



(51) International Patent Classification:

G06T 15/00 (2011.01) G06T 17/20 (2006.01)
G06T 17/00 (2006.01)

(21) International Application Number:

PCT/US2021/012073

(22) International Filing Date:

04 January 2021 (04.01.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/968,038 30 January 2020 (30.01.2020) US

(71) Applicant: **BOSTON POLARIMETRICS, INC.**
[US/US]; 380 Portage Avenue, Palo Alto, CA 94306 (US).

(72) Inventors: **VENKATARAMAN, Kartik**; c/o Boston Polarimetrics, Inc., 380 Portage Avenue, Palo Alto, CA 94306 (US). **KALRA, Agastya**; c/o Boston Polarimetrics, Inc., 380 Portage Avenue, Palo Alto, CA 94306 (US). **KADAMBI, Achuta**; c/o Boston Polarimetrics, Inc., 380 Portage Avenue, Palo Alto, CA 94306 (US).

(74) Agent: **LEE, Shaun P.**; Lewis Roca Rothgerber Christie LLP, P.O. Box 29001, Glendale, CA 91209-9001 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,

(54) Title: SYSTEMS AND METHODS FOR SYNTHESIZING DATA FOR TRAINING STATISTICAL MODELS ON DIFFERENT IMAGING MODALITIES INCLUDING POLARIZED IMAGES

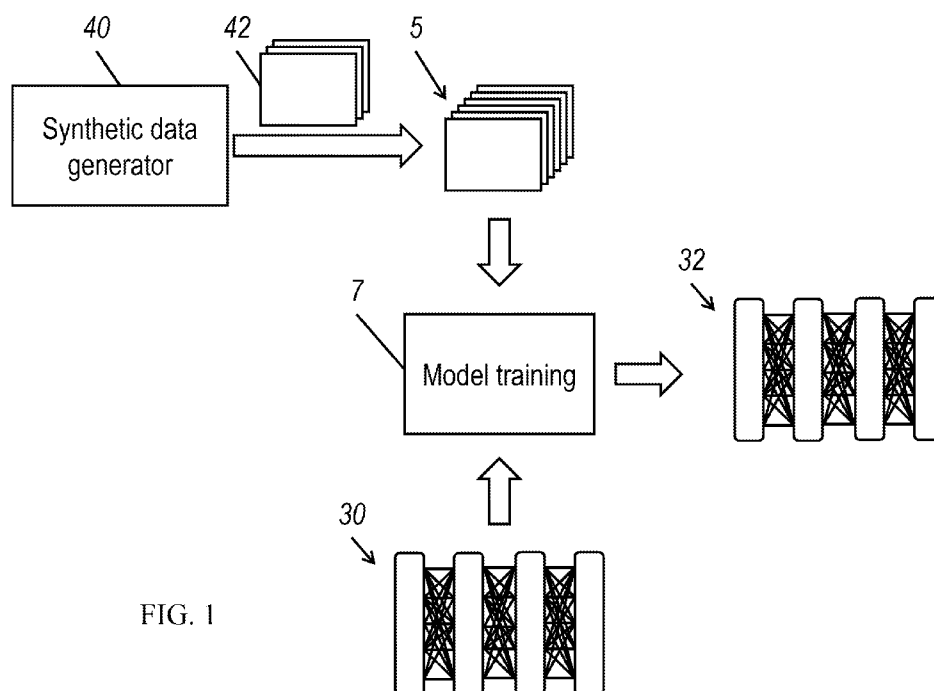


FIG. 1

(57) Abstract: A method of generating synthetic images of virtual scenes includes: placing, by a synthetic data generator implemented by a processor and memory, three-dimensional (3-D) models of objects in a 3-D virtual scene; adding, by the synthetic data generator, lighting to the 3-D virtual scene, the lighting including one or more illumination sources; applying, by the synthetic data generator, imaging modality-specific materials to the 3-D models of objects in the 3-D virtual scene in accordance with a selected multimodal imaging modality, each of the imaging modality-specific materials including an empirical model; setting a scene background in accordance with the selected multimodal imaging modality; and rendering, by the synthetic data generator, a two-dimensional image of the 3-D virtual scene based on the selected multimodal imaging modality to generate a synthetic image in accordance with the selected



SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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))*

**SYSTEMS AND METHODS FOR SYNTHESIZING DATA FOR TRAINING
STATISTICAL MODELS ON DIFFERENT IMAGING MODALITIES INCLUDING
POLARIZED IMAGES**

CROSS-REFERENCE TO RELATED APPLICATION(S)

[0001] This application claims priority to and the benefit of U.S. Provisional Patent Application No. 62/968,038, filed in the United States Patent and Trademark Office on January 30, 2020, the entire disclosure of which is incorporated by reference herein.

FIELD

[0002] Aspects of embodiments of the present disclosure relate to machine learning techniques, in particular the synthesis or generation of data for training machine learning models.

BACKGROUND

[0003] Statistical models such as machine learning models are generally trained using large amounts of data. In the field of computer vision, the training data generally includes labeled images, which are used to train deep learning models, such as convolutional neural networks, to perform computer vision tasks such as image classification and instance segmentation. However, manually collecting photographs of various scenes and labeling the photographs is time consuming and expensive. Some techniques for augmenting these training data sets include generating synthetic training data. For example, three-dimensional (3-D) computer graphics rendering engines (e.g., scanline rendering engines and ray tracing rendering engines) are capable of generating photorealistic two-dimensional (2-D) images of virtual environments of arrangements of 3-D models of objects that can be used for training deep learning models.

SUMMARY

[0004] Aspects of embodiments of the present disclosure relate to machine learning techniques, in particular the synthesis or generation of data for training machine learning models. In particular, aspects of embodiments of the present disclosure relate to synthesizing images for training machine learning models to perform computer vision tasks on input images that are captured based on imaging modalities other than images of the intensity of visible light in a scene.

1 **[0005]** According to one embodiment of the present disclosure, a method of
generating synthetic images of virtual scenes includes: placing, by a synthetic data
generator implemented by a processor and memory, three-dimensional (3-D) models
of objects in a 3-D virtual scene; adding, by the synthetic data generator, lighting to
5 the 3-D virtual scene, the lighting including one or more illumination sources;
applying, by the synthetic data generator, imaging modality-specific materials to the
3-D models of objects in the 3-D virtual scene in accordance with a selected imaging
modality, each of the imaging modality-specific materials including an empirical
model; setting a scene background in accordance with the selected imaging
10 modality; and rendering, by the synthetic data generator, a two-dimensional image of
the 3-D virtual scene based on the selected imaging modality to generate a synthetic
image in accordance with the selected imaging modality.

15 **[0006]** The empirical model may be generated based on sampled images
captured of a surface of a material using an imaging system configured to capture
images using the selected imaging modality, the sampled images may include
images captured of the surface of the material from a plurality of different poses with
respect to a normal direction of the surface of the material.

20 **[0007]** The selected imaging modality may be polarization and the imaging
system includes a polarization camera.

25 **[0008]** The selected imaging modality may be thermal and the imaging system
may include a thermal camera. The thermal camera may include a polarizing filter.

30 **[0009]** Each of the sampled images may be stored in association with the
corresponding angle of its pose with respect to the normal direction of the surface of
the material.

35 **[0010]** The sampled images may include: a first plurality of sampled images
captured of the surface of the material illuminated by light having a first spectral
profile; and a second plurality of sampled images captured of the surface of the
material illuminated by light having a second spectral profile different from the first
spectral profile.

40 **[0011]** The empirical model may include a surface light field function computed by
interpolating between two or more of the sampled images.

45 **[0012]** The empirical model may include a surface light field function computed by
a deep neural network trained on the sampled images.

50 **[0013]** The empirical model may include a surface light field function computed by
a generative adversarial network trained on the sampled images.

55 **[0014]** The empirical model may include a surface light field function computed by
a mathematical model generated based on the sampled images.

1 **[0015]** The method may further include applying style transfer to the synthetic image.

5 **[0016]** According to one embodiment of the present disclosure, a method for generating tensors in polarization feature space for a 3-D virtual scene, includes: rendering, by a synthetic data generator implemented by a processor and memory, an image of surface normals of a 3-D virtual scene including a plurality of 3-D models of objects, the surface normals including an azimuth angle component and a zenith angle component; determining, by the synthetic data generator, for a surface of a 3-D model of an object of the 3-D virtual scene, a material of the object; and
10 computing, by the synthetic data generator, the tensors in polarization feature space in accordance with the azimuth angle component and the zenith angle component of the surface normal, the tensors in polarization feature space including: a degree of linear polarization; and an angle of linear polarization at the surface of the object.

15 **[0017]** The method may further include: determining whether the surface of the 3-D model of the object is specular dominant; computing the tensors in polarization feature space based on specular polarization equations in response to determining that the surface of the 3-D model of the object is specular dominant; and computing the tensors in polarization feature space based on diffuse polarization equations in response to determining that the surface of the 3-D model of the object is specular
20 dominant.

25 **[0018]** The method may further include: computing the tensors in polarization feature space based on diffuse polarization equations.

30 **[0019]** The method may further include applying style transfer to the tensors in polarization feature space.

35 **[0020]** According to one embodiment of the present disclosure, a method for synthesizing a training data set based on generating a plurality of synthetic images generated in accordance with any of the above methods.

40 **[0021]** According to one embodiment of the present disclosure, a method for training a machine learning model includes: generating a training data set in accordance with any of the above methods; and computing parameters of the machine learning model based on the training data set.

45 **[0022]** According to one embodiment of the present disclosure, a system for generating synthetic images of virtual scenes includes: a processor; and memory storing instructions that, when executed by the processor, cause the processor to implement a synthetic data generator to: place three-dimensional (3-D) models of objects in a 3-D virtual scene; add lighting to the 3-D virtual scene, the lighting including one or more illumination sources; apply imaging modality-specific materials

1 to the 3-D models of objects in the 3-D virtual scene in accordance with a selected
imaging modality, each of the imaging modality-specific materials including an
empirical model; set a scene background in accordance with the selected imaging
modality; and render a two-dimensional image of the 3-D virtual scene based on the
5 selected imaging modality to generate a synthetic image in accordance with the
selected imaging modality.

[0023] The empirical model may be generated based on sampled images
captured of a surface of a material using an imaging system configured to capture
images using the selected imaging modality, and the sampled images may include
10 images captured of the surface of the material from a plurality of different poses with
respect to a normal direction of the surface of the material.

[0024] The selected imaging modality may be polarization and the imaging
system may include a polarization camera.

[0025] The selected imaging modality may be thermal and the imaging system
15 may include a thermal camera. The thermal camera may include a polarizing filter.

[0026] Each of the sampled images may be stored in association with the
corresponding angle of its pose with respect to the normal direction of the surface of
the material.

[0027] The sampled images may include: a first plurality of sampled images
20 captured of the surface of the material illuminated by light having a first spectral
profile; and a second plurality of sampled images captured of the surface of the
material illuminated by light having a second spectral profile different from the first
spectral profile.

[0028] The empirical model may include a surface light field function computed by
25 interpolating between two or more of the sampled images.

[0029] The empirical model may include a surface light field function computed by
a deep neural network trained on the sampled images.

[0030] The empirical model may include a surface light field function computed by
a generative adversarial network trained on the sampled images.

30 **[0031]** The empirical model may include a surface light field function computed by
a mathematical model generated based on the sampled images.

[0032] The memory may further store instructions that, when executed by the
processor, cause the synthetic data generator to apply style transfer to the synthetic
image.

35 **[0033]** According to one embodiment of the present disclosure, a system for
generating tensors in polarization feature space for a 3-D virtual scene includes: a
processor; and memory storing instructions that, when executed by the processor,

1 cause the processor to implement a synthetic data generator to: render an image of
surface normals of a 3-D virtual scene including a plurality of 3-D models of objects,
the surface normals including an azimuth angle component and a zenith angle
component; determine for a surface of a 3-D model of an object of the 3-D virtual
5 scene, a material of the object; and compute the tensors in polarization feature
space in accordance with the azimuth angle component and the zenith angle
component of the surface normal, the tensors in polarization feature space including:
a degree of linear polarization; and an angle of linear polarization at the surface of
the object.

10 **[0034]** The memory may further store instructions that, when executed by the
processor, cause the synthetic data generator to: determine whether the surface of
the 3-D model of the object is specular dominant; compute the tensors in polarization
feature space based on specular polarization equations in response to determining
that the surface of the 3-D model of the object is specular dominant; and compute
15 the tensors in polarization feature space based on diffuse polarization equations in
response to determining that the surface of the 3-D model of the object is specular
dominant.

[0035] The memory may further store instructions that, when executed by the
processor, cause the synthetic data generator to: compute the tensors in polarization
20 feature space based on diffuse polarization equations.

[0036] The memory may further store instructions that, when executed by the
processor, cause the synthetic data generator to: apply style transfer to the tensors
in polarization feature space.

[0037] According to one embodiment of the present disclosure, a system for
25 synthesizing a training data set is configured to synthesize the training data set using
the system of any of the above systems.

[0038] According to one embodiment of the present disclosure, a system for
training a machine learning model includes: a processor; and memory storing
instructions that, when executed by the processor, cause the processor to: receive a
30 training data set generated by any of the above systems; and compute parameters
of the machine learning model based on the training data set.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] The accompanying drawings, together with the specification, illustrate
35 exemplary embodiments of the present invention, and, together with the description,
serve to explain the principles of the present invention.

- 1 **[0040]** FIG. 1 is a block diagram depicting a system for training a statistical model to perform computer vision tasks based on images in various modalities, where the training is performed using data generated in accordance with embodiments of the present disclosure.
- 5 **[0041]** FIG. 2 is a schematic block diagram of a computer vision system configured to use polarization imaging and that can be trained based on synthetic polarization imaging data generated according to one embodiment of the present invention.
- 10 **[0042]** FIG. 3A is an image or intensity image of a scene with one real transparent ball placed on top of a printout of photograph depicting another scene containing two transparent balls ("spoofs") and some background clutter.
- 15 **[0043]** FIG. 3B depicts the intensity image of FIG. 3A with an overlaid segmentation mask as computed by a comparative Mask Region-based Convolutional Neural Network (Mask R-CNN) identifying instances of transparent balls, where the real transparent ball is correctly identified as an instance, and the two spoofs are incorrectly identified as instances.
- 20 **[0044]** FIG. 3C is an angle of polarization image computed from polarization raw frames captured of the scene according to one embodiment of the present invention.
- 25 **[0045]** FIG. 3D depicts the intensity image of FIG. 3A with an overlaid segmentation mask as computed using polarization data in accordance with an embodiment of the present invention, where the real transparent ball is correctly identified as an instance and the two spoofs are correctly excluded as instances.
- 30 **[0046]** FIG. 4 is a high-level depiction of the interaction of light with transparent objects and non-transparent (e.g., diffuse and/or reflective) objects.
- 35 **[0047]** FIG. 5 is a graph of the energy of light that is transmitted versus reflected over a range of incident angles to a surface having a refractive index of approximately 1.5.
- [0048]** FIG. 6 is a flowchart depicting a pipeline for generating synthetic mages according to one embodiment of the present disclosure.
- [0049]** FIG. 7 is a schematic diagram of the sampling a real material from multiple angles using a polarization camera system according to one embodiment of the present disclosure.
- [0050]** FIG. 8 is a flowchart depicting a method for capturing images of a material from different perspectives using a particular imaging modality to be modeled according to one embodiment of the present disclosure.

1 **[0051]** FIG. 9 is a flowchart depicting a method for rendering a portion of a virtual object based on the empirical model of a material according to one embodiment of the present disclosure.

5 **[0052]** FIG. 10 is a flowchart depicting a method for computing synthetic features or tensors in polarization representation spaces for a virtual scene according to one embodiment of the present disclosure.

[0053] FIG. 11 is a flowchart depicting a method for generating a training data set according to one embodiment of the present disclosure.

10 DETAILED DESCRIPTION

[0054] In the following detailed description, only certain exemplary embodiments of the present invention are shown and described, by way of illustration. As those skilled in the art would recognize, the invention may be embodied in many different forms and should not be construed as being limited to the embodiments set forth
15 herein. Like reference numerals designate like elements throughout the specification.

[0055] Aspects of embodiments of the present disclosure relate to systems and methods for synthesizing or generating data for training machine learning models for performing computer vision tasks on images captured based on modalities other than standard modalities such as color or monochrome cameras configured to
20 capture images based on the intensity of visible light. Examples of other modalities include images captured based on polarized light (e.g., images captured with a polarizing filter or polarization filter in an optical path of the camera for capturing circularly and/or linearly polarized light), non-visible or invisible light (e.g., light in the infrared or ultraviolet ranges), and combinations thereof (e.g., polarized infrared
25 light), however, embodiments of the present disclosure are not limited thereto and may be applied to other multi-spectral imaging techniques.

[0056] In more detail, aspects of embodiments of the present disclosure relate to generating synthetic images of objects in different imaging modalities for training a machine learning model to perform a computer vision task.

30 **[0057]** Generally, a computer vision system for computing segmentation maps that classify objects depicted in a scene may include a trained convolutional neural network that takes two-dimensional images (e.g., as captured by a color camera) as input and outputs segmentation maps based on those images. Such a convolutional neural network may be pre-trained on an existing data set such as ImageNet (see,
35 e.g., see, e.g., J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database. *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2009). However, these existing data sets may contain images

1 that are not representative of the images that are expected to be encountered in the
particular application of the computer vision system, and therefore these pre-trained
models may have poor performance on the particular computer vision task that the
computer vision system is intended to perform. For example, a computer vision
5 system for a manufacturing environment is more likely to encounter images of tools,
partially assembled products, manufacturing components, and the like, rather than
images of people, animals, household objects, and outdoors environments that may
be found in more "general purpose" data sets.

[0058] As such, "retraining" relates to updating the parameters (e.g., connection
10 weights) of a pre-trained model based on additional training data from a particular
target domain associated with the task to be performed by the re-trained model.
Continuing the above example, labeled images of tools, partially assembled
products, components, and the like from a particular manufacturing environment may
be used as training data for retraining a pre-trained model (e.g., a pre-trained
15 convolutional neural network) to improve its performance in detecting and classifying
objects encountered in that manufacturing environment. However, manually
collecting different images of typical scenes in that manufacturing environment and
labeling these images based on their ground truth values (e.g., identifying pixels that
correspond to different classes of objects) is generally a time consuming and
20 expensive task.

