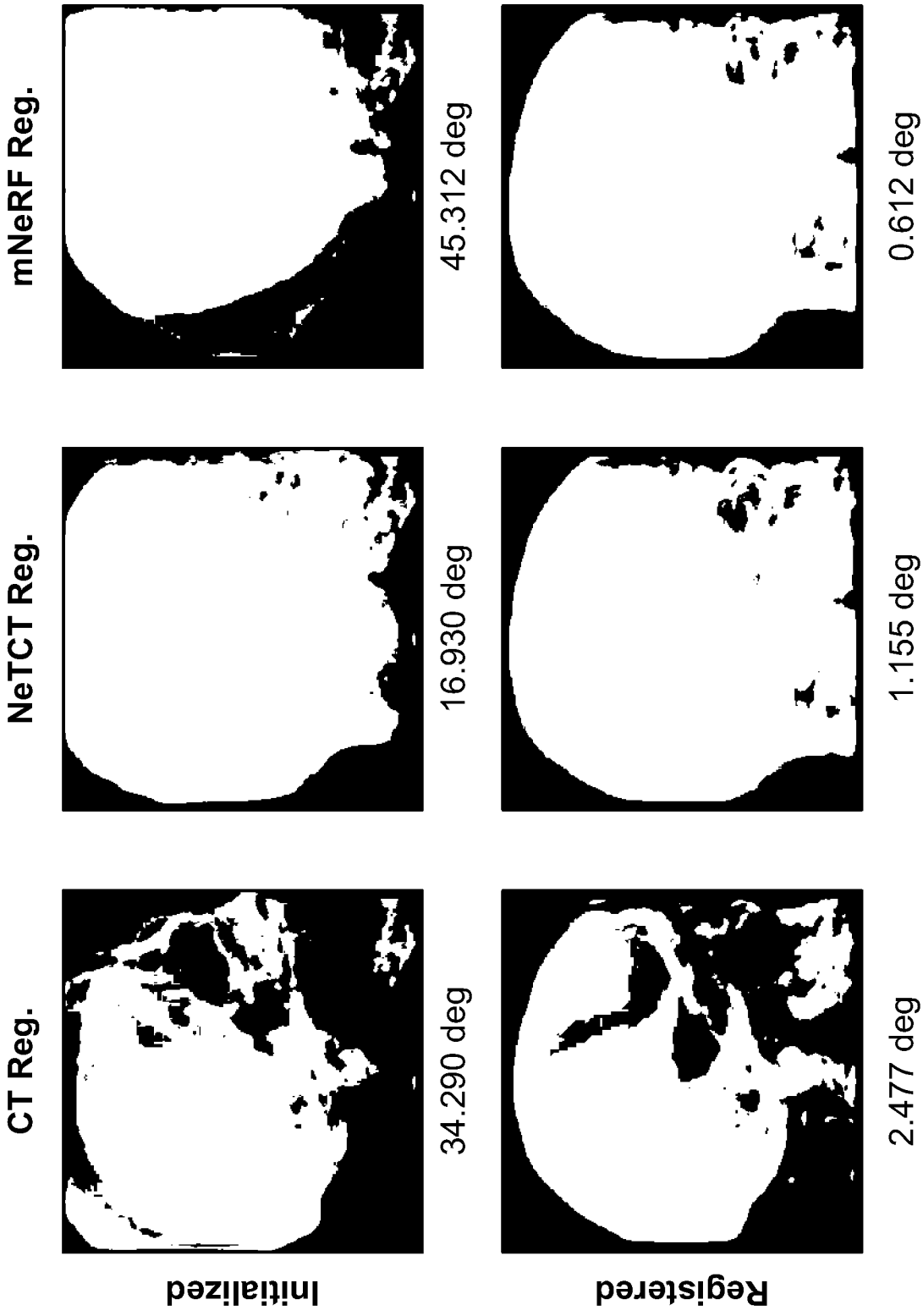

CT Reg.
NeTCT Reg.
mNeRF Reg.
Initialized
Registered
34.290 deg
16.930 deg
45.312 deg
2.477 deg
1.155 deg
0.612 deg

