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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
07 February 2019 (07.02.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/028472 A1

(51) International Patent Classification:
G06F 15/18 (2006.01) *G06T 7/246* (2017.01)
G06K 9/46 (2006.01) *G06T 7/60* (2017.01)
G06K 9/62 (2006.01) *H04N 1/387* (2006.01)
G06K 9/66 (2006.01)

(21) International Application Number:
PCT/US2018/045423

(22) International Filing Date:
06 August 2018 (06.08.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/541,603 04 August 2017 (04.08.2017) US

(71) Applicant: **OUTWARD, INC.** [US/US]; 1980 Zanker Road, Suite 20, San Jose, CA 95112 (US).

(72) Inventors: **CHUI, Clarence**; 1980 Zanker Road, Suite 20, San Jose, CA 95112 (US). **PARMAR, Manu**; 1980 Zanker Road, Suite 20, San Jose, CA 95112 (US).

(74) Agent: **PUGALIA, Amita**; Van Pelt, Yi & James LLP, 10050 N. Foothill Blvd., Suite 200, Cupertino, CA 95014 (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, 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, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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,

(54) Title: MACHINE LEARNING BASED IMAGE PROCESSING TECHNIQUES

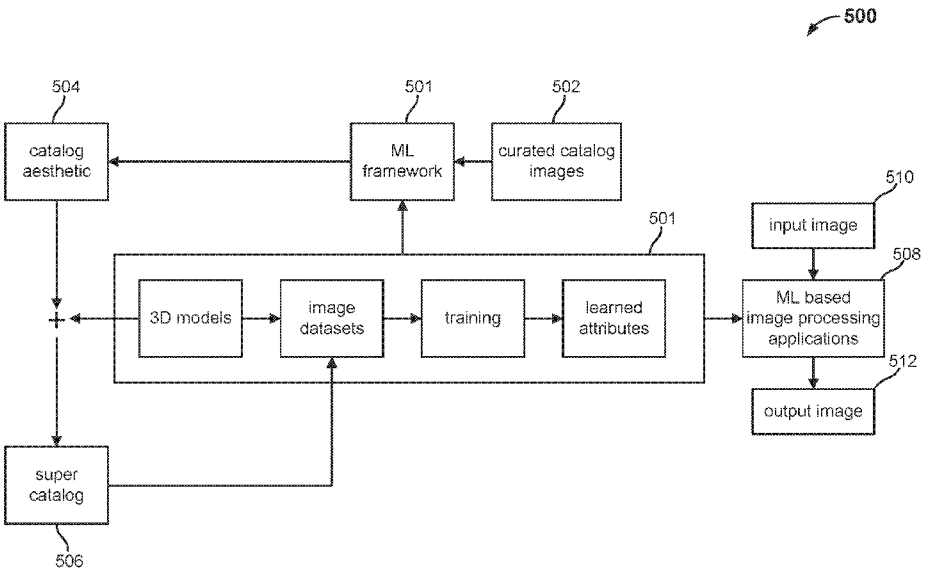

FIG. 5

(57) Abstract: A machine learning based image processing architecture and associated applications are disclosed herein. In some embodiments, a machine learning framework is trained to learn low level image attributes such as object/scene types, geometries, placements, materials and textures, camera characteristics, lighting characteristics, contrast, noise statistics, etc. Thereafter, the machine learning framework may be employed to detect such attributes in other images and process the images at the attribute level.

WO 2019/028472 A1

[Continued on next page]

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

MACHINE LEARNING BASED IMAGE PROCESSING TECHNIQUES

CROSS REFERENCE TO OTHER APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application No. 62/541,603 entitled DIRECT AND DERIVED 3D DATA FOR MACHINE LEARNING IN IMAGE BASED APPLICATIONS filed August 4, 2017, which is incorporated herein by reference for all purposes.

BACKGROUND OF THE INVENTION

[0002] Image processing techniques have classically operated at a pixel level, i.e., on pixel values or intensities. However, operating on low-level pixel data is not practical for applications such as altering the high-level visual appearance of an image. For such tasks, feature-based approaches are more effective. Feature-based techniques include first defining a set of specific features (e.g., edges, patches, SURF, SIFT, etc.) and then defining mathematical models on those features that can be used to analyze and manipulate image content.

[0003] Machine learning techniques may be employed to learn either or both features and mathematical model parameters based on pertinent application dependent cost functions. However, such artificial intelligence techniques require an exhaustive training data set that spans the space of all features for a particular application and that is labeled with relevant ground truth data. It has typically been prohibitive to gather exhaustive training data sets for most useful applications. Furthermore, it has been difficult to label images with more complex or sophisticated ground truth data. Thus, the use of machine learning techniques until lately has been limited to basic object recognition or classification applications.

[0004] An image processing architecture that overcomes such limitations and effectively leverages machine learning techniques is thus needed and disclosed herein as well as novel image-based applications resulting therefrom.

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] Various embodiments of the invention are disclosed in the following detailed description and the accompanying drawings.

[0006] Figure 1 is a high level block diagram of an embodiment of a machine learning based image processing framework for learning attributes associated with datasets.

[0007] Figure 2 illustrates examples of curated images from a catalog.

[0008] Figure 3 is a high level block diagram of an embodiment of an image processing framework for reducing or decomposing images belonging to a set into key attributes.

[0009] Figure 4 is a high level block diagram of an embodiment of an image processing framework for automatically generating curated images.

[0010] Figure 5 is a high level block diagram of an embodiment of a machine learning based image processing architecture.

[0011] Figure 6A is a high level block diagram of an embodiment of an image processing application for restyling.

[0012] Figure 6B illustrates an example of using a restyling application.

[0013] Figure 7A is a high level block diagram of an embodiment of an image processing application for replacing an object in an image with a different object.

[0014] Figure 7B illustrates an example of using an object replacement application.

[0015] Figure 8A is a high level block diagram of an embodiment of an image processing application for denoising an image.

[0016] Figure 8B illustrates an example of using a denoising application.

DETAILED DESCRIPTION

[0017] The invention can be implemented in numerous ways, including as a process; an apparatus; a system; a composition of matter; a computer program product embodied on a

computer readable storage medium; and/or a processor, such as a processor configured to execute instructions stored on and/or provided by a memory coupled to the processor. In this specification, these implementations, or any other form that the invention may take, may be referred to as techniques. In general, the order of the steps of disclosed processes may be altered within the scope of the invention. Unless stated otherwise, a component such as a processor or a memory described as being configured to perform a task may be implemented as a general component that is temporarily configured to perform the task at a given time or a specific component that is manufactured to perform the task. As used herein, the term 'processor' refers to one or more devices, circuits, and/or processing cores configured to process data, such as computer program instructions.

[0018] A detailed description of one or more embodiments of the invention is provided below along with accompanying figures that illustrate the principles of the invention. The invention is described in connection with such embodiments, but the invention is not limited to any embodiment. The scope of the invention is limited only by the claims, and the invention encompasses numerous alternatives, modifications, and equivalents. Numerous specific details are set forth in the following description in order to provide a thorough understanding of the invention. These details are provided for the purpose of example, and the invention may be practiced according to the claims without some or all of these specific details. For the purpose of clarity, technical material that is known in the technical fields related to the invention has not been described in detail so that the invention is not unnecessarily obscured.

[0019] An image processing architecture and image-based applications resulting therefrom are disclosed herein. Generally, images may comprise scenes including a single object, a plurality of objects, or rich (virtual) environments and furthermore may comprise stills or frames of animation or video sequences. Moreover, images may comprise (high quality) photographs or (photorealistic) renderings.

