



(51) International Patent Classification:

G01B 11/25 (2006.01) A61C 9/00 (2006.01)
A61B 5/00 (2006.01)

(21) International Application Number:

PCT/EP2024/066770

(22) International Filing Date:

17 June 2024 (17.06.2024)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

PA 2023 70299 16 June 2023 (16.06.2023) DK

(71) Applicant: 3SHAPE A/S [DK/DK]; Holmens Kanal 7,
1060 Copenhagen K (DK).

(72) Inventor: JENSEN, Janus Nørtoft; C/O 3Shape A/S, Hol-
mens Kanal 7, 1060 Copenhagen K (DK).

(74) Agent: GUARDIAN IP CONSULTING I/S; Diplomvej,
Building 381, 2800 Kgs. Lyngby (DK).

(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.

(54) Title: SYSTEM AND METHOD FOR INTRAORAL SCANNING

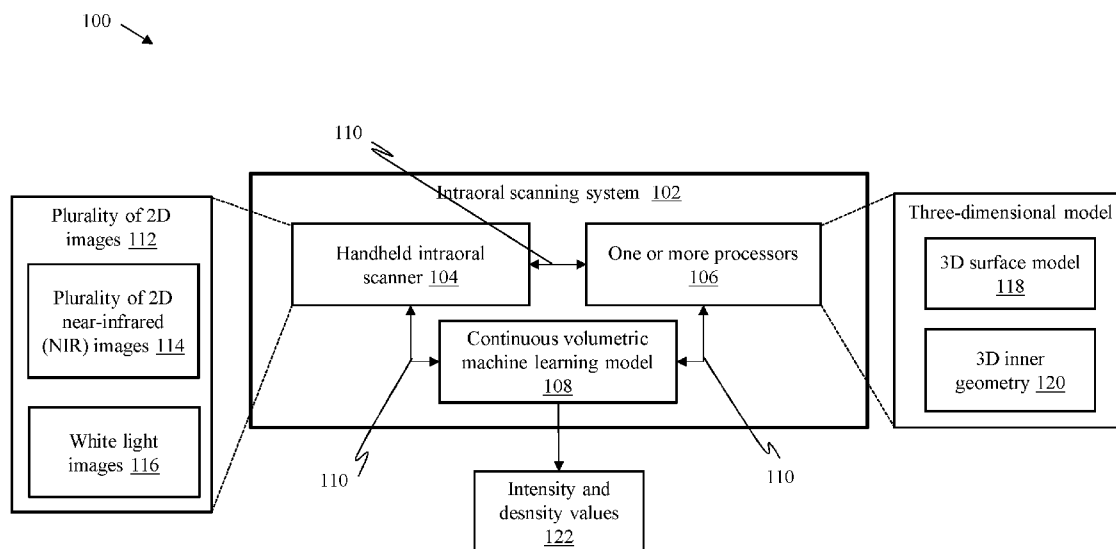


FIG. 1

(57) Abstract: An intraoral scanning system (102) including hand-held intraoral scanner (104), processors (106) and continuous volumetric ML model (108) for generating 3D inner geometry (120) of teeth (306) is provided. Intraoral scanner comprises sensors (214D) to detect NIR and visible light. The processors receive visible light and NIR information from the sensors, generate 3D surface model (118) of the teeth using visible light information, and determine input parameters for the continuous volumetric ML model by capturing 2D NIR images (114) of an internal region of the teeth from the NIR information. Each of the NIR images includes pixels (722) having a casting object. The input parameters comprise spatial location and viewing angle, and the casting object comprises point coordinates (726, 732). The continuous volumetric ML model processes the input parameters to determine an intensity value and a density value of pixels in 3D inner geometry of the 3D surface model.

(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, 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:

- *with international search report (Art. 21(3))*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

SYSTEM AND METHOD FOR INTRAORAL SCANNING

TECHNICAL FIELD

[1] An example embodiment of the present invention generally relates to intraoral scan registration and more particularly relates to an intraoral scanning system and a method for intraoral scanning to generate a three-dimensional inner geometry of a subject's teeth.

BACKGROUND OF THE INVENTION

[2] Intraoral scanners are electronic devices that may be used for, for example, capturing three-dimensional (3D) digital images. In an example, the intraoral scanners may include a light source that may project light rays onto an object to be scanned, such as teeth, gums, and other intraoral structures inside of patients' mouth. In an example, images captured by an intraoral scanner may be processed to generate digital impressions, such as a 3D surface model of oral cavity of a patient. The 3D surface model may be displayed on a screen for examination of the oral cavity.

[3] Typically, intraoral scanners are used for examination or treatments inside the oral cavities of patients. The use of intraoral scanners may eliminate the use of conventional impression material, plaster models and simplify clinical treatment procedures of teeth for the dentists while reducing patient discomfort. However, conventional intraoral scanners may not be suited to detect internal structures of teeth, for example, structures of enamel and dentin within a tooth of a patient as they are generally designed to detect the surface of the dentition only. In particular, the intraoral scanners may fail to determine general internal composition of the tooth inside oral cavity of the patient. To this end, the intraoral scanners are not suitable for detecting, for example, development of caries and cracks within enamel and underlying dentin of the tooth, bleeding or any other damage within the enamel and the underlying dentin of the tooth, and/or deep crack or margin lines or other error in prepared dental prostheses (such as, crown,

bridges, implants, inlays, onlays, etc.) for the patient. Particularly, information of internal structure of a natural tooth or a prosthetic implant in conjunction with surface information may be crucial for effective treatment of the patient and ensuring long span restorations with natural tooth or prosthetic implant.

5 [4] Conventionally methods for constructing 3D inner geometry of intraoral dental structure, such as teeth, using intraoral scanners are limited. In an example, a conventional method may cause an intraoral scanner to capture large number of images from various different viewpoints for reconstructing inner geometry of the dental structure in 3D. In this regard, based on several pixels of
10 the large number of captured images from different viewpoints and corresponding intensities, multiple artificial points are projected corresponding to the 3D inner geometry of the dental structure. Thereafter, only minimum casting object intensity or a minimum scattering intensity corresponding to each artificial point is considered for reconstructing the 3D inner geometry of the dental
15 structure in 3D.

[5] The conventional method of generating the 3D model of the 3D inner geometry of the dental structure possesses several disadvantages. In an example, the conventional methods fail to disambiguate different compositions of the 3D inner geometry of the dental structure reliably. In addition, the conventional
20 methods require very large number of images captured from large number of viewpoints as reliable scattering coefficients can only be determined if multiple view angles are covered. This may increase computational load and cause delay in generating the 3D model. As a result, real-time operations may not be performed on such 3D model of the 3D inner geometry of the dental structure. For example,
25 all points along a given ray of light is determined to be of the same material unless a point on the ray is also seen by another ray from another image with lower intensity, i.e., the another ray that also sees a point in the another image does not intersect with any other part of the dental structure beyond the point within the 3D inner geometry of the dental structure. For example, for a point to be determined
30 as enamel, it is necessary that at least one ray from the light rays that are projected

onto the dental structure and projected back to form an image that does not intersect with dentine or any other dental structure within the inner structure at any point. Therefore, there is a need of improved systems and methods of intraoral scanning to overcome the disadvantages of the conventional methods for
5 generating 3D model of inner geometry of the dental structure.

SUMMARY

[6] An intraoral scanning system, a method and a computer programmable product are provided for generating a three-dimensional (3D) inner geometry of a subject's teeth using a handheld intraoral scanner comprising sensors to detect
10 near-infrared (NIR) and visible light, render of three-dimensional (3D) surface model of the subject's teeth and a continuous volumetric machine learning model.

[7] Some embodiments are based on the understanding that the 3D inner geometry of the subject's teeth is determined using visible light information and
15 NIR information from the one or more sensors at the same time or with a delay in-between caused by an on-off switching of the visible light and the NIR light. The on-off switching may be needed as the intraoral scanning system is not configured to capture visible information and the NIR information at the same time. The period of the delay may be within 200 ms or within 500 ms.

[8] In one aspect, an intraoral scanning system to generate a three-dimensional (3D) inner geometry of a subject's teeth is disclosed. The intraoral scanning system may include a hand-held intraoral scanner configured to operate with one or more sensors to detect near-infrared (NIR) and visible light. The one or more sensors may comprise an image sensor. The intraoral scanning system
20 may include one or more processors operably connected to the hand-held intraoral scanner. The one or more processors may be configured to receive visible light information and near-infrared (NIR) information from the one or more sensors. The one or more processors may be configured to determine surface information from the visible light information in real time to generate a

three-dimensional (3D) surface model of the subject's teeth using the surface information. The one or more processors may be configured to capture a plurality of two-dimensional (2D) NIR images of an internal region of the subject's teeth from the NIR information in real time using the image sensor.

5 Each of the plurality of NIR images includes a plurality of corresponding pixels. The one or more processors may be configured to determine a set of input parameters for a casting object corresponding to each of the plurality of pixels, wherein the set of input parameters may comprise spatial location information and viewing angle information, and the casting object may comprise a plurality

10 of point coordinates. The intraoral scanning system may include a continuous volumetric machine learning model configured to receive and process the set of input parameters to determine an intensity value and a density value of a three-dimensional inner geometry of the 3D surface model. The continuous volumetric machine learning model may be configured to be trained using the plurality of

15 2D NIR images.

[9] In some embodiments, the continuous volumetric machine learning model may be trained by receiving the set of input parameters that corresponds to the casting object for each of the plurality of pixels corresponding to each of the plurality of 2D NIR images. The casting object includes the plurality of point

20 coordinates associated with the 3D surface model. The continuous volumetric machine learning model may be trained by generating, based on the continuous volumetric machine learning model using the set of input parameters, the intensity and the density value for each of the plurality of point coordinates. The continuous volumetric machine learning model may be trained by determining the

25 synthetic pixel value for the casting object based on the corresponding determined intensity value and density value for each of the plurality of point coordinates. Further, the continuous volumetric machine learning model may be trained by minimizing a loss function between the synthetic pixel value and a corresponding true pixel value of the plurality of pixels of the plurality of 2D NIR images.

[10] In some embodiments, the one or more processors may further be configured to determine a plurality of grid points within the 3D surface model, and determine the 3D inner geometry by arranging at least one of: the intensity value, or the density value at each of the plurality of grid points.

5 [11] In some embodiments, the intraoral scanning system may further comprise a display unit configured to display the 3D inner geometry based on at least one of: the the intensity value or the density value determined by the continuous volumetric machine learning model.

[12] In some embodiments, the display unit may be configured to display
10 the 3D inner geometry inside the 3D surface model.

[13] In some embodiments, the one or more processors may be further configured to determine a boundary between enamel and dentine in a dentition of the subject's teeth based on the the intensity values and the density values for the casting object of each of the plurality of pixels.

15 [14] In some embodiments, the casting object may be one of: a ray or a cone.

[15] In some embodiments, the hand-held intraoral scanner may comprise a projector unit that may be configured to illuminate the subject's teeth using one or more white colored wavelength pulses and one or more near infrared
20 (NIR) wavelength pulses. Furthermore, the hand-held intraoral scanner may comprise the one or more sensors that may be configured to generate a set of white light images and the plurality of 2D NIR images based on the illumination.

[16] In some embodiments, the 3D surface model is determined based on
25 the set of white light images.

[17] In some embodiments, the one or more processors may be further configured to estimate a relative position between the hand-held intraoral scanner

and the subject's teeth corresponding to each of the plurality of NIR images. The estimated relative position indicates the viewing angle information and the spatial location information for the corresponding casting object.

[18] In some embodiments, the casting object may be a cone. In this regard, the one or more processors may further be configured to determine one or more conical frustums for each of the cone corresponding to each of the plurality of pixels. The one or more conical frustums relate to the plurality of point coordinates. The one or more processors may further be configured to determine an integrated positional encoding for the one or more conical frustums for transforming the plurality of point coordinates of each of the cone. The integrated positional encoding comprises at least a Gaussian encoding and a sinusoidal encoding. The one or more processors may further be configured to generate the synthetic pixel value for each of the cone based on the integrated positional encoding for the corresponding one or more conical frustums and the continuous volumetric machine learning model.

[19] In some embodiments, the one or more processors may further be configured to determine a radius of each of the cone indicating the casting object corresponding to each of the plurality of pixels based on a size of the corresponding pixel from the plurality of pixels.

[20] In some embodiments, the one or more processors may further be configured to determine an average of content within a visible volume for each of the plurality of pixels, and render each of the cones indicating the casting object of each of the plurality of pixels based on the corresponding average of content. The content indicates a color intensity.

[21] In some embodiments, the continuous volumetric machine learning model may include a machine-learning based neural radiance field (NeRF) network.

[22] In another aspect, a method for generating a three-dimensional (3D) inner geometry of a subject's teeth using an intraoral scanning system is disclosed. The intraoral scanning system may comprise a hand-held intraoral scanner configured to operate with one or more sensors to detect near-infrared (NIR) and visible light, one or more processors operably connected to the hand-held intraoral scanner, and a continuous volumetric machine learning model. The one or more sensors may comprise an image sensor. The method comprises receiving visible light information and near-infrared (NIR) information from the one or more sensors. The method may further comprise determining surface information from the visible light information to generate a three-dimensional (3D) surface model of the subject's teeth using the surface information in real time. The method may further comprise capturing a plurality of two-dimensional (2D) NIR images of an internal region of the subject's teeth from the NIR information in real time using the image sensor. Each of the plurality of NIR images includes a plurality of corresponding pixels. The method may further comprise determining a set of input parameters for a casting object corresponding to each of the plurality of pixels. The set of input parameters may comprise spatial location information and viewing angle information, and the casting object may comprise a plurality of point coordinates. The method may further comprise processing, using the continuous volumetric machine learning model, the set of input parameters to determine an intensity value and a density value of a three-dimensional inner geometry of the 3D surface model. The continuous volumetric machine learning model is configured to be trained using the plurality of NIR images.

[23] In yet another aspect, a computer programmable product comprising a non-transitory computer readable medium having stored thereon computer executable instructions, which when executed by a processing circuitry, cause the processing circuitry to carry out operations. The operations may comprise receiving visible light information and near-infrared (NIR) information from one or more sensors, the one or more sensors being installed within a hand-held

intraoral scanner. The one or more sensors may be configured to detect near-infrared (NIR) and visible light, wherein the one or more sensors may comprise an image sensor. The operations may further comprise determining surface information from the visible light information to generate a three-dimensional (3D) surface model of the subject's teeth using the surface information in real time. The operations may further comprise capturing a plurality of two-dimensional (2D) NIR images of an internal region of the subject's teeth from the NIR information in real time using the image sensor. Each of the plurality of NIR images may include a plurality of corresponding pixels. The operations may further comprise determining a set of input parameters for a casting object corresponding to each of the plurality of pixels. The set of input parameters may comprise spatial location information and viewing angle information, and the casting object may comprise a plurality of point coordinates. The operations further comprise processing, using a continuous volumetric machine learning model, the set of input parameters to determine an intensity value and a density value of a three-dimensional inner geometry of the 3D surface model. The continuous volumetric machine learning model may be configured to be trained using the plurality of NIR images.

[24] The foregoing summary is illustrative only and is not intended to be in any way limiting. In addition to the illustrative aspects, embodiments, and features described above, further aspects, embodiments, and features will become apparent by reference to the drawings and the following detailed description.

[25] According to the present disclosure, an intraoral scanning system, a method and a computer programmable product are provided. One of the purposes of the present disclosure is to provide an enhanced processing capability in a handheld intraoral scanner to generate a three-dimensional (3D) inner geometry of a subject's teeth.

[26] Conventional systems may include intraoral scanners that are utilized to capture two-dimensional (2D) intraoral scans and/or three-dimensional (3D)

information of dental arches of patients. However, the conventional intraoral scanners may possess limited processing capability, that may only be utilized to capture 2D intraoral scans and/or 3D model of surface of the dental arches of the patients. However, 3D surface models of the dental arches may fail to detect
5 internal structures of teeth, for example, structures of enamel and dentine within the dental arches of the patients. As a result, diseases or anomalies within the internal structure of the teeth may not be identified, unless such diseases grow till the surface of the teeth or the dentist make use of other devices such as X-ray machines, however -rays are ionized radiations which can cause mutations in
10 human cells upon exposure which should be avoided. For example, conventional intraoral scanners may fail to identify early occurrence of an anomaly or a disease occurring inside a tooth, such as dentin erosion, enamel erosion, cracks or caries inside tooth, etc. unless a bone of the tooth or a surface of the tooth is affected. Due to delayed diagnosis of diseases or anomalies inside the tooth, irreparable
15 damage may occur in the tooth causing use of surgical method to remove such tooth. This may cause great discomfort to patients. Moreover, the conventional intraoral scanners may fail to identify an anomaly occurring inside a dental prosthetic implant. The identification of any anomaly occurring inside the dental prosthetic implant may be crucial to ensure long life of such dental prosthetic
20 implant. To this end, the information relating to inner geometry of teeth or dental arches of the patients in conjunction with surface information may be crucial for early diagnosis, effective treatment and ensuring long span restorations with natural tooth or prosthetic implant.

[27] In certain conventional method, inner geometry of teeth may be
25 reconstructed using intraoral scans. The conventional methods for generating inner geometry of teeth may be practically infeasible and possess other limitations owing to requirement of large number of viewpoints and angles of the intraoral 2D images. To project a point indicative of a part in inner geometry of teeth or a tooth, values of several pixels corresponding to the part of the 3D inner geometry

in large number of NIR 2D images from different viewing angles may be processed to identify intensities for the part in each of the several pixels.

[28] Thereafter, a minimum intensity value corresponding to the part based on the processing of several pixels may be used to project the 3D point indicative of the part. For example, the minimum intensity value for a pixel of a 2D image
5 may indicate that a reflection for the part is originating from a material of the part. Therefore, for a 3D point to be determined as of a part of enamel, it is necessary to receive at least one ray forming a pixel on a 2D image, such that the ray corresponding to the part of the enamel does not intersect with dentine at any
10 point. In other words, when the ray is reflected back from the part of the enamel for forming the 2D image, the pixel corresponding to the part of the enamel may have minimum intensity. Further, the 3D point may then be projected as indicative of the part of the enamel based on the minimum intensity.

[29] To this end, such processing of a pixel from large number of NIR 2D
15 images corresponding to the part may be processing intensive. Large amount of computing power may be required that may increase the size and price of the intraoral scanning system. In addition, a 3D model for the 3D inner geometry generated using the conventional methods may not be accurate due to dependence on the minimum intensity value for projecting 3D points. For example, all points
20 along a given ray, for example, a ray reflected from dentine, may be determined to be relating to a same material, such as the dentin, unless some point (corresponding to a part of inner geometry) on the ray is also seen by another ray, such as a ray reflected from enamel, from another image with lower intensity. As may be understood that reflections in the enamel are minimal, and in dentin it will
25 be somewhat higher as the light is scattered much more, therefore, conventional methods may fail to depict a boundary between the enamel the dentin within inner geometry of the teeth or tooth. Thus, usage of the conventional system for the intraoral scanning for generating 3D model of inner geometry may be inaccurate, time consuming, computing intensive, cumbersome, and, in some cases,
30 practically infeasible.