FIG. 11

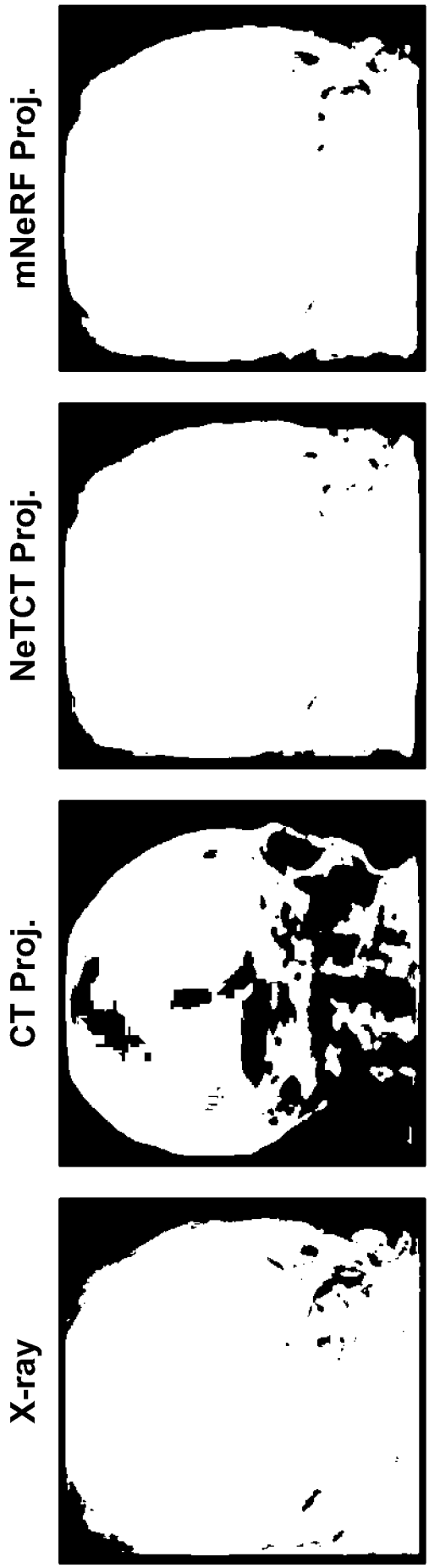


FIG. 12

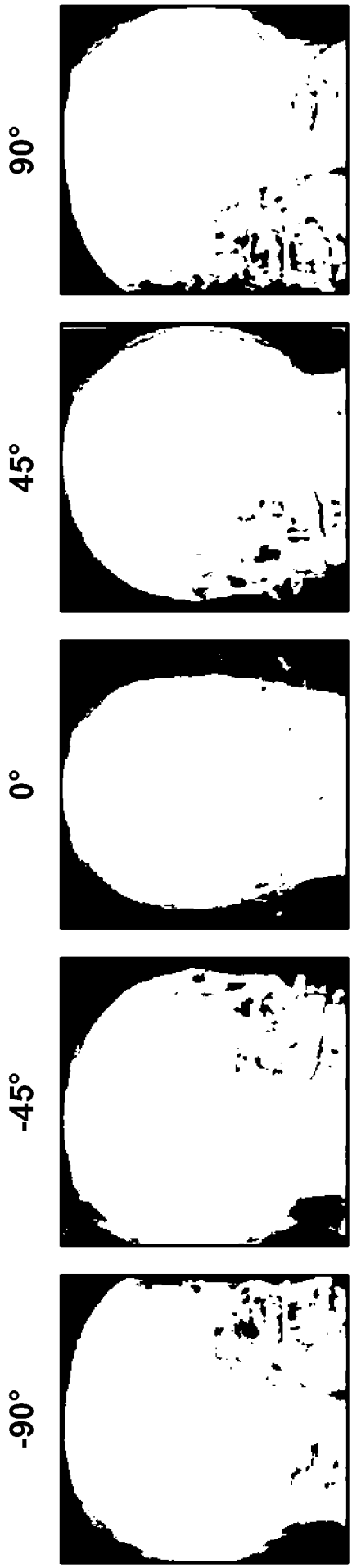


FIG. 13

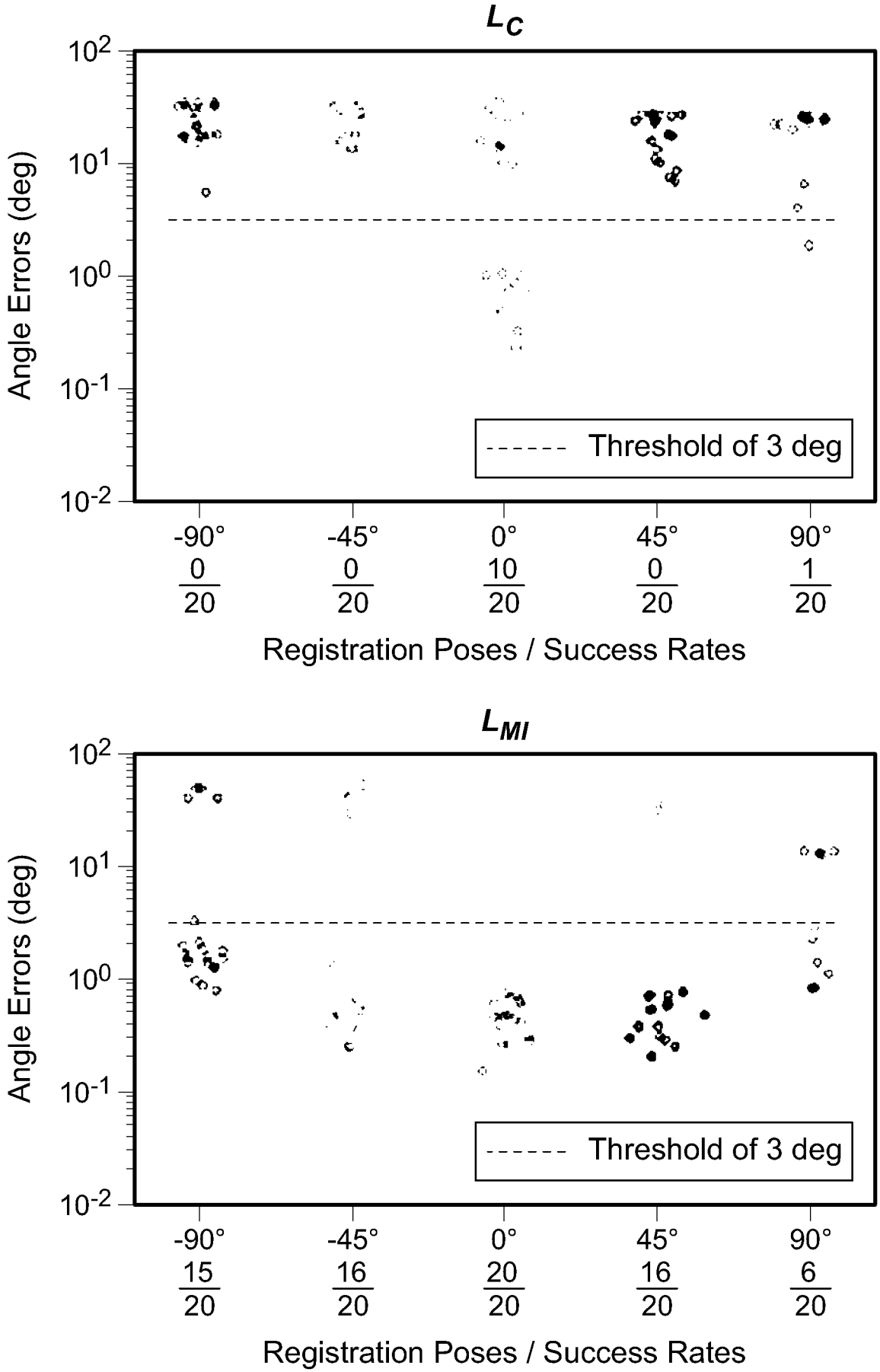

L_C
10^2
10^1
10^0
10^{-1}
10^{-2}
Angle Errors (deg)
Threshold of 3 deg
-90° 0/20
-45° 0/20
0° 10/20
45° 0/20
90° 1/20
Registration Poses / Success Rates
L_{MI}
10^2
10^1
10^0
10^{-1}
10^{-2}
Angle Errors (deg)
Threshold of 3 deg
-90° 15/20
-45° 16/20
0° 20/20
45° 16/20
90° 6/20
Registration Poses / Success Rates