[0020] Artificial intelligence techniques including machine learning techniques are data driven and therefore require exhaustive training data sets to be effective for most practical applications. Thus, the basis for using machine learning techniques for image-based applications relies on access to comprehensive training data sets that have been labeled or tagged with appropriate metadata that is relevant for desired applications.

[0021] In this description, machine learning techniques are generally described and in various embodiments may comprise the use of any combination of one or more machine learning architectures appropriate for a given application, such as deep neural networks and convolutional neural networks. Moreover, the terms "labels", "tags", and "metadata" are interchangeably used in this description to refer to properties or attributes associated and persisted with a unit of data or content, such as an image.

[0022] A fundamental aspect of the disclosed image processing architecture comprises a content generation platform for generating comprehensive image data sets labeled with relevant ground truth data. Such data sets may be populated, for example, with images rendered from three-dimensional (polygon mesh) models and/or captured from imaging devices such as cameras or scanning devices such as 3D scanners. Image data sets encompassing exhaustive permutations of individual and/or combinations of objects and arrangements, camera perspectives or poses, lighting types and locations, materials and textures, etc., may be generated, for example, during offline processes. Image assets may moreover be obtained from external sources. Metadata may be generated and associated with (e.g., used to label or tag) images when the images are generated at the time of generation and/or may be added or modified afterwards and furthermore may be automatically generated and/or may at least in part be determined or defined manually.

[0023] Figure 1 is a high level block diagram of an embodiment of a machine learning based image processing framework 100 for learning attributes associated with datasets. Image datasets 104 are collected or generated. In many cases, an image dataset 104 comprises high quality (i.e., high definition or resolution) photographs or photorealistic renderings. An image dataset 104 may at least in part be populated by images rendered from three-dimensional models 102 as depicted in the example of Figure 1, images captured from imaging or scanning devices, images sourced from external entities, or images generated from other existing images using various processing techniques. Generally, datasets 104 includes any arbitrary number of permutations of different perspectives or viewpoints, materials and textures, lighting sources and locations, camera configurations, object combinations and placements, etc.

[0024] Images comprising datasets 104 are tagged with comprehensive sets of labels or metadata. A set of labels defined and/or selected for images of a prescribed dataset may at least in part be application dependent. The set of labels may include one or more hierarchical

high level labels that provide classification of the object(s) and/or scene comprising an image. The set of labels may furthermore include lower level labels comprising ground truth data associated with rendering an image or parts thereof from underlying three-dimensional model(s) or capturing the image using a physical device such as a camera or a scanner. Examples of such labels include labels associated with the (three-dimensional) geometry of the scene comprising the image such as object types and locations, material properties of objects comprising the scene, surface normal vectors, lighting types and positions (e.g., direct sources as well as indirect sources such as reflective surfaces that contribute to substantial higher order bounces), camera characteristics (e.g., perspective or pose, orientation, rotation, depth information, focal length, aperture, zoom level), etc. The labels may include absolute and/or relative location or position, orientation, and depth information amongst various scene objects, light sources, and the (virtual) camera that captures the scene. Other labels that are not scene-based but instead image-based such as noise statistics (e.g., the number of samples of rays used in ray tracing when rendering) may be associated with an image. Furthermore, more complex or sophisticated labels may be defined for various applications by combining a plurality of other labels. Examples include defining labels for specific object poses relative to light sources, presence of special materials (e.g., leather) with specific light sources, etc.

[0025] As described, knowledge of ground truth data associated with generating images comprising a dataset 104 facilitates detailed as well as custom (e.g., application specific) labels to be associated with the images comprising the dataset 104, including many types and classes of labels that otherwise could not be manually identified and associated with images such as lighting type and location. Extensive, labeled datasets 104 are perfectly qualified for artificial intelligence based learning. Training 106 on a dataset 104, for example, using any combination of one or more appropriate machine learning techniques such as deep neural networks and convolutional neural networks, results in a set of one or more low level properties or attributes 110 associated with the dataset 104 to be learned. Such attributes may be derived or inferred from labels of the dataset 104. Examples of attributes that may be learned include attributes associated with object/scene types and geometries, materials and textures, camera characteristics, lighting characteristics, noise statistics, contrast (e.g., global and/or local image contrast defined by prescribed metrics, which may be based on, for instance, maximum and minimum pixel intensity values), etc. Other more intangible attributes that are (e.g., some unknown nonlinear) combinations or functions of a plurality of low level attributes may also be learned. Examples of such

intangible attributes include attributes associated with style, aesthetic, noise signatures, etc. In various embodiments, different training models may be used to learn different attributes. Furthermore, image processing framework 100 may be trained with respect to a plurality of different training datasets. After training on large sets of data to learn various attributes, image processing framework 100 may subsequently be employed to detect similar attributes or combinations thereof in other images for which such attributes are unknown as further described below.

[0026] Many applications include the creation of collections or portfolios of curated images that share one or more attributes or properties that are difficult to quantify or define but that collectively contribute to a distinct signature style or visual appearance. For example, images comprising product catalogs published by high end merchants or retailers typically have a specific branded style or aesthetic, and similarly frames of scenes of animation or video sequences are all often constrained to prescribed visual characteristics. In such applications, captured photographs or generated renderings are often subjected to manual post-processing (e.g., retouching and remastering) by artists to create publishable imagery having visual characteristics that conform to a desired or sanctioned style or aesthetic or theme. Many of the properties that contribute to style or aesthetic or theme are the result of artist manipulation beyond anything that can be achieved from photography/rendering or global post-processing. Thus, an artist imparted aesthetic on an image is an intangible quality that has conventionally been difficult to isolate, model, and replicate. As a result, many existing applications still require artists to manually post-process each image or frame to obtain a desired curated look.

[0027] Figure 2 illustrates examples of curated images from a catalog of a home furnishings retailer. As depicted, each image comprises different products as well as product placements, but the entire set of images shares the same high level style or aesthetic defining an inspired, branded look. The curated images of Figure 2 result from post-processing by an artist and are not specifically labeled with low level attributes or ground truth data. However, as can be seen, all images belong to a set having common characteristics other than just having similar types of content (i.e., home furnishings). In the given example, all images share the same aesthetic particular to a prescribed brand and thus visually appear similar and are recognizable as belonging to the same set. In some embodiments, the high level attribute of the set of images that results in the shared similar look may be modeled by a nonlinear

function of one or more lower level attributes that collectively define an aesthetic as further described next.

[0028] Figure 3 is a high level block diagram of an embodiment of an image processing framework 300 for reducing or decomposing images belonging to a set into key attributes, one or more of which may collectively define a high level aesthetic attribute of the set. Framework 300 operates on images 302 belonging to a set. Images 302 may comprise photographs or renderings. In some embodiments, the set of images 302 comprises a curated set or catalog of images (e.g., such as the images of Figure 2) that results from artist post-processing or manipulation to achieve a prescribed style or aesthetic or theme. In some cases, the images 302 are not labeled or tagged, e.g., with ground truth data. However, images 302 belong to the same set and, thus, share one or more common properties or attributes such as content type as well as visual characteristics and appearance. The set of images 302 are processed by a machine learning based framework 304 (e.g., which may comprise framework 100 of Figure 1) to detect or identify a set of one or more high and low level attributes 306 associated with the set of images 302.

