



(51) International Patent Classification:

G06T 19/00 (2011.01) G06T 7/50 (2017.01)
G06T 7/20 (2006.01)

(21) International Application Number:

PCT/US2017/031839

(22) International Filing Date:

09 May 2017 (09.05.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/447,128 17 January 2017 (17.01.2017) US
15/489,503 17 April 2017 (17.04.2017) US

(71) Applicant: FACEBOOK, INC. [US/US]; 1601 Willow
Road, Menlo Park, CA 94025 (US).

(72) Inventor: KOPF, Johannes, Peter; Facebook, INC., 1601
Willow Road, Menlo Park, CA 94025 (US).

(74) Agent: HULSE, Robert, A. et al.; Fenwick & West LLP,
Silicon Valley Center, 801 California Street, Mountain
View, CA 94041 (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, 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.

(54) Title: THREE-DIMENSIONAL SCENE RECONSTRUCTION FROM SET OF TWO-DIMENSIONAL IMAGES FOR
CONSUMPTION IN VIRTUAL REALITY

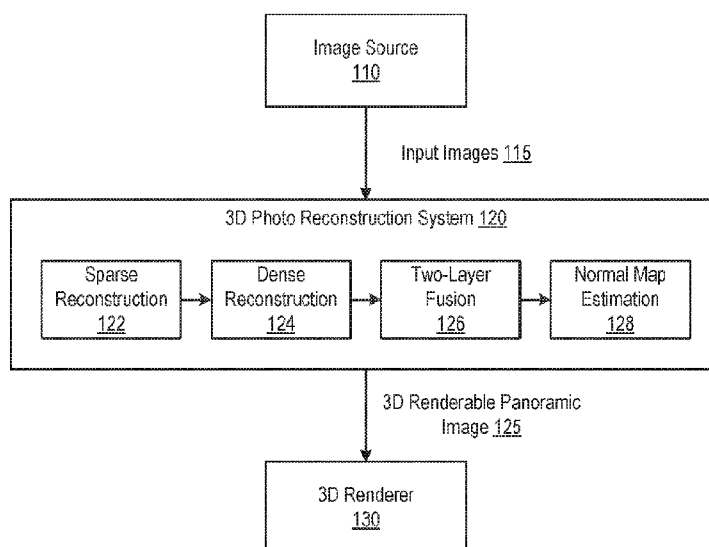


Fig. 1

(57) Abstract: To enable better sharing and preservation of immersive experiences, a graphics system reconstructs a three-dimensional scene from a set of images of the scene taken from different vantage points. The system processes each image to extract depth information therefrom and then stitches the images (both color and depth information) into a multi-layered panorama that includes at least front and back surface layers. The front and back surface layers are then merged to remove redundancies and create connections between neighboring pixels that are likely to represent the same object, while removing connections between neighboring pixels that are not. The resulting layered panorama with depth information can be rendered using a virtual reality (VR) system, a mobile device, or other computing and display platforms using standard rendering techniques, to enable three-dimensional viewing of the scene.

(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))*

THREE-DIMENSIONAL SCENE RECONSTRUCTION FROM SET OF TWO-DIMENSIONAL IMAGES FOR CONSUMPTION IN VIRTUAL REALITY

BACKGROUND

[0001] This disclosure relates generally to rendering three-dimensional images and more specifically to reconstructing a three-dimensional scene from a set of two-dimensional images.

[0002] People take pictures of scenes that they experience so they can share their experiences with others or to re-experience the scene at a later time. Unfortunately, technical limitations limit how well the subsequent experience is relived. Two-dimensional images do not provide the full three-dimensional experience of being there, and people usually do not carry around expensive and bulky three-dimensional cameras. Accordingly, it would be useful to enable people to capture a scene and digitally preserve it in a way that allows the person or friends to virtually immerse themselves in the scene and re-experience the sensation of being there at a later time. Preferably, this would be as easy as taking a picture today, using a standard phone or camera.

SUMMARY

[0003] A method, non-transitory computer-readable storage medium, and image reconstruction system generates a three-dimensional image from a plurality of two-dimensional input images. A plurality of input images of a scene is received in which the input images are taken from different vantage points. The input images may have varying amounts of overlap with each other and varying camera orientations. The plurality of input images is processed to generate a sparse reconstruction representation of the scene. The sparse reconstruction representation including a sparse point cloud specifying locations of a plurality of points that correspond to three-dimensional locations of surfaces of objects in the scene. Based in part on the sparse reconstruction representation, the plurality of input images is processed to generate respective dense reconstruction representations of each of the plurality of input images. Here, each of the respective dense reconstruction representations include a respective depth image for a corresponding input image in which the depth image includes both color and depth information. Front surfaces of the depth images are projected using a forward depth test to generate a plurality of front-warped images. Back surfaces of the depth images are projected using an inverted depth test to generate a plurality of back-

warped images. The front-warped images and the back-warped images are stitched to generate a two-layer panorama having a front surface panorama and a back surface panorama. The front surface panorama and the back surface panorama are then fused to generate the three-dimensional image comprising a multi-layered geometric mesh suitable for rendering the scene in the three-dimensional space.

[0004] Embodiments according to the invention are in particular disclosed in the attached claims directed to a method, a storage medium, a system and a computer program product, wherein any feature mentioned in one claim category, e.g. method, can be claimed in another claim category, e.g. system, as well. The dependencies or references back in the attached claims are chosen for formal reasons only. However any subject matter resulting from a deliberate reference back to any previous claims (in particular multiple dependencies) can be claimed as well, so that any combination of claims and the features thereof is disclosed and can be claimed regardless of the dependencies chosen in the attached claims. The subject-matter which can be claimed comprises not only the combinations of features as set out in the attached claims but also any other combination of features in the claims, wherein each feature mentioned in the claims can be combined with any other feature or combination of other features in the claims. Furthermore, any of the embodiments and features described or depicted herein can be claimed in a separate claim and/or in any combination with any embodiment or feature described or depicted herein or with any of the features of the attached claims.