[0059] As noted above, three-dimensional (3-D) rendering computer graphics
software may be used to generate training data for training machine learning models
for performing computer vision tasks. For example, existing 3-D models of those
tools, partially assembled products, and manufacturing components may be
25 arranged in a virtual scene in accordance with the variety of ways in which such
objects may be encountered in the real-world (e.g., including lighting conditions and
3-D models of support surfaces and equipment in the environment). For example,
partially assembled products may be placed on a 3-D model of a conveyor belt,
components may be located in a parts bin, and tools may be placed on a tool bench
30 and/or within a scene in the process of positioning a component within a partially
assembled product. Accordingly, a 3-D computer graphics rendering system is used
to generate photorealistic images of the range of typical arrangements of objects in a
particular environment. These generated images can also be automatically labeled.
In particular, when the particular 3-D models used to depict each of the different
35 types of objects are already associated with class labels (e.g., screws of various
sizes, pre-assembled components, products at various stages of assembly,

1 particular types of tools, etc.), segmentation maps can be automatically generated (e.g., by mapping surfaces of objects to their particular class labels).

5 **[0060]** However, 3-D rendering computer graphics software systems are generally tailored for generating images that represent typical imaging modalities based on the intensity of visible light (e.g., the intensities of red, green, and blue light). Such 3-D rendering software, such as Blender® by the Blender Foundation, generally does not account for behaviors of electromagnetic radiation that may be invisible or otherwise negligible when rendering photorealistic scenes. Examples of these additional behaviors include the polarization of light (e.g., as polarized light
10 interacts with transparent objects and reflective objects in a scene, as detected by a camera with a polarizing filter in its optical path), thermal or infrared radiation (e.g., as emitted by warm objects in a scene and as detected by a camera system sensitive to detect infrared light), ultraviolet radiation (e.g., as detected by a camera system sensitive to ultraviolet light), combinations thereof (e.g., polarization with
15 thermal radiation, polarization with visible light, polarization with ultraviolet light, etc.), and the like.

[0061] Therefore, aspects of embodiments of the present disclosure relate to systems and methods for modeling the behavior of various materials when imaged based on polarization or other imaging modalities. The data (e.g., images) generated
20 in accordance with embodiments of the present disclosure may then be used as training data for training deep learning models such as deep convolutional neural networks to compute predictions based on imaging modalities other than standard imaging modalities (e.g., the intensity of visible light or light in a visible portion of the electromagnetic spectrum).

25 **[0062]** As a motivating example, embodiments of the present disclosure will be described in the context of generating synthetic images of objects captured through a polarizing filter (referred to herein as “polarization raw frames”), where these images may be used in training a deep neural network such as a convolutional neural network to perform a task based on polarization raw frames. However,
30 embodiments of the present disclosure are not limited to generating synthetic polarization raw frames for training a convolutional neural network that takes polarization raw frames (or features extracted therefrom) as input data.

[0063] FIG. 1 is a block diagram depicting a system for training a statistical model to perform computer vision tasks based on images in various modalities, where the
35 training is performed using data generated in accordance with embodiments of the present disclosure. as shown in FIG. 1, training data 5 is supplied to a model training system 7, which takes a model 30 (e.g., a pre-trained model or a model architecture

1 with initialized weights) and uses the training data 5 to generate a trained model (or
re-trained model) 32. The model 30 and the trained model 32 may be a statistical
model such as a deep neural network (deep neural networks include convolutional
neural networks). A synthetic data generator 40 according to embodiments of the
5 present disclosure generates synthesized data 42, which may be included with the
training data 5 for generating the trained model 32. The model training system 7 may
apply an iterative process for updating the parameters of the model 30 to generate
the trained model 32 in accordance with the supplied training data 5 (e.g., including
the synthesized data 42). The updating of the parameters of the model 30 may
10 include, for example, applying gradient descent (and, in the case of a neural
network, backpropagation) in accordance with a loss function measuring a difference
between the labels and the output of the model in response to the training data. The
model training system 7 and the synthetic data generator 40 may be implemented
using one or more electronic circuits.

15 **[0064]** According to various embodiments of the present disclosure, the model
training system 7 and/or the synthetic data generator 40 are implemented using one
or more electronic circuits configured to perform various operations as described in
more detail below. Types of electronic circuits may include a central processing unit
(CPU), a graphics processing unit (GPU), an artificial intelligence (AI) accelerator
20 (e.g., a vector processor, which may include vector arithmetic logic units configured
efficiently perform operations common to neural networks, such dot products and
softmax), a field programmable gate array (FPGA), an application specific integrated
circuit (ASIC), a digital signal processor (DSP), or the like. For example, in some
circumstances, aspects of embodiments of the present disclosure are implemented
25 in program instructions that are stored in a non-volatile computer readable memory
where, when executed by the electronic circuit (e.g., a CPU, a GPU, an AI
accelerator, or combinations thereof), perform the operations described herein to
compute a segmentation map 20 from input polarization raw frames 18. The
operations performed by the model training system 7 and the synthetic data
30 generator 40 may be performed by a single electronic circuit (e.g., a single CPU, a
single GPU, or the like) or may be allocated between multiple electronic circuits (e.g.,
multiple GPUs or a CPU in conjunction with a GPU). The multiple electronic circuits
may be local to one another (e.g., located on a same die, located within a same
package, or located within a same embedded device or computer system) and/or
35 may be remote from one other (e.g., in communication over a network such as a
local personal area network such as Bluetooth®, over a local area network such as a
local wired and/or wireless network, and/or over wide area network such as the

1 internet, such a case where some operations are performed locally and other
operations are performed on a server hosted by a cloud computing service). One or
more electronic circuits operating to implement the model training system 7 and the
synthetic data generator 40 may be referred to herein as a computer or a computer
5 system, which may include memory storing instructions that, when executed by the
one or more electronic circuits, implement the systems and methods described
herein.

[0065] FIG. 2 is a schematic block diagram of a computer vision system
configured to use polarization imaging and that can be trained based on synthetic
10 polarization imaging data generated according to one embodiment of the present
invention.

[0066] For context, FIG. 2 is a schematic depiction of a system in which a
polarization camera images a scene and supplies polarization raw frames to a
computer vision system that includes a model that is trained to perform computer
15 vision tasks based on polarization raw frames or polarization features computed
based on polarization raw frames

[0067] A polarization camera 10 has a lens 12 with a field of view, where the lens
12 and the camera 10 are oriented such that the field of view encompasses the
scene 1. The lens 12 is configured to direct light (e.g., focus light) from the scene 1
20 onto a light sensitive medium such as an image sensor 14 (e.g., a complementary
metal oxide semiconductor (CMOS) image sensor or charge-coupled device (CCD)
image sensor).

[0068] The polarization camera 10 further includes a polarizer or polarizing filter
or polarization mask 16 placed in the optical path between the scene 1 and the
25 image sensor 14. According to various embodiments of the present disclosure, the
polarizer or polarization mask 16 is configured to enable the polarization camera 10
to capture images of the scene 1 with the polarizer set at various specified angles
(e.g., at 45° rotations or at 60° rotations or at non-uniformly spaced rotations).

[0069] As one example, FIG. 2 depicts an embodiment where the polarization
30 mask 16 is a polarization mosaic aligned with the pixel grid of the image sensor 14 in
a manner similar to a red-green-blue (RGB) color filter (e.g., a Bayer filter) of a color
camera. In a manner similar to how a color filter mosaic filters incoming light based
on wavelength such that each pixel in the image sensor 14 receives light in a
particular portion of the spectrum (e.g., red, green, or blue) in accordance with the
35 pattern of color filters of the mosaic, a polarization mask 16 using a polarization
mosaic filters light based on linear polarization such that different pixels receive light
at different angles of linear polarization (e.g., at 0°, 45°, 90°, and 135°, or at 0°, 60°

1 degrees, and 120°). Accordingly, the polarization camera 10 using a polarization
mask 16 such as that shown in FIG. 2 is capable of concurrently or simultaneously
capturing light at four different linear polarizations. One example of a polarization
camera is the Blackfly® S Polarization Camera produced by FLIR® Systems, Inc. of
5 Wilsonville, Oregon.

[0070] While the above description relates to some possible implementations of a
polarization camera using a polarization mosaic, embodiments of the present
disclosure are not limited thereto and encompass other types of polarization
cameras that are capable of capturing images at multiple different polarizations. For
10 example, the polarization mask 16 may have fewer than or more than four different
polarizations, or may have polarizations at different angles (e.g., at angles of
polarization of: 0°, 60° degrees, and 120° or at angles of polarization of 0°, 30°, 60°,
90°, 120°, and 150°). As another example, the polarization mask 16 may be
implemented using an electronically controlled polarization mask, such as an electro-
15 optic modulator (e.g., may include a liquid crystal layer), where the polarization
angles of the individual pixels of the mask may be independently controlled, such
that different portions of the image sensor 14 receive light having different
polarizations. As another example, the electro-optic modulator may be configured to
transmit light of different linear polarizations when capturing different frames, e.g., so
20 that the camera captures images with the entirety of the polarization mask set to,
sequentially, to different linear polarizer angles (e.g., sequentially set to: 0 degrees;
45 degrees; 90 degrees; or 135 degrees). As another example, the polarization
mask 16 may include a polarizing filter that rotates mechanically, such that different
polarization raw frames are captured by the polarization camera 10 with the
25 polarizing filter mechanically rotated with respect to the lens 12 to transmit light at
different angles of polarization to image sensor 14.

[0071] A polarization camera may also refer to an array of multiple cameras
having substantially parallel optical axes, such that each of the cameras captures
images of a scene from substantially the same pose. The optical path of each
30 camera of the array includes a polarizing filter, where the polarizing filters have
different angles of polarization. For example, a two-by-two (2×2) array of four
cameras may include one camera having a polarizing filter set at an angle of 0°, a
second camera having a polarizing filter set at an angle of 45°, a third camera having
a polarizing filter set at an angle of 90°, and a fourth camera having a polarizing filter
35 set at an angle of 135°.

[0072] As a result, the polarization camera captures multiple input images 18 (or
polarization raw frames) of the scene 1, where each of the polarization raw frames

1 18 corresponds to an image taken behind a polarization filter or polarizer at a
different angle of polarization ϕ_{pol} (e.g., 0 degrees, 45 degrees, 90 degrees, or 135
degrees). Each of the polarization raw frames is captured from substantially the
same pose with respect to the scene 1 (e.g., the images captured with the
5 polarization filter at 0 degrees, 45 degrees, 90 degrees, or 135 degrees are all
captured by a same polarization camera located at a same location and orientation),
as opposed to capturing the polarization raw frames from disparate locations and
orientations with respect to the scene. The polarization camera 10 may be
10 configured to detect light in a variety of different portions of the electromagnetic
spectrum, such as the human-visible portion of the electromagnetic spectrum, red,
green, and blue portions of the human-visible spectrum, as well as invisible portions
of the electromagnetic spectrum such as infrared and ultraviolet.

[0073] FIGS. 3A, 3B, 3C, and 3D provide background for illustrating the
segmentation maps computed by a comparative approach and semantic
15 segmentation or instance segmentation according to embodiments of the present
disclosure. In more detail, FIG. 3A is an image or intensity image of a scene with one
real transparent ball placed on top of a printout of photograph depicting another
scene containing two transparent balls ("spoofs") and some background clutter. FIG.
3B depicts an segmentation mask as computed by a comparative Mask Region-
20 based Convolutional Neural Network (Mask R-CNN) identifying instances of
transparent balls overlaid on the intensity image of FIG. 3A using different patterns of
lines, where the real transparent ball is correctly identified as an instance, and the
two spoofs are incorrectly identified as instances. In other words, the Mask R-CNN
algorithm has been fooled into labeling the two spoof transparent balls as instances
25 of actual transparent balls in the scene.

[0074] FIG. 3C is an angle of linear polarization (AOLP) image computed from
polarization raw frames captured of the scene according to one embodiment of the
present invention. As shown in FIG. 3C, transparent objects have a very unique
texture in polarization space such as the AOLP domain, where there is a geometry-
30 dependent signature on edges and a distinct or unique or particular pattern that
arises on the surfaces of transparent objects in the angle of linear polarization. In
other words, the *intrinsic texture* of the transparent object (e.g., as opposed to
extrinsic texture adopted from the background surfaces visible through the
transparent object) is more visible in the angle of polarization image of FIG. 3C than
35 it is in the intensity image of FIG. 3A.

[0075] FIG. 3D depicts the intensity image of FIG. 3A with an overlaid
segmentation mask as computed using polarization data in accordance with an

embodiment of the present invention, where the real transparent ball is correctly identified as an instance using an overlaid pattern of lines and the two spoofs are correctly excluded as instances (e.g., in contrast to FIG. 3B, FIG. 3D does not include overlaid patterns of lines over the two spoofs). While FIGS. 3A, 3B, 3C, and 3D illustrate an example relating to detecting a real transparent object in the presence of spoof transparent objects, embodiments of the present disclosure are not limited thereto and may also be applied to other optically challenging objects, such as transparent, translucent, and non-matte or non-Lambertian objects, as well as non-reflective (e.g., matte black objects) and multipath inducing objects.

[0076] *Polarization feature representation spaces*

[0077] Some aspects of embodiments of the present disclosure relate to systems and methods for extracting features from polarization raw frames, where these extracted features are used by the processing system 100 for the robust detection of optically challenging characteristics in the surfaces of objects. In contrast, comparative techniques relying on intensity images alone may fail to detect these optically challenging features or surfaces (e.g., comparing the intensity image of FIG. 3A with the AOLP image of FIG. 3C, discussed above). The term “first tensors” in “first representation spaces” will be used herein to refer to features computed from (e.g., extracted from) polarization raw frames 18 captured by a polarization camera, where these first representation spaces include at least polarization feature spaces (e.g., feature spaces such as AOLP and DOLP that contain information about the polarization of the light detected by the image sensor) and may also include non-polarization feature spaces (e.g., feature spaces that do not require information regarding the polarization of light reaching the image sensor, such as images computed based solely on intensity images captured without any polarizing filters).

[0078] The interaction between light and transparent objects is rich and complex, but the material of an object determines its transparency under visible light. For many transparent household objects, the majority of visible light passes straight through and a small portion (~4% to ~8%, depending on the refractive index) is reflected. This is because light in the visible portion of the spectrum has insufficient energy to excite atoms in the transparent object. As a result, the texture (e.g., appearance) of objects behind the transparent object (or visible through the transparent object) dominate the appearance of the transparent object. For example, when looking at a transparent glass cup or tumbler on a table, the appearance of the objects on the other side of the tumbler (e.g., the surface of the table) generally dominate what is seen through the cup. This property leads to some difficulties when

1 attempting to detect surface characteristics of transparent objects such as glass windows and glossy, transparent layers of paint, based on intensity images alone:

5 **[0079]** FIG. 4 is a high-level depiction of the interaction of light with transparent objects and non-transparent (e.g., diffuse and/or reflective) objects. As shown in FIG. 4, a polarization camera 10 captures polarization raw frames of a scene that includes a transparent object 402 in front of an opaque background object 403. A light ray 410 hitting the image sensor 14 of the polarization camera 10 contains polarization information from both the transparent object 402 and the background object 403. The small fraction of reflected light 412 from the transparent object 402 is heavily polarized, and thus has a large impact on the polarization measurement, in contrast to the light 413 reflected off the background object 403 and passing through the transparent object 402.

10 **[0080]** Similarly, a light ray hitting the surface of an object may interact with the shape of the surface in various ways. For example, a surface with a glossy paint may behave substantially similarly to a transparent object in front of an opaque object as shown in FIG. 4, where interactions between the light ray and a transparent or translucent layer (or clear coat layer) of the glossy paint causes the light reflecting off of the surface to be polarized based on the characteristics of the transparent or translucent layer (e.g., based on the thickness and surface normals of the layer), which are encoded in the light ray hitting the image sensor. Similarly, as discussed in more detail below with respect to shape from polarization (SfP) theory, variations in the shape of the surface (e.g., direction of the surface normals) may cause significant changes in the polarization of light reflected by the surface of the object. For example, smooth surfaces may generally exhibit the same polarization characteristics throughout, but a scratch or a dent in the surface changes the direction of the surface normals in those areas, and light hitting scratches or dents may be polarized, attenuated, or reflected in ways different than in other portions of the surface of the object. Models of the interactions between light and matter generally consider three fundamentals: geometry, lighting, and material. Geometry is based on the shape of the material. Lighting includes the direction and color of the lighting. Material can be parameterized by the refractive index or angular reflection/transmission of light. This angular reflection is known as a bi-directional reflectance distribution function (BRDF), although other functional forms may more accurately represent certain scenarios. For example, the bidirectional subsurface scattering distribution function (BSSRDF) would be more accurate in the context of materials that exhibit subsurface scattering (e.g. marble or wax).

[0081] A light ray 410 hitting the image sensor 16 of a polarization camera 10 has three measurable components: the intensity of light (intensity image/ I), the percentage or proportion of light that is linearly polarized (degree of linear polarization/DOLP/ ρ), and the direction of that linear polarization (angle of linear polarization/AOLP/ ϕ). These properties encode information about the surface curvature and material of the object being imaged, which can be used by the predictor 800 to detect transparent objects, as described in more detail below. In some embodiments, the predictor 800 can detect other optically challenging objects based on similar polarization properties of light passing through translucent objects and/or light interacting with multipath inducing objects or by non-reflective objects (e.g., matte black objects).

[0082] Therefore, some aspects of embodiments of the present invention relate synthesizing polarization raw frames that can be used to compute first tensors in one or more first representation spaces, which may include derived feature maps based on the intensity I , the DOLP ρ , and the AOLP ϕ . Some aspects of embodiments of the present disclosure also relate to directly synthesizing tensors in one or more representation spaces such as DOLP ρ and AOLP ϕ for use in training deep learning systems to perform computer vision tasks based on information regarding the polarization of light in a scene (and, in some embodiments, based on other imaging modalities such as thermal imaging and combinations of thermal and polarized imaging).