[30] On the other hand, the intraoral scanning system of the present disclosure includes a handheld intraoral scanner, processor and a continuous volumetric machine learning model for generating 3D inner geometry of teeth of a subject. The generation may in one embodiment be done in real time or as a subsequent processing step upon data acquisition. The intraoral scanning system may provide an enhanced processing capability within the handheld intraoral scanner, such that additional equipment or devices are not required to generate the 3D inner geometry of teeth. In particular, additional bulky devices or equipment may not be required for processing 2D images for generating 3D model of inner geometry along with 3D surface model of the teeth of the subject. Further, a user, such as a dentist may not have to carry and operate a bulky or cumbersome equipment for generating 3D model of inner geometry and/or surface of the teeth of the subject.

[31] The intraoral scanning system of the present disclosure includes the hand-held intraoral scanner that emits light rays, and generates 2D images using sensors. For example, the hand-held intraoral scanner may emit light rays having spectral range corresponding to visible white light wavelength range (for example, 400-750 nm) as well as near infra-red (NIR) wavelength range (for example, 750-2500 nm), continually. In this regard, the teeth of the subject need not be illuminated multiple times with white light and/or NIR separately for generating white light and 2D NIR image scans or images of the teeth. For example, based on scan of oral cavity or the teeth of the subject using the hand-held intraoral scanner, the white light images and 2D NIR images may be generated by on-off switching of a visible light source and a NIR light source in a sequence. The on-off switching may be needed as the intraoral scanner may not be configured to capture the visible information and the NIR information at the same time. In an example, the switching time interval may be within 200 ms or within 500 ms. Alternatively, the white light illumination and NIR illumination may be done simultaneously to avoid any relative motion between the intraoral scanner and the patient's teeth in between illumination. In such a

case, the intraoral scanner may include an appropriate optical filter to selectively expose individual pixels or groups of pixels with a subset of the illuminated light (in case of a single sensor). Alternatively, white light illumination and NIR illumination may be done simultaneously using different image sensors within the intraoral scanner, such that the different image sensors are configured to only be sensitive at a corresponding selective range relating to the white light or the NIR light.

[32] Further, intraoral scanning system of the present disclosure alleviates problems relating to inaccurate inner geometry of teeth by using continuous volumetric machine learning model that is based on Neural Radiance Fields (NeRFs). In particular, the processors of the intraoral scanning system does not base an estimate of a material of a 3D point, i.e., whether the 3D point corresponds to enamel or dentin, on minimum intensity value of pixels. Rather, point coordinates along rays are integrated through pixels in the 2D NIR images. Further, the integrated point coordinates of each pixel in the multiple 2D NIR images are optimized using a function defined by the continuous volumetric machine learning model to predict values indicating intensity and density values and at 3D points along the rays. The intensity value and the density value may describe a nature of material corresponding to the 3D point, thereby enabling a distinction between enamel and dentin, as well as identification of a boundary between the enamel and dentin. As such, point coordinates on a ray through a pixel may not belong to same material, and the optimization process is applied over all the rays. As a result, substantially less number of 2D NIR images is required to reconstruct the 3D inner geometry of the teeth in 3D. This is advantageous as a coverage of 2D NIR images from multiple angles and viewpoints obtained while performing the intraoral scanning session may be limited as the intraoral scanner is moved across the dental arch of the patient relatively fast and is the intraoral scanner only observes or illuminates a same region of the patient's teeth from a limited number of poses (view angles), such as occlusion, lingual and labial orientations.

[33] Further, the intraoral scanning system of the present disclosure includes or may be coupled to a display that may be communicatively coupled to the processors and the continuous volumetric machine learning model via the communication network. Based on the 3D inner geometry of the teeth and 3D surface model of the teeth generated using the white light images, a holistic interactive render of the teeth of the subject may be displayed on the display. For example, the display may be a display of a smart device, such as smart phone, monitor, television, tablet, laptop, etc. Hence, the intraoral scanning system may provide a user-friendly and a time efficient process for generating 3D model of inner geometry and rendering the holistic interactive view or 3D model of the teeth of the subject.

[34] To this end, the intraoral scanning system of the present disclosure makes it feasible to reconstruct inner geometry of the teeth of the subject in 3D by utilizing a limited number of images of the internal regions of the teeth obtained by acquiring 2D Near Infrared (NIR) images along with 3D surface information.

BRIEF DESCRIPTION OF THE DRAWINGS

[35] The present disclosure is illustrated by way of example and not by way of limitation in the figures of the accompanying drawings, in which the like reference numerals indicate like elements and in which:

FIG. 1 illustrates a network environment in which an intraoral scanning system for oral scanning and generating 3D inner geometry is implemented, in accordance with an example embodiment;

FIG. 2 illustrates a block diagram of a handheld intraoral scanner, in accordance with an example embodiment;

FIG. 3 is a schematic diagram of a scanning session, in accordance with an example embodiment;

FIG. 4 illustrates a sequence diagram that depicts generation of 3D surface model, in accordance with an example embodiment;

FIG. 5 illustrates a method for generating set of input parameters for casting objects, in accordance with an example embodiment;

5 FIG. 6 illustrates a method for training a continuous volumetric machine learning model, in accordance with an example embodiment;

FIG. 7A illustrates a sequence diagram that depicts generation of 3D inner geometry, in accordance with an example embodiment;

FIG. 7B shows a schematic illustration of a first type of casting
10 object, in accordance with an example embodiment;

FIG. 7C shows a schematic illustration of a second type of casting object, in accordance with an example embodiment;

FIG. 7D shows a schematic illustration of a generated 3D model of teeth of a subject, in accordance with an example embodiment;

15 FIG. 8 illustrates a method for generating synthetic pixel value using the continuous volumetric machine learning model, in accordance with an example embodiment;

FIG. 9 illustrates a method for generating three-dimensional (3D) inner geometry and 3D surface model of teeth of a subject, in accordance with an
20 example embodiment; and

FIG. 10 illustrates a schematic diagram that depicts an exemplary environment for generating three-dimensional (3D) inner geometry and 3D surface model of teeth and render of interactive 3D graphical representation in real-time, in accordance with an example embodiment.

25 **DETAILED DESCRIPTION**

[36] In the following description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present disclosure. It will be apparent, however, to one skilled in the art that the present disclosure may be practiced without these specific details. In other instances, systems and methods are shown in block diagram form only in order to avoid obscuring the present disclosure.

[37] Reference in this specification to “one embodiment” or “an embodiment” means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the present disclosure. The appearance of the phrase “in one embodiment” in various places in the specification are not necessarily all referring to the same embodiment, nor are separate or alternative embodiments mutually exclusive of other embodiments. Further, the terms “a” and “an” herein do not denote a limitation of quantity, but rather denote the presence of at least one of the referenced items. Moreover, various features are described which may be exhibited by some embodiments and not by others. Similarly, various requirements are described which may be requirements for some embodiments but not for other embodiments.

[38] Some embodiments of the present disclosure will now be described more fully hereinafter with reference to the accompanying drawings, in which some, but not all, embodiments of the disclosure are shown. Indeed, various embodiments of the disclosure may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will satisfy applicable legal requirements. Like reference numerals refer to like elements throughout. As used herein, the terms “data,” “content,” “information,” and similar terms may be used interchangeably to refer to data capable of being transmitted, received and/or stored in accordance with embodiments of the present disclosure. Further, the terms “processor,” “controller” and “processing circuitry” and similar terms may be used interchangeably to refer to the processor capable of processing

information in accordance with embodiments of the present disclosure. Further, the terms “electronic equipment”, “electronic devices” and “devices” are used interchangeably to refer to electronic equipment monitored by the system in accordance with embodiments of the present disclosure. Thus, use of any such terms should not be taken to limit the spirit and scope of embodiments of the present disclosure.

[39] The embodiments are described herein for illustrative purposes and are subject to many variations. It is understood that various omissions and substitutions of equivalents are contemplated as circumstances may suggest or render expedient but are intended to cover the application or implementation without departing from the spirit or the scope of the present disclosure. Further, it is to be understood that the phraseology and terminology employed herein are for the purpose of the description and should not be regarded as limiting. Any heading utilized within this description is for convenience only and has no legal or limiting effect.

[40] As used in this specification and claims, the terms “for example” “for instance” and “such as”, and the verbs “comprising,” “having,” “including” and their other verb forms, when used in conjunction with a listing of one or more components or other items, are each to be construed as open ended, meaning that that the listing is not to be considered as excluding other, additional components or items. Other terms are to be construed using their broadest reasonable meaning unless they are used in a context that requires a different interpretation.

[41] An intraoral scanning system, a method and a computer programmable product are provided for generating a three-dimensional (3D) inner geometry of a subject's teeth and rendering the 3D inner geometry along with 3D surface information of the teeth to provide a holistic view of the teeth.

[42] For instance, an exemplary network environment of the intraoral scanning system for oral scanning and generating 3D inner geometry is provided below with reference to FIG. 1.

[43] FIG. 1 illustrates an exemplary network environment 100 in which an intraoral scanning system 102 for oral scanning and generating 3D inner geometry is implemented, in accordance with an example embodiment. The intraoral scanning system 102 may be used to generate 3D inner geometry of teeth or dental arches within oral cavity of a subject. For example, the intraoral scanning system 102 may be used by a user, such as a person having knowledge of dentistry, for example, dentist, dental technician, and so forth. Further, it is possible that one or more components may be rearranged, changed, added, and/or removed without deviating from the scope of the present disclosure.

10 [44] The intraoral scanning system 102 may include a handheld intraoral scanner 104, one or more processors 106 and a continuous volumetric machine learning model 108. The network environment 100 may further include communication channels 110 that may be configured to establish communicative coupling between components of the intraoral scanning system 102, i.e., the handheld intraoral scanner 104, the one or more processors 106 (referred to as processors 106, hereinafter) and the continuous volumetric machine learning model 108.

[45] The network environment 100 may further include data generated by the intraoral scanning system 102. For example, the network environment 100 may include data captured by the handheld intraoral scanner 104, depicted as plurality of 2D images 112. As shown, the plurality of 2D images 112 generated by the handheld intraoral scanner 104 includes a plurality of two-dimensional (2D) near-infrared (NIR) images 114 and white light images 116. Moreover, data generated by the processors 106 may include three-dimensional (3D) models. The 3D models include three-dimensional (3D) surface model 118 of the teeth of the subject and 3D inner geometry 120 of the teeth of the subject. Moreover, data generated by the continuous volumetric machine learning model 108 may include intensity values and density values 122 for the 3D inner geometry 120.

[46] The intraoral scanning system 102 may be utilized for registration of intraoral scans and generating 3D inner geometry 120 and 3D surface model 118 of teeth. The intraoral scanning system 102 may include multiple components, such as the handheld intraoral scanner 104, the processors 106 and the continuous volumetric machine learning model 108 that may communicate with each other to register the intraoral scans having the 3D inner geometry 120 of teeth.

[47] In an example, the handheld intraoral scanner 104 may include the processors 106 and/or the continuous volumetric machine learning model 108. In another example, the handheld intraoral scanner 104 may be coupled to the processors 106 and/or the continuous volumetric machine learning model 108, wherein the processors 106 and/or the continuous volumetric machine learning model 108 may be located remotely and may perform operations associated with the intraoral scanning system 102. The intraoral scanning system 102 may have enhanced processing capabilities that may be required to process the plurality of 2D NIR images 114 and the visible light images 116 in real time to generate the 3D surface model and the 3D inner geometry.

[48] The handheld intraoral scanner 104 may be configured to capture the plurality of 2D images 112 during a scanning session of the teeth of a subject. The plurality of 2D images 112 may include the plurality of 2D NIR images 114 indicating images of the teeth of the subject captured using light that is emitted at wavelength corresponding to a range of NIR wavelengths. The plurality of 2D NIR images 114 may include images of the teeth of the subject from various viewing angles or viewing points. Further, the plurality of 2D images 112 may include the visible light images 116 indicating images of the teeth of the subject captured using light that is emitted at wavelength corresponding to a range of white light wavelengths or a selected sub spectrum of the visible light such as blue light (400-495 nm). The white light images 116 may also include images of the teeth of the subject from various viewing angles. In certain cases, the plurality of 2D images 112 may also include images captured at other visible light wavelengths.

[49] Based on the captured white light images 116 and visible light information, the processors 106 may be configured to determine 3D surface information of the teeth of the subject. The 3D surface information may be a digital representation of dental arch of the teeth of the subject depicted in a 3D space. The 3D surface information may include, for example, a 3D point cloud data corresponding to the white light images 116. The 3D point cloud data may correspond to 3D real-world coordinates. The 3D surface information may additionally include color texture reflected from the surface of the teeth.

[50] In an example, the handheld intraoral scanner 104 may include a web server that may be configured to communicate via a web network and establish a connection to the communication channels 110. The handheld intraoral scanner 104 may include one or more sensors. The handheld intraoral scanner 104 may be configured to execute the web server to provide visible light information and NIR information obtained from the sensors to the processors 106. In an example, the handheld intraoral scanner 104 may be configured to provide the 3D surface information based on the visible light information and/or the white light images 116 to the processors 106. The handheld intraoral scanning device 104 may further include a processing unit, a memory unit, a communication interface, and additional components. The processing unit, the memory unit, the communication interface, and the additional components may be communicatively coupled to each other. Details of the components of the handheld intraoral scanner 104 are further provided, for example, in FIG. 2.

[51] The processors 106 may include processing capabilities that may be required to process the visible light information or white light images 116 and NIR information or the plurality of 2D NIR images 114. The processors 106 may be configured to establish a connection to the communication channels 110. In an example, the processors 106 may be within the handheld intraoral scanner 104. For example, the processors 106 may receive the visible light information or the white light images 116 from the handheld intraoral scanner 104 or sensors of the handheld intraoral scanner 104 to generate 3D surface information for the teeth of

the subject. Further, based on the 3D surface information derived from the white light images 116, the 3D surface model 118 of the teeth of the subject is generated. The processors 106 may further render the 3D surface information or 3D surface model 118 of the teeth into an interactive 3D graphical representation.

5 **[52]** For example, the processors 106 may be embodied as one or more of various hardware processing means such as a coprocessor, a microprocessor, a controller, a digital signal processor (DSP), a processing element with or without an accompanying DSP, or various other processing circuitry including integrated circuits such as, for example, an ASIC (application specific integrated circuit), an
10 FPGA (field programmable gate array), a microcontroller unit (MCU), a hardware accelerator, a special-purpose computer chip, or the like. As such, in some embodiments, the processors 106 may include one or more processing cores configured to perform independently. A multi-core processor may enable multiprocessing within a single physical package. Additionally or alternatively,
15 the processors 106 may include one or more processors configured in tandem via the bus to enable independent execution of instructions, pipelining and/or multithreading. Additionally or alternatively, the processors 106 may include one or more processors capable of processing large volumes of workloads and operations to provide support for big data analysis. In an example embodiment,
20 the processors 106 may be in communication with the other components of the intraoral scanning system 102 (referred to as system 102, hereinafter) via a bus or the communication network 110 for passing information among components of the system 102.

25 **[53]** In an example, when the processors 106 are embodied as an executor of software instructions, the instructions may specifically configure the processors 106 to perform the algorithms and/or operations described herein when the instructions are executed. However, in some cases, the processors 106 may be a processor specific device (for example, a mobile terminal or a fixed computing device) configured to employ an embodiment of the present disclosure by further
30 configuration of the processors 106 by instructions for performing the algorithms

and/or operations described herein. The processors 106 may include, among other things, a clock, an arithmetic logic unit (ALU) and logic gates configured to support operation of the processors 106. The network environment, such as, 100 may be accessed using the communication network 110.

5 [54] The continuous volumetric machine learning model 108 may include a fully-connected neural network that may generate views or models of 3D scenes, based on a set of 2D images. The continuous volumetric machine learning model 108 may enhance capability of the system 102 in generating 3D inner geometry of the teeth of the subject. The continuous volumetric machine learning model 108
10 may be configured to process the NIR information or the plurality of 2D NIR images 114 to generate output indicating an intensity value and a density value of each point coordinates of each pixel of each of the plurality of 2D NIR images 114. The continuous volumetric machine learning model may be used to generate the 3D inner geometry 120 or inner volume of a tooth from a set of input
15 parameters. The continuous volumetric machine learning model 108 may be based on the concept that the 3D inner geometry 120 may be represented as a continuous field of light density and intensity values, where each pixel in the 2D NIR images 114 may correspond to the casting object in the continuous field of light density and intensity values. By using a deep neural network, the continuous volumetric
20 machine learning model 108 may be able to learn the mapping between the set of input parameters and the light density and intensity values. This may allow it to generate the inner volume 120 of the teeth from a given set of inputs. For example, the continuous volumetric machine learning model 108 may be trained to map values of viewing angle and spatial location to density and intensity using
25 volume rendering. The continuous volumetric machine learning model 108 may receive the plurality of 2D NIR images 114 representing inner region of the teeth of the subject and process the plurality of 2D NIR images 114 to generate the synthetic pixel value for pixels of the plurality of 2D NIR images 114. In an example, the continuous volumetric machine learning model 108 may include a
30 Neural Radiance Field (NeRF) – based model, or a multi-scale anti-aliasing based

NeRF (mip-NeRF) – based model or any other appropriate architecture or variation associated with the same family of Neural Radiance Field models.

[55] In an example, the processors 106 may be configured to generate and render the 3D inner geometry 120 of the teeth within 3D surface model 118 based on the trained continuous volumetric machine learning model. For example, the 3D inner geometry 120 of the teeth may be rendered inside the 3D surface model 118, as an interactive 3D graphical representation. The interactive 3D graphical representation may be rendered on a display unit of a device. The interactive 3D graphical representation may be viewed by the users, such as the dentists on the display to diagnose any disease or anomaly within the teeth and/or treat the subject. A view of the interactive 3D graphical representation may be modified by the users, based on a preference. For example, a perspective of the interactive 3D graphical representation may be changed or the interactive 3D graphical representation may be zoomed-in or zoomed-out as per the preference of the users.

[56] The display may be associated with any user accessible device such as a displaying unit, a monitor, a mobile phone, a smartphone, a tablet, a computer, an artificial reality (XR) device, and the like. In some examples, the display may be a part of the user accessible device. The display of may be, for example, a touch screen display. Additional, different, or fewer components may be provided. Further, it is possible that one or more components may be rearranged, changed, added, and/or removed without deviating from the scope of the present disclosure.