[0029] In some cases, machine learning based framework 304 is trained on large labeled image datasets comprising a substantial subset of, if not all, possible permutations of objects of a constrained set of possible objects that may appear in a prescribed scene type in order to learn associated attributes and combinations thereof and which may subsequently be employed to detect or identify such attributes in other images such as a corresponding set of curated or artist processed images 302 that include objects from the same constrained set of possible objects. Examples of attributes 306 that may be detected for the set of images 302 include object/scene types and geometries, materials and textures, camera characteristics, lighting characteristics, noise statistics, contrast, etc. In some embodiments, a high level aesthetic attribute that results in images 302 having a shared look is defined by a (e.g., unknown nonlinear) function or combination of a plurality of identified low level attributes 306. The described machine learning based framework 304 facilitates identifying shared properties of images belonging to a set 302 and identifying and isolating a high level shared style or aesthetic attribute that includes artist imparted characteristics. Attributes 306 identified for the set of images 302 may be employed to automatically label or tag images 302 that do not already have such labels or tags.

[0030] As previously described, many applications require artists to post-process images to impart prescribed styles or aesthetics. Thus, in such applications, curated imagery suitable for publication is often limited to a small number of viable shots that conform to a desired style or aesthetic. It would be useful to automatically generate more extensive sets of curated images, for example, that have prescribed styles or aesthetics but without requiring artists to impart the styles or aesthetics via post-processing. Figure 4 is a high level block diagram of an embodiment of an image processing framework 400 for automatically generating curated images without artist input. As depicted, a set of attributes 402, e.g., identified from a small set of artist created curated images such as using framework 300 of Figure 3, is applied to available three-dimensional object models 404 with configurable properties (geometries, lightings, materials, camera poses, etc.) to render any number of images 406 having attributes 402. Rendered images 406 automatically have the curated visual appearance that is defined by attributes 402 but without requiring artist input or post-processing.

[0031] In some embodiments, three-dimensional (polygon mesh) models for most, if not all, individual objects comprising a constrained set of possible objects that may be included in a prescribed scene type exist and may be employed to automatically render any number of additional curated shots of the scene type without artist input but having the same aesthetic or style as a relatively small set of artist created base imagery from which aesthetic or style attributes are identified, for example, using framework 300 of Figure 3. In some cases, a super catalog of curated images for a prescribed scene type may be generated of which only a small subset of shots are artist created and the rest are rendered automatically using framework 400. In some cases, a super catalog includes images comprising many, if not all, possible permutations and combinations of objects, materials, lightings, camera poses, placements, etc., associated with a particular scene space that all exhibit a prescribed style or aesthetic defined by attributes 402 but without requiring artist post-processing to impart the style or aesthetic. That is, aesthetic attributes isolated using framework 300 of Figure 3 may be used to render additional new images having that same aesthetic (i.e., the same low level attributes 306/402 that define the aesthetic) using the three-dimensional object models of framework 400. Images 406 generated using framework 400 may automatically be labeled or tagged with appropriate metadata or ground truth data since such images are rendered from well-defined three-dimensional models and known attributes. Moreover, large sets of labeled

images 406 that are generated may be employed to further train and build an associated machine learning based framework (e.g., framework 100 of Figure 1).

[0032] Figure 5 is a high level block diagram of an embodiment of a machine learning based image processing architecture 500. As depicted, image processing architecture 500 includes many components that have separately been described in detail with respect to Figures 1-4. Machine learning framework 501 (which may comprise framework 100 of Figure 1) is the foundation of image processing architecture 500 and trains on large datasets, e.g., which may at least in part be generated from available three-dimensional models, to learn attributes associated with the datasets. Machine learning framework 501 may then be used, for example, with other images to detect or identify such attributes or combinations thereof, which may not be known for the images prior to detection by machine learning framework 501.

[0033] In one example, machine learning framework 501 identifies attributes that collectively define an aesthetic 504 of a small set of curated catalog images 502 that have been post-processed by artists to have the aesthetic. In some cases, the isolated aesthetic (i.e., corresponding attributes) 504 may be applied to available three-dimensional models to generate a super catalog 506 of additional curated catalog images that all have that aesthetic but without requiring artist post-processing like the original set 502. Super catalog 506 may be used to further train and build machine learning framework 501. Thus, aesthetics or styles or themes may be identified using machine learning framework 501 and then applied to available three-dimensional models to generate additional datasets on which machine learning framework 501 may be further trained.

[0034] Machine learning framework 501 generally facilitates a variety of image processing applications 508 for modifying input images 510 or parts thereof to generate corresponding output images 512 having the desired modifications. Both high and low level attributes associated with images are detectable, and high level attributes are separable into constituent lower level attributes. Thus, independent decisions can be made on different attributes or combinations of attributes in various applications. Some example image processing applications 508 include restyling (e.g., changing aesthetic), object replacement, relighting (e.g., changing light source types and/or locations), etc., a few of which are further described next.

[0035] Figure 6A is a high level block diagram of an embodiment of an image processing application 600 for restyling, and Figure 6B illustrates an example of using restyling application 600. As depicted, machine learning framework 601 (e.g., framework 100 of Figure 1 or framework 501 of Figure 5) is employed to identify and isolate a catalog aesthetic 604 of an image catalog 602 and an image aesthetic 608 of an input image 606 so that the image aesthetic 608 can be removed or subtracted from image 606 and the catalog aesthetic 604 applied or added to image 606 to generate an output catalog version 610 of input image 606 that has the same aesthetic 604 as catalog 602. In this example, the types of attributes that comprise the catalog aesthetic 604 may be detected and isolated in input image 606 to determine image aesthetic 608. Restyling is illustrated in Figure 6B in which an input image 606 is restyled by restyling application 600 to generate a catalog image 610 that conforms to a prescribed catalog aesthetic 604. In this example, changing the aesthetic comprises modifying the lighting or brightness at different parts of the image in different manners.

[0036] Figure 7A is a high level block diagram of an embodiment of an image processing application 700 for replacing an object in an image with a different object, and Figure 7B illustrates an example of using object replacement application 700. In Figure 7A, a machine learning framework (e.g., framework 100 of Figure 1 or framework 501 of Figure 5) is employed to identify attributes 704-708 of an input image 702 that are associated with a prescribed object so that the object can be replaced with another object 710. New object 710 may be configured to have some of the same identified attributes 704-708 so that it can be consistently included in image 710 in place of the replaced object. Object replacement is illustrated in Figure 7B in which a sofa object in input image 702 is replaced with a different sofa object in output image 710.

[0037] Image processing applications 508 may furthermore comprise more complex enterprise applications such as aggregating objects from datasets having different aesthetic attributes into the same scene and styling to have a prescribed aesthetic. For example, home furnishings objects from one or more brands may be included in an image of a room but may be all styled to have the aesthetic of a prescribed brand. In such cases, the resulting image would have the curated look or aesthetic of a catalog image of the prescribed brand.

[0038] Generally, image processing applications 508 rely on attribute detection using machine learning framework 501. That is, actual attributes used to generate images 510 are

detected using machine learning framework 501 and modified to generate output images 512 having modified attributes. Thus, image modification based on attribute detection, manipulation, and/or modification as described herein is notably different from conventional image editing applications that operate at pixel levels on pixel values and do not have information on actual image content (e.g., objects) and underlying attributes associated with the physics of capturing or rendering the content such as geometry, camera, lighting, materials, textures, placements, etc., on which the disclosed techniques are based. The disclosed attribute detection and manipulation techniques are especially useful for photorealistic applications because conventional pixel manipulations are not sufficiently constrained to generate images that look real and consistent.