[0083] Measuring intensity I , DOLP ρ , and AOLP ϕ at each pixel requires 3 or more polarization raw frames of a scene taken behind polarizing filters (or polarizers) at different angles, ϕ_{pol} (e.g., because there are three unknown values to be determined: intensity I , DOLP ρ , and AOLP ϕ . For example, the FLIR® Blackfly® S Polarization Camera described above captures polarization raw frames with polarization angles ϕ_{pol} at 0 degrees, 45 degrees, 90 degrees, or 135 degrees, thereby producing four polarization raw frames $I_{\phi_{pol}}$, denoted herein as I_0 , I_{45} , I_{90} , and I_{135} .

[0084] The relationship between $I_{\phi_{pol}}$ and intensity I , DOLP ρ , and AOLP ϕ at each pixel can be expressed as:

$$I_{\phi_{pol}} = I \left(1 + \rho \cos \left(2(\phi - \phi_{pol}) \right) \right) \quad (1)$$

[0085] Accordingly, with four different polarization raw frames $I_{\phi_{pol}}$ (I_0 , I_{45} , I_{90} , and I_{135}), a system of four equations can be used to solve for the intensity I , DOLP ρ , and AOLP ϕ .

[0086] Shape from Polarization (SfP) theory (see, e.g., Gary A Atkinson and Edwin R Hancock. Recovery of surface orientation from diffuse polarization. IEEE

transactions on image processing, 15(6):1653–1664, 2006.) states that the relationship between the refractive index (n), azimuth angle (θ_a) and zenith angle (θ_z) of the surface normal of an object and the ϕ and ρ components of the light ray coming from that object follow the following characteristics when diffuse reflection is dominant:

$$\rho = \frac{\left(n - \frac{1}{n}\right)^2 \sin^2(\theta_z)}{2 + 2n^2 - \left(n + \frac{1}{n}\right)^2 \sin^2 \theta_z + 4 \cos \theta_z \sqrt{n^2 - \sin^2 \theta_z}} \quad (2)$$

$$\phi = \theta_a \quad (3)$$

and when the specular reflection is dominant:

$$\rho = \frac{2 \sin^2 \theta_z \cos \theta_z \sqrt{n^2 - \sin^2 \theta_z}}{n^2 - \sin^2 \theta_z - n^2 \sin^2 \theta_z + 2 \sin^4 \theta_z} \quad (4)$$

$$\phi = \theta_a - \frac{\pi}{2} \quad (5)$$

Note that in both cases ρ increases exponentially as θ_z increases and if the refractive index is the same, specular reflection is much more polarized than diffuse reflection.

[0001] Accordingly, some aspects of embodiments of the present disclosure relate to applying SfP theory to generate synthetic raw polarization frames 18 and/or AOLP and DOLP images based on the shapes of surfaces (e.g., the orientation of surfaces) in a virtual environment.

[0002] Light rays coming from a transparent objects have two components: a reflected portion including reflected intensity I_r , reflected DOLP ρ_r , and reflected AOLP ϕ_r and the refracted portion including refracted intensity I_t , refracted DOLP ρ_t , and refracted AOLP ϕ_t . The intensity of a single pixel in the resulting image can be written as:

$$I = I_r + I_t \quad (6)$$

[0003] When a polarizing filter having a linear polarization angle of ϕ_{pol} is placed in front of the camera, the value at a given pixel is:

$$I_{\phi_{pol}} = I_r \left(1 + \rho_r \cos \left(2(\phi_r - \phi_{pol})\right)\right) + I_t \left(1 + \rho_t \cos \left(2(\phi_t - \phi_{pol})\right)\right) \quad (7)$$

[0087] Solving the above expression for the values of a pixel in a DOLP ρ image and a pixel in an AOLP ϕ image in terms of I_r , ρ_r , ϕ_r , I_t , ρ_t , and ϕ_t :

$$\rho = \frac{\sqrt{(I_r \rho_r)^2 + (I_t \rho_t)^2 + 2 I_t \rho_t I_r \rho_r \cos(2(\phi_r - \phi_t))}}{I_r + I_t} \quad (8)$$

$$\phi = \arctan\left(\frac{I_r \rho_r \sin(2(\phi_r - \phi_t))}{I_t \rho_t + I_r \rho_r \cos(2(\phi_r - \phi_t))}\right) + \phi_r \quad (9)$$

[0088] Accordingly, equations (7), (8), and (9), above provide a model for forming first tensors 50 in first representation spaces that include an intensity image I , a DOLP image ρ , and an AOLP image ϕ according to one embodiment of the present disclosure, where the use of polarization images or tensor in polarization representation spaces (including DOLP image ρ and an AOLP image ϕ based on equations (8) and (9)) enables trained computer vision systems to reliably detect of optically challenging surface characteristics of objects that are generally not detectable by comparative systems that use only intensity I images as input.

[0089] In more detail, first tensors in polarization representation spaces (among the derived feature maps) such as the polarization images DOLP ρ and AOLP ϕ can reveal surface characteristics of objects that might otherwise appear textureless in an intensity I domain. A transparent object may have a texture that is invisible in the intensity domain I because this intensity is strictly dependent on the ratio of I_r / I_t (see equation (6)). Unlike opaque objects where $I_t = 0$, transparent objects transmit most of the incident light and only reflect a small portion of this incident light. As another example, thin or small deviations in the shape of an otherwise smooth surface (or smooth portions in an otherwise rough surface) may be substantially invisible or have low contrast in the intensity I domain (e.g., a domain in which polarization of light is not taken into account), but may be very visible or may have high contrast in a polarization representation space such as DOLP ρ or AOLP ϕ .

[0090] As such, one exemplary method to acquire surface topography is to use polarization cues in conjunction with geometric regularization. The Fresnel equations relate the AOLP ϕ and the DOLP ρ with surface normals. These equations can be useful for anomaly detection by exploiting what is known as *polarization patterns* of the surface. A polarization pattern is a tensor of size $[M, N, K]$ where M and N are horizontal and vertical pixel dimensions, respectively, and where K is the polarization data channel, which can vary in size. For example, if circular polarization is ignored and only linear polarization is considered, then K would be equal to two, because linear polarization has both an angle and a degree of polarization (AOLP ϕ and DOLP ρ). Analogous to a Moire pattern, in some embodiments of the present disclosure, the feature extraction module 700 extracts a polarization pattern in polarization representation spaces (e.g., AOLP space and DOLP space). In the example characterization output 20 shown in FIG. 1A and FIG. 1B shown above, the horizontal and vertical dimensions correspond to the lateral field of view of a narrow strip or patch of a surface of an object captured by the polarization camera 10.

1 However, this is one exemplary case: in various embodiments, the strip or patch of the surface may be vertical (e.g., much taller than wide), horizontal (e.g., much wider than tall), or have a more conventional field of view (FoV) that tends closer toward a square (e.g., a 4:3 ratio or 16:9 ratio of width to height).

5 **[0091]** While the preceding discussion provides specific examples of polarization representation spaces based on linear polarization in the case of using a polarization camera having one or more linear polarizing filters to capture polarization raw frames corresponding to different angles of linear polarization and to compute tensors in linear polarization representation spaces such as DOLP and AOLP, embodiments of
10 the present disclosure are not limited thereto. For example, in some embodiments of the present disclosure, a polarization camera includes one or more circular polarizing filters configured to pass only circularly polarized light, and where polarization patterns or first tensors in circular polarization representation space are further extracted from the polarization raw frames. In some embodiments, these additional
15 tensors in circular polarization representation space are used alone, and in other embodiments they are used together with the tensors in linear polarization representation spaces such as AOLP and DOLP. For example, a polarization pattern including tensors in polarization representation spaces may include tensors in circular polarization space, AOLP, and DOLP, where the polarization pattern may
20 have dimensions $[M, N, K]$, where K is three to further include the tensor in circular polarization representation space.

[0092] FIG. 5 is a graph of the energy of light that is transmitted versus reflected over a range of incident angles to a surface having a refractive index of approximately 1.5. As shown in FIG. 5, the slopes of the transmitted energy (shown
25 in FIG. 5 with a solid line) and reflected energy (shown in FIG. 5 with a dotted line) lines are relatively small at low incident angles (e.g., at angles closer to perpendicular to the plane of the surface). As such, small differences in the angle of the surface may be difficult to detect (low contrast) in the polarization pattern when the angle of incidence is low (e.g., close to perpendicular to the surface, in other
30 words, close to the surface normal). On the other hand, the slope of the reflected energy increases from flat, as the angle of incidence increases, and the slope of the transmitted energy decreases from flat (to have a larger absolute value) as the angle of incidence increases. In the example shown in FIG. 5 with an index of refraction of 1.5, the slopes of both lines are substantially steeper beginning at an incident angle
35 of around 60° , and their slopes are very steep at an incident angle of around 80° . The particular shapes of the curves may change for different materials in accordance with the refractive index of the material. Therefore, capturing images of surfaces

1 under inspection at incident angles corresponding to steeper portions of the curves
(e.g., angles close to parallel to the surface, such as around 80° in the case of a
refractive index of 1.5, as shown in FIG. 5) can improve the contrast and detectability
of variations in the surface shapes in the polarization raw frames 18 and may
5 improve the detectability of such features in tensors in polarization representation
spaces, because small changes in incident angle (due to the small changes in the
surface normal) can cause large changes in the captured polarization raw frames.

[0093] The use of polarization cameras to detect the presence and shape of
optically challenging objects and surfaces is described in more detail, for example, in
10 PCT Patent Application No. US/2020/048604, filed on August 28, 2020 and in PCT
Patent Application No. US/2020/051243, filed on September 17, 2020, the entire
disclosures of which are incorporated by reference herein. Such computer vision
systems may be trained to perform computer vision tasks on polarization data based
on training data generated in accordance with embodiments of the present
15 disclosure. In some embodiments, these computer vision systems use machine
learning models such as deep neural networks (e.g., convolutional neural networks)
to perform the computer vision tasks, where the deep learning models are configured
to take, as input, polarization raw frames and/or features in polarization
representation spaces.

20 **[0094]** Simulating the physics of polarization for different materials is a complex
task requiring an understanding of the material properties, the spectral profile and
polarization parameters of the illumination used, and the angle at which the reflected
light is observed by the observer. To truly simulate the physics of light polarization
and its impact on the illumination of objects would not only be a complex task but
25 would also be an intensely compute intensive task, such as by applying a complex
forward model that typically produces highly inaccurate (unrealistic) images.
Accordingly, various comparative 3-D computer graphics systems typically do not
accurately model the physics of light polarization and its impact on the illumination of
objects, and therefore are not capable of synthesizing or rendering images of virtual
30 environments in a manner that realistically represents how a corresponding real
environment would appear if imaged with a camera with a polarizing filter in its
optical path (e.g., a polarization camera). As such, comparative techniques for
generating synthetic data for training computer vision systems that operate on
standard imaging modalities (such as visible light images without polarizing filters)
35 are generally incapable of generating training data for training computer vision
systems that operate on other imaging modalities (e.g., polarization cameras,
thermal cameras, and the like).

1 **[0095]** As discussed above, aspects of embodiments of the present disclosure
relate to generating or synthesizing data for training machine learning models to
take, as input, data captured using imaging modalities other than images captured
by a standard camera (e.g., a camera configured to capture the intensity of visible
5 light without the use of filters such as polarizing filters), which will be referred to
herein as multimodal images or plenoptic images. The term “multimodal” refers to
the plenoptic theory of light, where each dimension of the plenoptic domain (e.g.,
wavelength, polarization, angle, etc.) is an example of a modality of light.
Accordingly, multimodal or plenoptic imaging includes, but is not limited to, the
10 concurrent use of multiple imaging modalities. For example, the term “multimodal”
may be used herein to refer to a single imaging modality, where that single imaging
modality is a modality different from the intensity of visible light without the use of
filters such as polarizing filters. Polarization raw frames captured by one or more
polarization cameras and/or tensors in polarization representation spaces are one
15 example of a class of inputs in a multimodal or plenoptic imaging modality (e.g.,
using multimodal imaging or plenoptic imaging).

[0096] Generally, various aspects of embodiments of the present disclosure
relate to four techniques may be used separately or in combination as part of a
pipeline for generating synthetic training data in accordance with multimodal or
20 plenoptic imaging modalities such as a polarized imaging modality. These
techniques include: domain randomization, texture mapping, normal mapping, and
style transfer, and will be discussed in more detail below.

[0097] FIG. 6 is a flowchart depicting a pipeline for generating synthetic mages
according to one embodiment of the present disclosure. In some embodiments of the
25 present disclosure, the operations of FIG. 6 are performed by the synthetic data
generator 40, for example, in special-purpose program instructions stored in a
memory of the synthetic data generator 40 that, when executed by the processor of
the synthetic data generator 40, cause the synthetic data generator 40 to perform the
special-purpose operations described herein for generating synthetic images based
30 on the physical simulation of optical phenomena. For the sake of convenience,
aspects of embodiments of the present disclosure will be described in the context of
applying polarization imaging in a manufacturing context to perform computer vision
tasks on optically challenging manufacturing components and tools, such as objects
having transparent, shiny metal, and/or dark matte surfaces.

35 **[0098]** In operation 610, the synthetic data generator 40 places 3-D models of
objects in a virtual scene. In the context of generating synthetic images of scenes in
a manufacturing environment, 3-D models of objects may be readily available from

1 computer aided design (CAD) models of components and partially or fully assembled
manufactured products. These CAD models may have previously been produced in
the product design phase and may be obtained from, for example, the vendor of the
component (e.g., from the vendor who supplied the components to the
5 manufacturer), publicly available information (e.g., data sheets), or from internal
product designers employed by the manufacturer. In some circumstances the CAD
models may be manually generated based on specifications of a component.

[0099] In some embodiments of the present disclosure, the 3-D models of objects
are placed in a virtual scene in a manner resembling the arrangement of those
10 objects as they would be expected to be encountered for the particular computer
vision task that the machine learning model will be trained to perform.

[00100] In the above example of computer vision in a manufacturing context, one
task is to perform instance segmentation on a bin of components, where the
components may be homogeneous (e.g., all the components in the bin are the same,
15 such as a bin of springs or screws) or heterogeneous (e.g., a mix of different types of
components, such as screws of different sizes or screws mixed with matching nuts).
The objects may be randomly arranged within the bin, where the components may
be oriented in many different directions in the bin, and where, in a bin of
heterogeneous components, the different types of components are mixed together,
20 as opposed to being separated in different parts of the bin. A computer vision system
may be trained to compute a segmentation map of the bin, to identify the location
and orientation of individual components within the bin (and, in the case of a bin of
heterogeneous components, the types of the objects). This segmentation map can
then be used by an actuator system, such that a robotic arm, to pick components out
25 of the bin and add the picked components to a partially assembled product.

[00101] Accordingly, in some embodiments of the present disclosure, the synthetic
data generator 40 generates a scene of components in a bin by placing a 3-D model
of a virtual bin in a scene, and dropping 3-D models of components into the virtual
bin, as simulated using a physics simulation engine, such as a physics engine
30 incorporated into a 3-D computer graphics rendering system. For example, the
Blender® 3-D rendering software includes a physics system that simulates various
physical real-world phenomena such as the movement, collision, and potential
deformation of rigid bodies, cloth, soft bodies, fluids, and the like, as affected by
gravity or other forces. Accordingly, a rigid body simulation may be used for
35 simulating the dropping of rigid components (e.g., screws, bolts, relatively stiff
springs) into a rigid virtual bin, and a soft body simulation may be used for elastic or

1 deformable components (e.g., string, wire, plastic sheeting, etc.) into a rigid virtual bin.

5 **[00102]** In more detail, a variety of difference scenes representing different potential states of the bin may be generated, such as by dropping various numbers of instances of the 3-D models of the components into a virtual bin. For example, if a typical bin has a maximum capacity of 1,000 screws, various scenes can be generated by dropping 1,000 screws, 900 screws, 500 screws, 100 screws, and 10 screws into a virtual bin to generate different scenes representing different potential fullness states of the virtual bin. In addition, multiple scenes may be generated for
10 any given number of screws (or the number of screws may be randomized between the generation of different scenes), where the arrangement of components within the bin is also randomized, such as by dropping components into the bin, one at a time, from different random locations above the bin.

15 **[00103]** Accordingly, in operation 610, the synthetic data generator 40 generates a scene containing an arrangement of representative objects.

20 **[00104]** In operation 630, the synthetic data generator 40 adds lighting to the virtual scene generated in operation 610. In particular, the synthetic data generator 40 adds one or more light sources to the virtual scene, where the light sources illuminate part or all of the surfaces of the objects in the bin. In some embodiments, the position of the one or more light sources is randomized, and multiple scenes are generated with light sources in different locations (e.g., different angles and distances) relative to the bin of parts in order to improve the robustness of the training. In some embodiments of the present disclosure, the virtual lighting includes virtual light sources that are representative of the light sources that would be found in
25 environments in which the computer vision system is trained to operate. Examples of potential representative light sources include different color temperatures corresponding to, for example, incandescent lights, fluorescent lights, light emitting diode (LED) bulbs, natural light from a simulated window in the environment, and other forms of lighting technology, where the shape of the virtual lights (e.g., the
30 direction of the rays emitted by the lights) may be in a range from direct light to diffuse light. In some embodiments of the present disclosure, the character of the light (e.g., color temperature and shape) is also randomized to generate different scenes with different types of lighting.

35 **[00105]** In operation 650, the synthetic data generator 40 applies modality-specific materials to the objects in the 3-D virtual scene. For example, in the case of generating synthesized polarization imaging data, polarization-specific materials are applied to the objects in the virtual scene, whereas in the case of generating

1 synthesized thermal imaging data, thermal imaging-specific materials may be
applied to the objects in the virtual scene. For the sake of illustration, polarization-
specific materials will be described in detail herein, but embodiments of the present
disclosure are not limited thereto and may also be applied to generating and
5 applying materials specific to multimodal imaging modalities and/or plenoptic
imaging modalities.