FIG. 14

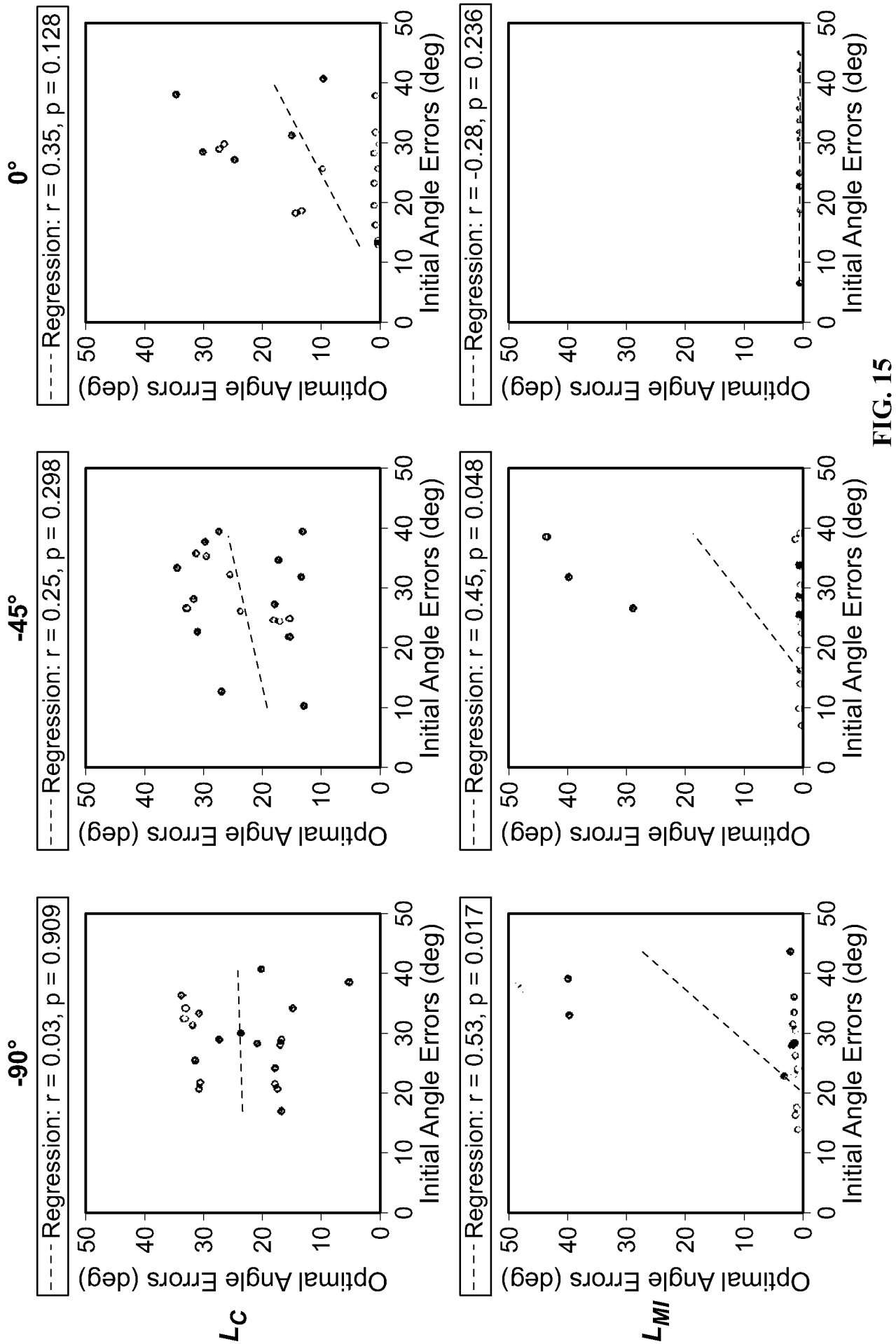

-90°
-45°
0°
L_C
L_{MI}
Regression: r = 0.03, p = 0.909
Regression: r = 0.25, p = 0.298
Regression: r = 0.35, p = 0.128
Regression: r = 0.53, p = 0.017
Regression: r = 0.45, p = 0.048
Regression: r = -0.28, p = 0.236
Optimal Angle Errors (deg)
Initial Angle Errors (deg)
0
10
20
30
40
50

FIG. 15

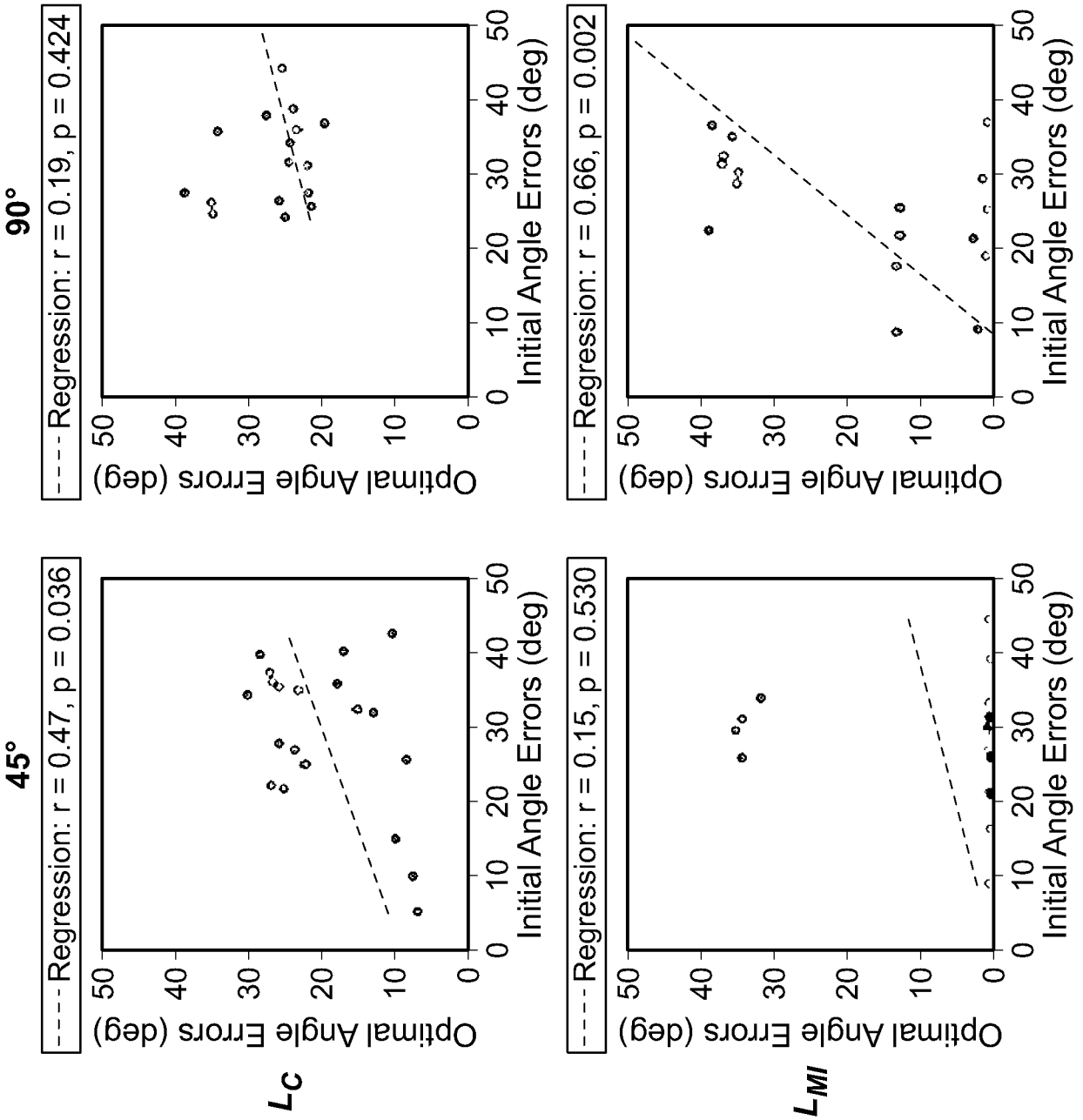

45°
90°
L_C
L_{MI}
Regression: r = 0.47, p = 0.036
Regression: r = 0.19, p = 0.424
Regression: r = 0.15, p = 0.530
Regression: r = 0.66, p = 0.002
Optimal Angle Errors (deg)
Initial Angle Errors (deg)
0
10
20
30
40
50

FIG. 15 (Cont.)