[0039] Another useful application 508 based on machine learning framework 501 comprises image denoising. One or more learned spatial filters may be applied to various parts of a noisy input image 510 (e.g., that is generated using a low number of samples of rays during ray tracing) to remove noise so that output image 512 has a noise profile or quality that is comparable to that achievable using a large number of samples of rays. That is, various parts of a sparsely sampled image are filtered using a set of filters identified by machine learning framework 501 to generate an output image equivalent to ray tracing with a much larger number of samples (e.g., a number of samples needed for complete convergence). As one example, a ten sample ray traced image can quickly be transformed into the equivalent of a corresponding thousand sample ray traced image by filtering the ten sample ray traced image with appropriate filters identified by machine learning framework 501. Thus, image render time can substantially be reduced by only ray tracing with a small number of samples and then using filters that predict pixel values that would result from larger numbers of samples. This technique effectively eliminates the need to ray trace with large numbers of samples while still producing images having qualities or noise profiles that ray tracing with large numbers of samples provides.

[0040] In some embodiments, training datasets for such a denoising application comprise ray traced snapshots of images at different sampling intervals, with each snapshot labeled with an attribute specifying the number of samples for that snapshot in addition to being labeled with other image attributes. Machine learning framework 501 trains on such datasets to learn spatial filters or parameters thereof for different numbers of samples. For example, filters may be learned for transforming from a low number (x) of samples to a high

number (y) of samples for many different values and combinations of x and y, where x << y. Noise signatures, however, are not only based on numbers of samples but also on one or more other image attributes that affect noise (e.g., during ray tracing) such as materials and lighting. Thus, different filter parameters may be learned for attribute combinations that result in different noise signatures, and a machine learning framework 501 that identifies filters for an input image may identify different filters or parameters for different portions of the image. For example, a different set of filter parameters may be identified for a portion of an image that has attribute combination "ten samples on leather with bright light" than a portion of the image that has attribute combination "ten samples on fabric in dim light".

[0041] Figure 8A is a high level block diagram of an embodiment of an image processing application 800 for denoising an image, and Figure 8B illustrates an example of using denoising application 800. In Figure 8A, a noisy (e.g., low sample count) input image 802 that violates a prescribed noise threshold is processed using one or more filters 804 that are identified by a machine learning based framework (e.g., framework 100 of Figure 1 or framework 501 of Figure 5) to generate a denoised output image 806 that satisfies the prescribed noise threshold. Thus, a noisy image 802 rendered using a low number of samples of rays is cleaned up or denoised 806 using learned filters 804, eliminating the need to actually ray trace with a large number of samples. Denoising is illustrated in Figure 8B in which a sparsely sampled input image 802 is processed by application 800 to generate a high quality output image 806.

[0042] Although the foregoing embodiments have been described in some detail for purposes of clarity of understanding, the invention is not limited to the details provided. There are many alternative ways of implementing the invention. The disclosed embodiments are illustrative and not restrictive.

CLAIMS

1. A method, comprising:

using a machine learning framework to detect a set of one or more attributes of an input image; and

outputting an output image comprising a modified version of the input image, wherein the input image is modified by modifying at least a subset of the detected set of attributes.

2. The method of claim 1, wherein the detected set of attributes is not known for the input image prior to detection by the machine learning framework.

3. The method of claim 1, wherein the detected set of attributes is associated with a style or aesthetic.

4. The method of claim 3, wherein the output image comprises a restyled version of the input image.

5. The method of claim 1, wherein the detected set of attributes is associated with a first style and a modified set of attributes of the output image is associated with a second style.

6. The method of claim 5, wherein modifying the input image comprises modifying the first style to the second style.

7. The method of claim 1, wherein the detected set of attributes is associated with an object in the input image.

8. The method of claim 7, wherein the object in the input image is replaced by a different object in the output image.

9. The method of claim 1, wherein the detected set of attributes is associated with lighting.

10. The method of claim 9, wherein the output image comprises a relit version of the input image.

11. The method of claim 1, wherein the detected set of attributes is associated with noise.

12. The method of claim 11, wherein the output image comprises a denoised version of the input image.

13. The method of claim 1, further comprising labeling or tagging the output image with a modified set of attributes.

14. The method of claim 1, wherein the detected set of attributes comprises one or more attributes associated with object/scene types, geometries, placements, materials, textures, camera characteristics, lighting characteristics, noise statistics, and contrast.

15. The method of claim 1, wherein the machine learning framework is trained on image datasets comprising permutations of a constrained set of objects associated with a prescribed scene type to which the input image belongs.

16. The method of claim 1, wherein the machine learning framework is trained on datasets comprising labeled images rendered from three-dimensional object models.

17. The method of claim 1, wherein the input image and the output image each comprises a photograph or a photorealistic rendering.

18. The method of claim 1, wherein the input image and the output image each comprises a frame of an animation or a video sequence.

19. A system, comprising:

a processor configured to:

use a machine learning framework to detect a set of one or more attributes of an input image; and

output an output image comprising a modified version of the input image, wherein the input image is modified by modifying at least a subset of the detected set of attributes; and

a memory coupled to the processor and configured to provide the processor with instructions.

20. A computer program product embodied in a non-transitory computer readable storage medium and comprising computer instructions for:

using a machine learning framework to detect a set of one or more attributes of an input image; and

outputting an output image comprising a modified version of the input image, wherein the input image is modified by modifying at least a subset of the detected set of attributes.

21. A method, comprising:

using a machine learning framework to detect an aesthetic attribute associated with a first set of images having a shared aesthetic; and

generating a second set of images having the same shared aesthetic by applying the detected aesthetic attribute to three-dimensional object models used to render the second set

of images.

22. The method of claim 21, wherein the shared aesthetic of the first set of images results from post-processing.

23. The method of claim 21, wherein the shared aesthetic is imparted to the first set of images by an artist.

24. The method of claim 21, wherein the second set of images is generated to have the same shared aesthetic without artist input or post-processing.

25. The method of claim 21, wherein the aesthetic attribute comprises a combination of one or more low level attributes.

26. The method of claim 25, wherein the aesthetic attribute is separable into its constituent low level attributes.

27. The method of claim 25, wherein the low level attributes comprise one or more attributes associated with object/scene types, geometries, placements, materials, textures, camera characteristics, lighting characteristics, noise statistics, and contrast.

28. The method of claim 21, wherein the shared aesthetic is associated with a prescribed brand.

29. The method of claim 21, wherein the shared aesthetic is associated with a prescribed scene type.

30. The method of claim 21, wherein the second set comprises more images than the first set.

31. The method of claim 21, wherein the first set and the second set collectively comprise a super catalog of images having the shared aesthetic of which only the first set is artist post-processed.

32. The method of claim 21, wherein the first set and the second set collectively comprise a super catalog of images having the shared aesthetic that is used to further train and build the machine learning framework.

33. The method of claim 21, further comprising labeling or tagging images of the first set with the detected aesthetic attribute.

34. The method of claim 21, further comprising labeling or tagging images of the second set with the aesthetic attribute.

35. The method of claim 21, wherein the machine learning framework is trained on image datasets comprising permutations of a constrained set of objects associated with a prescribed scene type to which the first set of images and the second set of images belong.

36. The method of claim 21, wherein the machine learning framework is trained on datasets comprising labeled images rendered from three-dimensional object models.

37. The method of claim 21, wherein images comprising the first set and the second set comprise photographs or photorealistic renderings.

38. The method of claim 21, wherein images comprising the first set and the second set comprise frames of an animation or a video sequence.

39. A system, comprising:

a processor configured to:

use a machine learning framework to detect an aesthetic attribute associated with a first set of images having a shared aesthetic; and

generate a second set of images having the same shared aesthetic by applying the detected aesthetic attribute to three-dimensional object models used to render the second set of images; and

a memory coupled to the processor and configured to provide the processor with instructions.

40. A computer program product embodied in a non-transitory computer readable storage medium and comprising computer instructions for:

using a machine learning framework to detect an aesthetic attribute associated with a first set of images having a shared aesthetic; and

generating a second set of images having the same shared aesthetic by applying the detected aesthetic attribute to three-dimensional object models used to render the second set of images.