[0005] In an embodiment according to the invention, a method comprises:
receiving a plurality of input images of a scene taken from different vantage points;
processing the plurality of input images to generate a sparse reconstruction representation of the scene, the sparse reconstruction representation including a sparse point cloud specifying locations of a plurality of points that correspond to three-dimensional locations of surfaces of objects in the scene;
based in part on the sparse reconstruction representation, processing the plurality of input images to generate respective dense reconstruction representations of each of the plurality of input images, the respective dense reconstruction representations each including respective depth images for the plurality of input images, the depth images including color and depth information;
projecting front surfaces of the depth images using a forward depth test to generate a plurality of front-warped images;
projecting back surfaces of the depth images using an inverted depth test to generate a

plurality of back-warped images;
stitching the front-warped images and the back-warped images to generate a two-layer panorama having a front surface panorama and a back surface panorama; and
fusing the front surface panorama and the back surface panorama in the two-layer panorama to generate the three-dimensional image comprising a multi-layered geometric mesh suitable for rendering the scene in the three-dimensional space.

[0006] Processing the plurality of input images to generate the sparse reconstruction representation may comprise:

applying a surface-from-motion algorithm to the plurality of input images.

[0007] Processing the plurality of input images to generate the respective dense reconstruction representations may comprise:

generating a near envelope prior that assigns a cost to estimated depth values in front of a near envelope; and

applying a multi-view stereo processing algorithm to estimate the depth values based on a cost function including the near envelope prior.

[0008] Generating the near envelope prior may comprise:

identifying anchor pixels in the plurality of input images that have high confidence depth estimates;

propagating the depth estimates of the anchor pixels to other pixels in the plurality of input images to generate approximate depth maps;

filtering the approximate depth maps to determine a near envelope; and

generating the near envelope prior based on the depth estimates and the near envelope.

[0009] Stitching the front-warped images and the back-warped images to generate a two-layer panorama may comprise:

stitching a depth panorama using depth values from the front-warped images;

stitching the front surface panorama using color values from the front-warped images and stitched depth values from the depth panorama;

stitching the back surface panorama using color values from the back-warped images and the stitched depth values from the depth panorama; and

combining the front surface panorama and the back surface panorama into the two-layer panorama.

[0010] Fusing the front surface panorama and the back surface panorama may comprise:

removing background pixels from the back surface panorama that match corresponding foreground pixels in the front surface panorama;

storing connections between neighboring pixels meeting a threshold similarity in depth and color information; and

hallucinating color and depth information in missing pixel locations.

[0011] In an embodiment according to the invention, a method may comprise: generating a normal map for the multi-layered geometric mesh, the normal map estimating for each pixel, an angle normal to a surface depicted by the pixel.

[0012] Generating the normal map may comprise: generating a base normal map from depth values in the three-dimensional image; generating a detailed normal from luminance values in the three-dimensional image; and transforming the detailed normal map onto the base normal map to generate a combined normal map.

[0013] The plurality of input images may have varying levels of overlap and orientation changes.

[0014] In an embodiment according to the invention, a non-transitory computer-readable storage medium may store instructions, the instructions when executed by a processor may cause the processor to perform steps including:

receiving a plurality of input images of a scene taken from different vantage points; processing the plurality of input images to generate a sparse reconstruction representation of the scene, the sparse reconstruction representation including a sparse point cloud specifying locations of a plurality of points that correspond to three-dimensional locations of surfaces of objects in the scene;

based in part on the sparse reconstruction representation, processing the plurality of input images to generate respective dense reconstruction representations of each of the plurality of input images, the respective dense reconstruction representations each including respective depth images for the plurality of input images, the depth images including color and depth information;

projecting front surfaces of the depth images using a forward depth test to generate a plurality of front-warped images;

projecting back surfaces of the depth images using an inverted depth test to generate a plurality of back-warped images;

stitching the front-warped images and the back-warped images to generate a two-layer panorama having a front surface panorama and a back surface panorama;

fusing the front surface panorama and the back surface panorama in the two-layer panorama to generate the three-dimensional image comprising a multi-layered geometric mesh suitable

for rendering the scene in the three-dimensional space.

[0015] Processing the plurality of input images to generate the sparse reconstruction representation may comprise:

applying a surface-from-motion algorithm to the plurality of input images.

[0016] Processing the plurality of input images to generate the respective dense reconstruction representations may comprise:

generating a near envelope prior that assigns a cost to estimated depth values in front of a near envelope; and

applying a multi-view stereo processing algorithm to estimate the depth values based on a cost function including the near envelope prior.

[0017] Generating the near envelope prior may comprise:

identifying anchor pixels in the plurality of input images that have high confidence depth estimates;

propagating the depth estimates of the anchor pixels to other pixels in the plurality of input images to generate approximate depth maps;

filtering the approximate depth maps to determine a near envelope; and

generating the near envelope prior based on the depth estimates and the near envelope.

[0018] Stitching the front-warped images and the back-warped images to generate a two-layer panorama may comprise:

stitching a depth panorama using depth values from the front-warped images;

stitching the front surface panorama using color values from the front-warped images and stitched depth values from the depth panorama;

stitching the back surface panorama using color values from the back-warped images and the stitched depth values from the depth panorama; and

combining the front surface panorama and the back surface panorama into the two-layer panorama.

[0019] Fusing the front surface panorama and the back surface panorama may comprise:

removing background pixels from the back surface panorama that match corresponding foreground pixels in the front surface panorama;

storing connections between neighboring pixels meeting a threshold similarity in depth and color information; and

hallucinating color and depth information in missing pixel locations.

[0020] The instructions when executed by processor further may cause the processor to perform steps including:

generating a normal map for the multi-layered geometric mesh, the normal map estimating for each pixel, an angle normal to a surface depicted by the pixel.

[0021] Generating the normal map may comprise:

generating a base normal map from depth values in the three-dimensional image;
generating a detailed normal from luminance values in the three-dimensional image; and
transforming the detailed normal map onto the base normal map to generate a combined normal map.

[0022] The plurality of input images may have varying levels of overlap and orientation changes.