[57] The communication channels 110 may be wired, wireless, or any combination of wired and wireless communication networks, such as cellular, wireless fidelity (Wi-Fi), internet, local area networks, or the like. In accordance with an embodiment, the communication channels 110 may be one or more wireless full-duplex communication channels. In one embodiment, the communication channels 110 may include one or more networks such as a data network, a wireless network, a telephony network, or any combination thereof. It

is contemplated that the data network may be any local area network (LAN), metropolitan area network (MAN), wide area network (WAN), a public data network (e.g., the Internet), short range wireless network, or any other suitable packet-switched network, such as a commercially owned, proprietary packet-switched network, e.g., a proprietary cable or fiber-optic network, and the like, or any combination thereof. In addition, the wireless network may be, for example, a cellular network and may employ various technologies including enhanced data rates for global evolution (EDGE), general packet radio service (GPRS), global system for mobile communications (GSM), Internet protocol multimedia subsystem (IMS), universal mobile telecommunications system (UMTS), etc., as well as any other suitable wireless medium, e.g., worldwide interoperability for microwave access (WiMAX), Long Term Evolution (LTE) networks (for e.g. LTE-Advanced Pro), 5G New Radio networks, ITU-IMT 2020 networks, code division multiple access (CDMA), wideband code division multiple access (WCDMA), wireless fidelity (Wi-Fi), wireless LAN (WLAN), Bluetooth, Internet Protocol (IP) data casting, satellite, mobile ad-hoc network (MANET), and the like, or any combination thereof. The handheld intraoral scanner 104 may be configured to communicate with the processors 106 and the continuous volumetric machine learning model 108 via the communication channels 110.

[58] For example, the subject may require a dental treatment. In such a case, the system 102 may be utilized by a user, such as a dentist to provide the dental treatment to the subject. In an embodiment, the subject may be present at a dental clinic. In such a case, the system 102 may be utilized in a treatment room of the dental clinic. In another embodiment, the subject may have requested for a home visit for the dental treatment. In such a case, the system 102 may be utilized in the home of the subject. To start the dental treatment, the handheld intraoral scanner 104 may be utilized by the user to capture the white light images 116 and the plurality of 2D NIR images 114 of the teeth of the subject, using one or more sensors.

[59] In operation, the one or more sensors of the handheld intraoral scanner 104 may be configured to detect NIR and/or visible light. For example, the NIR light and the visible light may be emitted by the one or more sensors, and the IR light and the visible light may be reflected from the inner region and the surface of the teeth. The characteristics, for example, wavelength, frequency, etc. of the emitted and received NIR light may be stored as NIR information and may be used to generate the plurality of images 2D NIR images 114. Further, the characteristics, for example, wavelength, frequency, etc. of the emitted and received visible light may be stored as visible light information and may be used to generate the white light images 116. The captured NIR information and the visible light information may be provided to the processors 106 of the system 102.

[60] On receiving the NIR information and the visible light information, the processors 106 may determine 3D surface information from the visible light information. Based on the 3D surface information, the processors 106 may be configured to generate the 3D surface model 118 of the teeth of the subject. Details of the capturing the visible light information for the white light images 116, determining the 3D surface information, and generating the 3D surface model 118 is further provided, for example, in FIG. 4.

[61] Once the 3D surface model 118 of the teeth is created, a 3D model of the 3D inner geometry of the teeth may have to be created. In this regard, the processor 106 may capture (or generate) the plurality of 2D NIR images 114 of the internal region of the teeth, using the received NIR information. For example, the processor 106 may filter the received NIR information to generate the plurality of 2D NIR images 114 of the internal region only. Each of the plurality of 2D NIR images 114 may include a plurality of pixels. That is, a 2D NIR image from the plurality of 2D NIR images 114 may be a collection of several pixels indicating different parts of the inner region captured by the 2D NIR image. Further, the processors 106 are configured to determine the set of input parameters for a casting object corresponding to each of the plurality of pixels of each of the

plurality of 2D NIR images 114. The set of input parameters may include spatial location information and viewing angle information corresponding to each of the plurality of pixels. A casting object corresponding to a pixel comprises a plurality of point coordinates along the casting object. The plurality of point coordinates may correspond to different materials, such as enamel or dentin, within the inner region of the teeth. For example, the set of input parameters may include five-dimensional (5D) information that may be fed to the continuous volumetric machine learning model 108 for generating the intensity values and density values 122 for generated grid points within the 3D surface model and a viewing directions for the plurality of 2D NIR images 114. Details of the generating the set of input parameters for the continuous volumetric machine learning model 108 is further provided, for example, in FIG. 5, FIG. 7A and FIG. 8.

[62] The continuous volumetric machine learning model 108 of the system 102 may be configured to receive and process the set of input parameters. The continuous volumetric machine learning model 108 is configured to be trained using the plurality of 2D NIR images 114. Based on the set of input parameters, the continuous volumetric machine learning model 108 may determine the synthetic pixel value 122 based on the intensity values and the density values along the casting object for corresponding pixel of the plurality of pixels. The intensity values and the density values are associated to each of the plurality of point coordinates along the casting object. In an example, the continuous volumetric machine learning model 108 may generate synthetic data, i.e., the synthetic pixel value 122 indicating color intensity value or density value of material corresponding to each of the plurality of pixels of each of the plurality of 2D NIR images 114. For example, synthetic data for a pixel may indicate integration and optimization of data relating to a plurality of point coordinates associated with the pixel. The synthetic data for the pixel may be used to determine a material (i.e., enamel or dentin) relating to the pixel and/or corresponding 3D point in the 3D inner geometry 120.

[63] The processors 106 may be configured to generate and render the 3D inner geometry 120 of the teeth within the 3D surface model 118 based on the trained continuous volumetric machine learning model. For example, the one or more processors is configured to generate grid points within the 3D surface model at different viewing directions of the 3D surface model, and these grid points along with the viewing directions are fed into the trained continuous volumetric machine learning model 108 which then generates intensity values and density values for the corresponding grid points. The trained continuous volumetric machine learning model includes a continuous definition of the 3D inner geometry which means that the trained continuous volumetric machine learning model is configured to evaluate and output intensity values and density values at any given grid points of the 3D inner geometry. A manner in which the continuous volumetric machine learning model 108 generates the synthetic pixel value 122 is described in detail in conjunction with following figures, for example, FIG. 7A and FIG. 8.

[64] After the generation of the 3D inner geometry 120 of the teeth of the subject, the system 102 may be connected with a display to render an interactive 3D graphical representation comprising the 3D inner geometry 120 and the 3D surface model 118. For example, the display may be any display unit capable of rendering interactive 3D graphical representation. The interactive 3D graphical representation of the teeth of the subject may be utilized as human-readable 3D data of the white light images 116 and the plurality of 2D NIR images 114 by the dentist. For example, the interactive 3D graphical representation of the teeth of the subject may be utilized for holistic review of condition of the teeth of the subject from inside as well as on the outside. The interactive 3D graphical representation may be modified or edited, for example, by use of gestures provided as an input by the dentist on the display. Details of the render of the 3D inner geometry 120 and the 3D surface model 118 into the interactive 3D graphical representation are further provided, for example, in FIG. 10.

[65] FIG. 2 illustrates a block diagram 200 of the handheld intraoral scanning device 104, in accordance with an example embodiment. FIG. 2 is explained in conjunction with elements of FIG. 1. The handheld intraoral scanning device 104 may include at least one processing unit (hereinafter, also referred to as “processing unit 202”), a memory unit 204, a web server 206, a monitoring unit 208, a temporary storage unit 210, a scanning feedback unit 212, an input/output (I/O) unit 214, and a communication interface 216.

[0001] The processing unit 202 may be embodied in a number of different ways. For example, the processing unit 202 may be embodied as one or more of various hardware processing means such as a coprocessor, a microprocessor, a controller, a digital signal processor (DSP), a processing element with or without an accompanying DSP, or various other processing circuitry including integrated circuits such as, for example, an ASIC (application specific integrated circuit), an FPGA (field programmable gate array), a microcontroller unit (MCU), a hardware accelerator, a special-purpose computer chip, or the like. In an embodiment, the processing unit 202 may be embodied as a high-performance microprocessor having series of System on Chip (SOCs) which includes relative powerful and power-efficient Graphics Processing Units (GPUs) and Central Processing Units (CPUs) and a small form factor. As such, in some embodiments, the processing unit 202 may include one or more processing cores configured to perform independently. A multi-core processor may enable multiprocessing within a single physical package. Additionally, or alternatively, the processing unit 202 may include one or more processors configured in tandem via the bus to enable independent execution of instructions, pipelining and/or multithreading.

[66] In some embodiments, the processing unit 202 may be configured to detect NIR and visible light, using the I/O unit 214 during the scanning session of teeth of a subject, such as a patient requiring a dental treatment. The detected NIR and visible light may be used to generate the plurality of 2D images 112, such as the plurality of 2D NIR images 114 and the white light images 116. The plurality of 2D images 112 may include images of the teeth of the subject from various

angles or viewing points. For example, the processing unit 202 may be configured to generate 3D surface information for the teeth based on the white light images 116, and the continuous volumetric machine learning model may generate the intensity values and density values 122 based on the plurality of 2D NIR images 114.

[67] In an example embodiment, the processing unit 202 may be in communication with the memory unit 204 via a bus for passing information among components of the handheld intraoral scanner 104.

[68] The memory unit 204 may be non-transitory and may include, for example, one or more volatile and/or non-volatile memories. In other words, for example, the memory unit 204 may be an electronic storage device (for example, a computer readable storage medium) comprising gates configured to store data (for example, bits) that may be retrievable by a machine (for example, a computing device like the processing unit 202). The memory unit 204 may be configured to store information, data, content, applications, instructions, or the like, for enabling the apparatus to carry out various functions in accordance with an example embodiment of the present disclosure. For example, the memory unit 204 may be configured to store the detected NIR and the detected visible light after the scanning session of the teeth is finished. The detected NIR and the visible light after the scanning session may be stored as NIR information and visible light information, respectively. In certain cases, the memory unit 204 may be configured to store compressed NIR information and visible light information. In some embodiments, the memory unit 204 may be configured to store calibration data required to measure the detected NIR and the visible light to generate the NIR information, the visible light information, the white light images 116 and/or the plurality of 2D NIR images 114. As exemplarily illustrated in FIG. 2, the memory unit 204 may be configured to store instructions for execution by the processing unit 202. As such, whether configured by hardware or software methods, or by a combination thereof, the processing unit 202 may represent an entity (for example, physically embodied in circuitry)

capable of performing operations according to an embodiment of the present disclosure while configured accordingly. Thus, for example, when the processing unit 202 is embodied as the microprocessor, the processing unit 202 may be specifically configured hardware for conducting the operations described herein.

5 Alternatively, as another example, when the processing unit 202 is embodied as an executor of software instructions, the instructions may specifically configure the processing unit 202 to perform the algorithms and/or operations described herein when the instructions are executed. The processing unit 202 may include, among other things, a clock, an arithmetic logic unit (ALU) and logic gates
10 configured to support operation of the processing unit 202.

[69] The web server 206 may be a software, a hardware or a combination thereof that may be configured to store and provide data to a web browser associated with the processor 106 and/or the continuous volumetric machine learning model 108. For example, the visible light information and the NIR
15 information may be provided to the web browser of the processors 106 via the web server 206. As the web server 206 may be accessed by any web browser, a need of installation of an additional software by the processors 106, to connect to the web server 206 may be eliminated. The web server 206 may communicate to one of the communication channels 110 via a web network. In an example, the
20 web server 206 and the processors 106 and/or the continuous volumetric machine learning model 108 may communicate to a common wireless full-duplex communication channel via the web network for transmission and reception of the visible light information and the NIR information. The web server 206 and the web browser may communicate via, for example, Hypertext Transfer Protocol
25 (HTTP), Simple Mail Transfer Protocol (SMTP), or File Transfer Protocol (FTP). Once the web server 206 and the web browser are connected, the web server 206 may provide a web application on the web browser.

[70] The monitoring unit 208 may be a software, a hardware or a combination thereof that may be configured to monitor a bandwidth of one of the
30 communication channels 110 (such as the wireless full-duplex communication

channel) via which the handheld intraoral scanning device 104 and the processors 106 and/or the continuous volumetric machine learning model 108 may be connected. Moreover, the monitoring unit 208 may be configured to monitor a connection of one of the communication channels 110 via which the handheld
5 intraoral scanning device 104 and the processors 106 and/or the continuous volumetric machine learning model 108 may be connected.

[71] In an embodiment, if the monitoring unit 208 determines that the bandwidth of the communication channels 110 is below a minimum bandwidth, the monitoring unit 208 may provide such information to the processing unit 202.
10 The processing unit 202 may down-sample the visible light information and the NIR information, based on the received information. In another embodiment, if the monitoring unit 208 may determine that the bandwidth of the communication channels 110 is below the minimum bandwidth for longer than a maximum period, the monitoring unit 208 may provide such information to the processing
15 unit 202. The processing unit 202 may compress and store the visible light information and the NIR information into the memory unit 204. In some embodiments, if the monitoring unit 208 may determine that the connection between the handheld intraoral scanning device 104 and the processors 106 and/or the continuous volumetric machine learning model 108 is lost, the monitoring
20 unit 208 may provide such information to the processing unit 202. In such a case, the processing unit 202 may compress and store the visible light information and the NIR information into the memory unit 204.

[72] The temporary storage unit 210 may be a software, a hardware or a combination thereof that may be configured to store the visible light information and the NIR information when the bandwidth of the communication channels 110
25 (such as the wireless full-duplex communication channel) is determined to be below the minimum bandwidth. The temporary storage unit 210 may further transmit the stored the visible light information and the NIR information to the processors 106 and/or the continuous volumetric machine learning model 108
30 when the bandwidth is determined to be above or equal the minimum bandwidth.

Examples of the temporary storage unit 210 may include, but may not be limited to, a random access memory (RAM), or a cache memory.

[73] The scanning feedback unit 212 may be a software, a hardware or a combination thereof that may be configured to receive status input from the monitoring unit 208. Based on the received status input, the scanning feedback unit 212 may provide a scanning feedback signal to the user, such as a dentist, of the handheld intraoral scanner 104. In an embodiment, the scanning feedback signal is used to provide guidance to the user regarding an area of the teeth where a scanning quality of the scanning session is low and adequate visible light information and/or the NIR information is not received. For example, the scanning feedback unit 212 may provide the scanning feedback signal as, for example, an acoustic feedback signal, a haptic feedback, or a visual feedback.

[74] The I/O unit 214 may include circuitry and/or software that may be configured to provide output to the user of the handheld intraoral scanning device 104 and receive, measure or sense input information. The I/O unit 214 may include a speaker 214A, a vibrator 214B, a projector unit 214C, and one or more sensors 214D. In an embodiment, the speaker 214A may be configured to output the acoustic feedback signal to guide the user. The vibrator 214B may be, for example, a transducer configured to convert the scanning feedback signal that may be an electrical signal into a mechanical output, such as the haptic feedback in form of vibrations to guide the user.

[75] As may be understood, the hand-held intraoral scanner 104 may be configured to detect the NIR and the visible light that may be reflected from the teeth of the subject. In this regard, the projector unit 214C may be configured to output one or more visible or white colored wavelength pulses, and one or more NIR wavelength pulses. For example, the white colored or visible wavelength pulses and the NIR wavelength pulses may be casted onto the teeth to illuminate the teeth of the subject, such as the patient. Further, the white colored wavelength pulses and the wavelength pulses may be reflected or refracted from the surface

and/or the inner region of the teeth. The one or more sensors 214D may be configured to detect visible or white colored wavelength pulses and NIR wavelength pulses that may be reflected and/or refracted from the surface or inner region of the teeth. In an example, the one or more sensors 214D may include one
5 or more image sensors, such as cameras. For example, the image sensors may be configured to generate the white light images 116 and the plurality of 2D NIR images 114 based on the illumination of the teeth using the NIR and the visible light.

[76] For example, the I/O unit 214 may include light emitting devices, such
10 as light emitting diodes (LEDs) to provide the scanning feedback signal in form of light to guide the user or the dentist. For example, plurality of LEDs may be divided into a left group of LEDs and a right group of LEDs to emit a flash of light based on which side of the teeth is not properly scanned. In an example, the I/O unit 214 may include addition input and output devices, such as buttons to
15 receive operational signals from the user, power sensor, battery charge sensor, etc.

[77] The communication interface 216 may comprise input interface and output interface for supporting communications to and from the handheld intraoral scanner 104. The communication interface 216 may be a device or circuitry embodied in either hardware or a combination of hardware and software that is
20 configured to receive and/or transmit data to/from the handheld intraoral scanner 104. In this regard, the communication interface 216 may include, for example, an antenna (or multiple antennae) and supporting hardware and/or software for enabling communications with a wireless communication network. Additionally, or alternatively, the communication interface 216 may include the circuitry for
25 interacting with the antenna(s) to cause transmission of signals via the antenna(s) or to handle receipt of signals received via the antenna(s). In some environments, the communication interface 216 may alternatively or additionally support wired communication. As such, for example, the communication interface 216 may include a communication modem and/or other hardware and/or software for

supporting communication via cable, digital subscriber line (DSL), universal serial bus (USB) or other mechanisms.

[78] In operation, the hand-held intraoral scanner 104 may be configured to capture 2D images or 2D scan of an oral cavity of the subject. In an example, a user of the hand-held intraoral scanner 104 may perform scanning session of the oral cavity of the subject. The hand-held intraoral scanner 104 may illuminate the oral cavity or the teeth of the subject using one or more white colored wavelength pulses and one or more NIR wavelength pulses, simultaneously. Such one or more white colored wavelength pulses and one or more NIR wavelength pulses may be generated by the projector unit 214C. For example, at a single time frame, pulses emitted by the projector unit 214C may include multiple wavelength, such as one or more white colored wavelength pulses and one or more NIR wavelength pulses. A number of the NIR wavelength pulses may be less than a number of the white colored wavelength pulses. In an embodiment, at a given time frame, a number of the white colored wavelength pulses may be three times of a number of the NIR wavelength pulses. It may be noted that such ratio of the number of the white colored wavelength pulses and the NIR wavelength pulses is only exemplary and should not be construed as a limitation. In other examples, the projector unit 214C may be configured to illuminate the teeth using different number of the white colored wavelength pulses and the NIR wavelength pulses in a single time frame, using only white colored wavelength pulses or the NIR wavelength pulses in a single time frame, and so forth. In certain cases, the projector unit 214C may also be configured to illuminate the teeth or oral cavity of the subject using visible light, such as corresponding to blue wavelength pulses, red wavelength pulses, and/or green wavelength pulses.

[79] After the illumination of the teeth or the oral cavity, the reflected or refracted wavelength pulses may be detected by the one or more sensors 214D. For example, the one or more sensors 214D may be image sensors. The one or more sensors 214D may be configured to generate the white light images 116

using the detected white light wavelength pulses, and the plurality of 2D NIR images 114 using the detected NIR wavelength pulses. The hand-held intraoral scanner 104 may be configured to transmit the detected visible light information (or the detected white light wavelength pulses) and the NIR information (or the detected NIR wavelength pulses) to the processors 106. Details of the operations performed by the processors 106 are further described in conjunction with FIG. 3.