41. A method, comprising:

using a machine learning framework to identify a set of one or more filters for removing noise from a sparsely ray traced input image; and

outputting an output image comprising a filtered version of the input image resulting from filtering the input image with the identified set of filters, wherein a quality of the output image is equivalent to ray tracing with more samples of rays than the sparsely ray traced input image.

42. The method of claim 41, wherein the output image is a denoised version of the input image.

43. The method of claim 41, wherein the sparsely ray traced input image violates a noise threshold and the output image satisfies the noise threshold.

44. The method of claim 41, wherein the sparsely ray traced input image is ray traced with a small number of samples of rays.

45. The method of claim 41, wherein the quality of the output image is equivalent to ray tracing with a large number of samples of rays.

46. The method of claim 41, wherein the quality of the output image is equivalent to ray tracing with a number of samples of rays needed for complete convergence.

47. The method of claim 41, wherein the set of filters predicts pixel values that would result from ray tracing with more samples of rays.

48. The method of claim 41, wherein the filters comprise spatial filters.

49. The method of claim 41, wherein to identify a set of one or more filters comprises to identify a set of one or more filter parameters.

50. The method of claim 41, wherein to identify a set of one or more filters comprises to identify different filters for different portions of the input image.

51. The method of claim 41, wherein a prescribed filter of the set of filters is associated with a corresponding noise signature.

52. The method of claim 51, wherein the noise signature is based on one or more attributes.

53. The method of claim 52, wherein the attributes comprise one or more attributes associated with object/scene types, geometries, placements, materials, textures, camera characteristics, lighting characteristics, numbers of samples, and contrast.

54. The method of claim 41, wherein the machine learning framework is trained to learn noise signatures and corresponding filters.

55. The method of claim 41, wherein the machine learning framework is trained on image datasets comprising ray traced snapshots at different sampling intervals.

56. The method of claim 41, wherein the machine learning framework is trained on image

datasets comprising permutations of a constrained set of objects associated with a prescribed scene type to which the input image belongs.

57. The method of claim 41, wherein the output image comprises a photorealistic rendering.

58. The method of claim 41, wherein the output image comprises a frame of an animation or a video sequence.

59. A system, comprising:

a processor configured to:

use a machine learning framework to identify a set of one or more filters for removing noise from a sparsely ray traced input image; and

output an output image comprising a filtered version of the input image resulting from filtering the input image with the identified set of filters, wherein a quality of the output image is equivalent to ray tracing with more samples of rays than the sparsely ray traced input image; and

a memory coupled to the processor and configured to provide the processor with instructions.

60. A computer program product embodied in a non-transitory computer readable storage medium and comprising computer instructions for:

using a machine learning framework to identify a set of one or more filters for removing noise from a sparsely ray traced input image; and

outputting an output image comprising a filtered version of the input image resulting from filtering the input image with the identified set of filters, wherein a quality of the output image is equivalent to ray tracing with more samples of rays than the sparsely ray traced input image.

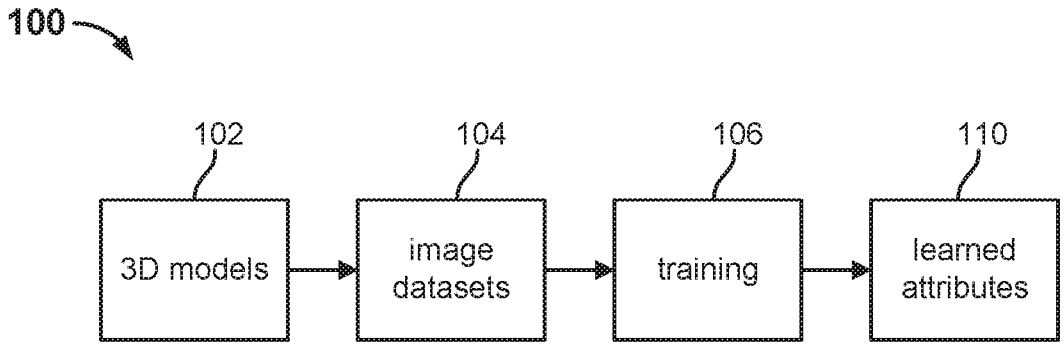
100
102
104
106
110
3D models
image datasets
training
learned attributes

FIG. 1

FIG. 2

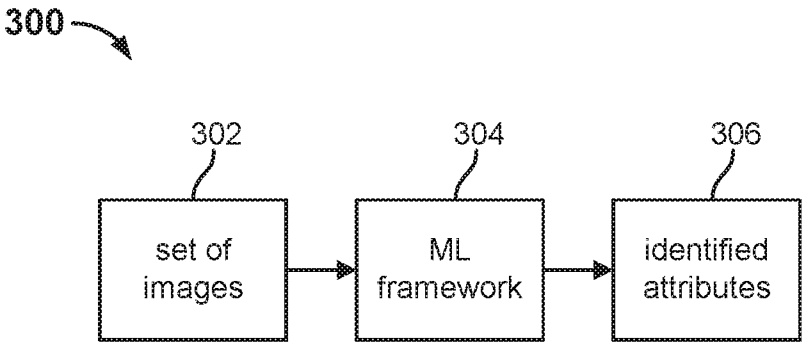
300
302
304
306
set of images
ML framework
identified attributes

FIG. 3

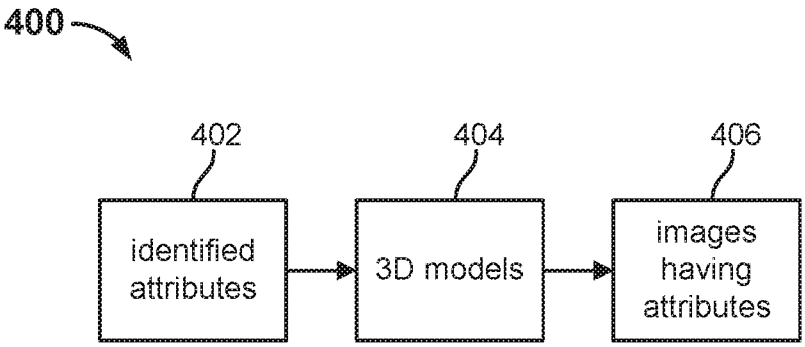
400
402
404
406
identified attributes
3D models
images having attributes