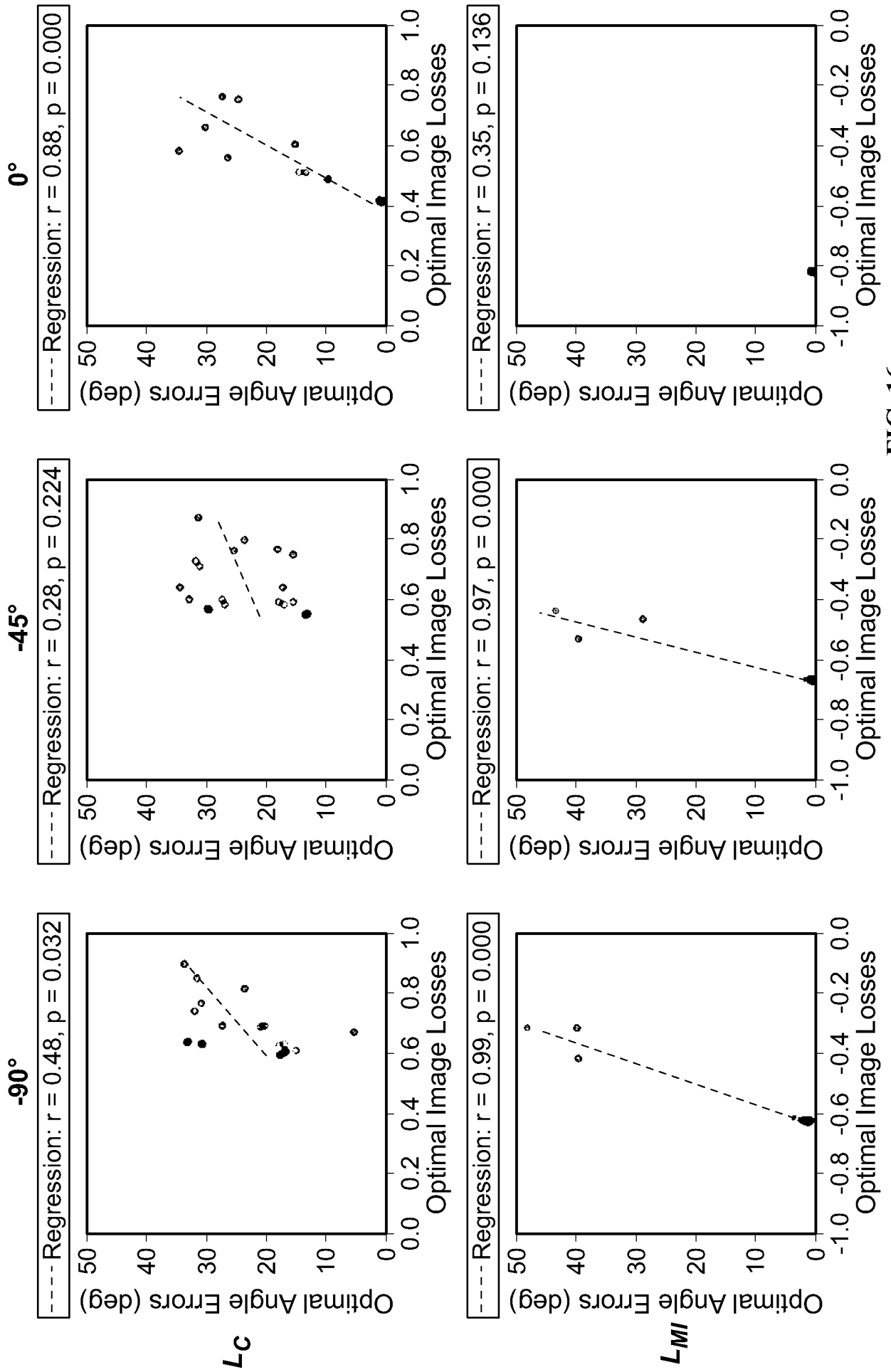

0°
-45°
-90°
L_C
L_{MI}
Regression: r = 0.88, p = 0.000
Regression: r = 0.28, p = 0.224
Regression: r = 0.48, p = 0.032
Regression: r = 0.35, p = 0.136
Regression: r = 0.97, p = 0.000
Regression: r = 0.99, p = 0.000
Optimal Angle Errors (deg)
Optimal Image Losses
0 10 20 30 40 50
0.0 0.2 0.4 0.6 0.8 1.0
-1.0 -0.8 -0.6 -0.4 -0.2 0.0

FIG. 16

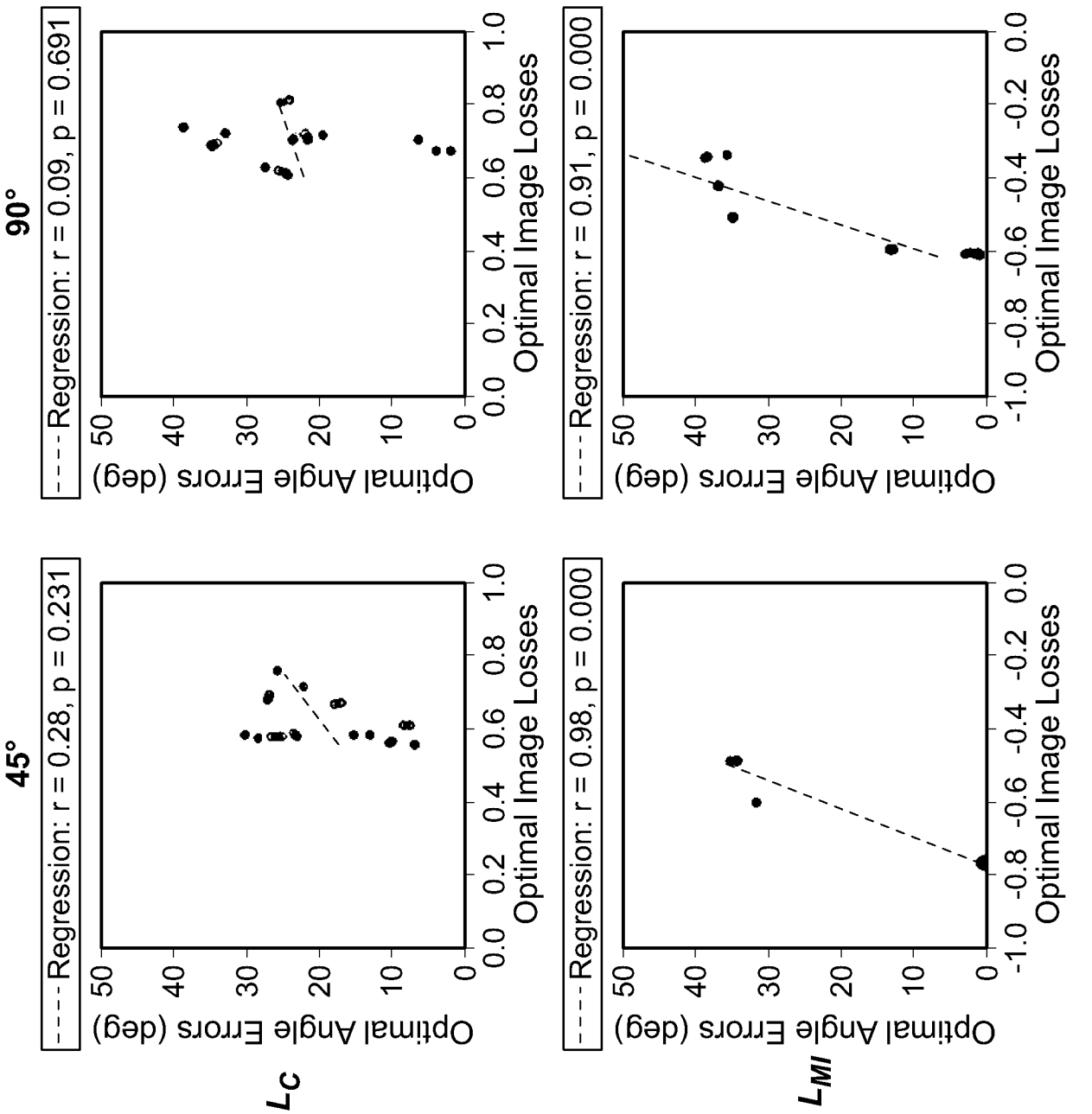

45°
90°
L_C
L_{MI}
Regression: r = 0.28, p = 0.231
Regression: r = 0.09, p = 0.691
Regression: r = 0.98, p = 0.000
Regression: r = 0.91, p = 0.000
Optimal Angle Errors (deg)
Optimal Image Losses
0
10
20
30
40
50
0.0
0.2
0.4
0.6
0.8
1.0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0

FIG. 16 (Cont.)