[80] FIG. 3 is a schematic diagram 300 of a scanning session, in accordance with an example embodiment. The schematic diagram 300 may include a user, such as a dentist 302 of the system 102, and a subject, such as a patient 304 whose teeth 306 may require treatment. The handheld intraoral scanner 104 may be utilized by the dentist 302 to capture a plurality of 2D images 308 of the teeth 306 of the patient 304. During the scanning session, the plurality of 2D images 308 including, for example, white light images, depicted as a white light image 308A, and 2D NIR images, depicted as a 2D NIR image 308B, of the teeth 306 of the patient 304 may be captured to generate 3D information 310 comprising, for example, the 3D surface model 118 and 3D inner geometry 120 of the teeth 306 of the patient 304.

[81] In an exemplary scenario, the dentist 302 and the patient 304 may be present at the dental clinic. The scanning session of the teeth 306 of the patient 304 may be initiated by the dentist 302 to perform a diagnosis of a condition of the teeth 306 of the patient 304. The handheld intraoral scanner 104 may include the projector unit 214C and the one or more sensors 214D. For example, the handheld intraoral scanner 104 is placed inside a mouth or oral cavity of the patient 304 and moved around the teeth 306 and gums of the patient 304 to record oral topography of the patient 304. In this regard, the projector unit 214C may illuminate the oral cavity of the patient 304 using visible light (or white colored light) wavelength pulses and NIR wavelength pulses. Furthermore, the one or more sensors, such as image sensors may capture white light images and 2D NIR images of the teeth 306.

[82] In an example, the handheld intraoral scanner 104 may record a size and a shape of each tooth, an interdental separation, an appearance of a surface of a palate, gums, implants, prostheses, and other elements that make up an interior of the oral cavity of the patient 304. In an embodiment, the handheld intraoral scanner 104 may be moved multiple times over the teeth 306 and the gums of the patient 304 to capture good-quality 2D images. For example, the plurality of 2D images 308 may include a white light image 308A detected based on illumination of the teeth 306 by visible light wavelength pulses, and a 2D NIR image 308B detected based on illumination of the teeth 306 by NIR wavelength pulses. It may be noted that the plurality of 2D images 308 may include plurality of white light images captured using the visible light wavelength pulses and plurality of NIR images captured using NIR wavelength pulses. The plurality of 2D images 308, such as the white light image 308A and the 2D NIR image 308B may be utilized to generate 3D information 310, particularly, 3D surface model 118 and 3D inner geometry 120 of the teeth 306.

[83] Once the scanning session starts, the plurality of 2D images 308, such as the white light image 308A and NIR image 308B may be captured. The handheld intraoral scanner 104 may be configured to process the plurality of 2D images 308 by applying calibration parameters and filtering noise from the plurality of 2D images 308. For example, the calibration parameters of the one or more sensors 214D may be applied to process the plurality of 2D images 308. Moreover, the noise may be filtered from the plurality of 2D images 308. Based on the processing of the plurality of 2D images 308, the handheld intraoral scanner 104 may be configured to provide high quality 2D images to the processors 106 of the system 102. Details of the generation of the 3D surface model 118 of the teeth 306 are further provided, for example, in FIG. 4.

[84] FIG. 4 illustrates a sequence diagram 400 that depicts generation of the 3D surface model 118, in accordance with an example embodiment. FIG. 4 is explained in conjunction with elements of FIG. 1, FIG. 2, and FIG. 3. The sequence diagram 400 may include the handheld intraoral scanner 104 and the one

or more processors 106. The sequence diagram 400 may depict operations performed by at least one of the handheld intraoral scanner 104 and the processors 106.

5 **[85]** At step 402, the projector unit 214C of the handheld intraoral scanner 104 may illuminate the oral cavity of the patient 304. The projector unit 214C illuminates the teeth 306 and other parts of the oral cavity using visible light or white light wavelength pulses and NIR wavelength pulses. Details of the illumination of the oral cavity and the teeth 306 of the patient 304 are provided in detail in, for example, in FIG. 2 and FIG. 3.

10 **[86]** At step 404, the one or more sensors 214D of the handheld intraoral scanner 104 may detect visible light and NIR. Such detected visible light and NIR may be reflected or refracted from surface or inner region of the teeth 306. In other words, the one or more sensors 214D may detect visible light information and NIR information reflected or refracted from the teeth 306.

15 **[87]** At step 406, the one or more sensors 214D of the handheld intraoral scanner 104 may capture white light images 116 and plurality of 2D NIR images 114 (referred to as NIR images 114, hereinafter). For example, the one or more sensors 214D may include image sensors that may be configured to capture the white light images 116 based on the detected visible light information and the
20 NIR images 114 based on the detected NIR information. In this manner, the one or more sensors 214D may be configured to generate the plurality of 2D images 112 of the teeth 306.

25 **[88]** At step 408, the plurality of 2D images 112 of the teeth 306 are received by the processors 106. As mentioned above, the plurality of 2D images 112 of the teeth 306 may include the white light images 116 and the NIR images 114.

[89] At step 410, 3D surface information is determined using the visible light information. In particular, the white light images 116 may be processed to

determine the 3D surface information. In an example, in-focus measurements in the white light images 116 may be determined and projected features of the teeth 306 may be tracked across the white light images 116. For example, a correspondence function may be performed or solved to triangulate depth information for the teeth 306 and determine the 3D surface information for the teeth 306.

[90] At step 412, the 3D surface model 118 is generated for the teeth 306 based on the 3D surface information. In an example, a 3D patch may be generated corresponding to a part of the surface of the teeth 306 by accessing calibration data of the one or more sensors 214D and transforming the 3D surface information corresponding to the part into real-world 3D coordinates and texture information. For example, different 3D patches may be generated for different parts of the surface of the teeth 306 using the white light images 116. In certain cases, there may be an overlap in parts covered by different 3D patches. For example, the 3D patch for the part of the surface of the teeth 306 may be registered or associated with one or more previously generated 3D patches for other parts and/or overlapping parts of the surface of the teeth 306 by locating corresponding data points. Thereafter, the 3D patch for the part may be fused with the one or more previously generated 3D patches relating to other parts of the surface of the teeth 306 to generate the 3D surface model 118. In an example, the 3D surface model 118 may include 3D points within voxels in a signed distance field, and the signed distance field may be converted into a 3D mesh for rendering the 3D surface model 118. The 3D surface model 118 may also include texture data of the surface of the teeth 306.

[91] After the 3D surface model 118 of the teeth 306 is generated, the processors 106 may be configured to generate the 3D inner geometry 120 of the teeth 306. Details of the generation of the 3D inner geometry may be

[92] FIG. 5 illustrates a method 500 for generating a set of input parameters for casting objects, in accordance with an example embodiment. The

set of input parameters are generated by the processors 106 using the NIR images 114. For example, the set of input parameters are generated for the continuous volumetric machine learning model 108. FIG. 5 is explained in conjunction with elements of FIG. 1, FIG. 2, FIG. 3 and FIG.4.

5 **[93]** At 502, a relative position between the hand-held intraoral scanner 104 and the teeth 306 is estimated. Once the processors 106 receive the NIR images 114, the processors 106 may be configured to estimate a relative pose between the teeth 306 and the hand-held intraoral scanner 104 for each of the NIR images 114.

10 **[94]** Further, the estimated relative pose indicates, for example, viewing angle information and spatial location information of the handheld intraoral scanner, i.e. the position of the image sensor unit within the handheld intraoral scanner in relation to the subject's teeth, while capturing a corresponding NIR
15 image from the NIR images 114. The viewing angle information for an NIR image indicates an angle formed between the hand-held intraoral scanner 104 and the teeth 306, at a time of capturing the NIR image. Moreover, the spatial location information indicates a location at which the hand-held intraoral scanner 104 is physically positioned relative to a position of the teeth 306.

20 **[95]** In an example, the projector unit 214C may illuminate the oral cavity and the teeth 306 using NIR wavelength pulses as well as white light wavelength pulses in a single time frame. As a result, the NIR images 114 may be captured with or between the white light images 116. In order to estimate relative position between the hand-held intraoral scanner 104 and the teeth 306 for a given NIR image, poses of one or more white light images associated with the given NIR
25 image may be interpolated. For example, the one or more white light images associated with the given NIR image may be captured during a same time frame as the given NIR image. Based on the poses of one or more white light images associated with the given NIR image, viewing angle information and spatial location information for the given image may be determined.

[96] At 504, the set of input parameters are determined based on the estimated relative pose. For example, the set of input parameters comprises spatial location information and viewing angle information for the casting objects that corresponds to each of the plurality of pixels for each of the NIR images 114. In this regard, the processors 106 may be configured to determine an object region for each of the NIR images 114. The object region in an NIR image may include a plurality of pixels indicating the teeth 306 or a part of the teeth 306 that may be captured within the corresponding NIR image. In an example, an object region or a plurality of pixels relating to the teeth 306 within an NIR image may be determined by projecting the 3D surface model 118 back to the NIR image and determining which pixels of the NIR image are hit by the projection. In such a case, the plurality of pixels of the object region may be identified to be pixels within the pixels that are hit by the projection of the 3D surface model 118 back to the NIR image. In another example, an object region or a plurality of pixels relating to the teeth 306 within an NIR image may be determined by performing edge detection on the NIR image to identify edges of the teeth 306 or part of the teeth 306 covered by the NIR image. In such a case, the plurality of pixels of the object region may be identified to be pixels within a detected edge in the NIR image.

[97] Once the plurality of pixels relating to the teeth 306 are determined for each of the NIR images 114, a casting object may be determined for each of the plurality of pixels. For example, the casting object may be a camera ray or a cone that may be projected through each of the plurality of pixels of each of the NIR images 114. Further, a casting object projected through a pixel may include a plurality of point coordinates. As such, the plurality of point coordinates of the casting object may correspond to different depth in the inner region of the teeth 306, and may relate to different material in the inner region of the teeth 306. Based on the plurality of point coordinates for a casting object, the set of input parameters corresponding to the casting object may be determined.

[98] In an example, the set of input parameters indicating the casting object may include the spatial location information (for example, values of (x, y, z) coordinates), and viewing angle information (for example, values of (θ, φ)). The set of input parameters indicating the casting object may be

5 determined based on the estimated relative position while capturing a corresponding NIR image. For example, the spatial location information for the casting object may indicate spatial location of each of the plurality of point coordinates (x, y, z) . Moreover, the viewing angle information for the casting object may indicate a viewing direction in polar coordinates θ , and φ , as seen by

10 the plurality of point coordinates or the casting object during capturing of the corresponding NIR image. In this manner, the set of input parameters may be determined for each casting object projected through each of the plurality of pixels of the object region within each of the NIR images 114.

[99] FIG. 6 illustrates a method 600 for training the continuous volumetric machine learning model 108, in accordance with an example embodiment. The

15 continuous volumetric machine learning model 108 may be trained based on the plurality of 2D NIR images 114. FIG. 6 is explained in conjunction with elements of FIG. 1, FIG. 2, FIG. 3, FIG. 4 and FIG. 5.

[100] For example, the plurality of NIR images 114 may be captured by the

20 hand-held intraoral scanner 104 from different positions. In an example, casting objects may be generated from each of the plurality of pixels of the plurality of NIR images 114 captured from different positions of the handheld intraoral scanner. For example, the image sensors, such as a camera of the hand-held intraoral scanner 104 may capture the 2D NIR images 114. Such 2D pixels of

25 the plurality of 2D NIR images 114 may contain part of tooth or teeth, for example, belonging to the patient 304. It may be noted that the casting objects may include corresponding plurality of point coordinates. Moreover, 3D surface model 118 may be generated for the subject's teeth based on white light images.

[101] In an example, the determined data, such as the plurality of NIR images 114, the positions, the casting objects, the plurality of point coordinates, the 3D surface model, etc. may form training data for training the continuous volumetric ML model 108. The training data may be generated for each of the patient so that the continuous volumetric ML model 108 is trained differently for every patient data.

[102] At step 602, the set of input parameters that corresponds to the casting objects for the plurality of 2D NIR images 114 is received. For example, the set of input parameters may be determined based on the known or calibrated positions of the sensors, such as the image sensors or the cameras of the hand-held intraoral scanner 104 that may be used to capture the plurality of 2D NIR images 114. It may be noted that the casting objects may be casted or projected through 2D pixels containing part of tooth within the plurality of 2D NIR images 114. Moreover, each of the casting objects includes the plurality of point coordinates.

[103] In an example, the 3D surface model 118 may be previously generated for the subject, for example, using the white light images 116 that may be captured along with the 2D NIR images 114. Further, based on the 3D surface model 118, grid points may be generated within which an artificial pixel indicative of the corresponding point coordinates may have to be plotted or rendered. Moreover, NIR camera rays or NIR wavelength rays are intersected with outer geometry, i.e., the 3D surface model 118 to determine which parts of the camera rays or NIR rays (and corresponding 2D pixels of the plurality of NIR images 114) lay inside the outer boundary of tooth or teeth.

[104] The continuous volumetric ML model 108 may be trained to determine synthetic pixel values of artificial pixels for rendering 3D inner geometry 120. Moreover, the set of input parameters such as, spatial location parameters (x, y, z) and viewing direction or viewing angle parameters (θ, ϕ) are determined based on known calibrations of the image sensors. In an

example, the several 3D positions (x_i, y_i, z_i) along the casting objects that lie inside the outer boundary or within the 3D training surface model 118 of corresponding tooth may be sampled to determine the plurality of point coordinates. Further, based on combining 3D positions (x_i, y_i, z_i) indicating spatial locations within viewing direction (θ, ϕ) for the plurality of point coordinates, the set of input parameters may be determined for the plurality of point coordinates as $(x_i, y_i, z_i, \theta, \phi)$.

[105] At step 604, an intensity value and a density value for each of the plurality of point coordinates are generated based on the continuous volumetric ML model 108. For example, the continuous volumetric ML model 108 may use the received set of input parameters. Based on the set of input parameters relating to the point coordinates of a casting object, the continuous volumetric ML model 108 may be trained to determine or output an intensity value and a density value (I_i, σ_i) for each of the point coordinates of the casting object based on corresponding input parameters $(x_i, y_i, z_i, \theta, \phi)$. As may be noted, the continuous volumetric ML model 108 may be trained to determine intensity values and density values for point coordinates of multiple different casting objects, belonging to the 2D NIR images 114 and relating to the subject.

[106] At step 606, a synthetic pixel value is determined for each of the casting objects based on the corresponding determined intensity values and density values for each of the plurality of point coordinates. For example, a synthetic pixel value for a casting object may be determined as an expected intensity $\hat{I}(r)$ along a camera ray casting object that captured the corresponding 2D pixel within the corresponding 2D NIR image. The expected intensity $\hat{I}(r)$ may be calculated based on an integration of the determined intensity value and

density values for the point coordinates relating to the casting object. Subsequently, synthetic pixel values is the expected intensity $\hat{I}(\mathbf{r})$ values.

[107] At step 608, a loss function between the synthetic pixel values and a corresponding true pixel value of the plurality of pixels of the plurality of 2D NIR images 114 is minimized. For example, the synthetic pixel values generated for the casting objects may indicate intensity values for a corresponding artificial pixel.

[108] In an example, given the synthetic pixel values for plurality of casting objects, indicating a batch or a set of computed expected intensities $\hat{I}(\mathbf{r})$, the loss function may be computed for a synthetic pixel value based on corresponding true pixel value as:

$$\mathcal{L} = \sum_{\mathbf{r} \in \mathcal{R}} \|\hat{I}(\mathbf{r}) - I(\mathbf{r})\|_2^2,$$

where $\mathcal{R}$ is the set casting objects and $I(\mathbf{r})$ is the ground truth intensities observed in the plurality of 2D NIR images 114 used for training.

[109] The loss function is generated based on the understanding that the casting objects in a set of casting objects $\mathcal{R}$ does not necessarily come from a same viewpoint of camera or image sensor. Further, the loss function corresponding to multiple synthetic values are used to update the weights of the continuous volumetric ML model 108 and iterate the steps of training as described in FIG. 8 until a convergence or a fixed amount of iterations is performed.

[110] To this end, the continuous volumetric ML model 108 may be trained each time a set of input parameters for a different patient is generated. Due to re-training of the continuous volumetric ML model 108, the synthetic pixel value is generated for the patient, such that the synthetic pixel value is user-specific.

Moreover, training data required for training the continuous volumetric ML model 108 is substantially less.

[111] The set of input parameters is provided to the continuous volumetric machine learning model 108 for solving a continuous volumetric scene function.
5 A manner in which the continuous volumetric machine learning model 108 operates is described in detail in conjunction with FIG. 7A.

[112] FIG. 7A is a sequence diagram 700 that depicts generation of the 3D inner geometry 120, in accordance with an example embodiment. The 3D inner geometry 120 is generated by the processors 106, using the continuous volumetric machine learning model 108. FIG. 7A is explained in conjunction with elements
10 of FIG. 1, FIG. 2, FIG. 3, FIG. 4, FIG. 5 and FIG. 6. The sequence diagram 700 may include the processors 106 and the continuous volumetric machine learning model 108. The sequence diagram 700 may depict operations performed by the processors 106 and the continuous volumetric machine learning model 108 of the
15 system 102 for generating the 3D inner geometry 120 and a 3D model of the teeth 306 of the subject 304.

[113] At step 702, the plurality of 2D NIR images 114 are captured. For example, the processors 106 may capture the NIR images 114 indicating internal region of the teeth 306 of the subject. The NIR images 114 may be captured
20 using the one or more sensors 214D, particularly, image sensors, from the NIR information received from the handheld intraoral scanner 104. In an example, based on calibration information of the image sensors and the NIR wavelength pulses detected by the image sensors, the NIR images 114 may be captured.

[114] At step 704, the set of input parameters for the continuous volumetric machine learning model 108 is determined. The set of input parameters may
25 include values corresponding to input variables, such as spatial location and viewing angle and direction, i.e., x, y, z, θ and φ . The processors 106 may determine the set of input parameters for each casting object corresponding to

each of the plurality of pixels of each of the NIR images 114. The set of input parameters may be generated to optimize a function of the continuous volumetric machine learning (ML) model 108. For example, a continuous volumetric scene function, $F(x, y, z, \theta, \phi)$, of the continuous volumetric ML model 108 may be optimized. Details of the generation of the set of input parameters are provided, for example, in FIG. 5.

[115] At step 706, the plurality of NIR images 114 and the set of input parameters are received by the continuous volumetric ML model 108. For example, the processors 106 may forward the NIR images 114 and the set of input parameters to the continuous volumetric ML model 108. The continuous volumetric ML model 108 may be implemented as the continuous volumetric scene function, where the continuous volumetric scene function or the continuous volumetric ML model 108 is a fully-connected deep neural network.