[0023] In an embodiment according to the invention, a system may comprise:

a processor; and

a non-transitory computer-readable storage medium storing instruction for generating a three-dimensional image, the instructions when executed by a processor causing the processor to perform steps including:

receiving a plurality of input images of a scene taken from different vantage points;

processing the plurality of input images to generate a sparse reconstruction representation of the scene, the sparse reconstruction representation including a sparse point cloud specifying locations of a plurality of points that correspond to three-dimensional locations of surfaces of objects in the scene;

based in part on the sparse reconstruction representation, processing the plurality of input images to generate respective dense reconstruction representations of each of the plurality of input images, the respective dense reconstruction representations each including respective depth images for the plurality of input images, the depth images including color and depth information;

projecting front surfaces of the depth images using a forward depth test to generate a plurality of front-warped images;

projecting back surfaces of the depth images using an inverted depth test to generate a plurality of back-warped images;

stitching the front-warped images and the back-warped images to generate a two-layer panorama having a front surface panorama and a back surface panorama;

fusing the front surface panorama and the back surface panorama in the two-layer panorama to generate the three-dimensional image comprising a multi-layered geometric mesh suitable for rendering the scene in the three-dimensional space.

[0024] Stitching the front-warped images and the back-warped images to generate a two-

layer panorama may comprise:

stitching a depth panorama using depth values from the front-warped images;

stitching the front surface panorama using color values from the front-warped images and stitched depth values from the depth panorama;

stitching the back surface panorama using color values from the back-warped images and the stitched depth values from the depth panorama; and

combining the front surface panorama and the back surface panorama into the two-layer panorama.

[0025] In an embodiment according to the invention, one or more computer-readable non-transitory storage media may embody software that is operable when executed to perform a method according to the invention or any of the above mentioned embodiments.

[0026] In an embodiment according to the invention, a system may comprise: one or more processors; and at least one memory coupled to the processors and comprising instructions executable by the processors, the processors operable when executing the instructions to perform a method according to the invention or any of the above mentioned embodiments.

[0027] In an embodiment according to the invention, a computer program product, preferably comprising a computer-readable non-transitory storage media, may be operable when executed on a data processing system to perform a method according to the invention or any of the above mentioned embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] FIG. 1 is a block diagram illustrating an example embodiment of a system for generating a three-dimensional image from a plurality of two-dimensional input images.

[0029] FIG. 2 is a flowchart illustrating an embodiment of a process for generating a sparse reconstruction representation of a scene.

[0030] FIG. 3 is a flowchart illustrating an embodiment of a process for generating a dense reconstruction representation of a scene.

[0031] FIG. 4 is a flowchart illustrating an embodiment of a process for fusing a plurality of depth images into a multi-layer panoramic image.

[0032] FIG. 5 is a flowchart illustrating an embodiment of a process for generating a normal map for a multi-layer panoramic image.

[0033] The figures depict embodiments of the present disclosure for purposes of illustration only. One skilled in the art will readily recognize from the following description

that alternative embodiments of the structures and methods illustrated herein may be employed without departing from the principles, or benefits touted, of the disclosure described herein.

DETAILED DESCRIPTION

OVERVIEW

[0034] To enable better sharing and preservation of immersive experiences, a graphics system reconstructs a three-dimensional scene from a set of images of the scene taken from different vantage points. The system processes each image to extract depth information therefrom and then stitches the images (both color and depth information) into a multi-layered panorama that includes at least front and back surface layers. The front and back surface layers are then merged to remove redundancies and create connections between neighboring pixels that are likely to represent the same object, while removing connections between neighboring pixels that are not. The resulting layered panorama with depth information can be rendered using a virtual reality (VR) system, a mobile device, or other computing and display platforms using standard rendering techniques, to enable three-dimensional viewing of the scene.

SYSTEM ARCHITECTURE

[0035] FIG. 1 illustrates a system for reconstructing a three-dimensional scene from a set of images, in accordance with one embodiment. As depicted, an image capture system 110 (e.g., a camera) is used to take a set of images 115 from different viewing positions in a scene and outputs the images 115 to a three-dimensional (3D) photo reconstruction system 120. The three-dimensional photo reconstruction system 120 processes the images 115 to generate a three-dimensional renderable panoramic image 125. The three-dimensional renderable panoramic image 125 is outputted to a three-dimensional renderer 130 that renders a three-dimensional image for display. In alternative embodiments, additional intermediate components may be included that are not expressly shown in FIG. 1. For example, an image storage (e.g., a database) may be included between the image capture system 110 and the three-dimensional photo reconstruction system 120 to store images 115 until selected for processing. Similarly, an image storage may exist between the three-dimensional photo reconstruction system 120 and the three-dimensional renderer 130 to store the renderable panoramic image 125. Additionally, conventional networking components may facilitate communications between the various systems 110, 120, 130 or any intermediate components (e.g., storage systems). Each of the illustrated systems 110, 120, 130 may include one or

more processors and a computer-readable storage medium that stores instructions that when executed cause the respective systems to carry out the processes and functions attributed to the systems 110, 120, 130 described herein.

[0036] The image capture system 110 may comprise any system capable of taking a set of images of a scene, such as a standalone consumer camera or a camera built into a phone or other mobile device. The image capture system 110 may also include a camera that is wearable on a human, such as outward facing cameras in a virtual reality (VR) or augmented reality (AR) headset.

[0037] The image capture system 110 may capture the set of images in an unstructured manner. For example, the particular positions from which each image is captured do not need to be at precise known positions and the images may have varying amounts of overlap or differences in orientation. Furthermore, the number of images captured may be arbitrary within reasonable upper and lower bounds. In one example, the set of images may be taken by moving the image capture system 110 sideways (e.g., at half arm's length) while taking a series of still images. In another embodiment, the images may be taken by multiple different cameras positioned at different locations. Generally, to form a cohesive representation of the captured scene, at least some of the images may overlap with one or more other images. Thus, the images captured by the image capture system 110 may be captured quickly in a simple user-friendly manner without requiring special equipment or specialized knowledge about how to position the camera.

[0038] The image capture system 110 may also include software coupled to the camera that assists the user in capturing images, such as by guiding the user to take a sufficient number of images. To this end, the software may use motion or position sensors in the image capture system to record approximate relative positions of the capture for each image and possibly to help guide the user when capturing the images.