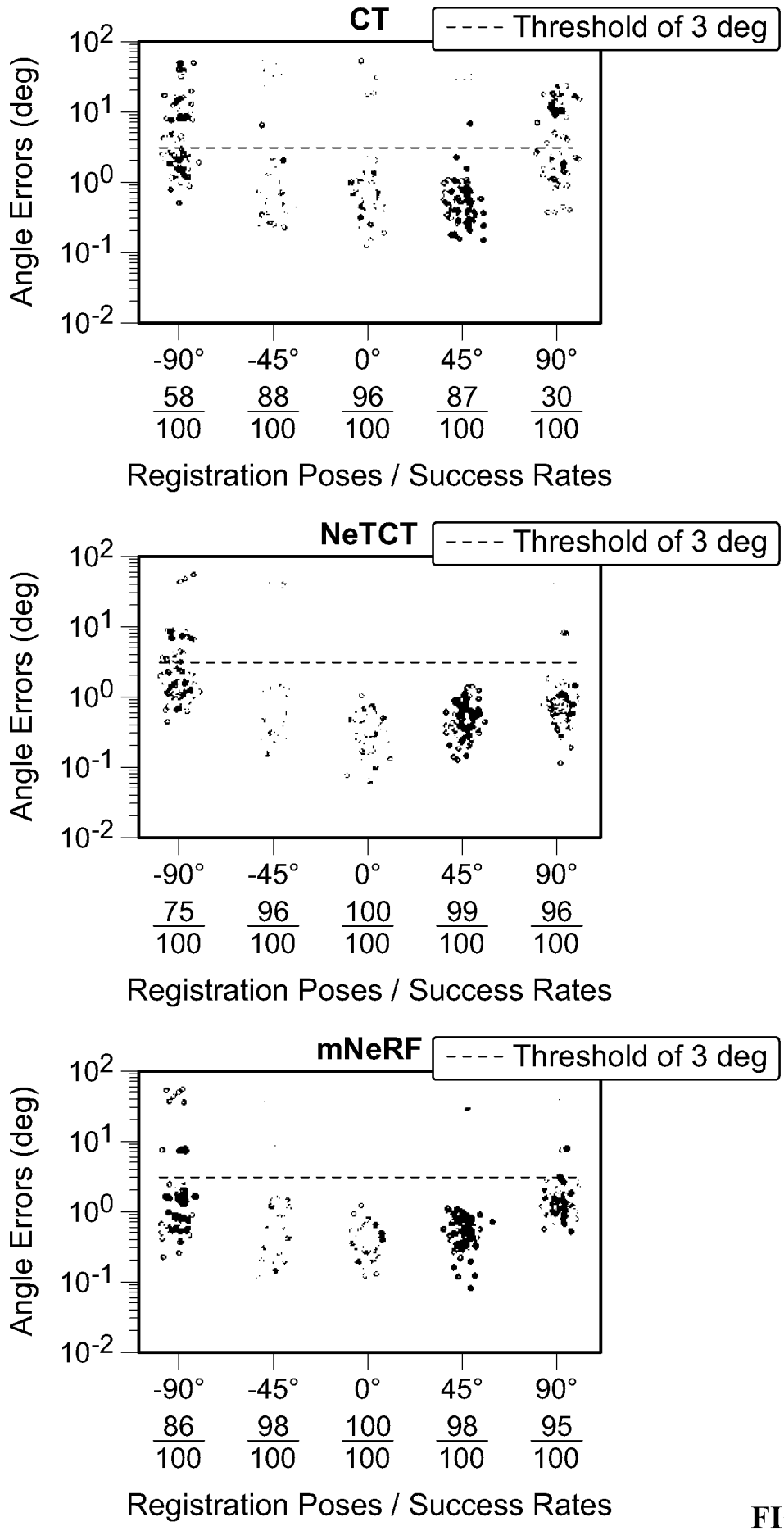

CT
---- Threshold of 3 deg
Angle Errors (deg)
10^2
10^1
10^0
10^{-1}
10^{-2}
-90° 58/100
-45° 88/100
0° 96/100
45° 87/100
90° 30/100
Registration Poses / Success Rates
NeTCT
---- Threshold of 3 deg
Angle Errors (deg)
-90° 75/100
-45° 96/100
0° 100/100
45° 99/100
90° 96/100
Registration Poses / Success Rates
mNeRF
---- Threshold of 3 deg
Angle Errors (deg)
-90° 86/100
-45° 98/100
0° 100/100
45° 98/100
90° 95/100
Registration Poses / Success Rates