FIG. 4

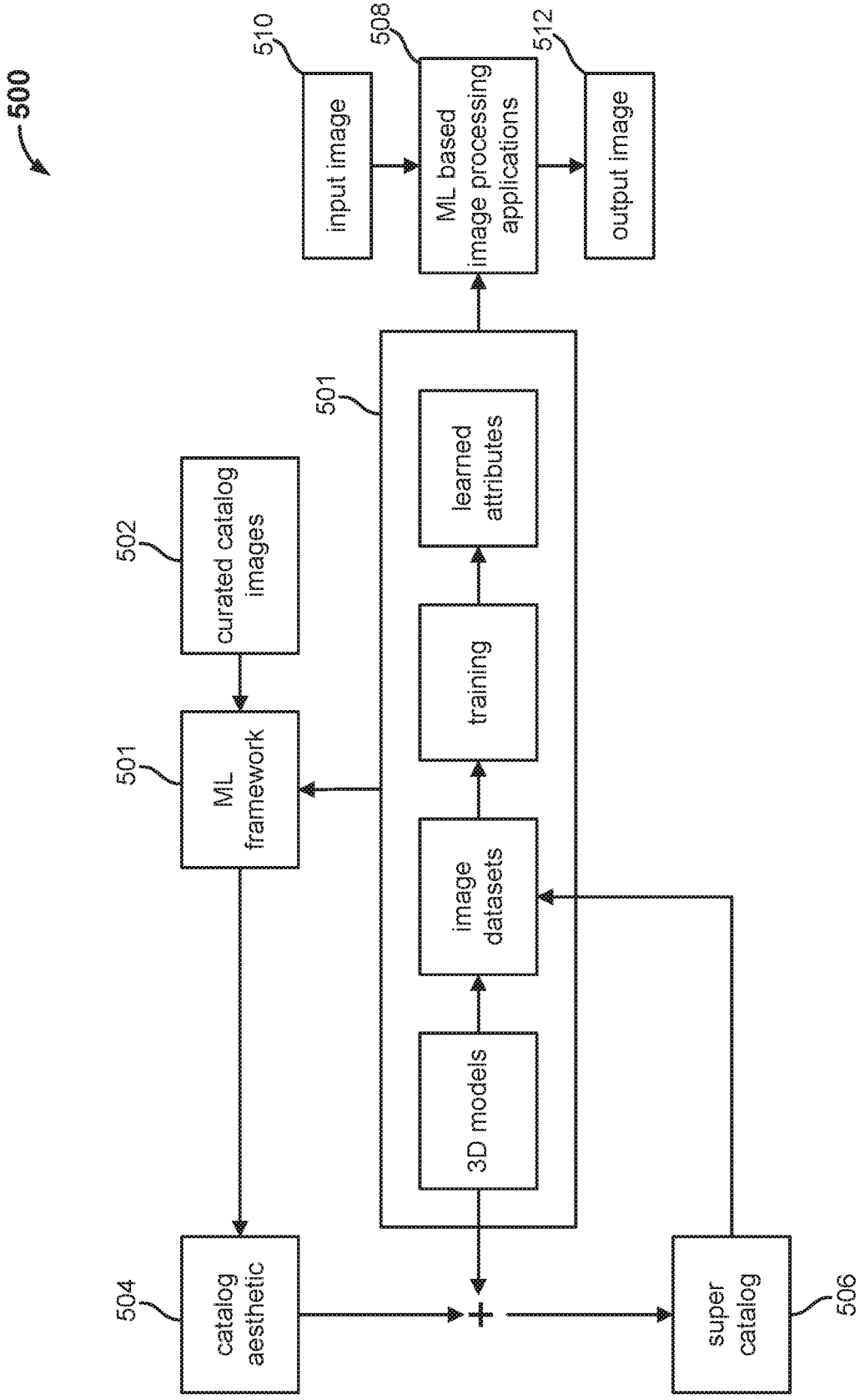
500
510
input image
508
ML based image processing applications
512
output image
501
learned attributes
training
image datasets
3D models
502
curated catalog images
501
ML framework
504
catalog aesthetic
super catalog
506

FIG. 5

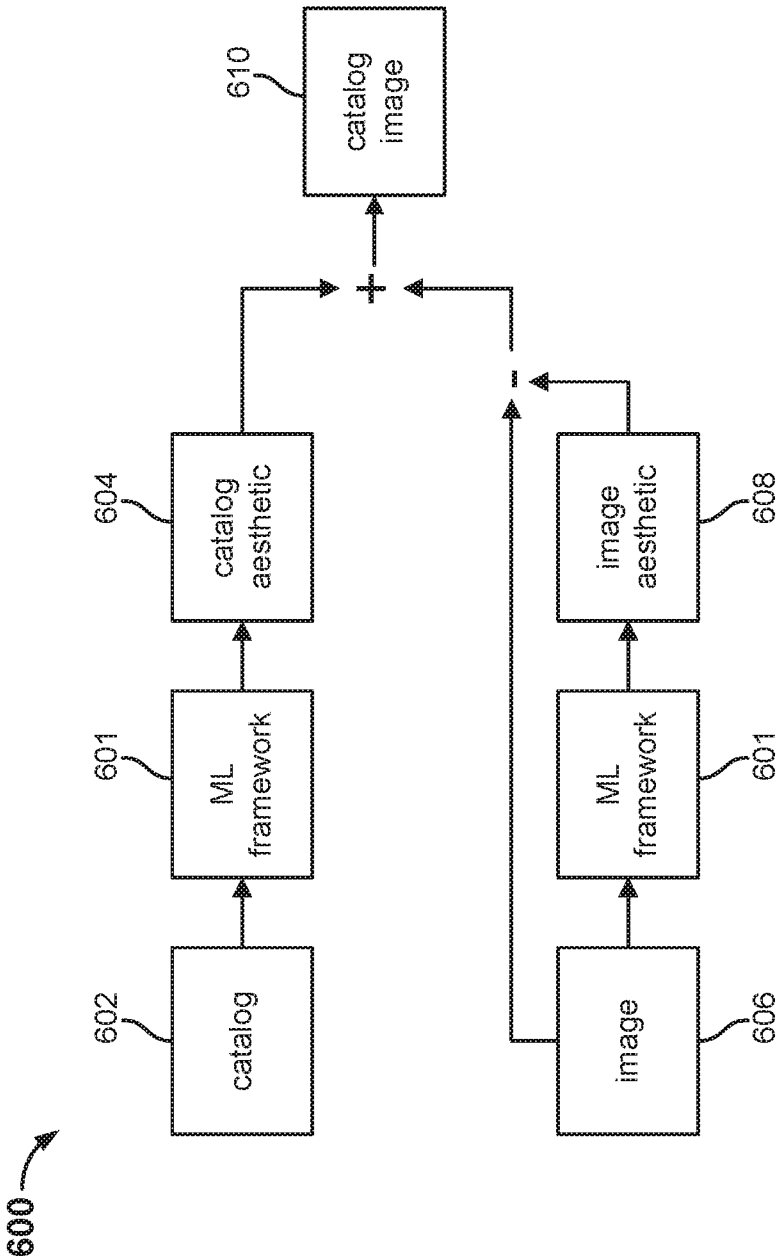
600
catalog
602
ML framework
601
catalog aesthetic
604
image
606
ML framework
601
image aesthetic
608
+
-
catalog image
610