[0039] The image capture system 110 may furthermore include software for selecting a set of images that are suitable for processing by the three-dimensional reconstruction system 120 from a larger collection of images. For example, in one embodiment, images that are intended for re-constructing into a three-dimensional renderable image may be tagged with metadata to indicate that they were captured as part of an image capture sequence specifically for this purpose. Alternatively, the image capture system 110 may selectively determine images that are suitable for reconstruction based on a number of criteria, such as being captured within a time threshold or within physical proximity to a certain location. The images may be selected from discrete images or from frames of video captured by the image

capture system 110.

[0040] The image capture system 110 may include multiple cameras that may be operated by separate users to capture images of a scene. The images may be stored to a common database and may be processed to select suitable images in the same way that images from a single camera may be processed. For example, in one scenario, multiple users may upload images so that several users can share images captured by others and reconstruct the scene using the shared images. In a social networking environment, social connections may be used to determine which images are available to others. For example, a user may set a privacy setting to indicate which other users can access and use their images, which may be set based on social connections between the users.

[0041] Before the set of images is used by the three-dimensional photo reconstruction system, one or more pre-processing tasks may be performed on the images. For example, the images may be transcoded into a predefined format and file quality.

[0042] The set of images is then provided to the three-dimensional (3D) photo reconstruction system 120, which converts the set of images into a three-dimensional renderable panoramic image with multiple layers and depth information. The three-dimensional photo reconstruction system 120 includes several functional modules for performing this conversion task, including a sparse reconstruction module 122, a dense reconstruction module 124, a two-layer fusion module 126, and a normal map estimation module 128. The sparse reconstruction module 122 processes the input images 115 to generate a sparse reconstruction representation of the scene. The dense reconstruction module 124 then uses the sparse reconstruction representation to generate depth maps for each of the input images 115. The dense reconstruction module 124 may apply a modified multi-view stereo algorithm that results in improved depth estimations relative to conventional algorithms executing in this context. Once the depth information is densified by the dense reconstruction module 124, the depth images including the original set of images and their corresponding depth maps are provided to the two-layer fusion module 126. The two-layer fusion module 126 merges the depth images into a multi-layer panorama comprising, for example, a two-layer panoramic mesh that is renderable by the three-dimensional renderer 130. The normal map estimation module 128 estimates normals for surfaces of the two-layer panoramic mesh that can be used in by the renderer 130 to add various effects such as lighting or flooding effects to the rendered image. Processes performed by each of these modules are described in further detail below.

[0043] Once the three-dimensional renderable panoramic image is constructed, it may be

rendered for viewing by the three-dimensional renderer 130. The three-dimensional renderer 130 may comprise, for example, a virtual reality (VR) system, a mobile device, or other computing and display platforms that can display rendered three-dimensional content using standard rendering techniques.

THREE-DIMENSIONAL PHOTO RECONSTRUCTION

Sparse Reconstruction

[0044] FIG. 2 illustrates an embodiment of a process performed by the sparse reconstruction module 122 for generating a sparse reconstruction representation from a plurality of input images. A plurality of input images are received 202. The input images comprise two-dimensional images that may be pre-selected as images that are suitable for reconstruction into the three-dimensional image. Generally, the input images include a set of images of a scene captured from different vantage points, with at least some of the input images overlapping other images of the scene.

[0045] The sparse reconstruction module applies 204 a structure-from-motion algorithm to the input images that reconstructs three-dimensional structure from its projections into the set of input images. An example of structure-from-motion algorithm that may be suitable for this purpose is COLMAP. From the result of the structure-from-motion algorithm, the sparse reconstruction module 122 outputs the sparse reconstruction representation that may include a set of camera poses for each image and a point cloud approximation of the scene depicted in the images. Here, the set of camera poses may include, for each image, intrinsic camera parameters such as focal length, image sensor format, and principal point, and extrinsic camera parameters denoting the position and orientation of the camera (relative to the scene) when the image was captured. [The description of the Dense Reconstruction does not state how the camera poses are used. Are these used for dense reconstruction or elsewhere, and can you add a brief explanation below in the steps where these are used?] The point cloud comprises a sparse set of data points in a three-dimensional space that represent external surfaces of objects detected in the scene.

Dense Reconstruction

[0046] FIG. 3 illustrates an embodiment of a process performed by the dense reconstruction module 124 for generating the dense reconstruction representations using a near envelope prior. The dense reconstruction representation provided by the process comprises a plurality of depth images corresponding to the original input images. The depth images include the color information from the respective original input images and include a

depth map for each image that indicates, for each pixel, an estimated distance to a surface of an object in the scene depicted by that pixel given the viewpoint of the image.

[0047] In an embodiment, generating the dense reconstruction representation may be achieved by first computing 310 the near envelope prior and then applying 320 a modified plane-sweep multi-view stereo (MVS) algorithm that includes the near envelope reconstruction prior as an additional cost function. In conventional MVS algorithms, near-depth hypotheses may be noisy because these points are seen in fewer images and the photo consistency measure is therefore less reliable. This unreliability makes existing MVS algorithms more likely fall victim to common stereo pitfalls, such as repeated structures in the scene, slight scene motion, or materials with view-dependent (shiny) appearance. The near envelope reconstruction prior addresses these limitations as described below.

[0048] The MVS algorithm treats depth estimation as an energy minimization problem, which optimizes the pixel depths d_i by solving the following problem:

$$\underset{d}{\operatorname{argmin}} \sum_i E_{data}(i) + \lambda_{smooth} \sum_{(i,j) \in \mathcal{N}} E_{smooth}(i,j), \quad (1)$$

[0049] where i is a pixel and c_i is its color. The smoothness term E_{smooth} is the product of a color- and a depth-difference cost,

$$E_{smooth}(i,j) = w_{color}(c_i, c_j) w_{depth}(d_i, d_j), \quad (2)$$

[0050] where w_{color} is a color difference cost between color c_i and c_j and w_{depth} is a depth difference cost between depths d_i and d_j . The smoothness term E_{smooth} encourages the depth map to be smooth wherever the image lacks texture. The data term combines three costs:

$$E_{data}(i) = E_{photo}(i) + E_{sky}(i) + E_{sfm}(i). \quad (3)$$

[0051] The photo-consistency term E_{photo} measures the agreement in appearance between a pixel and its projection into multiple other images. The sky prior E_{sky} encourages large depths for pixels that are classified by a sky detector as likely be part of sky. The structure from motion prior E_{sfm} encourages the result to stay close to the sparse reconstruction.