FIG. 17

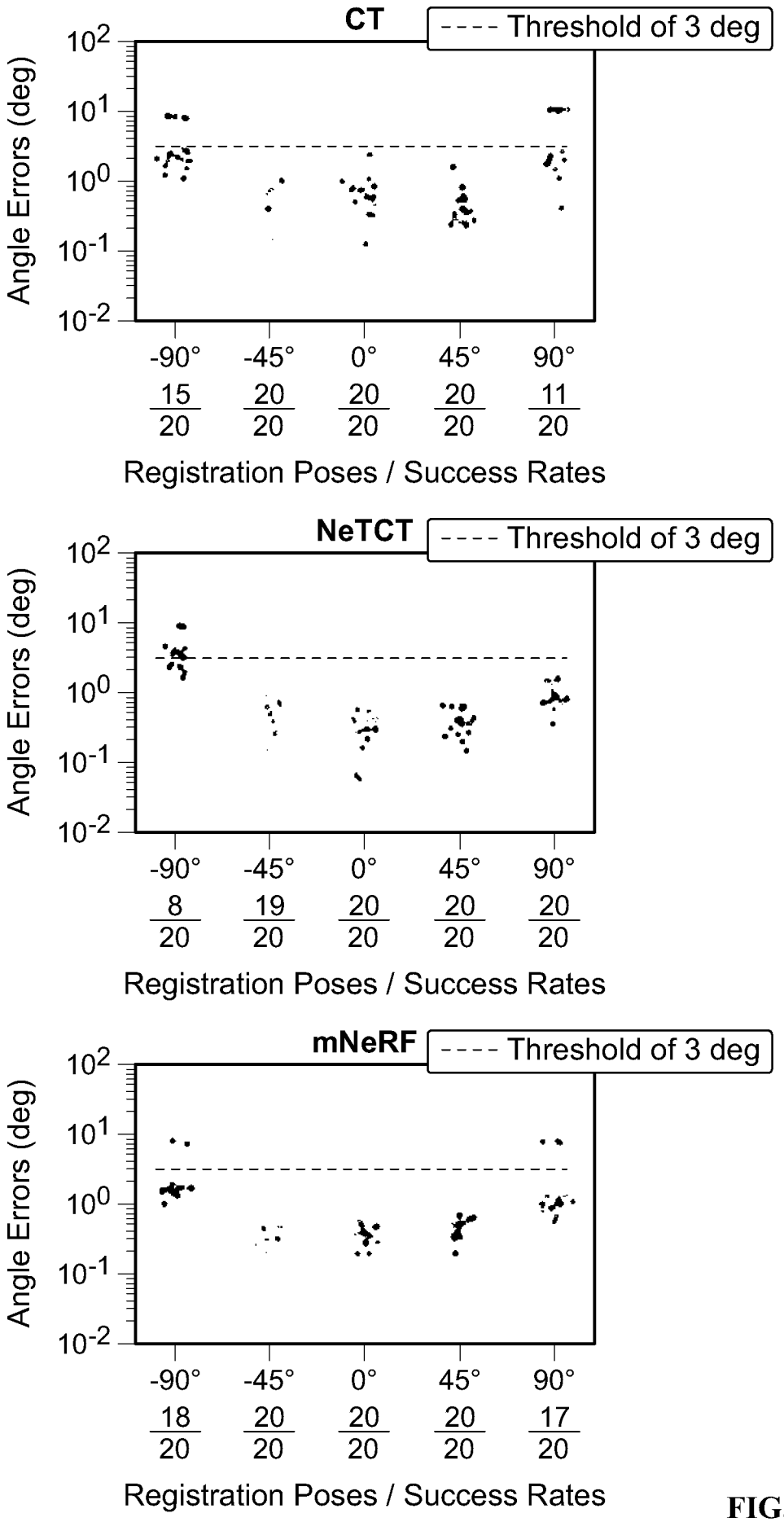
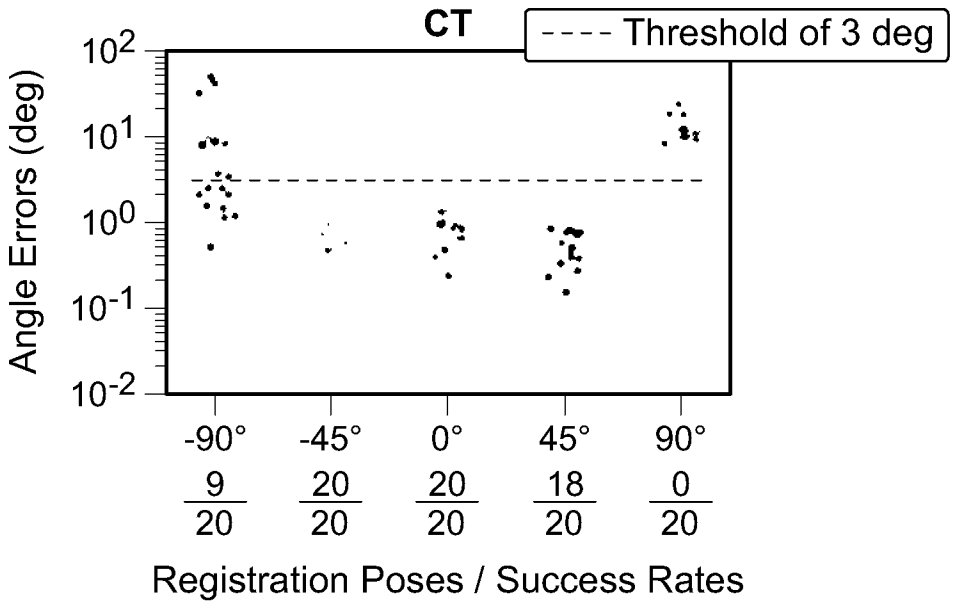
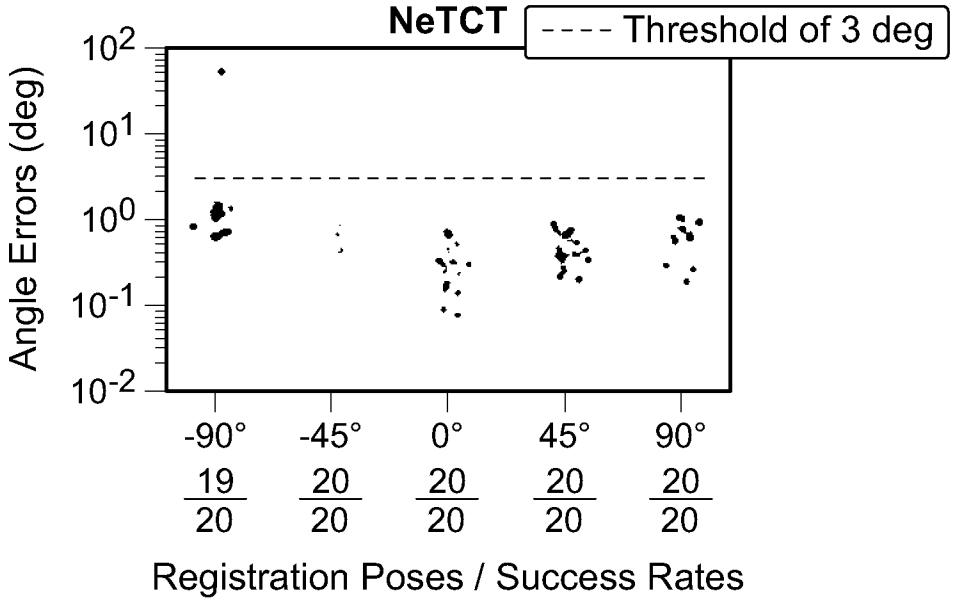

CT
---- Threshold of 3 deg
Angle Errors (deg)
10^2
10^1
10^0
10^{-1}
10^{-2}
-90° 15/20
-45° 20/20
0° 20/20
45° 20/20
90° 11/20
Registration Poses / Success Rates
NeTCT
---- Threshold of 3 deg
Angle Errors (deg)
10^2
10^1
10^0
10^{-1}
10^{-2}
-90° 8/20
-45° 19/20
0° 20/20
45° 20/20
90° 20/20
Registration Poses / Success Rates
mNeRF
---- Threshold of 3 deg
Angle Errors (deg)
10^2
10^1
10^0
10^{-1}
10^{-2}
-90° 18/20
-45° 20/20
0° 20/20
45° 20/20
90° 17/20
Registration Poses / Success Rates