FIG. 6A

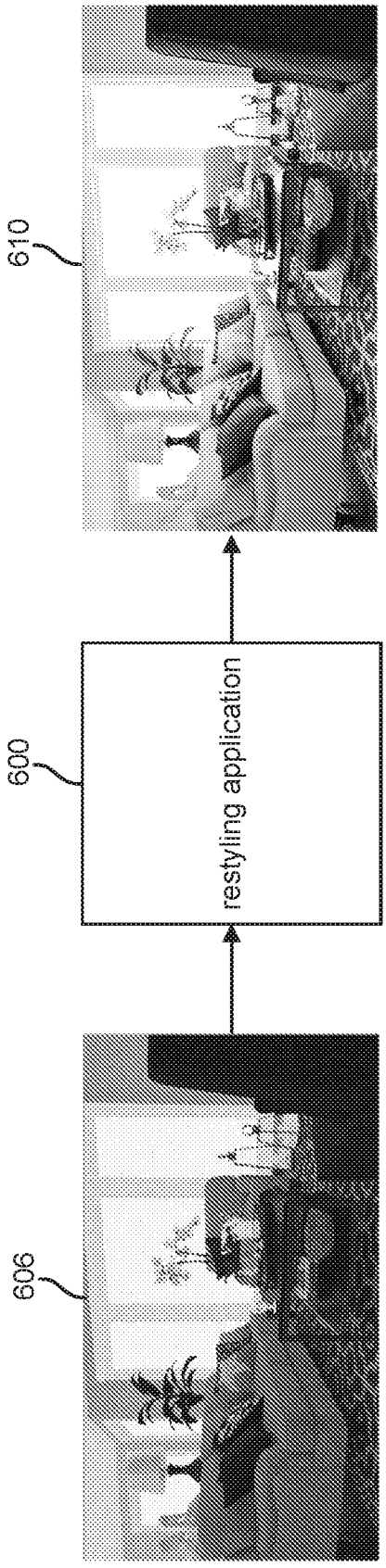

FIG. 6B

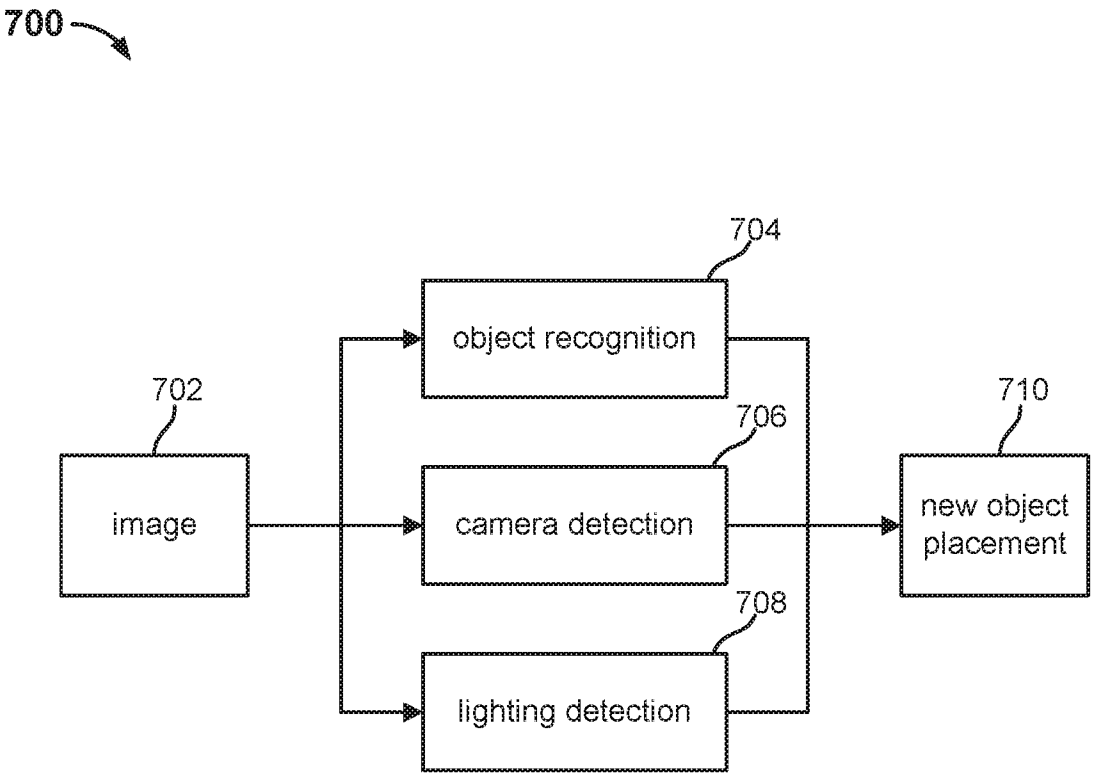
700
704
object recognition
702
image
706
camera detection
710
new object placement
708
lighting detection