[0052] The dense reconstruction module 124 discretizes the potential depth labels and uses the plane-sweep stereo algorithm to build a cost volume with depth hypotheses for each pixel. While this restricts the algorithm to reconstructing discrete depths without normal, it has the advantage that it can extract a globally optimized solution using a Markov random field (MRF) solver, which can often recover plausible depth for textureless regions using its smoothness term. The dense reconstruction module 124 optimizes the MRF at a reduced

resolution, e.g., using a FastPD library for performance reasons. The dense reconstruction module 124 then upscales the result to full resolution with a joint bilateral upsampling filter, e.g., using a weighted median filter instead of averaging to prevent introducing erroneous middle values at depths discontinuities.

[0053] The near envelope is an estimate of a conservative but tight lower bound n_i for the pixel depths at each pixel. This boundary is used to discourage nearby erroneous depths by augmenting the data term ($E_{data}(i)$) in Eq. (3) to include the following additional cost term in the sum:

$$E_{near}(i) = \begin{cases} \lambda_{near} & \text{if } d_i < n_i \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

[0054] where λ_{near} is a predefined parameter (e.g., $(\lambda_{near} = 1)$). The additional cost term E_{near} penalizes reconstructing depths closer than the near envelope, thus reducing or eliminating erroneously low depth estimates in the depth maps.

[0055] To compute the near envelope n the dense reconstruction module 124 first identifies 302 anchor pixels with reliable depth estimates. The dense reconstruction module 124 computes the anchors from two sources: (1) stereo matching where it is reliable, and (2) the point features already computed in the point cloud during the sparse reconstruction. Stereo matching is known to work well at strong image gradients. This is used by computing a lightweight MVS result and carefully considering the pixels around image edges for anchors. The dense reconstruction module 124 obtains a noisy depth map by independently computing the minimum of E_{data} for each pixel, e.g., dropping the smoothness term from Eq. 1. The dense reconstruction module 124 uses a geometric consistency filter with aggressive settings to discard a significant portion of the incorrectly estimated depth values. An edge detector is used to compute an edge mask. If the image edge coincides with a depth edge, it is desirable to ensure that the depth belonging to the front layer is selected. This may be achieved by dilating the edge mask by 1 pixel and applying a 5x5 morphological minimum filter to the masked pixels inside the detected edges. In addition to these anchors computed from stereo, the dense reconstruction module 124 also adds all the sparse point features that are observed in the image to the set of anchor pixels.

[0056] The dense reconstruction module 124 propagates 304 the depths of the anchor pixels to the remaining pixels to generate an approximate depth map. The sparse anchor depths are spread to the remaining pixels by solving a first-order Poisson system, similar to the one used in the colorization algorithm:

$$\operatorname{argmin}_x \sum_i w_i (x_i - x'_i)^2 + \sum_{(i,j) \in \mathcal{N}} w_{ij} (x_i - x_j)^2, \quad (5)$$

[0057] where x'_i are the depths of the anchor pixels (where defined), and x_i are the densely propagated depths solved for, $w_{ij} = e^{-\{c_i - c_j\}^2 / 2\sigma_{\text{c}}^2}$ is the color-based affinity term, and w_i is a unary weight term that represents confidence in each anchor:

$$w_i = \begin{cases} 0.5 & \text{if } x_i \text{ is a feature observed in 2 images,} \\ 2 & \text{if } x_i \text{ is a feature observed in 3 images,} \\ 10 & \text{if } x_i \text{ is a feature in } \geq 4 \text{ images, or an edge depth.} \end{cases} \quad (6)$$

[0058] The dense reconstruction module 124 then filters 306 the approximate depth maps to determine the near envelope. For example, the dense reconstruction module 124 makes the propagated depths more conservative by multiplying them with a constant factor (e.g., 0.6) and subsequently applying a morphological minimum filter with diameter set to about 10% of the image diagonal. They are further cleaned up by smoothing with a wide Gaussian kernel (σ set to about 5% of the diagonal). The near envelope prior $E_{\text{near}}(i)$ is then computed 308 from the near envelope n based on Eq. (4) above.

Two-Layer Fusion

[0059] FIG. 4 illustrates an embodiment of a process performed by the two-layer fusion module 126 to generate the two-layer panorama. Here, the two-layer fusion module 126 warps each depth map into a central panoramic image (using equirectangular projection) by triangulating it into a grid-mesh and rendering it with the normal rasterization pipeline, letting the depth test select the front surface when multiple points fall onto the same panorama. One problem with a simple approach to this task is that long stretched triangles at depth discontinuities connecting foreground and background pixels might obscure other good content, and it is undesirable to include them in the stitch in any case. The problem can be resolved by blending the z -values of the rasterized fragments with a stretch penalty $s \in [0, 1]$ before the depth test, $z' = (z + s) / 2$. The division by 2 keeps the value z' in normalized clipping space. The stretch penalty,

$$s = 1 - \min\left(\frac{\alpha}{\tau_{\text{stretch}}}, 1\right) \quad (7)$$

considers the grazing angle α from the original viewpoint and penalizes small values below τ_{stretch} (e.g., $\tau_{\text{stretch}} = 1.66^\circ$), i.e. rays that are nearly parallel to the triangle surface. This modification pushes highly stretched triangles back, so potentially less stretched back

surfaces can win over instead.

[0060] Since the system operates to reconstruct not just the first visible surface, but several depth layers, it generates a second back surface warp for each image. One possible way to generate these is depth peeling; however, this method is best suited for systems that can depend on having very accurate depth maps available. More robust results with less reliable depth maps can be achieved, instead, by assuming a depth complexity of two and projecting surfaces using different depth tests prior to stitching. [What is meant by “depth tests” here? Does this just dictate whether you choose the front or back depth values in overlapping regions of the stitch where a pixel has two different depth values?]