FIG. 18A





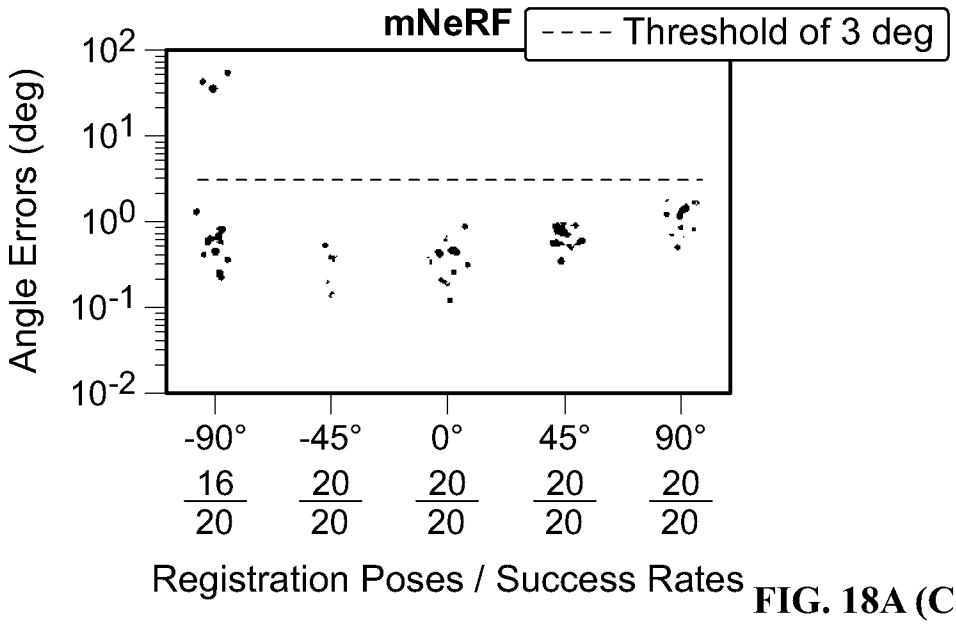
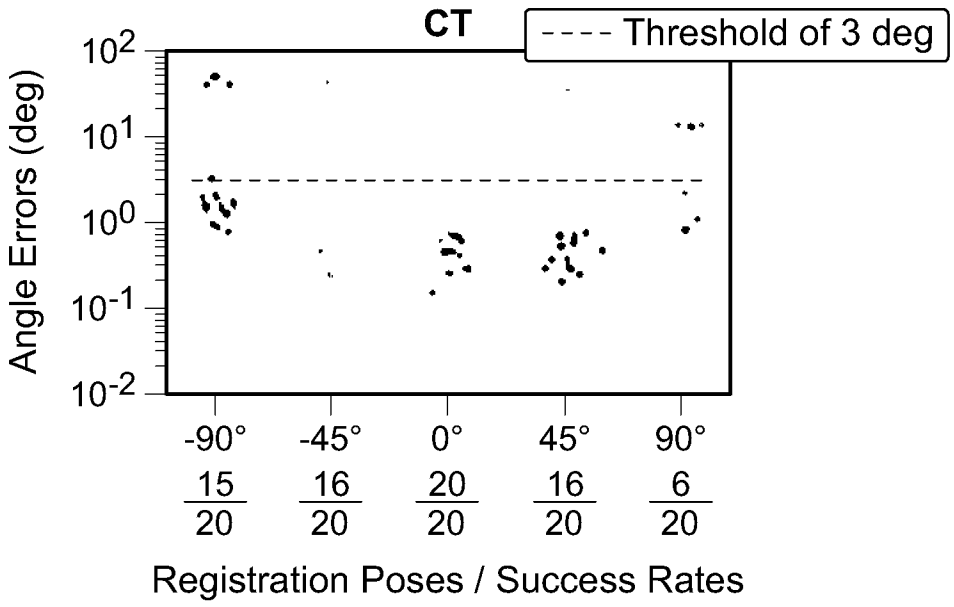
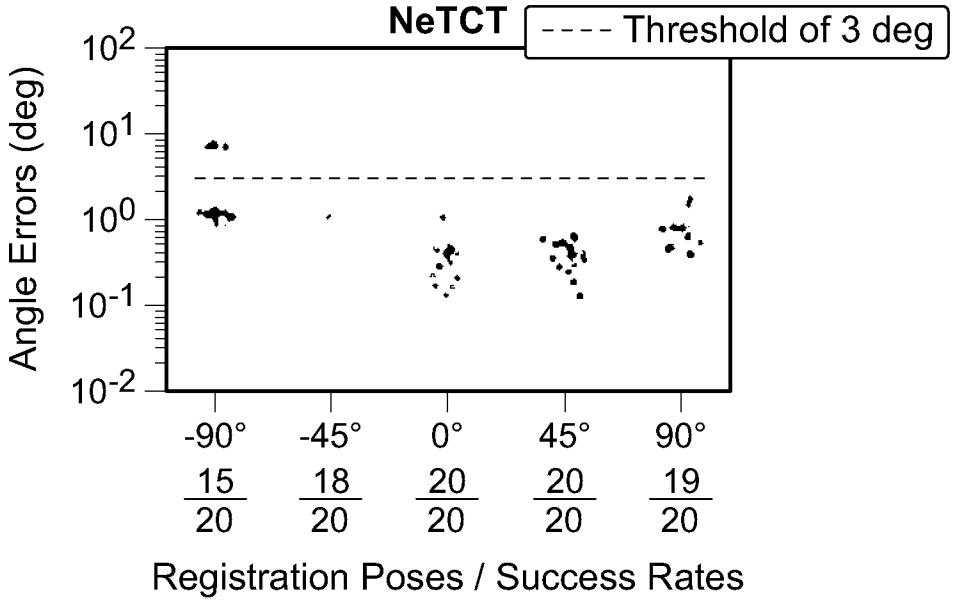


FIG. 18A (Cont. 1)





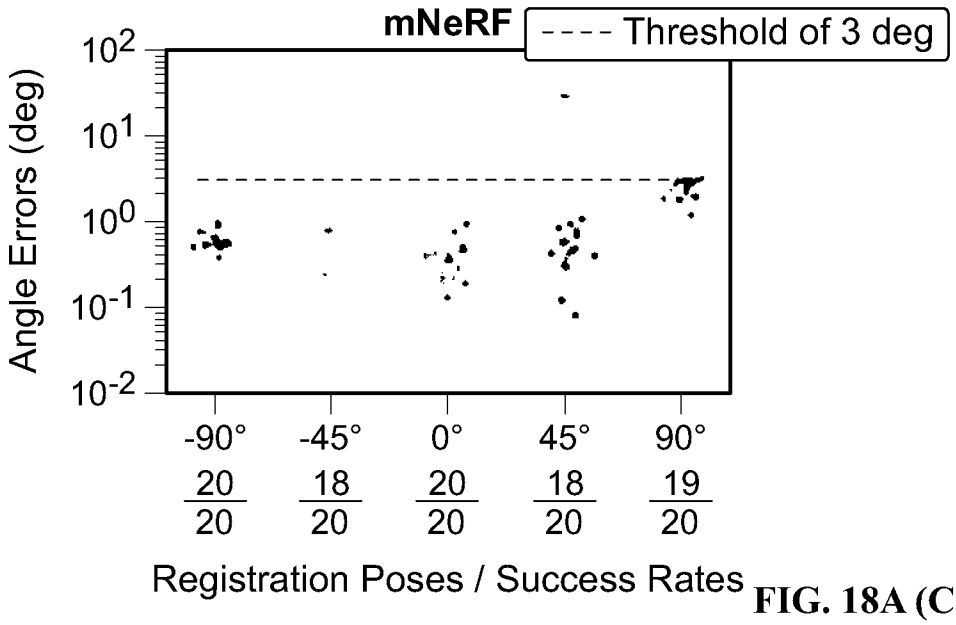


FIG. 18A (Cont. 2)