[00106] Some aspects of embodiments of the present disclosure relate to domain
randomization, in which the material appearance of objects in a scene are
randomized beyond the typical appearance of the objects. For example, in some
10 embodiments, a large number of materials with random colors (e.g., thousands of
different materials of different, randomly selected colors) are applied to the different
objects in the virtual scene. In a real-world environment, the objects in a scene
generally have well-defined colors (e.g., rubber washers generally all look matte
black and screws may be particular shades of shiny black, matte black, gold, or
15 shiny metal). However, real-world objects can often have different appearances due
to changes in lighting conditions, such as the color temperature of lights, reflections,
specular highlights, and the like. Accordingly, applying randomization to the colors of
the materials applied to the objects when generating training data expands the
domain of the training data to also encompass unrealistic colors, thereby increasing
20 diversity in the training data for training a more robust machine learning model that is
capable of making accurate predictions (e.g., more accurate instance segmentation
maps) in a wider variety of real-world conditions.

[00107] Some aspects of embodiments of the present disclosure relate to
performing texture mapping to generate models of materials that are dependent on
25 one or more parameters (parameterized materials) in accordance with the imaging
modality. For example, as discussed above, the appearance of a given surface in a
scene, as imaged by a polarization camera system, may change based on the
properties of the material of the surface, the spectral profile and polarization
parameters of the illumination source or illumination sources (light sources) in the
30 scene, the incident angle of light onto the surface, and the viewpoint angle of the
observer (e.g., the polarization camera system). As such, simulating the physics of
polarization for different materials is a complex and computationally-intensive task.

[00108] As such, some aspects of embodiments of the present disclosure relate to
emulating the physics of various imaging modalities based on empirical data, such
35 as real-world images captured of real-world materials. In more detail, an imaging
system implementing the particular imaging modality of interest (e.g., a polarization
camera system) is used to collect sample images from an object made of the

1 particular material of interest. In some embodiments, the collected sample images are used to compute an empirical model of the material, such as its surface light-field function (e.g., a bi-directional reflectance density function or BRDF).

5 **[00109]** FIG. 7 is a schematic diagram of the sampling a real material from multiple angles using a polarization camera system according to one embodiment of the present disclosure. FIG. 8 is a flowchart depicting a method 800 for capturing images of a material from different perspectives using a particular imaging modality to be modeled according to one embodiment of the present disclosure. As shown in FIG. 7, a surface 702 of a physical object (e.g., a washer, a screw, or the like) is made of a material of interest (e.g., respectively, black rubber, chrome plated stainless steel, or the like). In operation 810, this material is placed into a physical scene (e.g., on a laboratory benchtop). In operation 830, a physical illumination source 704, such as an LED lamp or a fluorescent lamp is placed in the scene and arranged to illuminate at least a portion of the surface 702. For example, as shown in FIG. 7, ray 706
15 emitted from the physical illumination source 704 is incident on a particular point 708 of the surface 702 at an incident angle α at a particular point 708 on the surface 702 with respect to the normal direction 714 of the surface 702 at that particular point 708.

20 **[00110]** In operation 850, an imaging system is used to capture images of the surface 702 of the object from multiple poses with respect to the normal direction of the surface. In the embodiment shown in FIG. 7, a polarization camera system 710 is used as the imaging system to capture images of the surface 702, including the portions illuminated by the physical illumination source 704 (e.g., including the particular point 708). The polarization camera system 710 captures images the surface 702 from different poses 712, such as by moving the polarization camera system 710 from one pose to the next, and capturing polarization raw frames from each pose. In the embodiment shown in FIG. 7, the polarization camera system 710 images the surface 702 at a fronto-parallel observer angle β of 0° in first pose 712A (e.g., a fronto-parallel view from directly above or aligned with the surface normal 714 at the point 708), at an intermediate observer angle β such as an angle of 45° with respect to the surface normal 714 in second pose 712B, and at a shallow observer angle β (e.g., slightly less than 90° , such as 89°) with respect to the surface normal 714 in third pose 712C.

35 **[00111]** As discussed above, a polarization camera system 710 is generally configured to capture polarization raw frames with polarization filters at different angles (e.g., with a polarization mosaic having four different angles of polarization in the optical path of a single lens and sensor system, with an array of four cameras,

1 each of the cameras having a linear polarization filter at a different angle, with a polarizing filter set at a different angle for different frames captured at different times from the same pose, or the like).

5 **[00112]** In operation 870, the images captured by the imaging system are stored in with the relative poses of the camera with respect to the normal direction of the surface (e.g., the observer angle β). For example, the observer angle β may be stored in the metadata associated with the images and/or the images may be indexed, in part, based on the observer angle β . In some embodiments, the images may be indexed by parameters which include: observer angle β (or angle of camera position with respect to surface normal), material type, and illumination type.

10 **[00113]** Accordingly, in the arrangement shown in FIG. 7 and using, for example, the method of FIG. 8, the polarization camera system 710 captures multiple images (e.g., four images at linear polarization angles of 0° , 45° , 90° , and 135°) of the material in given illumination conditions (e.g., where the spectral profile of the physical illumination source 704 is known) at different angles of reflection (e.g., at different poses 712).

15 **[00114]** Each of these perspectives or poses 712 gives a different polarization signal due to the nature of the physics of polarization. Accordingly, by capturing images of the surface 702 from different observer angles, a model of the BRDF of the material can be estimated based on interpolating between the images captured with the physical illumination source 704 at one or more closest corresponding incident angles α by the camera system at the one or more poses 712 having closest corresponding observer angles β .

20 **[00115]** While the embodiment of FIG. 7 merely depicts three poses 712 for convenience, embodiments of the present disclosure are not limited thereto and the material may be sampled at higher rates, such as with 5° spacing between adjacent poses, or smaller spacings. For example, in some embodiments, the polarization camera system 712 is configured to operate as a video camera system, where polarization raw frames are captured at a high rate, such as 30 frames per second, 60 frames per second, 120 frames per second, or 240 frames per second, thereby resulting in a high density of images captured at a large number of angles with respect to the surface normal.

25 **[00116]** Similarly, in some embodiments, the pose of the physical illumination source 704 with respect to the surface 702 is modified, such that rays of light emitted from the physical illumination source 704 are incident on the surface 702 at different angles α , where multiple images of the surface are similarly captured by the polarization camera system 710 from different poses 712.

1 **[00117]** The sampling rates of the different angles (e.g., the incident angle α and
the observer or polarization camera system angle β) can be chosen such that
intermediate perspectives can be interpolated (e.g., bilinearly interpolated) without
significant loss of realism. In various embodiments of the present disclosure, the
5 spacing of the intervals may depend on physical characteristics of the imaging
modality, where some imaging modalities exhibit more angle sensitivity than others,
and therefore high accuracy may be possible with fewer poses (spaced more widely
apart) for modalities that are less angle-sensitive, whereas modalities having higher
angle-sensitivity may use larger numbers of poses (spaced more closely together).
10 For example, in some embodiments, when capturing polarization raw frames for a
polarization imaging modality, the poses 712 of the polarization camera system 710
are set at interval angles of approximately five degrees (5°) apart, and the images of
the surface 702 may also be captured with the physical illumination source 704 at
various positions, similarly spaced at angles of approximately five degrees (5°) apart.

15 **[00118]** In some circumstances, the appearance of a material under the imaging
modality of the empirical model is also dependent on the type of illumination source,
such as incandescent lights, fluorescent lights, light emitting diode (LED) bulbs,
sunlight, and therefore parameters of the illumination source or illumination sources
used to light the real-world scene are included as parameters of the empirical model.
20 In some embodiments, different empirical models are trained for different illumination
sources (e.g., one model of a material under natural lighting or sunlight and another
model of the material under fluorescent lighting).

[00119] Referring back to FIG. 6, in some embodiments, in operation 670 the
synthetic data generator 40 sets a virtual background for the scene. In some
25 embodiments, the virtual background is an image captured using the same imaging
modality as the modality being simulated by the synthetic data generator 40. For
example, in some embodiments, when generating synthetic polarization images, the
virtual background is a real image captured using a polarization camera, and when
generating synthetic thermal images, the virtual background is a real image captured
30 using a thermal camera. In some embodiments, the virtual background is an image
of an environment similar to the environments in which the trained machine learning
model is intended to operate (e.g., a manufacturing facility or factory in the case of
computer vision systems for manufacturing robots). In some embodiments, the
virtual background is randomized, thereby increasing the diversity of the synthetic
35 training data set.

[00120] In operation 690, the synthetic data generator 40 renders the 3-D scene
based on the specified imaging modality (e.g., polarization, thermal, etc.) using one

1 or more of the empirically derived, modality-specific models of materials. Some
aspects of embodiments of the present disclosure relate to rendering images based
on an empirical model of a material according to one embodiment of the present
disclosure. The empirical model of the material may be developed as discussed
5 above, based on samples collected from images captured of real-world objects made
of the material of interest.

[00121] Generally, a 3-D computer graphics rendering engine generates 2-D
renderings of virtual scenes by computing the color of each pixel in the output image
in accordance with the color of a surface of the virtual scene that is depicted by that
10 pixel. For example, in a ray tracing rendering engine, a virtual ray of light is emitted
from the virtual camera into the virtual scene (in reverse of the typical path of light in
the real world), where the virtual ray of light interacts with the surfaces of 3-D models
of objects in the virtual scene. These 3-D models are typically represented using
geometric shapes such as meshes of points that define flat surfaces (e.g., triangles),
15 where these surfaces may be assigned materials that describe how the virtual ray of
light interacts with the surface, such as reflection, refraction, scattering, dispersion,
and other optical effects, as well as a texture that represents the color of the surface
(e.g., the texture may be a solid color or may be, for example, a bitmap image that is
applied to the surface). The path of each virtual ray of light is followed (or "traced")
20 through the virtual scene until it reaches a light source in the virtual scene (e.g., a
virtual light fixture) and the accumulated modifications of the textures encountered
along the path from the camera to the light source are combined with the
characteristics of the light source (e.g., color temperature of the light source) to
compute the color of the pixel. This general process may be modified as understood
25 by those skilled in the art, such as performing anti-aliasing (or smoothing) by tracing
multiple rays through different parts of each pixel and computing the color of the
pixel based on a combination (e.g., average) of the different colors computed by
tracing the different rays interacting with the scene.

[00122] FIG. 9 is a flowchart depicting a method 900 for rendering a portion of a
30 virtual object based on the empirical model of a material according to one
embodiment of the present disclosure. In particular, FIG. 9 describes one
embodiment relating to computing a color when tracing of one ray through one pixel
of a virtual scene as the ray interacts with a surface having a material modeled
according to one embodiment of the present disclosure. However, a person having
35 ordinary skill in the art before the effective filing date of the present application would
understand how the technique described herein may be applied as a part of a larger
rendering process, where multiple colors are computed for a given pixel of an output

1 image and combined or where a scanline rendering process is used instead of ray tracing.

5 **[00123]** In more detail, the embodiment of FIG. 9 depicts a method for rendering a surface of an object in a virtual scene based on a view from a virtual camera in the virtual scene, where the surface has a material modeled in accordance with embodiments of the present disclosure. Given that the objects are being rendered synthetically, and the synthetic data generator 40 has access to the ground truth geometry of each object being rendered, the per-pixel normal, material type, and illumination type are all known parameters that appropriately modulate the graphical rendering of the material. During the rendering process, camera rays are traced from the optical center of the virtual camera to each 3-D point on the object that is visible from the camera. Each 3-D point (e.g., having X-Y-Z coordinates) on the object is mapped to a 2-D coordinate (e.g., having U-V coordinates) on the surface of the object. Each U-V coordinate on the surface of the object has its own surface light-field function (e.g., a bi-directional reflectance function or BRDF) represented as a model that is generated based on the images of real materials, as described above, for example, with respect to FIGS. 7 and 8.

15 **[00124]** In operation 910, the synthetic data generator 40 (e.g., running a 3-D computer graphics rendering engine) determines a normal direction of the given surface (e.g., with respect to a global coordinate system). In operation 930, the synthetic data generator 40 determines the material of the surface of the object as assigned to the surface as part of the design of the virtual scene.

20 **[00125]** In operation 950, the synthetic data generator 40 determines an observer angle β of the surface, e.g., based on a direction from which the ray arrived at the surface (e.g., if the surface is the first surface reached by the ray from the camera, then the angle from the virtual camera to the surface, otherwise the angle from which the ray arrived at the surface from another surface in the virtual scene). In some embodiments, in operation 950, the incident angle α is also determined, based on the angle at which the ray leaves the surface (e.g., in a direction toward a virtual light source in the scene, due to the reversal of ray directions during ray tracing). In some circumstances, the incident angle α depends on characteristics of the material determined in operation 930, such as whether the material is transparent, reflective, refractive, diffuse (e.g., matte), or combinations thereof.

25 **[00126]** In operations 970 and 990, the synthetic data generator 40 configures a model of the material based on the observer angle β (and, if applicable, the incident angle α and other conditions, such as the spectral profile or polarization parameters of illumination sources in the scene) and computes a color of the pixel based, in part,

30

35

1 on the configured model of the material. The model of the material may be retrieved
from a collection or data bank of models of different standard materials (e.g., models
materials have been empirically generated based on types of materials expected to
be depicted in virtual scenes generated by the synthetic data generator 40 for
5 generating training data for a particular application or usage scenario, such as
materials of components used in manufacturing a particular electronic device in the
case of computer vision for supporting robotics in manufacturing the electronic
device), where the models are generated based on images of real materials
captured as described above in accordance with embodiments of the present
10 disclosure. For example, in operation 930, the synthetic data generator 40 may
determine that the surface of the object in the virtual scene is made of black rubber,
in which case a model of a material generated from captured images of a real
surface made of black rubber is loaded and configured in operation 970.

[00127] In some circumstances, the virtual scene includes objects having surfaces
15 made of materials that are not represented in the data bank or collection of models
of materials. Accordingly, some aspects of embodiments of the present disclosure
relate to simulating the appearance of materials that do not have exact or similar
matches in the data bank of models of materials by interpolating between the
predictions made by different real models. In some embodiments, the existing
20 materials are represented in an embedding space based on a set of parameters
characterizing the material. More formally, interpretable material embeddings M ,
such that $F(M_{glass}, \theta_{out}, \phi_{out}, x, y)$ would give the polarized surface light field for
glass with observer angle β represented by $(\theta_{out}, \phi_{out})$, and at a location (x, y) on
the surface (mapped to (u, v) coordinate space on the 3-D surface), and a similar
25 embedding may be performed for another material such as rubber
 $F(M_{rubber}, \theta_{out}, \phi_{out}, x, y)$. This embedding of the materials in embedding space can
then be parameterized in an interpretable way using, for example, beta variational
autoencoders (VAEs) and then interpolated to generate new materials that are not
based directly on empirically collected samples, but, instead, are interpolations
30 between multiple different models that were separately constructed based on their
own empirically collected samples. The generation of additional materials in this way
further extends the domain randomization of synthetic training data generated in
accordance with embodiments of the present disclosure and improves the
robustness of deep learning models trained based on this synthetic data.

35 **[00128]** Various embodiments of the present disclosure relate to different ways the
model of the material may be implemented.

1 **[00129]** In various embodiments of the present disclosure, the model representing
the surface-light field function or BRDF of the material is represented using, for
example, a deep learning based BRDF function (e.g., based on a deep neural
5 function (e.g., a set of one or more closed-form equations or one or more open-form
equations that can be solved numerically), or a data-driven BRDF function that uses
linear interpolation.

[00130] In operation 970, the synthetic data generator 40 configures a model of the
material identified in operation 950 based on the current parameters, such as the
10 incident angle α and the observer angle β . In the case of a data-driven BRDF
function using linear interpolation, in operation 970, the synthetic data generator 40
retrieves images of the material identified in operation 950 that are closest in
parameter space to the parameters of the current ray. In some embodiments, the
materials are indexed (e.g., stored in a database or other data structure) and
15 accessible in accordance with the material type, the illumination type, the incident
angle of the light, and the angle of the camera with respect to the surface normal of
the material (e.g., the observer angle). However, embodiments of the present
disclosure are not limited to the parameters listed above, and other parameters may
be used, depending on the characteristics of the imaging modality. For example, for
20 some materials, the incident angle and/or the illumination type may have no effect on
the appearance of the material, and therefore these parameters may be omitted and
need not be determined as part of method 900.

[00131] Accordingly, in the case of a data-driven BRDF function with linear
interpolation, in operation 970, the synthetic data generator 40 retrieves one or more
25 images that are closest to the given parameters of the current ray associated with
the current pixel being rendered. For example, the observer angle may be at 53°
from the surface normal of the surface of the object, and samples of the real-world
material may include images captured at observer angles spaced 5° apart, in this
example, images captured at 50° and at 55° with respect to the surface normal of the
30 real-world object made of the material of interest. Accordingly, the images of the
real-world material that were captured at 50° and 55° would be retrieved (in
circumstances where additional parameters are used, these parameters, such as
incident angle and illumination type, are further identify particular images to be
retrieved).

35 **[00132]** Continuing the example of a data-driven BRDF function with linear
interpolation, operation 990, the synthetic data generator 40 computes a color of the
surface for the pixel based on the closest images. In circumstances where there is

1 only one matching image (e.g., if observer angle in the virtual scene matches an
observer angle of one of the sampled images), then the sampled image is used
directly for computing the color of the surface. In circumstances where there are
multiple matching images, the synthetic data generator 40 interpolates the colors of
5 the multiple images. For example, in some embodiments, linear interpolation is used
to interpolate between the multiple images. More specifically, if the observer angle is
between four different sample images that have different observer angles in the
azimuthal angle with respect to the surface normal and polar angle with respect to an
incident angle of the illumination source, then bilinear interpolation may be used to
10 interpolate between the four images along the azimuthal and polar directions. As
another example, if the appearance of the material is further dependent on incident
angle, then further interpolation may be performed based on images captured at
different incident angles (along with interpolating between images captured at
different observer angles for each of the different incident angles). Accordingly, in
15 operation 990, a color of the surface of the scene is computed for the current pixel
based on combining one or more images captured of real-world materials.