[0061] In the process of FIG. 4, the front-most surfaces of each depth map are first projected 402 into an equirectangular projection using z' for the depth test to generate a front-warped image for each of the depth images. The back-most surfaces of each depth map are projected 404 into an equirectangular projection using $z'' = 1 - z'$ for the depth test, effectively inverting it.

[0062] Once the depth maps are warped into a common perspective (or camera pose), the two-layer fusion module 126 can stitch the warped images into front and back surface panoramas. For scenes where exposure brackets are captured, this is where all the exposures are fused. The colors are linearized using a naive gamma = 2.2 assumption, and multiplied with the appropriate exposure factor.

[0063] The stitching described above may be performed by solving a discrete pixel labeling problem, where each pixel i in the panorama chooses the label α_i from one of the warped sources. The labels may be chosen by minimizing the energy of a cost function that may depend on, for example, a stereo confidence weight, a triangle stretch weight, an exposure penalty, and a depth penalty, a pairwise color match, and a depth disparity.

[0064] For example, in one embodiment, the two-layer fusion module 126 stitches 406 the depth maps of the front-warped depth images to generate a depth-only panorama. To stitch the depth panorama, the two-layer fusion module 126 optimizes this objective:

$$\underset{\{\alpha\}}{\operatorname{argmin}} \underbrace{\lambda_1 E_{\text{stereo}} + \lambda_2 E_{\text{stretch}}}_{\text{Data terms}} + \underbrace{\lambda_3 E_{\text{color}} + \lambda_4 E_{\text{disp}}}_{\text{Smoothness terms}} \quad (8)$$

[0065] The stereo data term encourages selecting pixels whose depth was estimated with high confidence. The two-layer fusion module 126 uses the maximum-likelihood measure of:

$$E_{\text{stereo}} = \sum_i -\log MLM_i^{\alpha_i}, \quad (9)$$

[0066] but only for pixels that pass the geometric consistency test. For inconsistent pixels the two-layer fusion module 126 sets $MLM_i^{ai} = 0.001$. The triangle stretch term discourages selecting pixels from long “rubber sheet” triangles:

$$E_{stretch} = \sum_i -\log s_i^{ai}. \quad (10)$$

[0067] The color smoothness term is a truncated version of the seam-hiding pairwise cost from “GraphCut Textures”:

$$E_{color} = \sum_{i,j \in \mathcal{N}} \min(\|c_i^{ai} - c_j^{aj}\|_2^2, \tau_c) + \min(\|c_j^{aj} - c_i^{ai}\|_2^2, \tau_c), \quad (11)$$

[0068] and the disparity smoothness term is a similar truncated term:

$$E_{disp} = \sum_{i,j \in \mathcal{N}} \min\left(\left|\frac{1}{d_i^{ai}} - \frac{1}{d_j^{aj}}\right|, \tau_d\right) + \min\left(\left|\frac{1}{d_j^{aj}} - \frac{1}{d_i^{ai}}\right|, \tau_d\right). \quad (12)$$

[0069] After having obtained a depth stitch by solving Eq. 8, the two-layer fusion module 126 uses it to constrain the subsequent front and back color stitches. Here, the two-layer fusion module 126 stitches 408 the front-warped images using the color values into a foreground panorama. The two-layer fusion module 126 stitches 410 the back-warped images into a background panorama. In performing these stitches, the two-layer fusion module 126 adds two additional data terms that constrain color exposure and depth selection:

$$E_{exposure} = \lambda_5 \sum_i \mu_i^{ai}, \quad E_{depth} = \lambda_6 \sum_i v_i^{ai}. \quad (13)$$

[0070] All penalties μ_i^{ai} , v_i^{ai} are 0, except in the following conditions. The exposure penalty μ_i^{ai} is 1 if a pixel is over-exposed (except in the darkest bracket) or under-exposed (except in the brightest bracket). The depth penalty v_i^{ai} can be set differently for the front and back color stitch. For example, for the front color stitch, it is set to 1 only if a pixel’s depth is not within a factor of [0.95, 1/0.95] of the depth stitch. For the back color stitch, it is set to 1 only if a pixel’s depth is less than a factor of 1/0.95 of the depth stitch.

[0071] In an embodiment, the two-layer fusion module 126 sets the following balancing coefficients:

$$\lambda_1 = 5, \quad \lambda_2 = 50, \quad \lambda_3 = \lambda_4 = 25, \quad \lambda_5 = 100, \quad \lambda_6 = 75, \quad (14)$$

and solves the labeling problems at a reduced 512x256 resolution using an alpha expansion algorithm, and upsamples the resulting label map to full resolution (8192x4096 pixels) using a simple PatchMatch-based upsampling algorithm.

[0072] The stitched depth panoramas occasionally still contain small floating elements.

The two-layer fusion module 126 “despeckles” them by first identifying strong depth discontinuities between neighboring pixels using a depth ratio test, then finding small disconnected components with fewer than 4096 pixels, and removing them by filling their depths in with a median filter. In addition, the depth map is smoothed with a joint bilateral filter whose kernel is cuts across the discontinuities computed above.

[0073] Once the front and back panoramas are generated, the two-layer fusion module 126 then fuses 412 the foreground panorama and the background panorama into the two-layer representation. To do this, the two-layer fusion module 126 represents the two-layer panorama as a graph. Each pixel i in the panorama has up to two nodes that represent the foreground and background layers. If they exist, they are denoted f_i and b_i respectively. Each node n has a depth value $d(n)$ and a foreground / background label $l(n) \in \{F, B\}$. The two-layer fusion module 126 generates for both front and back panoramas, fully 4-connected but disjoint grid-graphs independently. Each node is assigned a depth and label according to the panorama from which it is drawn. These graphs contain redundant coverage of some scene objects, but this is fully intentional and will be useful to remove color fringes around depth continuities. The two-layer fusion module 126 removes the redundancies by removing all the b_i nodes that are too similar to their f_i counterparts, i.e., $d(f_i)/d(b_i) < \tau_{\text{dratio}} = 0.75$.