[116] In an example, the continuous volumetric ML model 108 may include a neural radiance field (NeRF) based neural network model. In certain cases, the continuous volumetric ML model 108 may include an enhanced version of NeRF, such as mip-NeRF. The mip-NeRF may be used to reduce aliasing, show fine details in the 3D inner geometry, and reduces error rates. It may be understood that when the casting object is projected as a ray, the continuous volumetric ML model 108 may include the NeRF based neural network model, whereas when the casting object is projected as a cone, the continuous volumetric ML model 108 may include the mip-NeRF based neural network model. An advantage of using a cone shape- based casting object is to incorporate a model consideration which represents a tele-centric perspective of the intraoral scanner when acquiring the 2D NIR images. This may improve the end result.

[117] The set of input parameters relates to casting objects projected through each of plurality of pixels of the NIR images 114. The plurality of pixels of the NIR images 114 may correspond to the teeth 306, such as a part of

the teeth 306 or a point on the teeth 306. Further, a casting object, i.e. a ray or a cone may be projected through each of the plurality of pixels of each of the NIR images 114. Details of types of a casting object projected through a pixel are further provided, for example in FIG. 7B and FIG. 7C.

5 [118] At step 708, an intensity value and a density value for each of the plurality of point coordinates along a casting object are generated using the set of input parameters. For example, casting objects are projected from each of the plurality of pixels of each of the NIR images 114. Each casting object may include a plurality of point coordinates that may have corresponding value of,
10 for example, spatial location and viewing angle or viewing direction. Further, each pixel projecting the casting object may also have a corresponding value of, for example, color intensity and texture information.

[119] The continuous volumetric ML model 108 may be configured to determine the intensity value and the density value of each of the point
15 coordinates based on obtained NIR images 114 and the set of input parameters. For example, the continuous volumetric scene function $F(x, y, z, \theta, \phi)$ may be optimized for sampled point coordinates of a casting object based on input parameters (x, y, z) corresponding to spatial location information of the point coordinate and the input parameters (θ, ϕ) corresponding to viewing angle
20 information of the point coordinate. The optimized continuous volumetric scene function $F(x, y, z, \theta, \phi)$ may be used to generate an output. The one or more processors may be configured to generate grid points within the 3D surface model at different viewing directions of the 3D surface model. The 3D surface model is determined based on the received visible light information. The model
25 108 receives the generated grid points and the viewing directions of the grid points and determines intensity values and density values for each of the grid points and for the different viewing directions. Based on the intensity values and density values, the one or more processors is configured to generate a 3D inner

geometry of the 3D surface model by mapping the intensity values and density values onto the corresponding grid points for the corresponding viewing directions of the 3D surface model. The 3D surface model including the 3D inner geometry may then be displayed. Based on a user input, the one or more processors is configured to zoom into the 3D surface model to closely inspect the 3D inner geometry of the 3D surface model. Based on a user input, the one or more processors is configured to rotate the 3D surface model in such a way that the 3D inner geometry is similar rotated.

[120] At step 710, a synthetic pixel value for a casting object is determined. For example, an intensity value and a density value of each of the plurality of point coordinates of a casting object may be integrated to determine a synthetic pixel value corresponding to the casting object. The synthetic pixel values may be determined for plurality of casting objects corresponding to plurality of pixels of each of the NIR images 114. The synthetic pixel value of the casting object may indicate pixel information, such as color, depth, transparency, location, texture, etc. corresponding to an artificial pixel. Such artificial pixel may be projected as an artificial point in the 3D inner geometry of the teeth 306 of the subject. For example, for a 2D pixel from the NIR images 114, a plurality of artificial points or artificial pixels may be determined based on the processing of the NIR images 114 using the continuous volumetric ML model 108.

[121] At step 712, the processors 106 may be configured to generate grid points within 3D surface model 118 at different viewing directions of the 3D surface model 118 based. It may be noted that the processors 106 may generate the 3D surface model 118 for the teeth 306 of the subject using visible light information and/or white light images 116 received from the handheld intraoral scanner 104. The generated grid points and viewing directions are fed 714 into the continuous volumetric machine learning model 108, and the model 108 is then configured to determine intensity values and density values for the corresponding grid points and forward 716 the values to the one or more processors 106. The one or more processors 106 is then configured to generate

718 an 3D inner geometry of the 3D surface model 118 based on the intensity values and the density values for the corresponding grid points.

[122] In an embodiment, the intensity value and the density values of the plurality of point coordinates may indicate nature of material at the plurality of point coordinates within a pixel of an NIR image.

[123] Further, the processors 106 may be configured to identify a distinction between enamel and dentin and a boundary between the enamel and the dentin within the 3D inner geometry 120 of teeth 306 based on determined intensity values and the density values that have been determined based on the grid points within the 3D surface model and the viewing of the grid points. The processors 106 may identify change in material in the 3D inner geometry based on a change in the intensity values and the density values between the grid points. A change in the intensity values and the density values that may be above a certain change threshold may correspond to a change in material. The one or more processors may be configured to determine whether an intensity value and a density value corresponds to a dental feature, such as an anatomy feature, a disease feature or a mechanical feature. The anatomy feature may be an enamel, a dentine, or a pulp. The disease feature may be a crack or a caries. The mechanical feature may be a filling and/or a composite restoration.

[124] To this end, the 3D inner geometry 120 may be generated within the 3d surface model 118, such that the 3D inner geometry 120 identifies different materials within the dentition of teeth 306 of the subject. The different materials may correspond to, but may not be limited to, enamel and dentin. In some cases, the different materials may correspond to, any prosthetic implant or any filling that may inserted within a tooth or material of an artificial tooth.

[125] Referring to FIG. 7B, there is shown a schematic illustration 720 of a first type of casting object, in accordance with an example embodiment. Referring to FIG. 7C, there is shown a schematic illustration 730 of a second type of casting object, in accordance with another example embodiment. For

example, the first type of the casting object is a ray, and the second type of casting object is a cone. For example, an NIR image 718 may be an image from the NIR images 114 that may be captured by the processors 106 using the image sensors and NIR information. After receiving the NIR image 718, the processors
5 106 may be configured to determine object region, i.e., pixels depicting parts of the teeth 306 of the subject. For example, a plurality of pixels indicating the object region, i.e., the teeth 306 within the NIR image 718 may be identified.

[126] Further, the processors 106 may be configured to project or render a casting object from each of the plurality of pixels of the NIR image 718 as well
10 as other NIR images 114 captured by the processors 106. The casting object may be projected from the plurality of pixels to determine the set of input parameters for optimizing the continuous volumetric ML model 108. Pursuant to present example, the whole NIR image 718 may be rendered by rendering or projecting casting objects through each of the plurality of pixels.

15 [127] Referring particularly to FIG. 7B, a ray-based casting object 724 (referred to as ray 724, hereinafter) may be casted or projected through a pixel 722 of the NIR image 718. The ray 724 may include a plurality of point coordinates, depicted as point coordinates 726A, 726B, 726C, 726D and 726E (collectively referred to as point coordinates 726, hereinafter). For example,
20 each of the point coordinates 726 along the ray 724 may have to be transformed with a positional encoding, for example, using a gamma function. The point coordinates 726 may be sampled along the ray 724. Spatial location information and viewing angle information corresponding to each transformed point coordinate is used to determine the set of input parameters for optimizing the
25 continuous volumetric scene function of the continuous volumetric ML model 108.

[128] Referring to particularly to FIG. 7C, a cone-based casting object 728 (referred to as cone 728, hereinafter) may be casted or projected through the pixel 722 of the NIR image 718. For example, a radius of the cone 728

indicating the casting object corresponding to the pixel 722 may be determined from the based on a size of the corresponding pixel 722 on an image plane.

[129] In an example, the processors 106 may be configured to determine an average of content within a visible volume 734 for the pixel 722. For example, the content within the visible volume 734 of the pixel 722 may indicate a color intensity of the pixel 722. Furthermore, the average of the color intensity within the visible volume 734 may enable to identify presence of, for example, part or point of teeth 306, caries, lesion, cracks, dentin, enamel junction, filling, or any other object present on or inside the point of teeth 306 captured by the pixel 722. For example, the processors 106 may be configured to render or project the cone 728 as the casting object corresponding to the pixel 722 based on the corresponding average of content. Further, the cone 728 may be rendered such that the cone 728 models a whole volume of space that may be visible through the pixel 722 based on the average of content or color intensity within the visible volume 734.

[130] The cone 728 may include a plurality of point coordinates, depicted as point coordinates 732A, 732B, 732C, 732D and 732E (collectively referred to as point coordinates 732, hereinafter). Further, the cone 728 may be sliced into conical frustums corresponding to the point coordinates 732. For example, a conical frustum 736 may correspond to the point coordinate 732D. To this end, each of the point coordinates 732 along the cone 728 may be transformed with a positional encoding of a volume of the corresponding conical frustums. The point coordinates 732 may be sampled along the cone 728. Spatial location information and viewing angle information corresponding to each transformed point coordinate is used to determine the set of input parameters for optimizing the continuous volumetric scene function of the continuous volumetric ML model 108, wherein the continuous volumetric ML model 108 is the mip-NeRF based neural network. Details of generating the set of input parameters for the cone-based casting objects are further described, for example, in FIG. 8.

[131] A manner in which the set of input parameters corresponding to casting objects are used by the continuous volumetric ML model 108 is described in detail in conjunction with FIG. 7A and FIG. 8.

[132] FIG. 7D shows a schematic illustration of a generated 3D model 740 of the teeth 306 of the subject, in accordance with an example embodiment. For example, the processors 106 may be configured to generate the 3D model 740. The 3D model 740 comprises the 3D surface model 118 and the 3D inner geometry 120 of the teeth 306. It may be noted that the 3D inner geometry 120 may be generated by rendering grid points within the 3D surface model at different viewing directions of the 3D surface model 118. Moreover, the intensities and densities of the corresponding grid points are determined based on the continuous volumetric machine learning model.

[133] The 3D model 740 of the teeth 306 shows geometry of inner region of the teeth 306. This may enable the user or the dentist to identify any disease, anomaly or abnormality inside the surface of the teeth 306. In particular, the 3D model 740 provides accurate representation of condition of different layers inside the surface of the teeth 306. The 3D model may be used to identify, for example, caries, cracks, lesions, abnormal growth, fillings, boundary between enamel and dentin, etc. inside the surface of the teeth 306.

[134] For example, the 3D model 740 may include the 3D inner geometry 120 rendered within the 3D surface model 118. The 3D model 740, based on the 3D inner geometry 120, may represent the part of the teeth 306 corresponding to enamel 742, part of the teeth 306 corresponding to a dentine 744, as well as a boundary 746 or junction between the enamel 742 and the dentine 744 within the dentition of the teeth 306.

[135] FIG. 8 illustrates a method 800 for generating the synthetic pixel values 122 using the continuous volumetric machine learning model 108, in accordance with an example embodiment. The method 800 describes generating the synthetic pixel values 122 when the projected casting objects correspond to

the second type of casting objects, such as the cone 728. In such a case, the continuous volumetric ML model 108 may be implemented using, for example, mip-NeRF based neural network model. FIG. 7 is explained in conjunction with elements of FIG. 1, FIG. 2, FIG. 3, FIG. 4, FIG. 5, FIG. 6, FIG. 7A, FIG. 7B, FIG. 7C and FIG. 7D.

[136] At step 802, one or more conical frustums are determined for each of the cone corresponding to each of the plurality of pixels. For example, the conical frustum 736 may be determined for the cone 728 corresponding to the pixel 722. The conical frustum 736 may relate to the point coordinate 732D from the plurality of point coordinates 732 of the cone 728. In an example, the processors 106 may determine the conical frustum 736 such that the conical frustum covers a visible volume around the point coordinate 732D. It may be noted that similar conical frustums may be determined for each of the point coordinates 732 of the cone 728.

[137] At step 804, an integrated positional encoding for the one or more conical frustums is determined for transforming the point coordinates 732 of the cone 728. The integrated positional encoding of the conical frustums for the point coordinates 732 may include a Gaussian encoding and/or a sinusoidal encoding.

[138] For example, fitting a multivariate Gaussian in a positional encoding of the conical frustum 736 may reduce complexity in manipulation of the positional encoding analytically. Further, an expected integrated positional encoding may be determined based on Gaussian fitted conical frustum 736 and each of other Gaussian fitted conical frustums corresponding to, for example, the point coordinates 732 for the cone 728. Similarly, expected integrated positional encoding may be determined for the cones projected from the each of the plurality of pixels of the plurality of NIR images 114. The integrated positional encoding of conical frustums of the cone 728 may have properties of sinusoids as well as Gaussians.

[139] At 806, the synthetic pixel value for each of the cone is generated based on the integrated positional encoding for the corresponding one or more conical frustums and the continuous volumetric ML model 108.

[140] For example, the continuous volumetric ML model 108 may be configured to change a width of the Gaussian fittings of the conical frustums in the integrated positional encodings at higher frequencies to determine the synthetic pixel values 122. For example, the integrated positional encodings may approach zero when the width gets wider indicating small volume of cone and the integrated positional encodings may increase towards non-zero value when the width gets narrower indicating large volume of the cones. Thereby, the MIP-Nerf based continuous volumetric ML model 108 is able to reason about a scale of the input NIR images 114 by looking at a scale of the integrated positional encodings of the casted cones. The continuous volumetric ML model 108 may distinguish between small visible volume and large visible volumes corresponding to the cones.

[141] FIG. 9 illustrates a method 900 for generating the 3D inner geometry 120 and the 3D surface model 118 of the teeth 306 of the subject, in accordance with an example embodiment. FIG. 9 is explained in conjunction with elements of FIG. 1, FIG. 2, FIG. 3, FIG. 4, FIG. 5, FIG. 6, FIG. 7A, FIG. 7B, FIG. 7C, FIG. 7D and FIG. 8. The steps of the method 900 may be performed by the system 102. As described, the system 102 includes the handheld intraoral scanner 104, the processors 106 and the continuous volumetric ML model 108. The hand-held intraoral scanner 104 comprises the projector unit 214C and the one or more sensors 214D. The projector unit 214C may illuminate the teeth 306 of the subject with visible light wavelength pulses and NIR wavelength pulses. The one or more sensors 214D may detect NIR and visible light. For example, the one or more sensors 214D may include image sensors that may be configured to capture white light images 116 based on detecting visible light and NIR images 114 based on detecting NIR. Further, the processors 106 may be configured to

generate the 3D inner geometry 120 of the teeth 306 using the continuous volumetric ML model 108.

[142] At step 902, visible light information and NIR information are received from the one or more sensors 214D. The visible light information and the NIR information may indicate visible light wavelength pulses and NIR wavelength pulses, respectively that may be detected by the one or more sensors 214D. For example, the detected visible light wavelength pulses and the NIR wavelength pulses may be reflected from the teeth 306 of the subject. In certain cases, the visible light information and the NIR information may also include calibration information relating to the one or more sensors 214D based on which the visible light wavelength pulses and the NIR wavelength pulses are detected. In an example, the visible light information and the NIR information may also include white light images 116 and NIR images 114, respectively.

[143] At step 904, surface information is determined in real time. In an example, the processors 106 may be configured to generate the surface information of the teeth 306 from the visible light information or the white light images 116. Based on the surface information, the processors 106 may be configured to generate the 3D surface model 118 of the teeth 306 of the subject. Details of generating the 3D surface model 118 are provided in conjunction with FIG. 4.

[144] At step 906, the plurality of 2D NIR images 114 of an internal region of the teeth 306 of the subject may be captured from the NIR information. For example, the processors 106 may capture the NIR images 114 using the image sensors from the one or more sensors 214D. As may be noted that each of the NIR images 114 may include a corresponding set of pixels. In an example, a plurality of pixels of each of the NIR images 114 corresponding to parts or points on the teeth 306 may be determined for further processing.

[145] At step 908, a set of input parameters is determined. In this regard, a casting object may be projected from each of the plurality of pixels of each of

the NIR images 114. The casting objects may include corresponding plurality of point coordinates. The plurality of point coordinates correspond to different depth within a corresponding casting object and a corresponding pixel. For example, the plurality of point coordinates may correspond to different materials inside the teeth 306 of the subject. As such, the plurality of point coordinates of the corresponding casting object may not belong to same material. Further, the processors 106 may be configured to determine the set of input parameters for the plurality of point coordinates of the casting objects projected from the plurality of pixels of the NIR images 114. For example, the set of input parameters may include spatial location information, such as values of (x, y, z) and viewing angle information, such as values of (θ, φ) . Based on the set of input parameters, optimization process is applied over all the casting objects of the plurality of pixels and parameters of optimized continuous volumetric scene function, $I'(x, y, z, \theta, \varphi)$, may be defined.

[146] At step 910, the set of input parameters is processed using the continuous volumetric machine learning model 108. For example, the continuous volumetric machine learning model may process the set of input parameters to determine a synthetic pixel value for a casting object based on an intensity value and a density value for point coordinates. The continuous volumetric machine learning model 108 is configured to be trained using the plurality of NIR images 114. Details of training of the continuous volumetric machine learning model 108 are described in conjunction with, for example, FIG. 8.

[147] It will be understood that each step of the sequence diagram 900 may be implemented by various means, such as hardware, firmware, processor, circuitry, and/or other communication devices associated with execution of software including one or more computer program instructions. For example, one or more of the steps described above may be embodied by computer program instructions. In this regard, the computer program instructions which embody the

steps described above may be stored by a memory of the system 102, employing an embodiment of the present disclosure. As will be appreciated, any such computer program instructions may be loaded onto a computer or other programmable apparatus (for example, hardware) to produce a machine, such that the resulting computer or other programmable apparatus implements the functions specified in the sequence diagram 1000. These computer program instructions may also be stored in a computer-readable memory that may direct a computer or other programmable apparatus to function in a particular manner, such that the instructions stored in the computer-readable memory produce an article of manufacture the execution of which implements the function specified in the sequence diagram 900. The computer program instructions may also be loaded onto a computer or other programmable apparatus to cause a series of operations to be performed on the computer or other programmable apparatus to produce a computer-implemented process such that the instructions which execute on the computer or other programmable apparatus provide operations for implementing the functions specified in the sequence diagram 900.

[148] Accordingly, the steps of the sequence diagram 900 support combinations of means for performing the specified functions and combinations of operations for performing the specified functions for performing the specified functions. It will also be understood that one or more steps of the sequence diagram 900, and combinations of steps in the sequence diagram 900, may be implemented by special purpose hardware-based computer systems which perform the specified functions, or combinations of special purpose hardware and computer instructions. The sequence diagram 900 of FIG. 9 is used for the generation and rendering of 3D model of teeth comprising 3D inner geometry of the teeth. Fewer, more, or different steps may be provided.