[00133] In some embodiments of the present disclosure where the model is a deep
learning network, a surface light field function of the material is implemented with a
model includes training a deep neural network to predict the value of a bi-directional
20 reflectance function directly from a set of parameters. In more detail, the images
captured of a real material from a plurality of different poses, as discussed above, for
example, with respect to FIGS. 7 and 8, are used to generate training data relating
parameters such as observer angle β , incident angle α , spectral properties of the
illumination source, and the like, to the observed appearance of a portion of the
25 material (e.g., at the center of the image). As such, in some embodiments, a deep
neural network is trained (e.g., applying backpropagation) to estimate a BRDF
function based on training data collected from images collected of real materials. In
these embodiments, a model is configured in operation 970 by, if there are multiple
deep neural networks, selecting a deep neural network from a plurality of deep
30 neural networks associated with the model (e.g., selected based on matching
parameters of the virtual scene with parameters of the data used to train the deep
neural network, such as parameters of an illumination source) and supplying
parameters to the input of the selected deep neural network (or only deep neural
network, if there is only one deep neural network associated with the model), such
35 as the observer angle β , the incident angle α , and the like. In operation 990, the
synthesized data generator 40 computes the color of the surface of the virtual object
from the configured model by forward propagating through the deep neural network

1 to compute a color at the output, where the computed color is the color of the surface of the virtual scene, as predicted by the deep neural network of the configured model.

[00134] In some embodiments of the present disclosure where the model is a deep
5 learning network, a surface light field function of the material is implemented with a model includes one or more conditional generative adversarial networks (see, e.g., Goodfellow, Ian, et al. "Generative adversarial nets." *Advances in Neural Information Processing Systems*. 2014.). Each conditional generative adversarial network may be trained to generate images of a material based on a random input and one or
10 more conditions, where the conditions include the current parameters of the viewing of a surface (e.g., observer angle β , incident angle α of each illumination source, polarization states of each illumination source, and material properties of the surface). According to some embodiments, a discriminator is trained in an adversarial manner to determine, based on an input image and a set of conditions
15 associated with the image, whether the input image was a real image captured under the given conditions or generated by the conditional generator conditioned based on the set of conditions. By alternately retraining the generator to generate images that can "fool" the discriminator and training the discriminator to discriminate between generated images and real images, the generator is trained to generate
20 realistic images of materials captured under various capture conditions (e.g., at different observer angles), thereby enabling the trained generator to represent the surface light-field function of the material. In some embodiments of the present disclosure, different generative adversarial networks are trained for different conditions for the same material, such as for different types of illumination sources,
25 different polarization states of the illumination sources, and the like. In these embodiments, a model is configured in operation 970 by, if there are multiple conditional generative adversarial networks (GANs) associated with the model, selecting a conditional GAN from a plurality of conditional GANs associated with the model (e.g., selected based on matching parameters of the virtual scene with
30 parameters of the data used to train the deep neural network, such as parameters of an illumination source) and supplying the parameters of the virtual scene as the conditions of the conditional GAN, such as the observer angle β , the incident angle α , and the like. In operation 990, the synthesized data generator 40 computes the color of the surface of the virtual object from the configured model by forward
35 propagating through the conditional GAN to compute a color at the output (e.g., a synthesized image of the surface of the object based on the current parameters),

1 where the computed color is the color of the surface of the virtual scene, as generated by the conditional GAN of the configured model.

5 **[00135]** In some embodiments of the present disclosure, the surface light field function is modeled using a closed-form mathematically derived bidirectional reflectance distribution function (BRDF) configured by the empirically collected samples (e.g., pictures or photographs) of the real material, as captured from different angles, such as in accordance with the method described with respect to FIG. 8. Examples of techniques for configuring a BRDF based on collected pictures or photographs of a material from different angles or poses are described, e.g., in 10 Ramamoorthi, Ravi, and Pat Hanrahan. "A Signal-Processing Framework for Inverse Rendering." *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*. 2001. and Ramamoorthi, Ravi. *A Signal-Processing Framework for Forward and Inverse Rendering*. Stanford University, 2002, 52–79. Accordingly, in some embodiments, a closed-form mathematically derived BRDF is 15 configured using empirically collected samples of a real material and included as a component of a material model for modeling multimodal and/or plenoptic characteristics of a material for the computer rendering of multimodal and/or plenoptic images of virtual scenes.

20 **[00136]** In some embodiments, virtual objects made of multiple materials that are either bi-modal or multimodal would have a similar set of images for each type of material used in the virtual object in question. The appearances of the different materials are then combined in the final rendering of the image (e.g., additively combined in accordance with weights associated with each material in the virtual model). In some embodiments, this same approach of combining multiple materials 25 is applied to multi-layered materials such as transparent coatings on shiny materials, and the like. In some embodiments, multi-layered materials are modeled by separately sampling (e.g., capturing images of) the multi-layered material.

30 **[00137]** The end effect of the rendering process using an empirical model of materials according to embodiments of the present disclosure is that the final rendering has an emulated polarization signal that is close to the real polarization signal in a real environment. The accuracy of the empirical model in rendering the materials depicted in the virtual environment depends on how closely the conditions of the virtual environment match the conditions under which samples were captured of the real-world material (e.g., how closely the spectral profile of the illumination 35 source in the virtual scene matched the real-world illumination source, how closely the observer angles in the virtual scene matched the observer angles used when performing sampling, and the like).

1 **[00138]** As noted above, while aspects of embodiments of the present disclosure
are described herein in the context of simulating or emulating the appearance of
polarization, embodiments of the present disclosure are not limited thereto. The
appearance of materials under multimodal imaging modalities and/or plenoptic
5 imaging modalities, such as thermal, thermal with polarization, and the like, can also
be captured in accordance with embodiments of the present disclosure. For
example, the behavior of materials in a thermal imaging modality (e.g., infrared
imaging) can similarly be modeled by capturing images of a material from multiple
poses using a thermal camera in an arrangement similar to that shown in FIG. 7 and
10 a method as described in FIG. 8. Based on these captured images, an appearance
of a material under thermal imaging can then be simulated in a 3-D rendering engine
by retrieving corresponding images and interpolating images, as necessary, in a
manner similar to that shown in FIG. 9.

15 **[00139]** Accordingly, some aspects of embodiments of the present disclosure
relate to systems and methods for generating synthetic image data of virtual scenes
as they would appear under various imaging modalities such as polarization imaging
and thermal imaging, by rendering images of the virtual scenes using empirical
models of materials in the virtual scene. In some embodiments, these empirical
models may include images captured of real-world objects using one or more
20 imaging modalities such as polarization imaging and thermal imaging. These
synthetic image data can then be used to train machine learning models to operate
on image data captured by imaging systems using these imaging modalities.

25 **[00140]** Some aspects of embodiments of the present disclosure relate to
generating synthetic data relating to image features that would typically be generated
from imaging data. As a specific example, some aspects of embodiments of the
present disclosure relate to generating synthetic features or tensors in polarization
representation spaces (e.g., in degree of linear polarization or DOLP ρ and angle of
linear polarization or AOLP ϕ). As discussed above, shape from polarization (SfP)
provides a relationship between the DOLP ρ and the AOLP ϕ and the refractive
30 index (n), azimuth angle (θ_a) and zenith angle (θ_z) of the surface normal of an object.

[00141] Accordingly, some aspects of embodiments relate to generating a
synthetic degree of linear polarization or DOLP ρ and angle of linear polarization or
AOLP ϕ for surfaces of a virtual scene that are visible to the virtual camera based on
the refractive index n , azimuth angle (θ_a) and zenith angle (θ_z) of the surface of the
35 virtual scene, all of which are known parameters of the virtual 3-D scene.

[00142] FIG. 10 is a flowchart depicting a method 1000 for computing synthetic
features or tensors in polarization representation spaces for a virtual scene

1 according to one embodiment of the present disclosure. In operation 1010, the
synthetic data generator 40 renders a normal image (e.g., an image where every
pixel corresponds to the direction of the surface normal of the virtual scene at that
5 component and a zenith angle θ_z component. In operation 1030, the synthetic data
generator 40 separates the normal vector at each point of the normal image into two
components: the azimuth angle θ_a and the zenith angle θ_z at that pixel. As noted
above, these components can be used to compute an estimate of the DOLP ρ and
AOLP ϕ by using the shape from polarization equations (2) and (3) for the diffuse
10 case and equations (4) and (5) for the specular case. To simulate realistic
polarization error, in some embodiments of the present disclosure, the synthetic data
generator 40 applies a semi-global perturbation to the normal map before applying
the polarization equations (e.g., equations (2), (3), (4), and (5)). This perturbation
changes the magnitude of the normals while preserving the gradient of the normals.
15 This simulates errors caused by the material properties of the object and their
interactions with polarization. In operation 1050, for a given pixel, the synthetic data
generator 40 determines the material of the surface of the object based on the
parameters of the objects in the virtual scene (e.g., the material associated with the
surface at each pixel of the normal map), and the material is used in conjunction with
20 the geometry of the scene, in accordance with 3-D rendering techniques, to
determine whether the given pixel is specular dominant. If so, then the synthetic data
generator 40 computes the DOLP ρ and AOLP ϕ based on the specular equations
(4) and (5) in operation 1092. If not, then the synthetic data generator 40 computes
the DOLP ρ and AOLP ϕ based on the diffuse equations (2) and (3) in operation
25 1094.

[00143] In some embodiments, it is assumed that all of the surfaces are diffuse,
and therefore operations 1050 and 1070 can be omitted and the synthetic DOLP ρ
and AOLP ϕ are computed based on the shape from polarization equations (2) and
(3) for the diffuse case.

30 **[00144]** In some embodiments, the synthesized DOLP ρ and AOLP ϕ data are
rendered into color images by applying a colormap such as the “viridis” colormap or
the “Jet” colormap (see, e.g., Liu, Yang, and Jeffrey Heer. “Somewhere over the
rainbow: An empirical assessment of quantitative colormaps.” *Proceedings of the*
2018 CHI Conference on Human Factors in Computing Systems. 2018.). These
35 color mapped versions of the synthesized tensors in polarization space may be more
easily supplied as input for retraining a pre-trained machine learning model such as
convolutional neural network. In some embodiments, when synthesizing DOLP ρ and

1 AOLP ϕ data, random colormaps are applied to the various synthesized data such
that the synthesized training data set includes color images representing DOLP ρ
and AOLP ϕ data in a variety of different colormaps, such that, at inference time, the
network will be able to perform predictions without regard to the particular colormap
5 used to encode the real DOLP ρ and AOLP ϕ data. In other embodiments of the
present disclosure, the same colormap is applied to all of the synthetic DOLP ρ and
AOLP ϕ data (or a first colormap is used for DOLP ρ and a second, different
colormap is used for AOLP ϕ), and at inference time, colormaps are applied to the
extracted tensors in polarization representation space to match the synthetic training
10 data (e.g., the same first colormap is used for encoding the DOLP ρ extracted from
the captured real polarization raw frames and the same second colormap is used for
encoding AOLP ϕ extracted from the captured real polarization raw frames).

[00145] Accordingly, some aspects of embodiments of the present disclosure
relate to synthesizing features in representation spaces specific to particular imaging
15 modalities, such by synthesizing DOLP ρ and AOLP ϕ in polarization representation
spaces for a polarization imaging modality.

[00146] Some aspects of embodiments of the present disclosure relate to
combinations of the above techniques for generating synthetic images for training
machine learning models. FIG. 11 is a flowchart depicting a method for generating a
20 training data set according to one embodiment of the present disclosure. One or
more virtual scenes representative of the target domain may be generated as
discussed above (e.g., for generating images of bins of components, by selecting
one or more 3-D models of components and dropping instances of the 3-D models
into a container). For example, some aspects of embodiments of the present
25 disclosure relate to forming a training data set based on: (1) images generated
purely by domain randomization in operation 1110, (2) images generated purely
through texture mapping (e.g., generated in accordance with embodiments of FIG. 9)
in operation 1112, and (3) images generated purely through normal mapping (e.g.,
generated in accordance with embodiments of FIG. 10) in operation 1114.

30 **[00147]** In addition, the training data set may include images generated using
models of materials generated by interpolating between different empirically
generated models, as parameterized in embedding space, as discussed above.

[00148] In some embodiments of the present disclosure, the images generated in
accordance with (1) domain randomization, (2) texture mapping, and (3) normal
35 mapping are further processed by applying style transfer or other filter to the
generated image in operations 1120, 1122, and 1124, respectively, before adding
the image to the training data set. Applying style transfer causes images that appear

1 somewhat different, as generated using the three techniques described above, to
have a more consistent appearance. In some embodiments, the style transfer
process transforms the synthesized input images to appear more similar to an image
captured based on the imaging modality of interest (e.g., causing images generated
5 using (1) domain randomization and feature maps generated using (3) normal
mapping to appear more like polarization raw frames) or by causing the synthesized
input images to appear more artificial, such as by applying an unrealistic painterly
style to the input images (e.g., causing images generated using (1) domain
randomization, (2) renderings using texture mapping, and feature maps generated
10 using (3) normal mapping to appear like a painting made with a paintbrush on
canvas).

[00149] In some embodiments, a neural style transfer network is trained and used
to perform the style transfer in operation 1122 on the images selected for the training
data set, such as SytleGAN (see, e.g., Karras, Tero, et al. "Analyzing and improving
15 the image quality of stylegan." *Proceedings of the IEEE/CVF Conference on
Computer Vision and Pattern Recognition*. 2020.) for complex global style transfers;
patched based networks (see, e.g., Chen, Tian Qi, and Mark Schmidt. "Fast patch-
based style transfer of arbitrary style." *arXiv preprint arXiv:1612.04337* (2016).) for
local style transfers; and networks using domain adaptation (see, e.g., Dundar,
20 Aysegul, et al. "Domain stylization: A strong, simple baseline for synthetic to real
image domain adaptation." *arXiv preprint arXiv:1807.09384* (2018).). As a result, all
of the images in the training data set may have a similar style or appearance
regardless of the method by which the images were obtained (e.g., whether through
(1) domain randomization, (2) texture mapping, (3) normal mapping, or other sources
25 such as real images of objects as captured using an imaging system implementing
the modality of interest, such as polarization imaging or thermal imaging), as
transformed by a style transfer operation.

[00150] In some embodiments of the present disclosure, the images for the
training data set are sampled from the synthesized data sets (1), (2), and (3) based
30 on hard example mining (see, e.g., Smirnov, Evgeny, et al. "Hard example mining
with auxiliary embeddings." *Proceedings of the IEEE Conference on Computer
Vision and Pattern Recognition Workshops*. 2018.) Using hard example mining to
sample the synthesized data sets can improve the efficiency of the training process
by reducing the size of the training set to remove substantially redundant images that
35 would not have much impact on the training process while keeping the "hard
examples" that have more of an impact on the resulting trained model.

1 **[00151]** As briefly mentioned above, when generating training data for supervised
learning, the synthetic data generator 40 also automatically generates labels (e.g.,
desired outputs) for the synthesized images. For example, when generating training
data for training a machine learning model to perform an image classification task,
5 the generated label for a given image may include the classes of the objects
depicted in the image. These classification label may be generated by identifying
each unique type of object that is visible in the virtual scene. As another example,
when generating training data for training a machine learning model to perform an
instance segmentation task, the generated label may include a segmentation map
10 where each instance of each object is uniquely identified (e.g., with a different
instance identifier) along with its class (e.g., where objects of the same type have the
same class identifier). For example, a segmentation map may be generated by
tracing rays from the camera into the virtual scene, where each ray may intersect
with some first surface of the virtual scene. Each pixel of the segmentation map is
15 labeled accordingly based on the instance identifier and class identifier of the object
containing the surface that was struck by the ray emitted from the camera through
the pixel.

[00152] As discussed above, and referring to FIG. 1 the resulting training data set
of synthesized data 42 generated by the synthetic data generator 40 is then used as
20 training data 5 by a model training system 7 to train a model 30, such as a pre-
trained model or a model initialized with random parameters, to produce a trained
model 32. Continuing the example presented above in the case of generating
training data in accordance with a polarization imaging modality, the training data set
5 may be used to train the model 30 to operate on polarization input features such as
25 polarization raw frames (e.g., the images generated through texture mapping) and
tensors in polarization representation spaces (e.g., images generated through
normal mapping).

[00153] Accordingly, the training data 5 including the synthesized data 42 is used
to train or retrain a machine learning model 30 to perform a computer vision task
30 based on a particular imaging modality. For example, synthetic data in accordance
with a polarization imaging modality may be used to retrain a convolutional neural
network that may have been pre-trained to perform instance segmentation based on
standard color images to perform instance segmentation based on polarization input
features.

35 **[00154]** In deployment, a trained model 32 trained based on training data
generated in accordance with embodiments of the present disclosure is then
configured to take input similar to the training data such as polarization raw frames

1 and/or tensors in polarization representation spaces (where these input images are further modified by the same style transfer, if any, that was applied when generating the training data) to generate predicted outputs such as segmentation maps.

5 **[00155]** While some embodiments of the present disclosure are described herein with respect to a polarization imaging modality, embodiments of the present disclosure are not limited thereto and include with multimodal imaging modalities and/or plenoptic imaging modalities such as thermal imaging, thermal imaging with polarization (e.g., with a polarizing filter), and ultraviolet imaging. In these
10 embodiments using different modalities, real-world image samples captured from real-world materials using imaging systems implementing those modalities are used to generate models of materials as they would appear under those imaging modalities, and the surface light field functions of the materials with respect to those modalities are modeled as described above (e.g., using a deep neural network, a generative network, linear interpolation, explicit mathematical model, or the like) and
15 used to render images in accordance with those modalities using a 3-D rendering engine. The rendered images in the modality may then be used to train or retrain one or more machine learning models, such as convolutional neural networks, to perform computer vision tasks based on input images captured using those modalities.