[0074] The result is not redundant, but now the b graph contains many isolated components, and the f graph contains long connections across discontinuities. The connectivity is then recomputed. For each pair of neighboring pixels, the two-layer fusion module 126 considers all f and b nodes if they exist. The two-layer fusion module 126 also consider all pairs of nodes from i and j and sorts them by their depth ratio, most similar first. Then, it connects the most similar pair if the depth ratio is above τ_{dratio} . If there is another pair that can be connected, it is connected unless such a connection would cross an edge.

[0075] To this point, the two-layer fusion module 126 has generated a well-connected two-layer graph. The back layer contains some extra content, but large translations can still reveal holes. The back layer is next expanded to further hallucinating depth and color. The two-layer fusion module 126 expands the back layer in an iterative fashion, one pixel ring at a time. In each iteration, the two-layer fusion module 126 identifies b and f nodes that are not connected in one direction. It tentatively creates a new candidate neighbor node for these pixels and sets their depth and colors to the average of the nodes that spawned them. Candidate nodes are kept only if they are not colliding with already existing nodes (using τ_{dratio}), and if they become connected to the nodes that spawned them.

Normal Map Estimation

[0076] An optional step in the three-dimensional reconstruction pipeline is to compute a normal map. In one embodiment, the goal is merely to compute a normal that looks plausible and good enough for simple artist-driven lighting effects, rather than an accurate normal map of the scene. A process for generating the normal map is illustrated in FIG. 5. In this embodiment, the normal map estimation module 128 generates 502 a first (base) normal map from the depth values in the two-layer panorama that is accurate with respect to surface slopes but not textures on each surface. The normal map estimation module 128 also generates 504 a second (details) map from the luminance values that has surface details that are artistically-driven to produce desired effects in response to lighting. The normal map estimation module 128 then transforms 506 the second normal map (with the texture details) onto the first normal map (with accurate orientation) to get a combined normal map that serves as a good approximation for the scene.

[0077] In one embodiment, the base normal map is piece-wise smooth but discontinuous at depth edges, and contain the correct surface slopes. This normal map is computed by filtering the depth normals with a guided filter, guided by the depth map. A wide window size may be used that corresponds to about a solid angle of about 17.5 degrees.

[0078] To generate the details map, the normal map estimation module 128 hallucinates from the luminance image, estimating a normal map by hallucinating a depth map just from the image data, assuming surface depth is inversely related to image intensity. While the depth generated in this manner is highly approximate, it is fully consistent with the image data and provides a surprisingly effective means for recovering geometric detail variations.

[0079] The two normal maps are then combined. For a given pixel with polar and azimuthal angles (θ, ϕ) , n_f is the normal obtained by applying the guided filter to the depth normals, and n_i is the normal obtained from the image-based estimation method. R_s is a local coordinate frame for the image-based normal. It is obtained by setting the first row to a vector pointing radially outward,

$$R_{s,0} = (\sin\theta \cos\phi, \cos\theta, \sin\theta \sin\phi), \quad (15)$$

[0080] and the other rows through cross products with the world up vector w_{up} .

$$R_{s,1} = \frac{R_{s,0} \times w_{up}}{\|R_{s,0} \times w_{up}\|}, \quad R_{s,2} = R_{s,0} \times R_{s,1} \quad (16)$$

[0081] A similar coordinate frame R_f is defined for the filtered depth normal, by setting $R_{f,0} = -n_f$ and the other rows analog to Eq. 16. Then, the normal map estimation module 128 transfers the details as follows:

$$n_c = R_f^{-1} R_s n_i \quad (17)$$

[0082] The resulting normal map can then be provided with the two-layer mesh to a 3D rendering system 130 for viewing by another user.

SUMMARY

[0083] The foregoing description of the embodiments of the invention has been presented for the purpose of illustration; it is not intended to be exhaustive or to limit the invention to the precise forms disclosed. Persons skilled in the relevant art can appreciate that many modifications and variations are possible in light of the above disclosure.

[0084] Some portions of this description describe the embodiments of the invention in terms of algorithms and symbolic representations of operations on information. These algorithmic descriptions and representations are commonly used by those skilled in the data processing arts to convey the substance of their work effectively to others skilled in the art. These operations, while described functionally, computationally, or logically, are understood to be implemented by computer programs or equivalent electrical circuits, microcode, or the like. Furthermore, it has also proven convenient at times, to refer to these arrangements of operations as modules, without loss of generality. The described operations and their associated modules may be embodied in software, firmware, hardware, or any combinations thereof.

[0085] Any of the steps, operations, or processes described herein may be performed or implemented with one or more hardware or software modules, alone or in combination with other devices. In one embodiment, a software module is implemented with a computer program product comprising a computer-readable medium containing computer program code, which can be executed by a computer processor for performing any or all of the steps, operations, or processes described.

[0086] Embodiments of the invention may also relate to an apparatus for performing the operations herein. This apparatus may be specially constructed for the required purposes, and/or it may comprise a general-purpose computing device selectively activated or reconfigured by a computer program stored in the computer. Such a computer program may be stored in a tangible computer readable storage medium or any type of media suitable for storing electronic instructions, and coupled to a computer system bus. Furthermore, any computing systems referred to in the specification may include a single processor or may be architectures employing multiple processor designs for increased computing capability.

[0087] Embodiments of the invention may also relate to a computer data signal embodied

in a carrier wave, where the computer data signal includes any embodiment of a computer program product or other data combination described herein. The computer data signal is a product that is presented in a tangible medium or carrier wave and modulated or otherwise encoded in the carrier wave, which is tangible, and transmitted according to any suitable transmission method.

[0088] Finally, the language used in the specification has been principally selected for readability and instructional purposes, and it may not have been selected to delineate or circumscribe the inventive subject matter. It is therefore intended that the scope of the invention be limited not by this detailed description, but rather by any claims that issue on an application based hereon.