[149] FIG. 10 illustrates a schematic diagram 1000 that depicts an exemplary environment for generating 3D inner geometry 120 and 3D surface model 118 of the teeth and render of an interactive 3D graphical representation 1010 in real-time, in accordance with an example embodiment. FIG. 10 is

explained in conjunction with elements of FIG. 1, FIG. 2, FIG. 3, FIG. 4, FIG. 5, FIG. 6, FIG. 7A, FIG. 7B, FIG. 7C, FIG. 7D, FIG. 8 and FIG. 9. The schematic diagram 1000 may include a dentist 1002 and a patient 1004.

[150] The handheld intraoral scanning device 104 may be utilized by the dentist 1002 to capture the plurality of 2D images 112 comprising NIR images 114 and white light images 116 of teeth 1006 of the patient 1004. The handheld intraoral scanning device 104 may transmit the plurality of 2D images 112 to the processors 106. The processors 106 may process the white light images 116 and generate 3D surface information for the teeth 1006. Based on the 3D surface information, the processors 106 may generate the 3D surface model 118 of the teeth 1006 of the patient 1004.

[151] Further, the processors 106 may process the NIR images 114 to generate a set of input parameters for optimizing a function of the continuous volumetric machine learning model 108. The set of input parameters are generated for each casting object projected through each of plurality of pixels of each of the NIR images 114. The casting object may be a ray or a cone. For example, set of input parameters may be used to optimize a continuous volumetric scene function of the continuous volumetric machine learning model 108. Given the optimized continuous volumetric scene function, $F(x, y, z, \theta, \varphi)$, having the set of input parameters, the continuous volumetric machine learning model 108 may generate intensity value and density value for each of the plurality of point coordinates of the casting object. Further, the continuous volumetric machine learning model 108 may generate a synthetic pixel value for the casting object, for example, by integrating the intensity values and density values for the plurality of point coordinates. In this manner, synthetic pixel values 122 for plurality of casting objects corresponding to plurality of pixels in each of the NIR images 114 may be determined.

[152] The 3D model of the teeth 1006 of the patient 1004 may be displayed on the display unit 1008 as the interactive 3D graphical representation 1010, in

real-time. Further, the interactive 3D graphical representation 1010 may be generated in the real-time and the interactive 3D graphical representation 1010 may be manipulated, such as rotated or viewed in multiple perspectives by the dentist 1002 as required.

5 [153] Thus, the intraoral scanning system 102 may enable the processing of the plurality of 2D images 112 to generate 3D surface model 118 along with 3D inner geometry 120 of the teeth 1006. The 3D inner geometry 120 enables viewing of condition, shape, size, and other characteristics of inner regions of the teeth 1006. The generated 3D inner geometry 120 of the accurately identifies
10 different material, such as enamel and dentin within the inner region of the teeth 1006, based on comparison of intensity value and/or the density values of casting objects with the change threshold. For example, if a difference between a first intensity value or a density value of a first casting object and a second pixel value of a second casting object is greater than the change threshold, the first
15 casting object and the second casting object may be identified to belong to different materials. In this manner a boundary between the enamel and the dentin within the teeth 1006 of the patient 1004 may be identified in the 3D inner geometry 120.

[154] The users, such as the dentists may be able to access the interactive 3D
20 graphical representation 1010 of the 3D model of the teeth 1006 by using the display unit. Thus, the intraoral scanning system 102 enables real-time processing of the plurality of 2D images 112 and generating of the 3D model for real-time access by the users via the display unit.

[155] Many modifications and other embodiments of the inventions set
25 forth herein will come to mind of one skilled in the art to which these inventions pertain having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood that the inventions are not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the

appended claims. Moreover, although the foregoing descriptions and the associated drawings describe example embodiments in the context of certain example combinations of elements and/or functions, it should be appreciated that different combinations of elements and/or functions may be provided by
5 alternative embodiments without departing from the scope of the appended claims. In this regard, for example, different combinations of elements and/or functions than those explicitly described above are also contemplated as may be set forth in some of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes
10 of limitation.

Parts list:

- Intraoral scanning system – 102
- Hand-held intraoral scanner – 104
- 5 Processors – 106
 - Continuous volumetric machine learning model -108
 - 2D images – 112, 308
 - 2D NIR images – 114, 308B, 718
 - 2D white light images – 116, 308A
- 10 3D surface model – 118
 - 3D inner geometry – 120
 - Dentist – 302, 1002
 - Patient – 304, 1004
 - Subject's teeth – 306, 1006
- 15 3D information – 310
 - Pixel – 722
 - Ray casting object – 724
 - Point coordinates – 726A, 726B, 726C, 726D, 726E, 732A, 732B, 732C, 732D and 732E
- 20 Cone casting object – 728
 - Visible volume – 734
 - Conical frustum – 736
 - 3D model – 740
 - Enamel – 742
- 25 Dentine – 744

Boundary – 746

Display unit – 1008

Interactive 3D graphical representation – 1010

5 Item list:

1. An intraoral scanning system (102) configured to generate a three-dimensional (3D) inner geometry (120) of a subject's teeth (306), the system comprising:

10 a hand-held intraoral scanner (104) configured to operate with one or more sensors (214D) to detect near-infrared (NIR) and visible light, wherein the one or more sensors comprises an image sensor;

one or more processors (106) operably connected to the hand-held intraoral scanner, the one or more processors being configured to:

15 receive visible light information and near-infrared (NIR) information from the one or more sensors;

determine, in real time, surface information from the visible light information to generate a three-dimensional (3D) surface model (118) of the subject's teeth using the surface information;

20 capture, in real time using the image sensor, a plurality of two-dimensional (2D) NIR images (114) of an internal region of the subject's teeth from the NIR information, wherein each of the plurality of NIR images includes a plurality of corresponding pixels (722); and

25 determine a set of input parameters for a casting object corresponding to each of the plurality of pixels, wherein the set of input parameters comprises spatial location information and viewing angle information, and the casting object comprises a plurality of point coordinates (726, 732); and

30 a continuous volumetric machine learning model (108) configured to receive and process the set of input parameters to determine an intensity value and a density value of the three-dimensional inner geometry of the 3D surface

model, wherein the continuous volumetric machine learning model is configured to be trained using the plurality of 2D NIR images.

2. The intraoral scanning system (102) according to item 1, wherein
5 the continuous volumetric machine learning model (108) is trained by:

receiving (602) the set of input parameters that corresponds to the casting object for each of the plurality of 2D NIR images (114), wherein the casting object includes the plurality of point coordinates (726, 732) associated with the 3D surface model (118);

10 generating (804), based on the continuous volumetric machine learning model using the set of input parameters, the intensity value and the density value for each of the plurality of point coordinates;

determining (606) a synthetic pixel value for the casting object based on the corresponding determined intensity value and the density value for each of
15 the plurality of point coordinates; and

minimizing (608) a loss function between the synthetic pixel value and a corresponding true pixel value of the plurality of pixels (722) of the plurality of 2D NIR images by changing the intensity value and the density value for each of the plurality of point coordinates.

20

3. The intraoral scanning system (102) according to any of the previous items, wherein the one or more processors (106) are further configured to:

determine a plurality of grid points within the 3D surface model (118);
25 and

determine the 3D inner geometry (120) by arranging at least one of: the intensity value or the density value at each of the plurality of grid points.

4. The intraoral scanning system (102) according to any of the
30 previous items, further comprising a display unit (1008) configured to display the 3D inner geometry (120) based on at least one of: the intensity value or the

density value determined by the continuous volumetric machine learning model (108).

5 5. The intraoral scanning system (102) according to item 4, wherein the display unit (1008) is configured to display the 3D inner geometry (120) inside the 3D surface model (118).

6. The intraoral scanning system (102) according to any of the previous items, wherein the one or more processors (106) are further configured
10 to:

 determine a boundary (746) between enamel (742) and dentine (744) in a dentition of the subject's teeth (306) based on the determined intensity value and the density value for the casting object of each of the plurality of pixels (722).

15

7. The intraoral scanning system (102) according to item 6, wherein the one or more processors (106) are further configured to:

 determine the boundary (746) based on a change in at least one of: the intensity values or the density values for at least two casting objects of the plurality of pixels (722) being above a change threshold.

20

8. The intraoral scanning system (102) according to any of the previous items, wherein the casting object is one of a ray (724) or a cone (728).

25 9. The intraoral scanning system (102) according to any of the previous items, wherein

 the hand-held intraoral scanner (104) comprises a projector unit (214C) configured to illuminate the subject's teeth (306) using one or more white colored wavelength pulses and one or more near infrared (NIR) wavelength
30 pulses; and

the one or more sensors (214D) are configured to generate a set of white light images (116) and the plurality of 2D NIR images (114) based on the illumination.

5 10. The intraoral system (102) according to item 9, wherein the 3D surface model (118) is determined based on the set of white light images (116).

11. The intraoral scanning system (102) according to any of the previous items, wherein the one or more processors (106) are further configured
10 to:

estimate a relative position between the hand-held intraoral scanner (104) and the subject's teeth (306) corresponding to each of the plurality of NIR images (114), wherein the estimated relative position indicates the viewing angle information and the spatial location information for the corresponding
15 casting object.

12. The intraoral scanning system (102) according to any of the previous items, wherein the casting object is a cone (728), the one or more processors are further configured to:

20 determine (802) one or more conical frustums (736) for each of the cone corresponding to each of the plurality of pixels (722), wherein the one or more conical frustums relate to the plurality of point coordinates (732);

determine (804) an integrated positional encoding for the one or more conical frustums for transforming the plurality of point coordinates of each of
25 the cone, the integrated positional encoding comprising at least a Gaussian encoding and a sinusoidal encoding; and

generate (806) the synthetic pixel value for each of the cone based on the integrated positional encoding for the corresponding one or more conical frustums and the continuous volumetric machine learning model (108).

13. The intraoral scanning system (102) according to item 12, wherein the one or more processors (106) are further configured to:

5 determine a radius of each of the cone (728) indicating the casting object corresponding to each of the plurality of pixels (722) based on a size of the corresponding pixel from the plurality of pixels.

14. The intraoral scanning system (102) according to item 12, wherein the one or more processors (106) are further configured to:

10 determine an average of content within a visible volume (734) for each of the plurality of pixels (722), wherein the content indicates a color intensity; and

render each of the cone (728) indicating the casting object of each of the plurality of pixels based on the corresponding average of content.

15 15. The intraoral scanning system (102) according to any of the previous items, wherein the continuous volumetric machine learning model (108) is a machine-learning based neural radiance field (NeRF) network.

20 16. A method (900) for generating a three-dimensional (3D) inner geometry (120) of a subject's teeth (306) using an intraoral scanning system (102), the intraoral scanning system comprising:

25 a hand-held intraoral scanner (104) configured to operate with one or more sensors (214D) to detect near-infrared (NIR) and visible light, wherein the one or more sensors comprises an image sensor,

one or more processors (106) operably connected to the hand-held intraoral scanner, and

a continuous volumetric machine learning model (108),

wherein the method comprises:

30 receiving (902) visible light information and near-infrared (NIR) information from the one or more sensors;

determining (904), in real time, surface information from the visible light information to generate a three-dimensional (3D) surface model (118) of the subject's teeth using the surface information;

capturing (906), in real time using the image sensor, a plurality of two-dimensional (2D) NIR images (114) of an internal region of the subject's teeth from the NIR information, wherein each of the plurality of NIR images includes a plurality of corresponding pixels (722);

determining (908) a set of input parameters for a casting object corresponding to each of the plurality of pixels, wherein the set of input parameters comprises spatial location information and viewing angle information, and the casting object comprises a plurality of point coordinates (726, 732); and

processing (910), using the continuous volumetric machine learning model, the set of input parameters to determine an intensity value and a density value of a three-dimensional inner geometry of the 3D surface model, wherein the continuous volumetric machine learning model is configured to be trained using the plurality of 2D NIR images.

17. The method according to item 17, the method further comprising:
determining a plurality of grid points within the 3D surface model (118);

determining the 3D inner geometry (120) by arranging at least one of: the intensity value or the density value at each of the plurality of grid points based on the set of input parameters; and

displaying the 3D inner geometry inside the 3D surface model.

18. The method according to any of the previous items, the method further comprising:

determining a boundary (746) between enamel (742) and dentine (744) in a dentition of the subject's teeth (306), based on a change in at least one of:

the intensity values or the density values for at least two casting objects of the plurality of pixels being above a change threshold.

19. A computer programmable product comprising a non-transitory
5 computer readable medium having stored thereon computer executable instructions, which when executed by a processing circuitry, cause the processing circuitry to carry out operations, the operations comprising:

receiving visible light information and near-infrared (NIR) information
from one or more sensors (214D), the one or more sensors being installed within
10 a hand-held intraoral scanner (104), the one or more sensors being configured to detect near-infrared (NIR) and visible light, wherein the one or more sensors comprises an image sensor;

determining, in real time, surface information from the visible light
information to generate a three-dimensional (3D) surface model (118) of a
15 subject's teeth (306) using the surface information;

capturing, in real time using the image sensor, a plurality of two-
dimensional (2D) NIR images (114) of an internal region of the subject's teeth
from the NIR information, wherein each of the plurality of NIR images includes
a plurality of corresponding pixels (722);

20 determining a set of input parameters for a casting object corresponding to each of the plurality of pixels, wherein the set of input parameters comprises spatial location information and viewing angle information, and the casting object comprises a plurality of point coordinates (726, 732); and

processing, using a continuous volumetric machine learning model
25 (118), the set of input parameters to determine an intensity value and a density value of a three-dimensional inner geometry (120) of the 3D surface model, wherein the continuous volumetric machine learning model is configured to be trained using the plurality of 2D NIR images.

CLAIMS

1. An intraoral scanning system (102) configured to generate a three-dimensional (3D) inner geometry (120) of a subject's teeth (306), the
5 system comprising:
a hand-held intraoral scanner (104) configured to operate with one or more sensors (214D) to detect near-infrared (NIR) and visible light, wherein the one or more sensors comprises an image sensor;
one or more processors (106) operably connected to the hand-held
10 intraoral scanner, the one or more processors being configured to:
receive visible light information and near-infrared (NIR) information from the one or more sensors;
determine, in real time, surface information from the visible light information to generate a three-dimensional (3D) surface model
15 (118) of the subject's teeth using the surface information;
capture, in real time using the image sensor, a plurality of two-dimensional (2D) NIR images (114) of an internal region of the subject's teeth from the NIR information, wherein each of the plurality of NIR images includes a plurality of corresponding pixels (722); and
20 determine a set of input parameters for a casting object corresponding to each of the plurality of pixels, wherein the set of input parameters comprises spatial location information and viewing angle information, and the casting object comprises a plurality of point coordinates (726, 732); and
25 a continuous volumetric machine learning model (108) configured to receive and process the set of input parameters to determine an intensity value and a density value of the three-dimensional inner geometry of the 3D surface model, wherein the continuous volumetric machine learning model is configured to be trained using the plurality of 2D NIR images.

2. The intraoral scanning system (102) according to claim 1, wherein the continuous volumetric machine learning model (108) is trained by:

receiving (602) the set of input parameters that corresponds to the casting object for each of the plurality of 2D NIR images (114), wherein the casting object includes the plurality of point coordinates (726, 732) associated with the 3D surface model (118);

generating (804), based on the continuous volumetric machine learning model using the set of input parameters, the intensity value and the density value for each of the plurality of point coordinates;

determining (606) a synthetic pixel value for the casting object based on the corresponding determined intensity value and the density value for each of the plurality of point coordinates; and

minimizing (608) a loss function between the synthetic pixel value and a corresponding true pixel value of the plurality of pixels (722) of the plurality of 2D NIR images by changing the intensity value and the density value for each of the plurality of point coordinates.

3. The intraoral scanning system (102) according to any of the previous claims, wherein the one or more processors (106) are further configured to:

determine a plurality of grid points within the 3D surface model (118); and

determine the 3D inner geometry (120) by arranging at least one of: the intensity value or the density value at each of the plurality of grid points.

25

4. The intraoral scanning system (102) according to any of the previous claims, further comprising a display unit (1008) configured to display the 3D inner geometry (120) based on at least one of: the intensity value or the density value determined by the continuous volumetric machine learning model (108).

30

5. The intraoral scanning system (102) according to claim 4, wherein the display unit (1008) is configured to display the 3D inner geometry (120) inside the 3D surface model (118).

5 6. The intraoral scanning system (102) according to any of the previous claims, wherein the one or more processors (106) are further configured to:

determine a boundary (746) between enamel (742) and dentine (744) in a dentition of the subject's teeth (306) based on the determined intensity value
10 and the density value for the casting object of each of the plurality of pixels (722).

7. The intraoral scanning system (102) according to claim 6, wherein the one or more processors (106) are further configured to:

15 determine the boundary (746) based on a change in at least one of: the intensity values or the density values for at least two casting objects of the plurality of pixels (722) being above a change threshold.

8. The intraoral scanning system (102) according to any of the
20 previous claims, wherein the casting object is one of a ray (724) or a cone (728).

9. The intraoral scanning system (102) according to any of the previous claims, wherein

the hand-held intraoral scanner (104) comprises a projector unit (214C)
25 configured to illuminate the subject's teeth (306) using one or more white colored wavelength pulses and one or more near infrared (NIR) wavelength pulses; and

the one or more sensors (214D) are configured to generate a set of white light images (116) and the plurality of 2D NIR images (114) based on the
30 illumination.

10. The intraoral system (102) according to claim 9, wherein the 3D surface model (118) is determined based on the set of white light images (116).

11. The intraoral scanning system (102) according to any of the previous claims, wherein the one or more processors (106) are further configured to:

estimate a relative position between the hand-held intraoral scanner (104) and the subject's teeth (306) corresponding to each of the plurality of NIR images (114), wherein the estimated relative position indicates the viewing angle information and the spatial location information for the corresponding casting object.

12. The intraoral scanning system (102) according to any of the previous claims, wherein the casting object is a cone (728), the one or more processors are further configured to:

determine (802) one or more conical frustums (736) for each of the cone corresponding to each of the plurality of pixels (722), wherein the one or more conical frustums relate to the plurality of point coordinates (732);

determine (804) an integrated positional encoding for the one or more conical frustums for transforming the plurality of point coordinates of each of the cone, the integrated positional encoding comprising at least a Gaussian encoding and a sinusoidal encoding; and

generate (806) the synthetic pixel value for each of the cone based on the integrated positional encoding for the corresponding one or more conical frustums and the continuous volumetric machine learning model (108).

13. The intraoral scanning system (102) according to claim 12, wherein the one or more processors (106) are further configured to:

determine a radius of each of the cone (728) indicating the casting object corresponding to each of the plurality of pixels (722) based on a size of the corresponding pixel from the plurality of pixels.

14. The intraoral scanning system (102) according to claim 12, wherein the one or more processors (106) are further configured to:

5 determine an average of content within a visible volume (734) for each of the plurality of pixels (722), wherein the content indicates a color intensity; and

render each of the cone (728) indicating the casting object of each of the plurality of pixels based on the corresponding average of content.