[00156] As such, aspects of embodiments of the present disclosure relate to
20 systems and methods for generating simulated or synthetic data representative of image data captured by imaging systems using a variety of different imaging modalities such as polarization, thermal, ultraviolet, and combinations thereof. The simulated or synthetic data can be used to as a training data set and/or to augment a training data set for training a machine learning model to perform tasks, such as
25 computer vision tasks, on data captured using an imaging modality corresponding to the imaging modality of the simulated or synthetic data.

[00157] While the present invention has been described in connection with certain exemplary embodiments, it is to be understood that the invention is not limited to the disclosed embodiments, but, on the contrary, is intended to cover various
30 modifications and equivalent arrangements included within the spirit and scope of the appended claims, and equivalents thereof.

[00158] In some embodiments of the present disclosure, the order in which operations are performed may differ from those as depicted in the figures and as described herein. For example, while FIG. 6 depicts one example of a
35 method for generating synthetic images, embodiments of the present disclosure are not limited thereto. For example, some of the operations shown in FIG. 6 may be performed in a different order, or may be performed concurrently. As specific

1 examples, in various embodiments of the present disclosure, the operations of
placing 3-D models of objects in a virtual scene 610, adding lighting to the virtual
scene 630, applying modality-specific materials to objects in the virtual scene 650,
and setting the scene background 670 may be performed in various orders before
5 rendering the 3-D scene based on a specified imaging modality in operation 690. As
another example, while FIG. 8 depicts an embodiment in which lighting is placed in a
real-world scene after placing a real-world object into a scene, embodiments of the
present disclosure are not limited thereto and lighting may be added to a scene
before placing real-world objects in the scene.

10 **[00159]** In some embodiments of the present disclosure, some operations may be
omitted or not performed, and in some embodiments, additional operations not
described herein may be performed before, after, or between various operations
described herein.

15

20

25

30

35

1 WHAT IS CLAIMED IS:

1. A method of generating synthetic images of virtual scenes comprising:
placing, by a synthetic data generator implemented by a processor and
memory, three-dimensional (3-D) models of objects in a 3-D virtual scene;
5 adding, by the synthetic data generator, lighting to the 3-D virtual scene, the
lighting comprising one or more illumination sources;
applying, by the synthetic data generator, imaging modality-specific materials
to the 3-D models of objects in the 3-D virtual scene in accordance with a selected
multimodal imaging modality, each of the imaging modality-specific materials
10 comprising an empirical model;
setting a scene background in accordance with the selected multimodal
imaging modality; and
rendering, by the synthetic data generator, a two-dimensional image of the 3-D
virtual scene based on the selected multimodal imaging modality to generate a
15 synthetic image in accordance with the selected multimodal imaging modality.

2. The method of claim 1, wherein the empirical model is generated based
on sampled images captured of a surface of a material using an imaging system
configured to capture images using the selected multimodal imaging modality, and
20 wherein the sampled images comprise images captured of the surface of the
material from a plurality of different poses with respect to a normal direction of the
surface of the material.

3. The method of claim 2, wherein the selected multimodal imaging
25 modality is polarization and the imaging system comprises a polarization camera.

4. The method of claim 2, wherein the selected multimodal imaging
modality is thermal and the imaging system comprises a thermal camera.

30 5. The method of claim 4, wherein the thermal camera comprises a
polarizing filter.

6. The method of claim 2, wherein each of the sampled images is stored in
association with the corresponding angle of its pose with respect to the normal
35 direction of the surface of the material.

1 7. The method of claim 2, wherein the sampled images comprise:
 a first plurality of sampled images captured of the surface of the material
 illuminated by light having a first spectral profile; and
 a second plurality of sampled images captured of the surface of the material
5 illuminated by light having a second spectral profile different from the first spectral
 profile.

 8. The method of claim 2, wherein the empirical model comprises a surface
light field function computed by interpolating between two or more of the sampled
10 images.

 9. The method of claim 2, wherein the empirical model comprises a surface
light field function computed by a deep neural network trained on the sampled images.

15 10. The method of claim 2, wherein the empirical model comprises a surface
light field function computed by a generative adversarial network trained on the
sampled images.

20 11. The method of claim 2, wherein the empirical model comprises a surface
light field function computed by a mathematical model generated based on the
sampled images.

25 12. The method of claim 1, further comprising applying style transfer to the
synthetic image.

 13. A method for generating tensors in polarization feature space for a 3-D
virtual scene, comprising:

 rendering, by a synthetic data generator implemented by a processor and
memory, an image of surface normals of a 3-D virtual scene comprising a plurality of
30 3-D models of objects, the surface normals comprising an azimuth angle component
and a zenith angle component;

 determining, by the synthetic data generator, for a surface of a 3-D model of an
object of the 3-D virtual scene, a material of the object; and

35 computing, by the synthetic data generator, the tensors in polarization feature
space in accordance with the azimuth angle component and the zenith angle
component of the surface normal, the tensors in polarization feature space comprising:

1 a degree of linear polarization; and
 an angle of linear polarization at the surface of the object.

14. The method of claim 13 further comprising:
5 determining whether the surface of the 3-D model of the object is specular
 dominant;
 computing the tensors in polarization feature space based on specular
 polarization equations in response to determining that the surface of the 3-D model of
 the object is specular dominant; and
10 computing the tensors in polarization feature space based on diffuse
 polarization equations in response to determining that the surface of the 3-D model of
 the object is specular dominant.

15. The method of claim 13 comprising:
15 computing the tensors in polarization feature space based on diffuse
 polarization equations.

16. The method of claim 13, further comprising applying style transfer to the
 tensors in polarization feature space.
20

17. A method for synthesizing a training data set based on generating a
 plurality of synthetic images generated in accordance with any of claims 1–16.

18. A method for training a machine learning model comprising:
25 generating a training data set in accordance with claim 17; and
 computing parameters of the machine learning model based on the training
 data set.

19. A system for generating synthetic images of virtual scenes comprising:
30 a processor; and
 memory storing instructions that, when executed by the processor, cause the
 processor to implement a synthetic data generator to:
 place three-dimensional (3-D) models of objects in a 3-D virtual scene;
 add lighting to the 3-D virtual scene, the lighting comprising one or more
35 illumination sources;
 apply imaging modality-specific materials to the 3-D models of objects in
 the 3-D virtual scene in accordance with a selected multimodal imaging

1 modality, each of the imaging modality-specific materials comprising an
empirical model;
set a scene background in accordance with the selected multimodal
imaging modality; and
5 render a two-dimensional image of the 3-D virtual scene based on the
selected multimodal imaging modality to generate a synthetic image in
accordance with the selected multimodal imaging modality.

20. The system of claim 19, wherein the empirical model is generated based
10 on sampled images captured of a surface of a material using an imaging system
configured to capture images using the selected multimodal imaging modality, and
wherein the sampled images comprise images captured of the surface of the
material from a plurality of different poses with respect to a normal direction of the
surface of the material.

15 21. The system of claim 20, wherein the selected multimodal imaging
modality is polarization and the imaging system comprises a polarization camera.

20 22. The system of claim 20, wherein the selected multimodal imaging
modality is thermal and the imaging system comprises a thermal camera.

23. The system of claim 22, wherein the thermal camera comprises a
polarizing filter.

25 24. The system of claim 20, wherein each of the sampled images is stored
in association with the corresponding angle of its pose with respect to the normal
direction of the surface of the material.

30 25. The system of claim 20, wherein the sampled images comprise:
a first plurality of sampled images captured of the surface of the material
illuminated by light having a first spectral profile; and
a second plurality of sampled images captured of the surface of the material
illuminated by light having a second spectral profile different from the first spectral
profile.

35

1 26. The system of claim 20, wherein the empirical model comprises a surface light field function computed by interpolating between two or more of the sampled images.

5 27. The system of claim 20, wherein the empirical model comprises a surface light field function computed by a deep neural network trained on the sampled images.

10 28. The system of claim 20, wherein the empirical model comprises a surface light field function computed by a generative adversarial network trained on the sampled images.

15 29. The system of claim 20, wherein the empirical model comprises a surface light field function computed by a mathematical model generated based on the sampled images.

20 30. The system of claim 19, wherein the memory further stores instructions that, when executed by the processor, cause the synthetic data generator to apply style transfer to the synthetic image.

25 31. A system for generating tensors in polarization feature space for a 3-D virtual scene comprising:

 a processor; and

 memory storing instructions that, when executed by the processor, cause the processor to implement a synthetic data generator to:

 render an image of surface normals of a 3-D virtual scene comprising a plurality of 3-D models of objects, the surface normals comprising an azimuth angle component and a zenith angle component;

 determine for a surface of a 3-D model of an object of the 3-D virtual scene, a material of the object; and

 compute the tensors in polarization feature space in accordance with the azimuth angle component and the zenith angle component of the surface normal, the tensors in polarization feature space comprising:

 a degree of linear polarization; and

 an angle of linear polarization at the surface of the object.

1 32. The system of claim 31, wherein the memory further stores instructions that, when executed by the processor, cause the synthetic data generator to:

 determine whether the surface of the 3-D model of the object is specular dominant;

5 compute the tensors in polarization feature space based on specular polarization equations in response to determining that the surface of the 3-D model of the object is specular dominant; and

 compute the tensors in polarization feature space based on diffuse polarization equations in response to determining that the surface of the 3-D model of the object is specular dominant.

 33. The system of claim 31, wherein the memory further stores instructions that, when executed by the processor, cause the synthetic data generator to:

 compute the tensors in polarization feature space based on diffuse polarization equations.

 34. The system of claim 31, wherein the memory further stores instructions that, when executed by the processor, cause the synthetic data generator to:

 apply style transfer to the tensors in polarization feature space.

 35. A system for synthesizing a training data set is configured to synthesize the training data set using a system according to any of claims 19–34.

 36. A system for training a machine learning model comprising:

 a processor; and

 memory storing instructions that, when executed by the processor, cause the processor to:

 receive a training data set generated by a system according to claim 35; and

 compute parameters of the machine learning model based on the training data set.