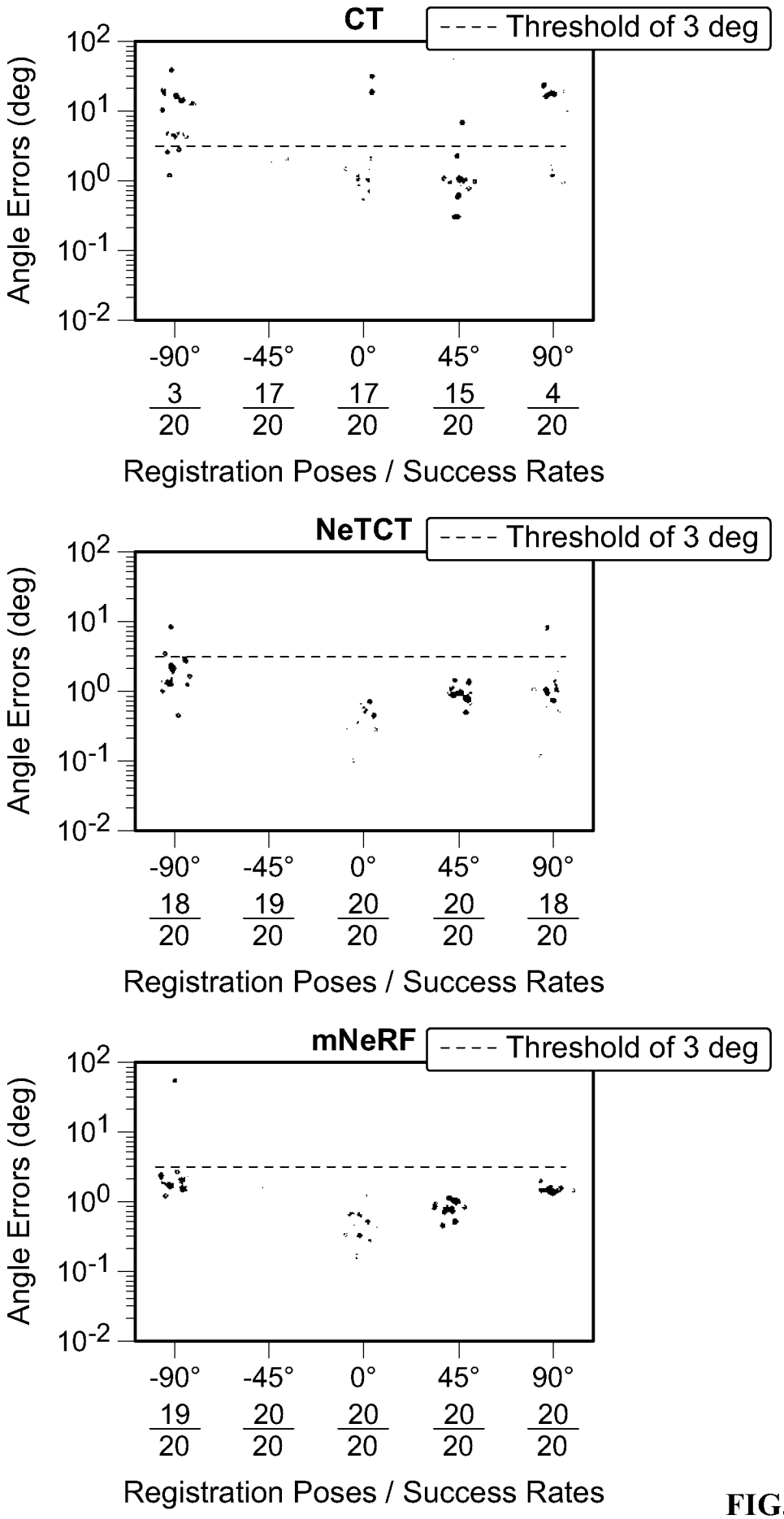

CT
---- Threshold of 3 deg
Angle Errors (deg)
10^2
10^1
10^0
10^{-1}
10^{-2}
-90° 3/20
-45° 17/20
0° 17/20
45° 15/20
90° 4/20
Registration Poses / Success Rates
NeTCT
---- Threshold of 3 deg
Angle Errors (deg)
10^2
10^1
10^0
10^{-1}
10^{-2}
-90° 18/20
-45° 19/20
0° 20/20
45° 20/20
90° 18/20
Registration Poses / Success Rates
mNeRF
---- Threshold of 3 deg
Angle Errors (deg)
10^2
10^1
10^0
10^{-1}
10^{-2}
-90° 19/20
-45° 20/20
0° 20/20
45° 20/20
90° 20/20
Registration Poses / Success Rates

FIG. 18B

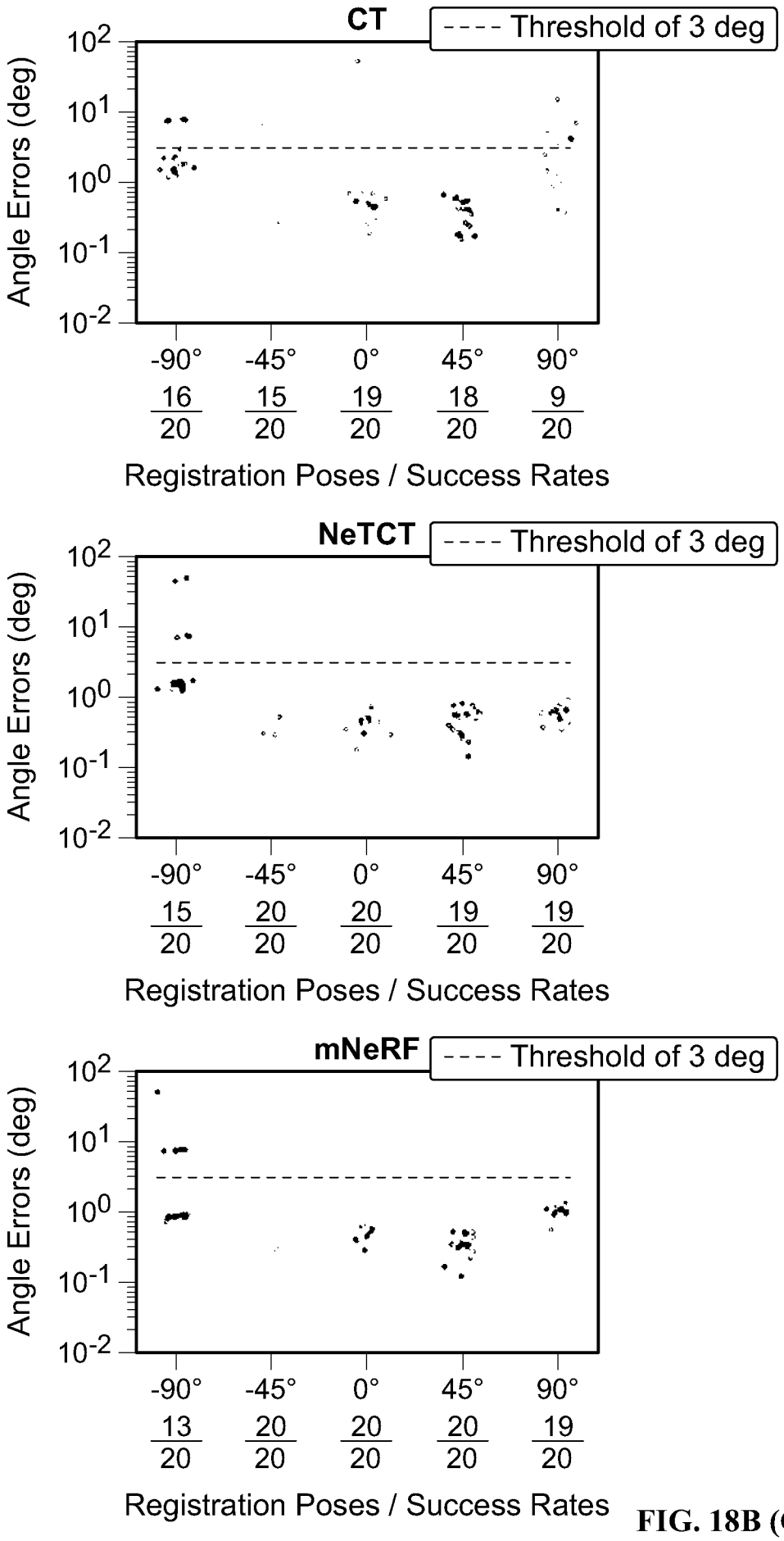

CT
---- Threshold of 3 deg
Angle Errors (deg)
10^2
10^1
10^0
10^{-1}
10^{-2}
-90° 16/20
-45° 15/20
0° 19/20
45° 18/20
90° 9/20
Registration Poses / Success Rates
NeTCT
---- Threshold of 3 deg
Angle Errors (deg)
-90° 15/20
-45° 20/20
0° 20/20
45° 19/20
90° 19/20
Registration Poses / Success Rates
mNeRF
---- Threshold of 3 deg
Angle Errors (deg)
-90° 13/20
-45° 20/20
0° 20/20
45° 20/20
90° 19/20
Registration Poses / Success Rates

FIG. 18B (Cont.)