10 15. The intraoral scanning system (102) according to any of the previous claims, wherein the continuous volumetric machine learning model (108) is a machine-learning based neural radiance field (NeRF) network.

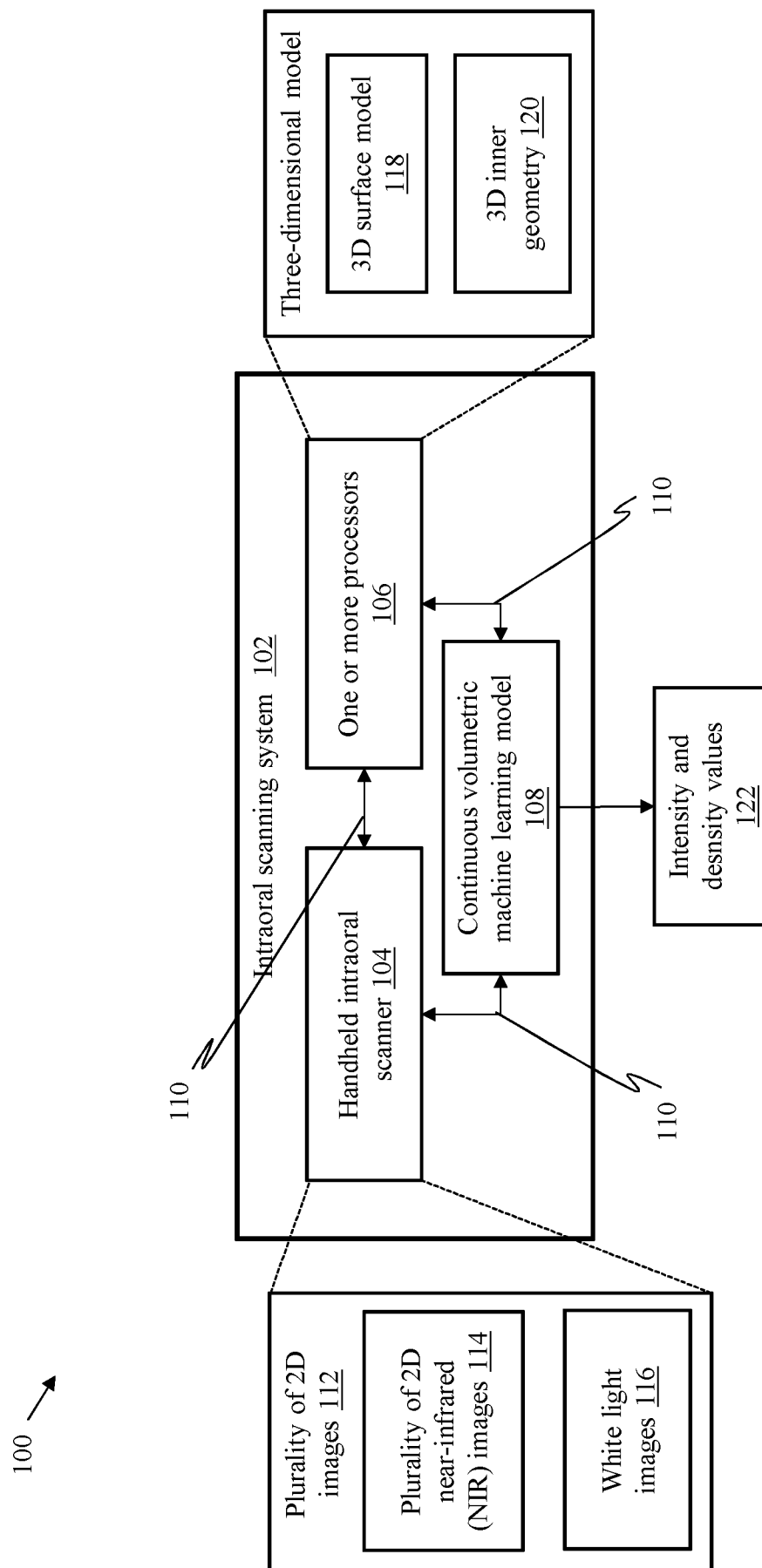


FIG. 1

200 →

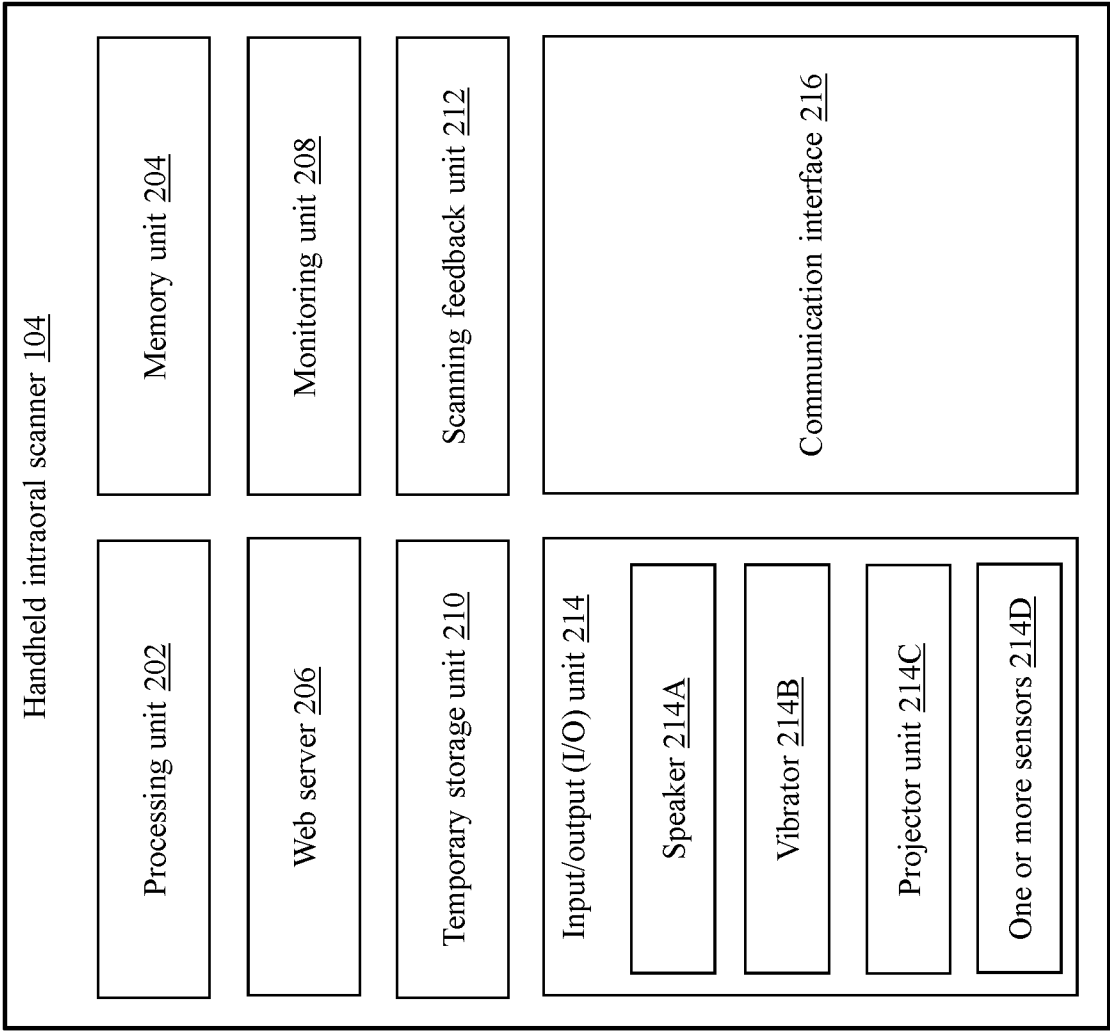


FIG. 2

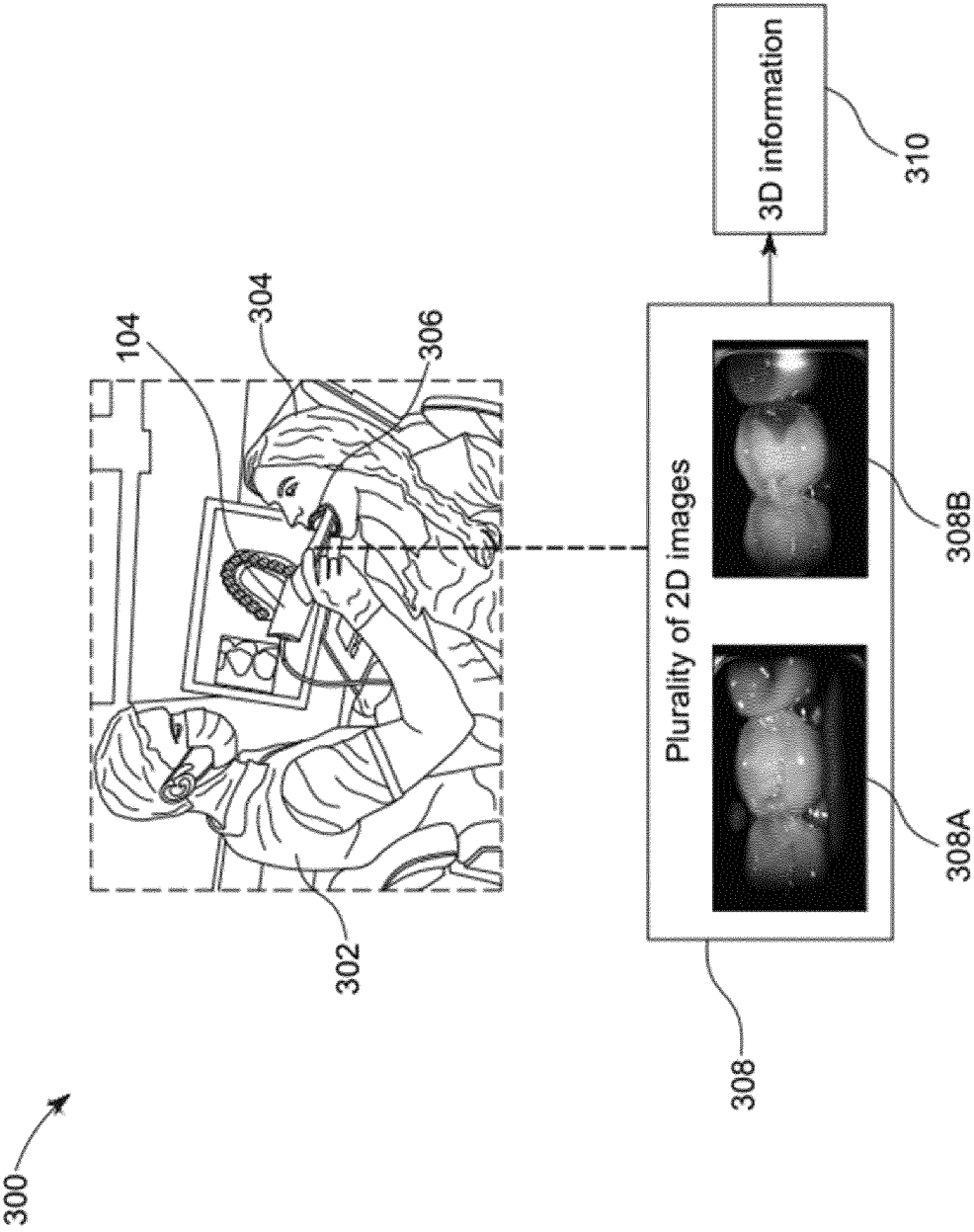


FIG. 3

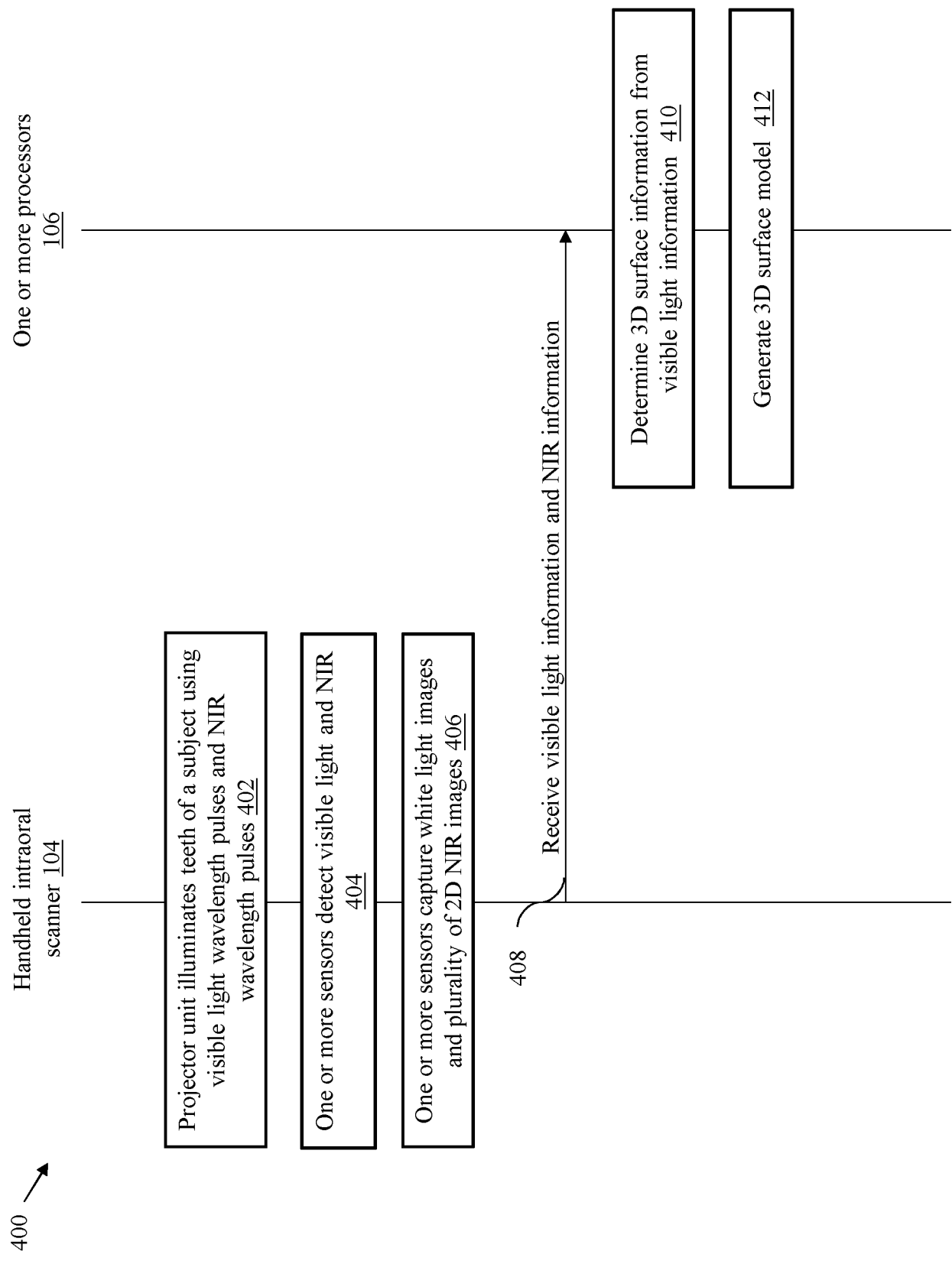


FIG. 4

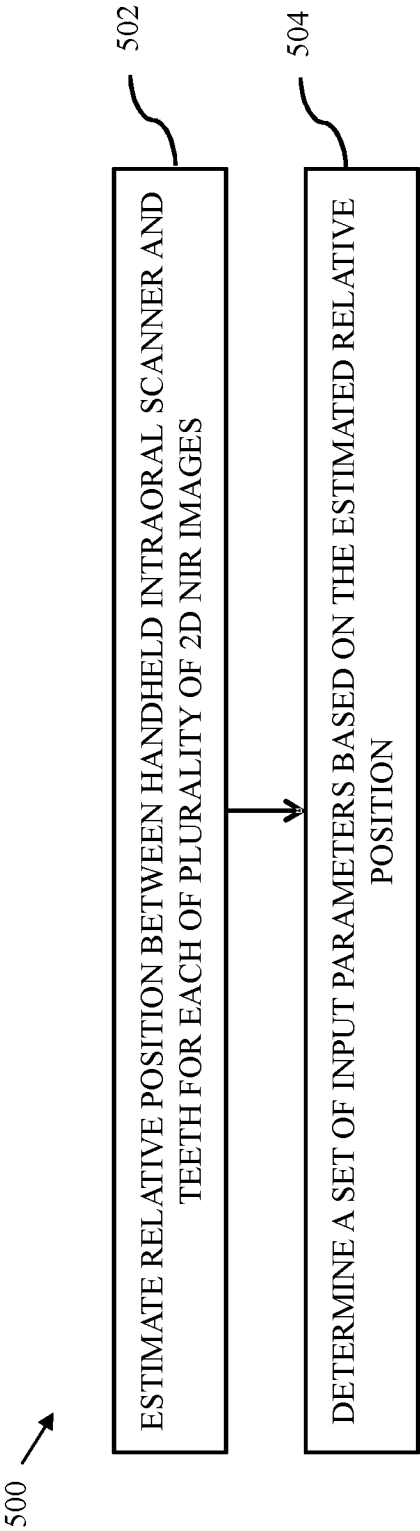


FIG. 5

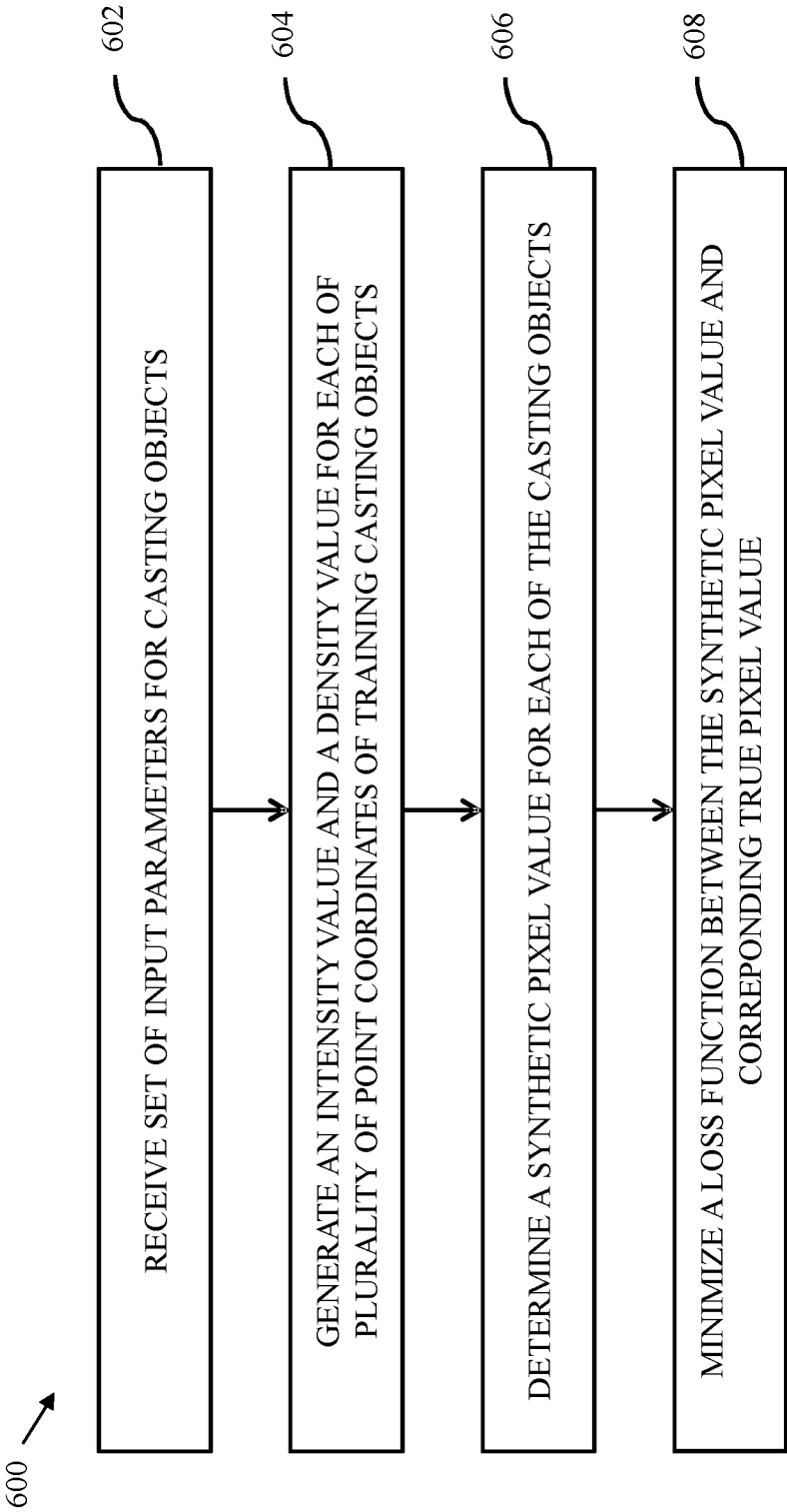


FIG. 6

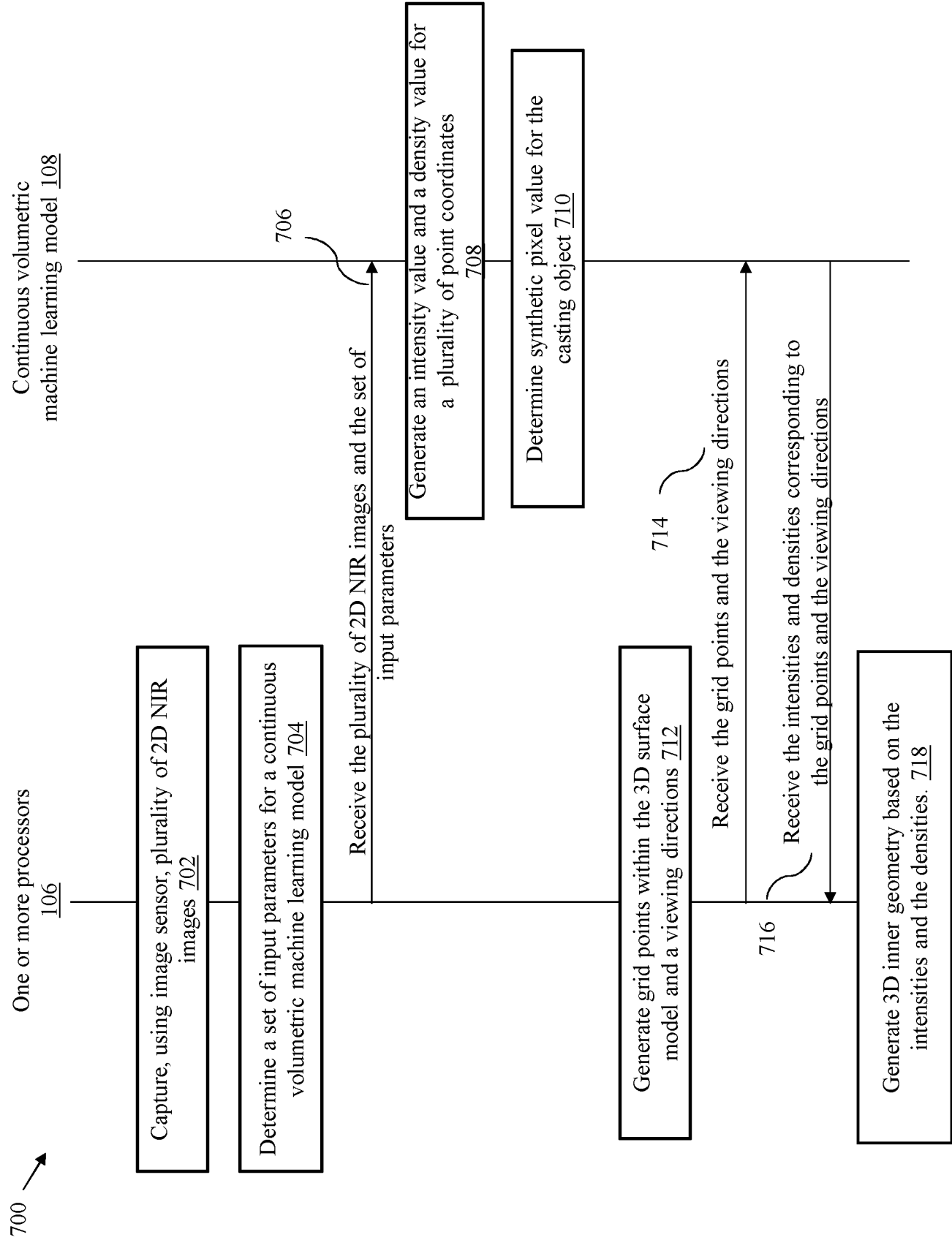
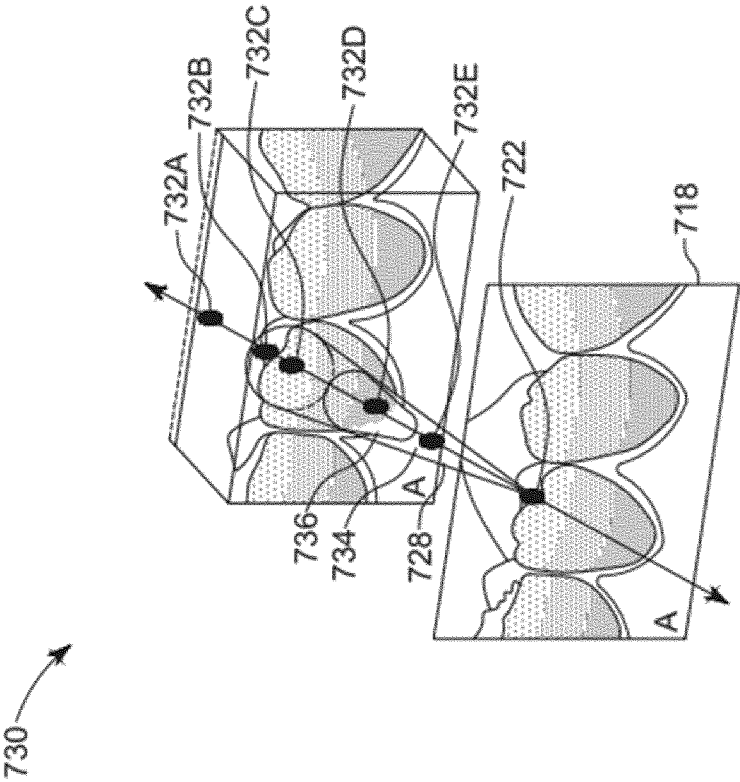


FIG. 7A



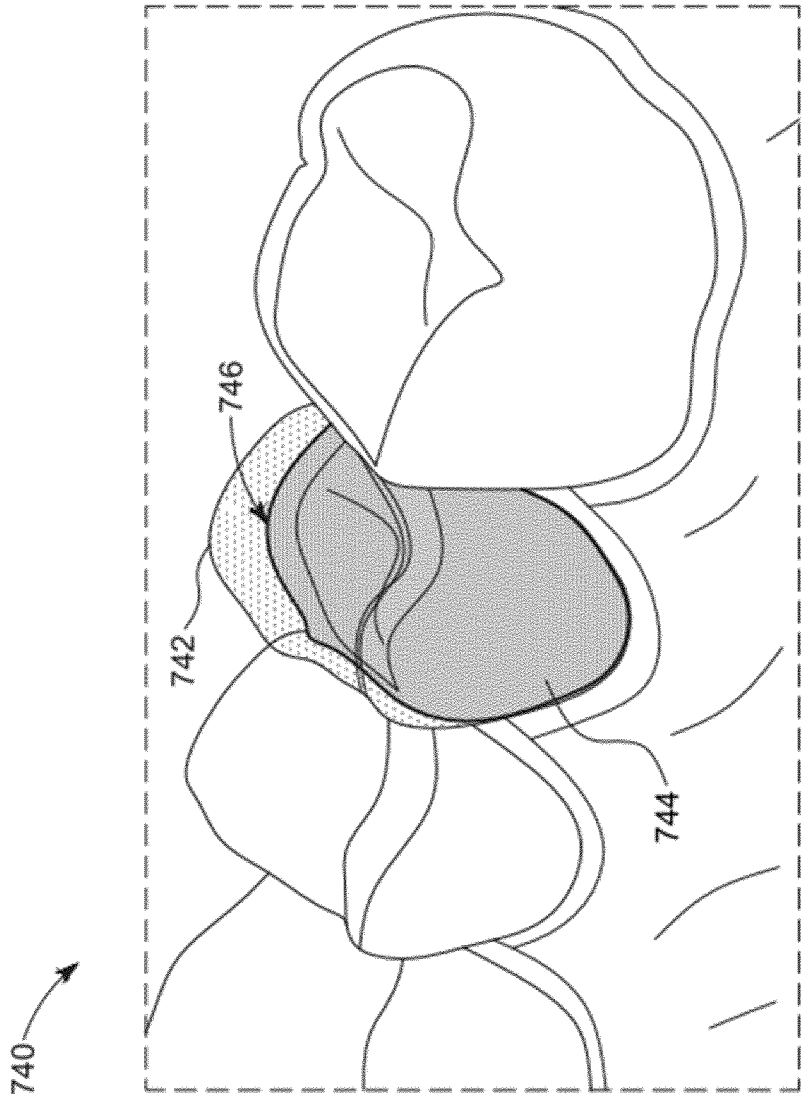


FIG. 7D

800 →

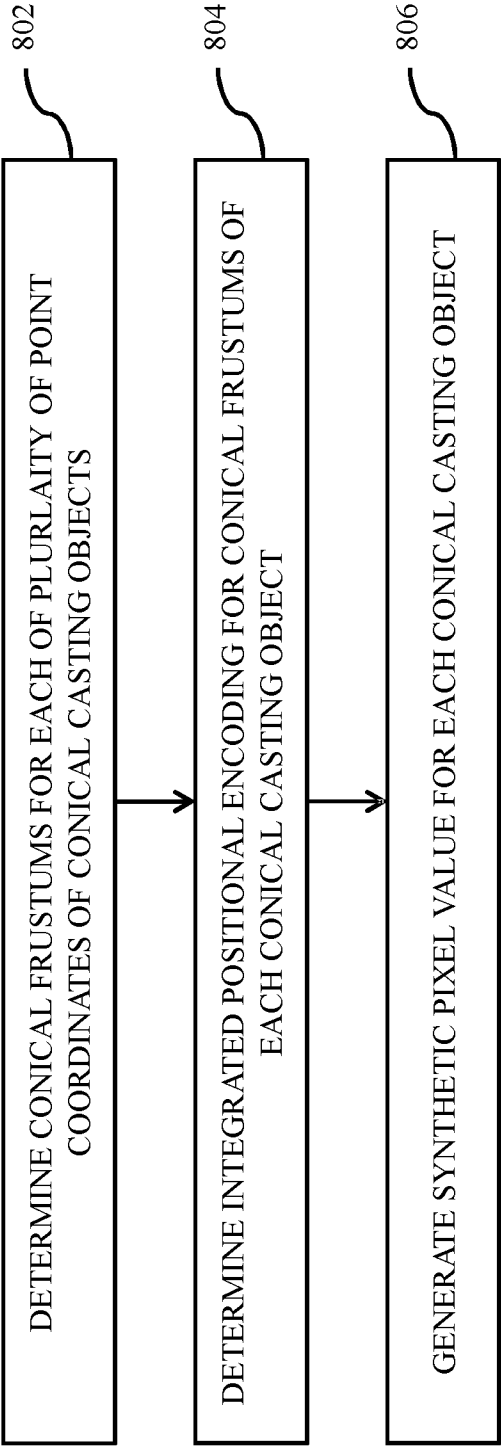


FIG. 8

900 →

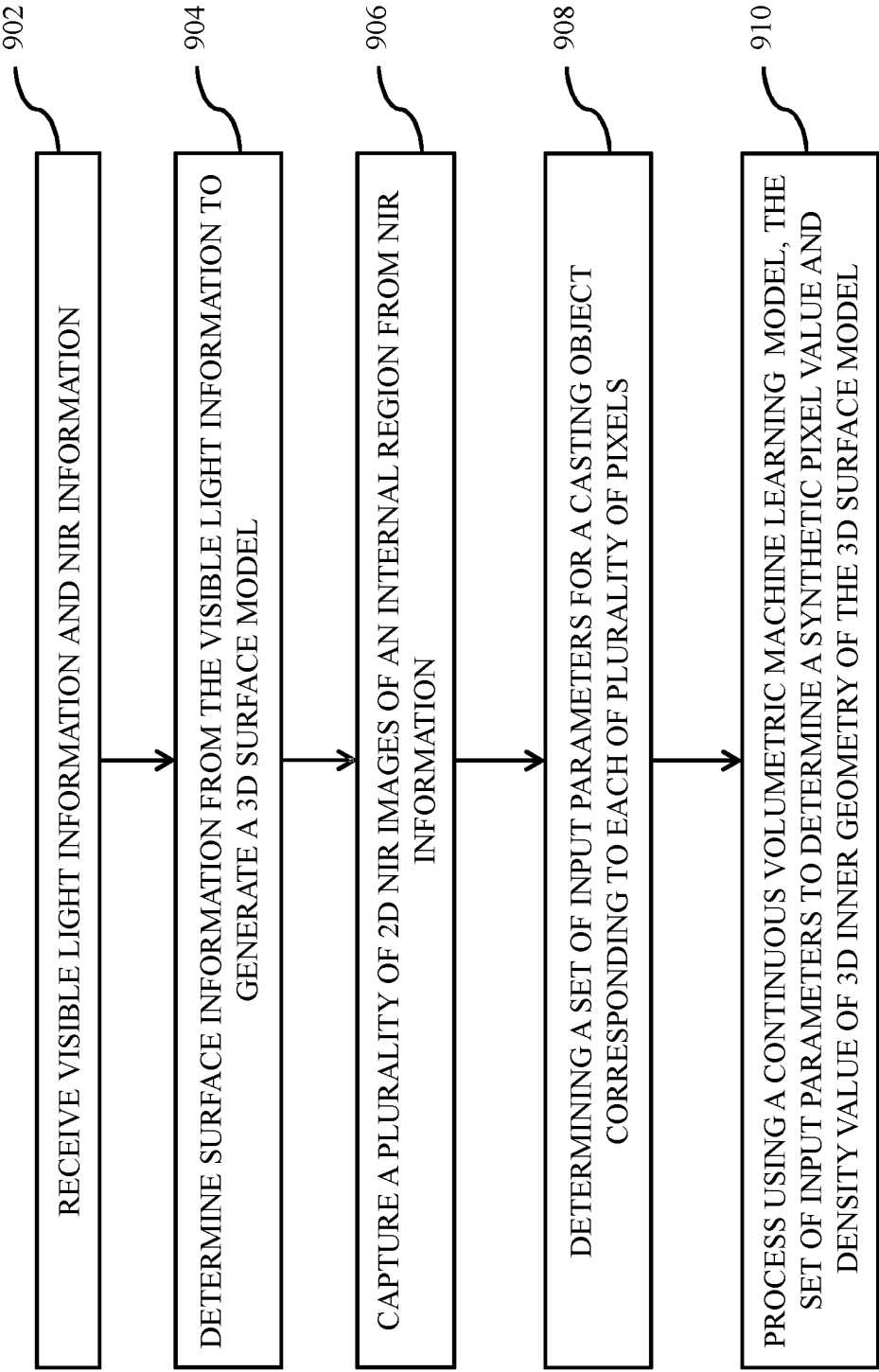


FIG. 9



FIG. 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2024/066770

A. CLASSIFICATION OF SUBJECT MATTER INV. G01B11/25 A61B5/00 A61C9/00 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) A61B G01B A61C Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) EPO-Internal		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2020/205943 A1 (ELBAZ GILAD [IL] ET AL) 2 July 2020 (2020-07-02)	1 - 11, 15
Y	paragraphs [0008], [0231], [0233], [0259], [0326], [0379], [0446], [0307] - [0309], [0380], [0071] paragraphs [0073], [0261], [0235], [0292], [0325], [0287], [0326]; claims 1,9; figures 27, 1a, 24-25, 2E -----	12 - 14
Y	ZHOU YAXUAN ET AL: "Towards AR-assisted visualisation and guidance for imaging of dental decay", HEALTHCARE TECHNOLOGY LETTERS, vol. 6, no. 6, 1 December 2019 (2019-12-01), pages 243-248, XP006088054, DOI: 10.1049/HTL.2019.0082 paragraphs [0001], [03.2] ----- - / - -	12 - 14
☒ Further documents are listed in the continuation of Box C. ☒ See patent family annex.		
* Special categories of cited documents : "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 9 October 2024		Date of mailing of the international search report 23/10/2024
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Fazio, Valentina

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2024/066770

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2022/296184 A1 (KOPELMAN AVI [US] ET AL) 22 September 2022 (2022-09-22) paragraph [0003] -----	12 - 14

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2024/066770

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2020205943 A1	02-07-2020	US 2019269485 A1	05-09-2019
		US 2020188068 A1	18-06-2020
		US 2020205943 A1	02-07-2020
		US 2021128282 A1	06-05-2021
		US 2023240819 A1	03-08-2023
		US 2024115361 A1	11-04-2024

US 2022296184 A1	22-09-2022	US 2022296184 A1	22-09-2022
		WO 2022204091 A1	29-09-2022