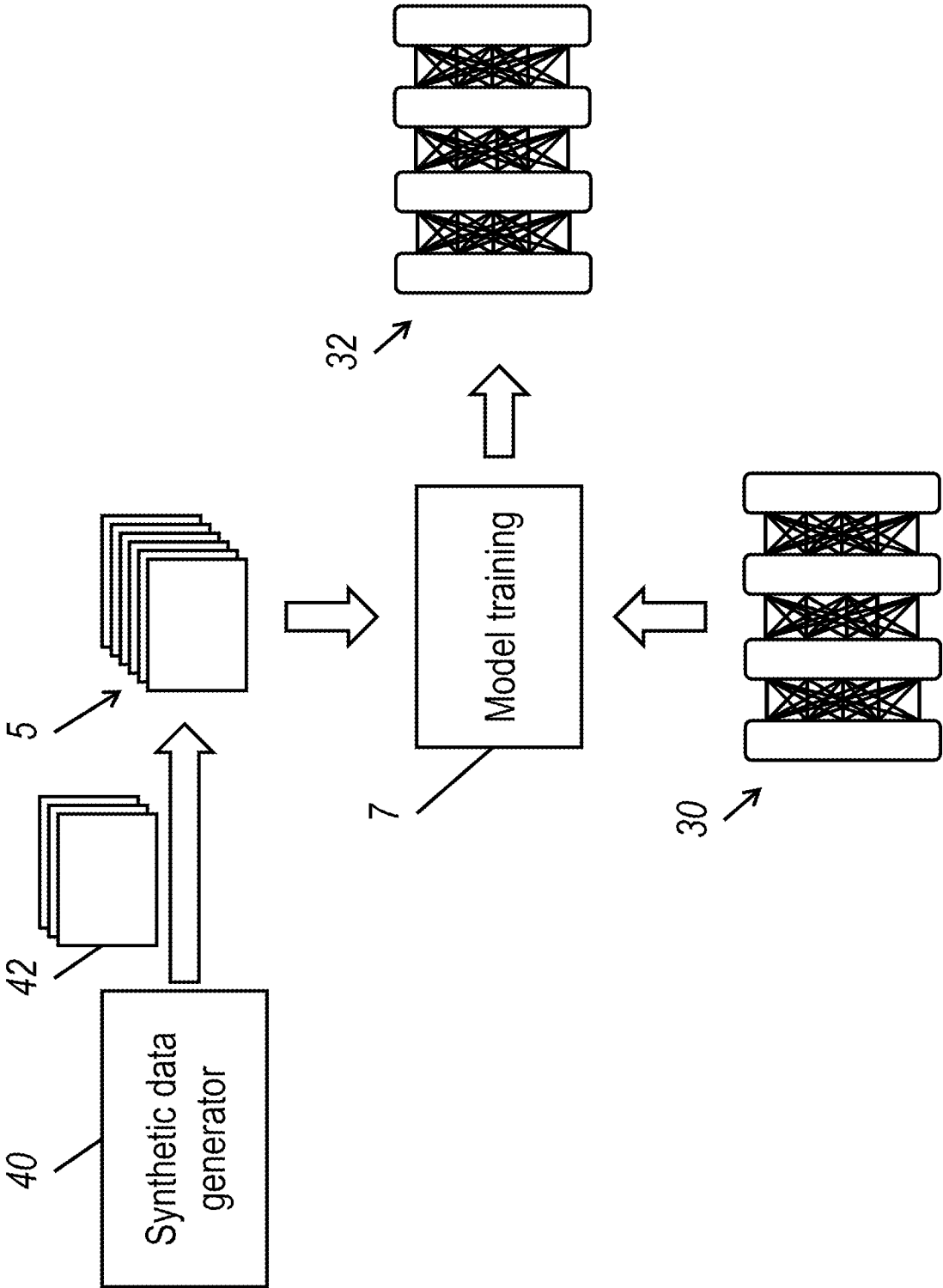


FIG. 1

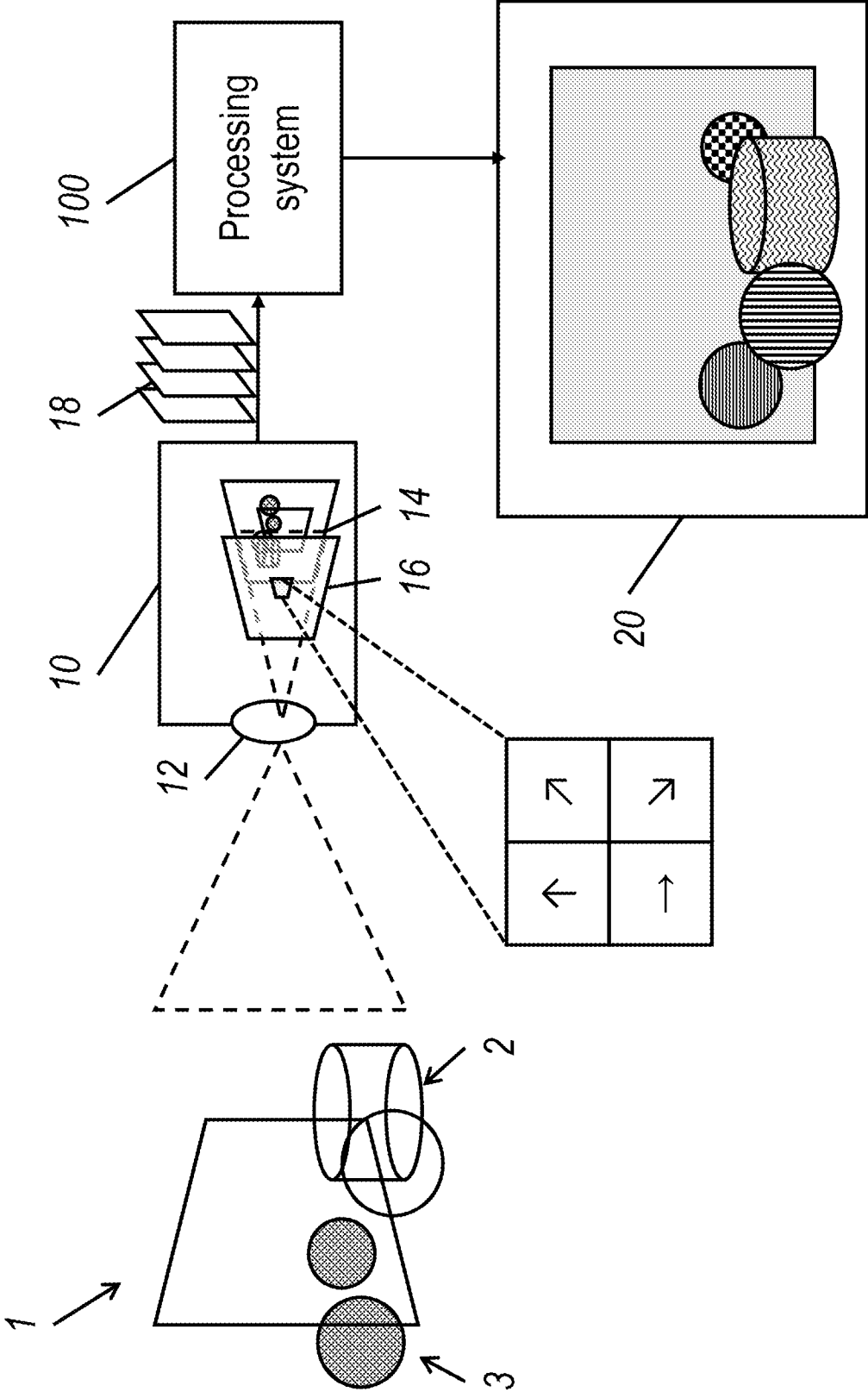


FIG. 2

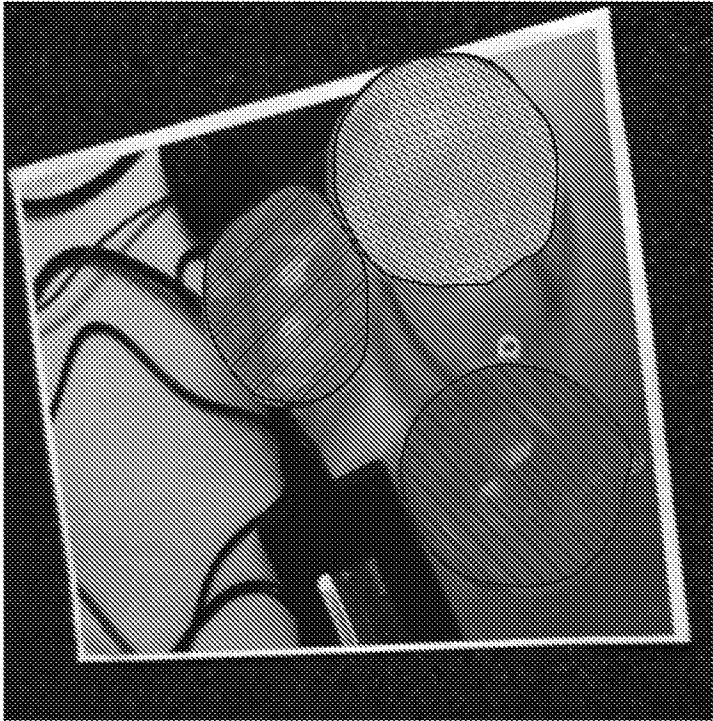


FIG. 3B



FIG. 3A

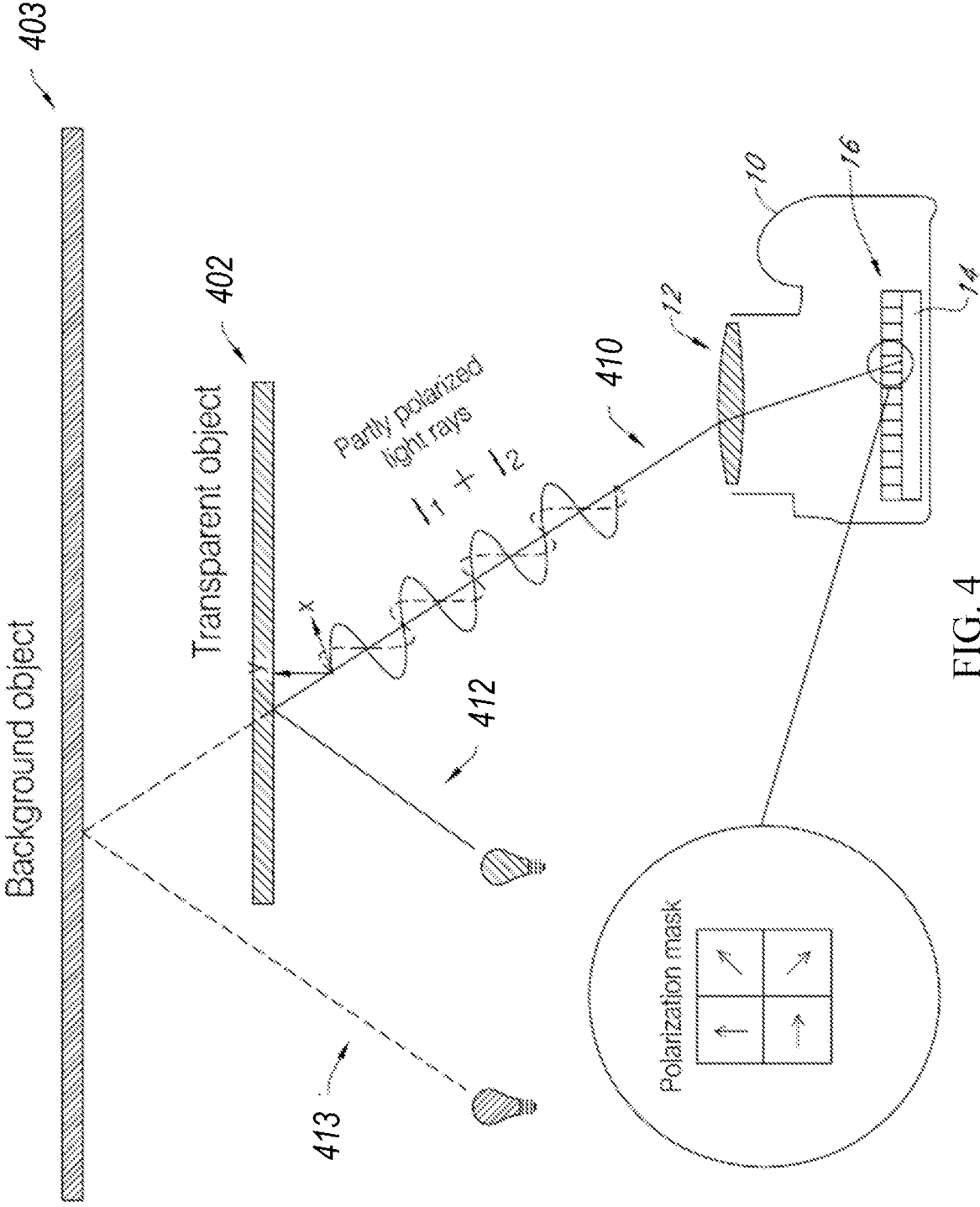
4/12



FIG. 3D



FIG. 3C



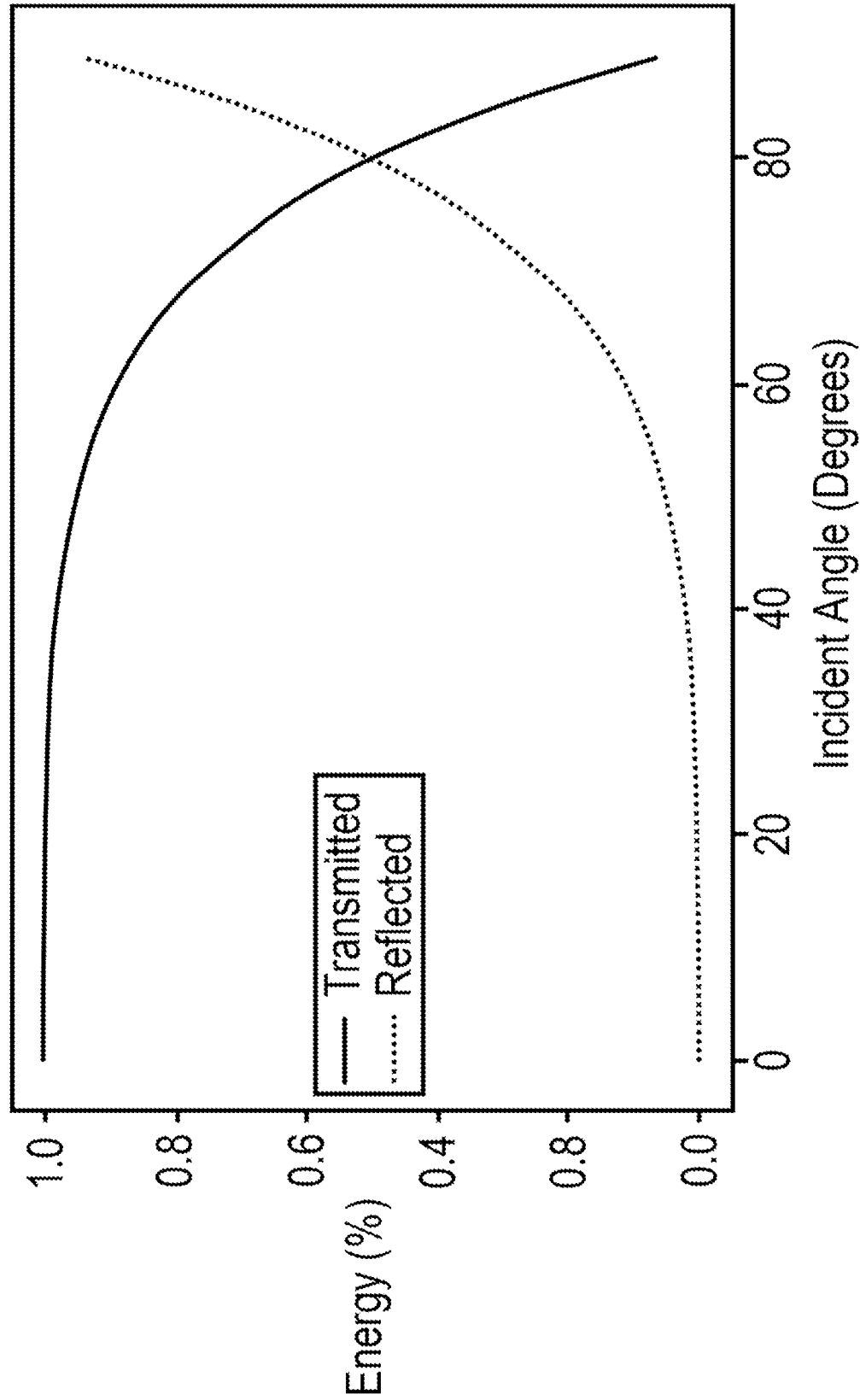


FIG. 5

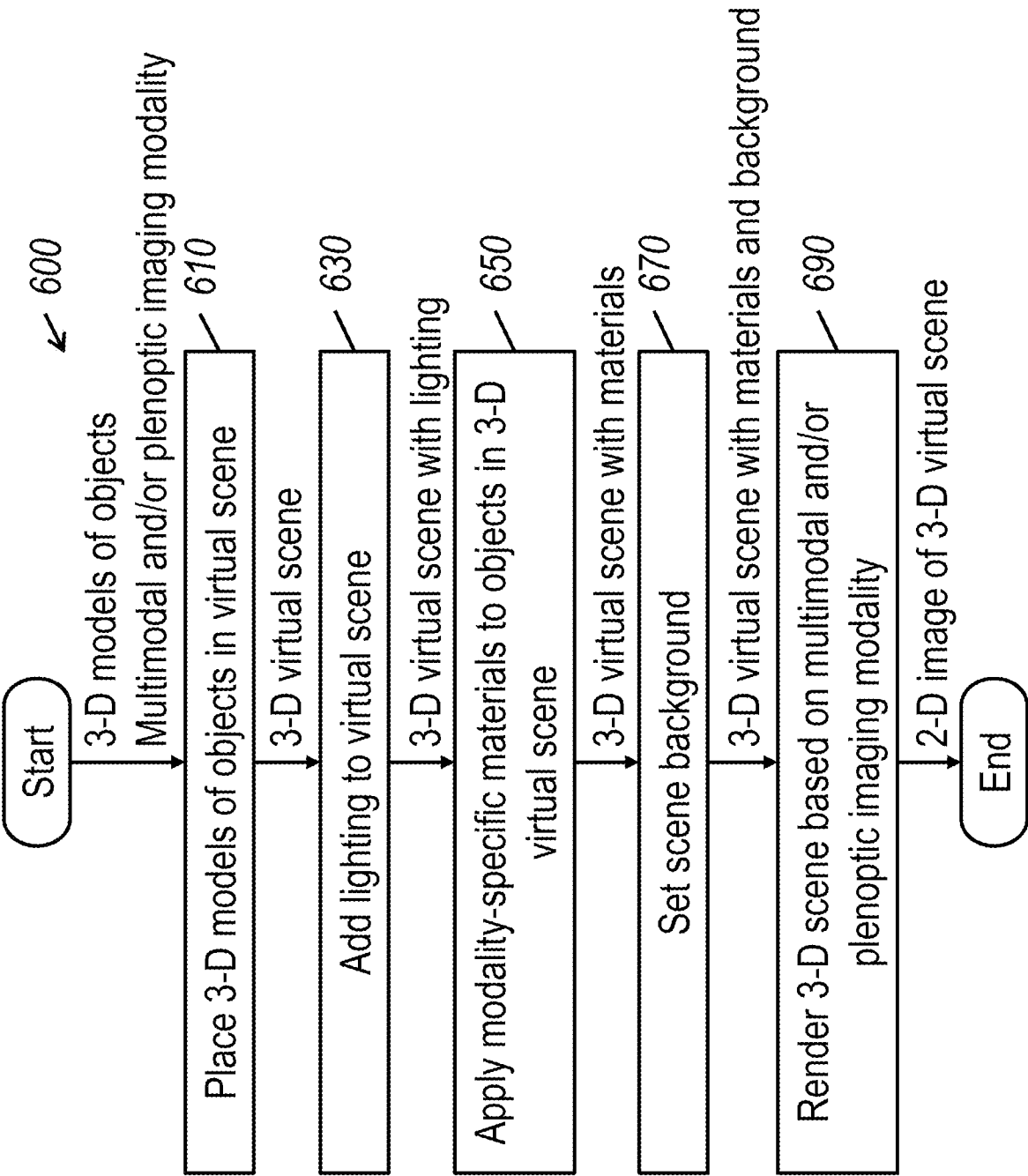


FIG. 6

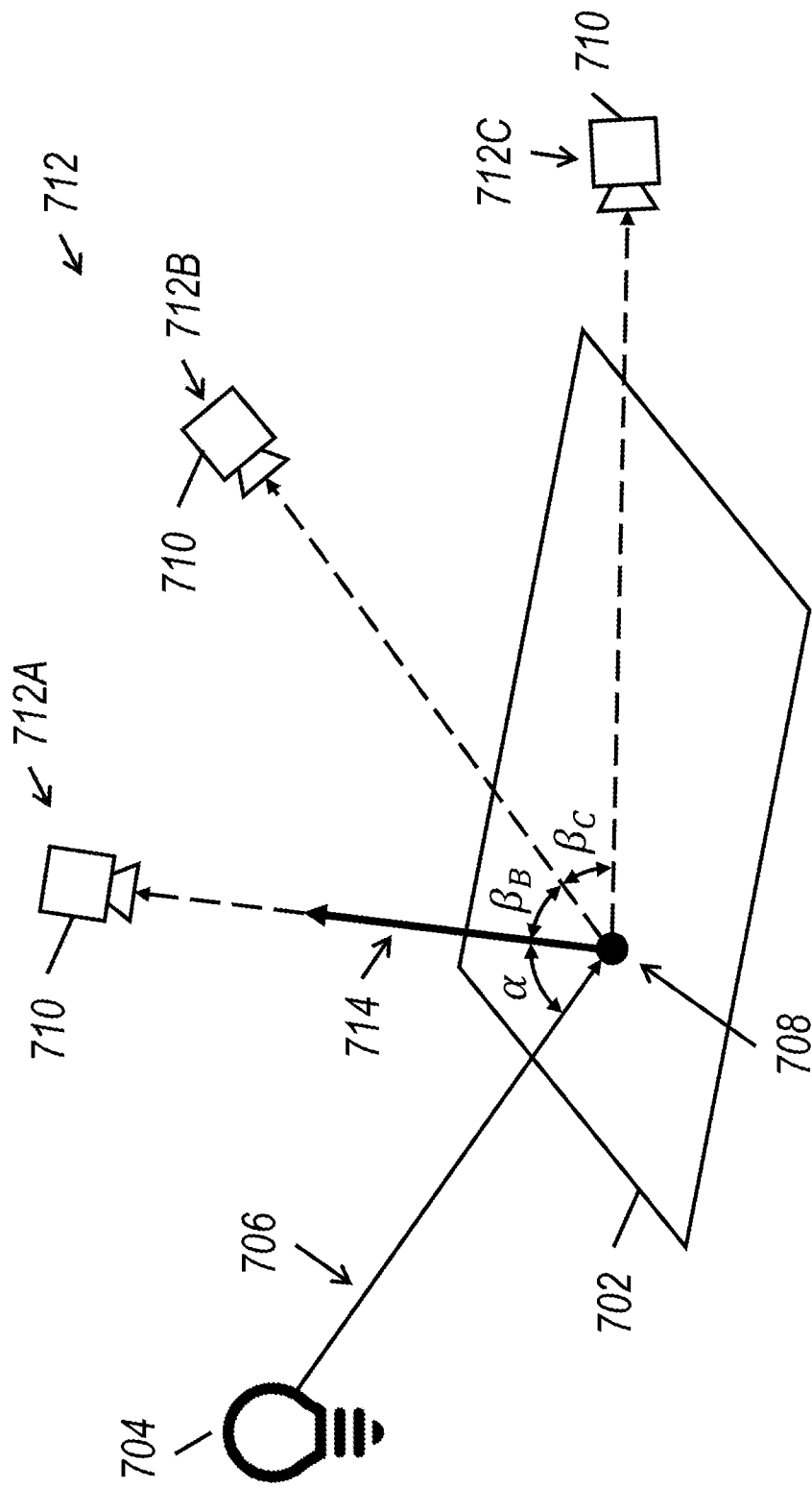


FIG. 7

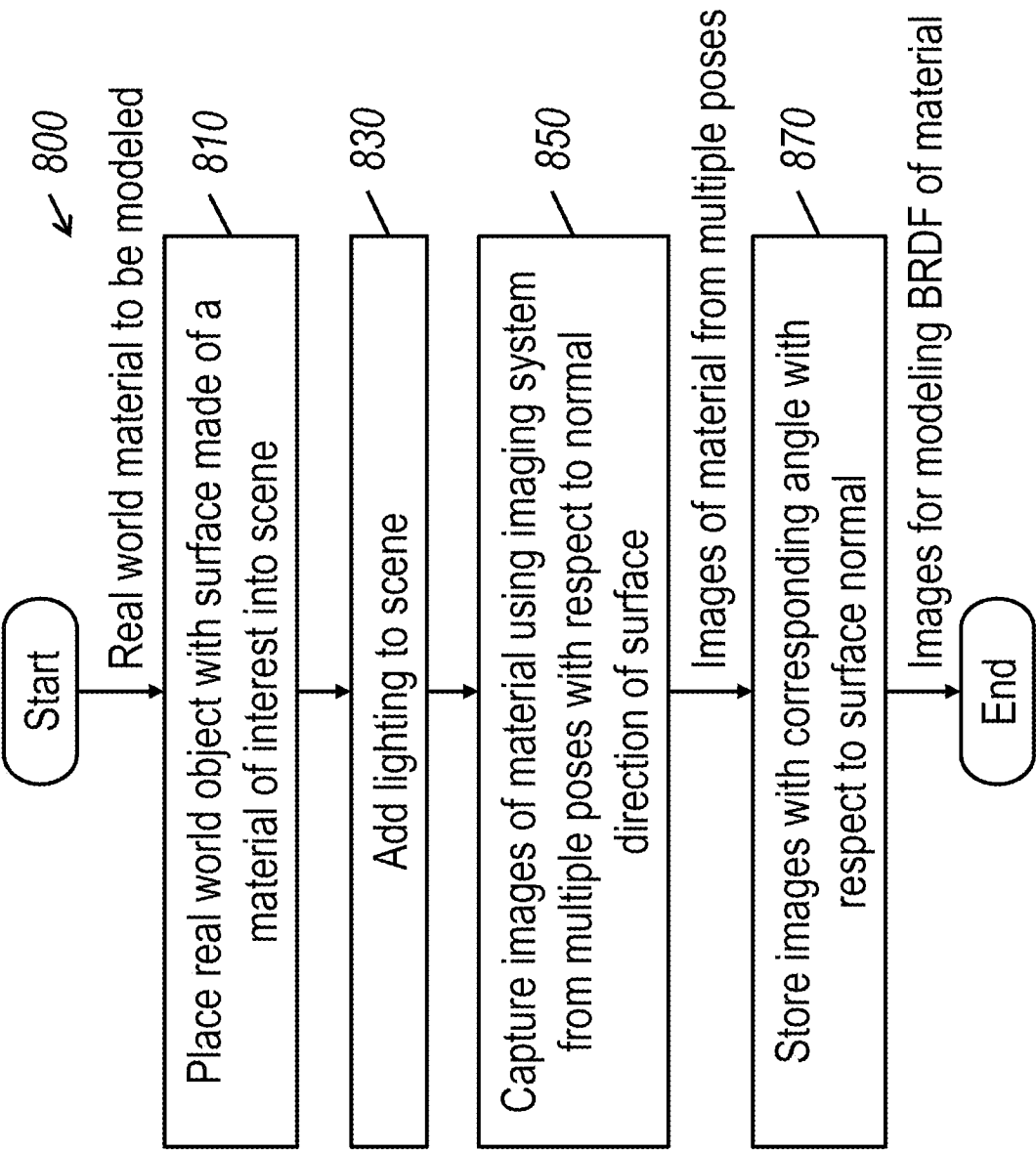


FIG. 8

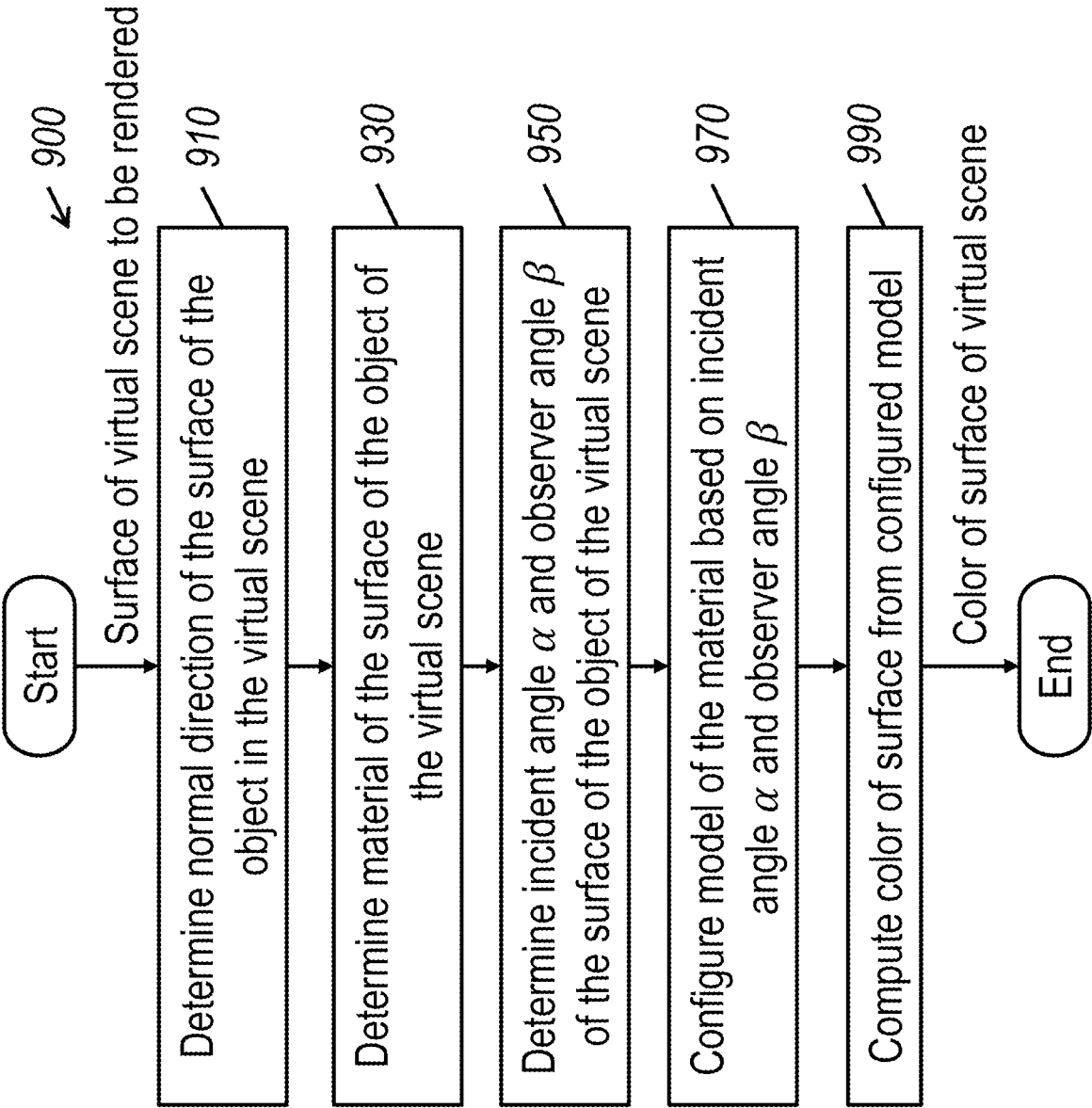


FIG. 9

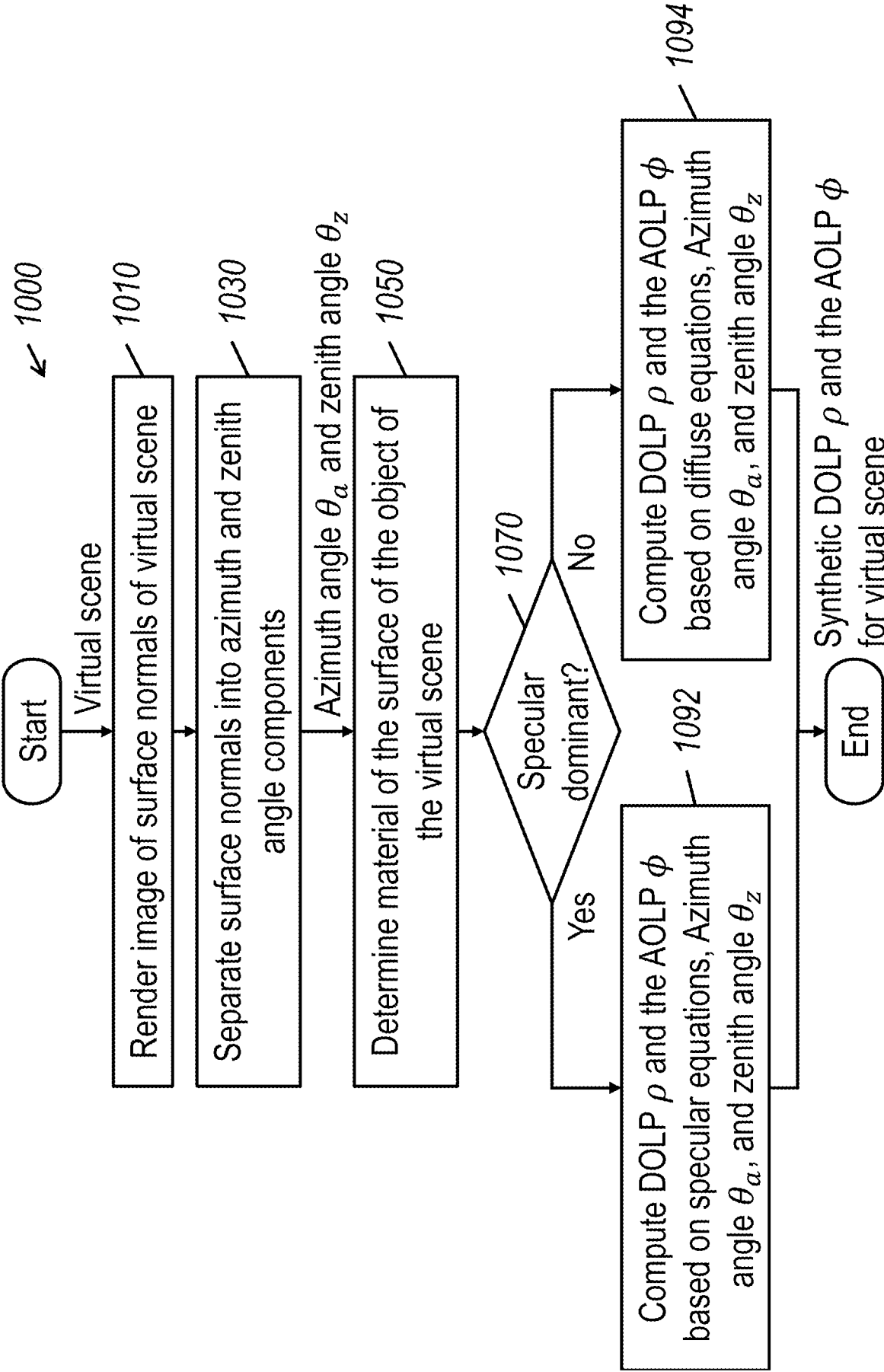


FIG. 10

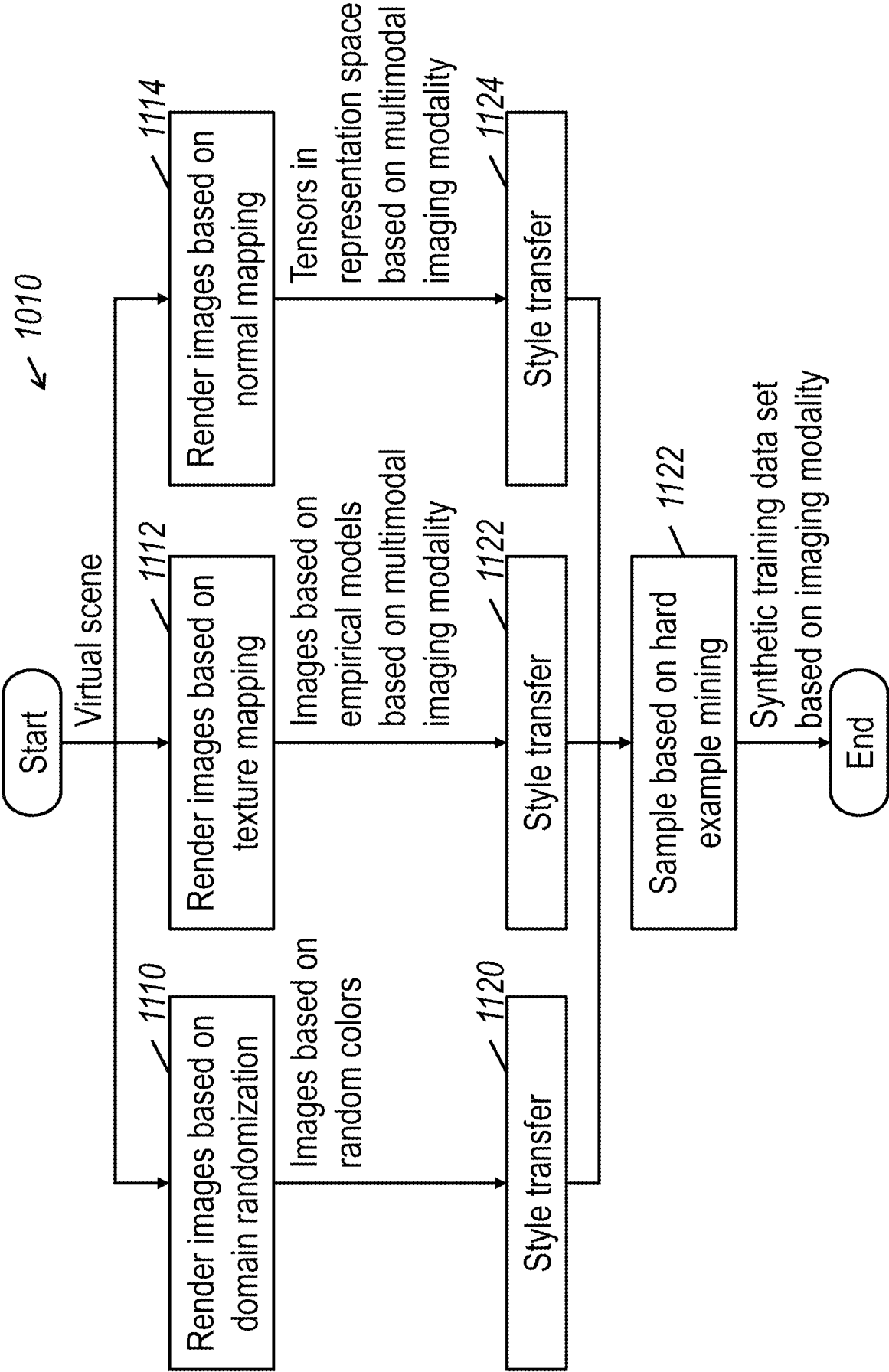


FIG. 11

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 21/12073

A. CLASSIFICATION OF SUBJECT MATTER

IPC - G06T 15/00; G06T 17/00; G06T 17/20 (2021.01)

CPC - G06F 3/04815; G06K 9/34; G06T 17/00; G06T 7/00

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)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2015/0043806 A1 (Adobe Systems Incorporated) 12 February 2015 (12.02.2015), entire document	1-12, 17/(1-12), 18/(17/1-12), 19-30, 35/(19-30), 36/(35/19-30)
A	US 2018/0308281 A1 (draw, Inc.) 25 October 2018 (25.10.2018), entire document	1-12, 17/(1-12), 18/(17/1-12), 19-30, 35/(19-30), 36/(35/19-30)
A	US 2015/0325048 A1 (Mobile R&D Inc.) 12 November 2015 (12.11.2015), entire document	1-12, 17/(1-12), 18/(17/1-12), 19-30, 35/(19-30), 36/(35/19-30)
A	US 2002/0128060 A1 (Belhumeur et al.) 12 September 2002 (12.09.2002), entire document	1-12, 17/(1-12), 18/(17/1-12), 19-30, 35/(19-30), 36/(35/19-30)
A	US 2011/0227922 A1 (Shim) 22 September 2011 (22.09.2011), entire document	1-12, 17/(1-12), 18/(17/1-12), 19-30, 35/(19-30), 36/(35/19-30)

☐ 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

"D" document cited by the applicant in the international application

"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

20 April 2021 (20.04.2021)

Date of mailing of the international search report

APR 29 2021

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Lee Young

Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 21/12073

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1.

Group I: Claims 1-12, 17/(1-12), 18/(17/1-12), 19-30, 35/(19-30), 36/(35/19-30) directed to a method/system for generating synthetic images of virtual scenes.

Group II: Claims 13-16, 17/(13-16), 18/(17/13-16), 31-34, 35/(31-34), 36/(35/31-34) directed to a method/system for generating tensors in polarization feature space for a 3-D virtual scene.

(See extra sheet)

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-12, 17/(1-12), 18/(17/1-12), 19-30, 35/(19-30), 36/(35/19-30)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

In continuation of Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet):