FIG. 7A

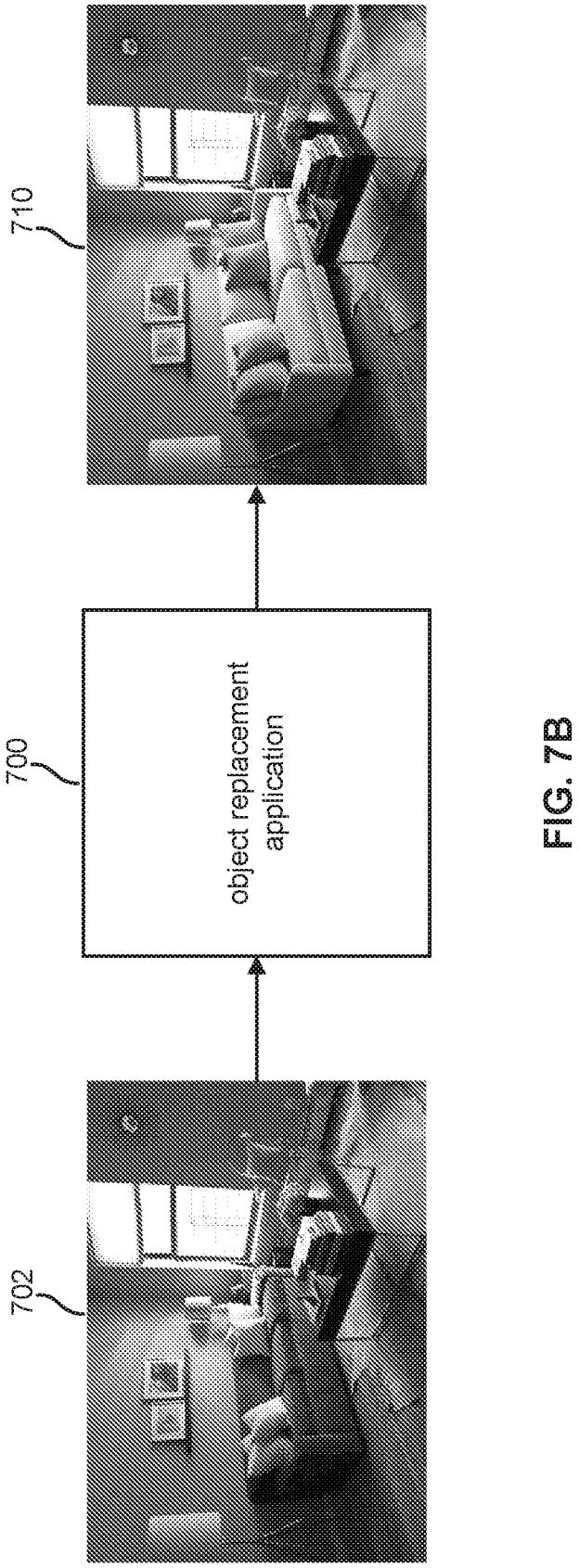

FIG. 7B

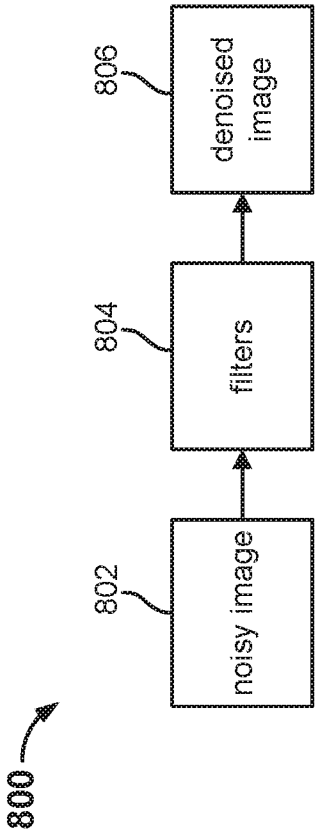

FIG. 8A

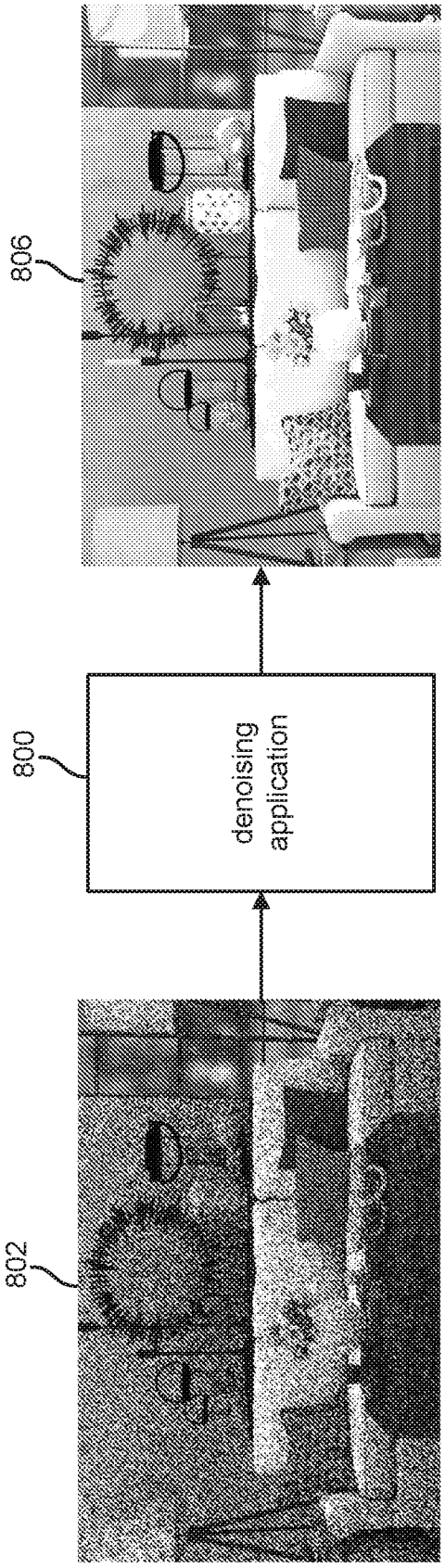

FIG. 8B

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 18/45423

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. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I: Claims 1-20, drawn to a method of and a computer program product for modifying a subset of the detected set of attributes.

Group II: Claims 21-40, drawn to a method of and a computer program product for applying detected aesthetic attribute to three-dimensional object models.

Group III: Claims 41-60, drawn to a method of, a system of, and a computer program product for ray tracing with more samples.

-- Please See Supplemental Box --

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.:

1-20

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.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 18/45423

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 15/18, G06K 9/46, G06K 9/62, G06K 9/66, G06T 7/246, G06T 7/60, H04N 1/387 (2018.01)
CPC - G01N 21/8851, G01N 2021/8887, G01N 2223/401, G05B 2219/36414, G06F 15/18, G06F 17/30247, G06K 9/00067, G06K 9/00973, G06K 9/036, G06K 9/3233, G06K 9/3241, G06K 9/46, G06K 9/62, G06K 9/66, G06K 9/4604, G06T 5/001, G06T 7/0002, G06T 7/246, G06T 7/33, G06T 7/60, G06T 7/62, G06T 7/70, G06T 2207/20081, G06T 2207/20084, G06T 2207/30168, G06T 2219/004, H04N 1/387, H04N 1/40093, H04N 2201/3

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.
X -- Y -- A	US 2017/0103512 A1 (SIEMENS HEALTHCARE GMBH) 13 April 2017 (13.04.2017), para [0008], [0015]-[0018], [0020], [0021], [0024], [0026], [0031], [0034], [0035]	1, 2, 9, 11, 13, 14, 19, 20 --- 3-8, 10, 12, 17, 18 --- 15, 16
Y -- A	US 2017/0097948 A1 (ADOBE SYSTEMS INCORPORATED) 06 April 2017 (06.04.2017), para [0005], [0006], [0018], [0025], [0041], [0042], [0044], [0045], [0049], [0054]-[0056]	3-8 --- 15, 16
Y -- A	US 2013/0235067 A1 (APPLE INC.) 12 September 2013 (12.09.2013), para [0004], [0008], [0015], [0073], [0096],, [0119], [0121], [0322], [0397]	10, 17, 18 --- 15, 16
Y -- A	US 5,526,446 A (ADELSON et al.) 11 June 1996 (11.06.1996), col 6, ln 39-67; col 20, ln 12-46	12 --- 15, 16
A, P	US 2018/0121762 A1 (SNAP INC.) 03 May 2018 (03.05.2018), para [0017]-[0110]	1-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
14 November 2018	29 NOV 2018
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 W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Continued from Box No. III, Observations where unity of invention is lacking,

The inventions listed as Groups I, II, and III 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:

Special Technical Features

Groups II and III do not require a method, comprising: using a machine learning framework to detect a set of one or more attributes of an input image; and outputting an output image comprising a modified version of the input image, wherein the input image is modified by modifying at least a subset of the detected set of attributes; and a computer program product embodied in a non-transitory computer readable storage medium and comprising computer instructions for: using a machine learning framework to detect a set of one or more attributes of an input image; and outputting an output image comprising a modified version of the input image, wherein the input image is modified by modifying at least a subset of the detected set of attributes, as required by Group I.

Groups I and III do not require a method, comprising: using a machine learning framework to detect an aesthetic attribute associated with a first set of images having a shared aesthetic; and generating a second set of images having the same shared aesthetic by applying the detected aesthetic attribute to three-dimensional object models used to render the second set of images; and a computer program product embodied in a non-transitory computer readable storage medium and comprising computer instructions for: using a machine learning framework to detect an aesthetic attribute associated with a first set of images having a shared aesthetic; and generating a second set of images having the same shared aesthetic by applying the detected aesthetic attribute to three-dimensional object models used to render the second set of images, as required by Group II.

Groups I and II do not require a method, comprising: using a machine learning framework to identify a set of one or more filters for removing noise from a sparsely ray traced input image; and outputting an output image comprising a filtered version of the input image resulting from filtering the input image with the identified set of filters, wherein a quality of the output image is equivalent to ray tracing with more samples of rays than the sparsely ray traced input image; a system, comprising: a processor configured to: use a machine learning framework to identify a set of one or more filters for removing noise from a sparsely ray traced input image; and output an output image comprising a filtered version of the input image resulting from filtering the input image with the identified set of filters, wherein a quality of the output image is equivalent to ray tracing with more samples of rays than the sparsely ray traced input image; and a memory coupled to the processor and configured to provide the processor with instructions; and a computer program product embodied in a non-transitory computer readable storage medium and comprising computer instructions for: using a machine learning framework to identify a set of one or more filters for removing noise from a sparsely ray traced input image; and outputting an output image comprising a filtered version of the input image resulting from filtering the input image with the identified set of filters, wherein a quality of the output image is equivalent to ray tracing with more samples of rays than the sparsely ray traced input image, as required by Group III.

Shared Common Features

The only feature shared by Groups I, II, and III that would otherwise unify the groups is using a machine learning framework, outputting an output image, and a computer program product embodied in a non-transitory computer readable storage medium and comprising computer instructions for using a machine learning framework. However, this shared technical feature does not represent a contribution over prior art, because the shared technical feature is anticipated by US 2017/0103512 A1 (Siemens Healthcare GmbH). Siemens Healthcare GmbH discloses using a machine learning framework (para [0004], [0006], [0043], learning-based framework... deep learning network... machine instructions.), outputting an output image (para [0026], [0031], [0044], scanners used to output images... generating images.), and a computer program product embodied in a non-transitory computer readable storage medium (para [0007], [0034], [0038]) and comprising computer instructions for using a machine learning framework (para [0004], [0006], [0043], learning-based framework... deep learning network... machine instructions.).

As the technical features were known in the art at the time of the invention, this cannot be considered a special technical feature that would otherwise unify the groups.

Groups I, II, and III therefore lack unity under PCT Rule 13 because they do not share a same or corresponding special technical feature.