The groups of inventions listed above do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The special technical feature of the Group I invention is adding, by the synthetic data generator, lighting to the 3-D virtual scene, the lighting comprising one or more illumination sources; applying, by the synthetic data generator, imaging modality-specific materials to the 3-D models of objects in the 3-D virtual scene in accordance with a selected multimodal imaging modality, each of the imaging modality-specific materials comprising an empirical model; setting a scene background in accordance with the selected multimodal imaging modality; and rendering, by the synthetic data generator, a two-dimensional image of the 3-D virtual scene based on the selected multimodal imaging modality to generate a synthetic image in accordance with the selected multimodal imaging modality, not required by Group II. (See Figure 8).

The special technical feature of the Group II invention is rendering, by a synthetic data generator implemented by a processor and memory, an image of surface normals of a 3-D virtual scene comprising a plurality of 3-D models of objects, the surface normals comprising an azimuth angle component and a zenith angle component; determining, by the synthetic data generator, for a surface of a 3-D model of an object of the 3-D virtual scene, a material of the object; and computing, by the synthetic data generator, the tensors in polarization feature space in accordance with the azimuth angle component and the zenith angle component of the surface normal, the tensors in polarization feature space comprising: a degree of linear polarization; and an angle of linear polarization at the surface of the object., not required by Group I. (See Figure 10).

Groups I-II share the technical features of, in various combination, utilizing a synthetic data generator implemented by a processor and memory, in conjunction with three-dimensional (3-D) models of objects in a 3-D virtual scene. However, these shared technical features fail to represent a contribution over the prior art of US 2018/0308281 A1 to draw, Inc., which discloses utilizing a synthetic data generator implemented by a processor and memory, in conjunction with three-dimensional (3-D) models of objects in a 3-D virtual scene (para [0086], [0094], [0131] - "the 3D graphic generation system in accordance with the present embodiment is composed mainly of a material photographing apparatus 10 which photographs an actual site scenery 3 in a real world as a photographic material which is a still image or a motion picture for use in the background of a virtual space, and a 3D application system 2 for presenting interactive video content such as games"; "while actually photographing a real place as a model of the background of a virtual space with the material photographing apparatus 10 and acquiring the turntable environment data D1 containing any of a lighting position, a lighting type, a lighting amount and the number of light sources in the real place, three-dimensional objects D3 drawn as computer graphics are synthesized with the photographic material photographed by the material image photographing unit").

Thus, the inventions listed as Groups I-II lack unity of invention because they do not share a same or corresponding special technical feature providing a contribution over the prior art.