What is claimed is:

1. A method comprising:

receiving a plurality of input images of a scene taken from different vantage points;
processing the plurality of input images to generate a sparse reconstruction representation of the scene, the sparse reconstruction representation including a sparse point cloud specifying locations of a plurality of points that correspond to three-dimensional locations of surfaces of objects in the scene;
based in part on the sparse reconstruction representation, processing the plurality of input images to generate respective dense reconstruction representations of each of the plurality of input images, the respective dense reconstruction representations each including respective depth images for the plurality of input images, the depth images including color and depth information;
projecting front surfaces of the depth images using a forward depth test to generate a plurality of front-warped images;
projecting back surfaces of the depth images using an inverted depth test to generate a plurality of back-warped images;
stitching the front-warped images and the back-warped images to generate a two-layer panorama having a front surface panorama and a back surface panorama; and
fusing the front surface panorama and the back surface panorama in the two-layer panorama to generate the three-dimensional image comprising a multi-layered geometric mesh suitable for rendering the scene in the three-dimensional space.

2. The method of claim 1, wherein processing the plurality of input images to generate the sparse reconstruction representation comprises:

applying a surface-from-motion algorithm to the plurality of input images.

3. The method of claim 1, wherein processing the plurality of input images to generate the respective dense reconstruction representations comprises:

generating a near envelope prior that assigns a cost to estimated depth values in front of a near envelope; and

applying a multi-view stereo processing algorithm to estimate the depth values based on a cost function including the near envelope prior.

4. The method of claim 3, wherein generating the near envelope prior comprises:

identifying anchor pixels in the plurality of input images that have high confidence depth estimates;

- propagating the depth estimates of the anchor pixels to other pixels in the plurality of input images to generate approximate depth maps;
- filtering the approximate depth maps to determine a near envelope; and
- generating the near envelope prior based on the depth estimates and the near envelope.
5. The method of claim 1, wherein stitching the front-warped images and the back-warped images to generate a two-layer panorama comprises:
- stitching a depth panorama using depth values from the front-warped images;
- stitching the front surface panorama using color values from the front-warped images and stitched depth values from the depth panorama;
- stitching the back surface panorama using color values from the back-warped images and the stitched depth values from the depth panorama; and
- combining the front surface panorama and the back surface panorama into the two-layer panorama.
6. The method of claim 1, wherein fusing the front surface panorama and the back surface panorama comprises:
- removing background pixels from the back surface panorama that match corresponding foreground pixels in the front surface panorama;
- storing connections between neighboring pixels meeting a threshold similarity in depth and color information; and
- hallucinating color and depth information in missing pixel locations.
7. The method of claim 1, further comprising:
- generating a normal map for the multi-layered geometric mesh, the normal map estimating for each pixel, an angle normal to a surface depicted by the pixel.
8. The method of claim 7, wherein generating the normal map comprises:
- generating a base normal map from depth values in the three-dimensional image;
- generating a detailed normal from luminance values in the three-dimensional image;
- and
- transforming the detailed normal map onto the base normal map to generate a combined normal map.
9. The method of claim 1, wherein the plurality of input images having varying levels of overlap and orientation changes.
10. A non-transitory computer-readable storage medium storing instructions, the instructions when executed by a processor causing the processor to perform steps including:

receiving a plurality of input images of a scene taken from different vantage points;
processing the plurality of input images to generate a sparse reconstruction
representation of the scene, the sparse reconstruction representation including
a sparse point cloud specifying locations of a plurality of points that
correspond to three-dimensional locations of surfaces of objects in the scene;
based in part on the sparse reconstruction representation, processing the plurality of
input images to generate respective dense reconstruction representations of
each of the plurality of input images, the respective dense reconstruction
representations each including respective depth images for the plurality of
input images, the depth images including color and depth information;
projecting front surfaces of the depth images using a forward depth test to generate a
plurality of front-warped images;
projecting back surfaces of the depth images using an inverted depth test to generate a
plurality of back-warped images;
stitching the front-warped images and the back-warped images to generate a two-layer
panorama having a front surface panorama and a back surface panorama;
fusing the front surface panorama and the back surface panorama in the two-layer
panorama to generate the three-dimensional image comprising a multi-layered
geometric mesh suitable for rendering the scene in the three-dimensional
space.

11. The non-transitory computer-readable storage medium of claim 10, wherein processing the plurality of input images to generate the sparse reconstruction representation comprises:
applying a surface-from-motion algorithm to the plurality of input images.
12. The non-transitory computer-readable storage medium of claim 10, wherein processing the plurality of input images to generate the respective dense reconstruction representations comprises:
generating a near envelope prior that assigns a cost to estimated depth values in front of a near envelope; and
applying a multi-view stereo processing algorithm to estimate the depth values based on a cost function including the near envelope prior.
13. The non-transitory computer-readable storage medium of claim 12, wherein generating the near envelope prior comprises:

identifying anchor pixels in the plurality of input images that have high confidence depth estimates;
propagating the depth estimates of the anchor pixels to other pixels in the plurality of input images to generate approximate depth maps;
filtering the approximate depth maps to determine a near envelope; and
generating the near envelope prior based on the depth estimates and the near envelope.

14. The non-transitory computer-readable storage medium of claim 10, wherein stitching the front-warped images and the back-warped images to generate a two-layer panorama comprises:
stitching a depth panorama using depth values from the front-warped images;
stitching the front surface panorama using color values from the front-warped images and stitched depth values from the depth panorama;
stitching the back surface panorama using color values from the back-warped images and the stitched depth values from the depth panorama; and
combining the front surface panorama and the back surface panorama into the two-layer panorama.
15. The non-transitory computer-readable storage medium of claim 10, wherein fusing the front surface panorama and the back surface panorama comprises:
removing background pixels from the back surface panorama that match corresponding foreground pixels in the front surface panorama;
storing connections between neighboring pixels meeting a threshold similarity in depth and color information; and
hallucinating color and depth information in missing pixel locations.
16. The non-transitory computer-readable storage medium of claim 10, wherein the instructions when executed by processor further cause the processor to perform steps including:
generating a normal map for the multi-layered geometric mesh, the normal map estimating for each pixel, an angle normal to a surface depicted by the pixel.
17. The non-transitory computer-readable storage medium of claim 16, wherein generating the normal map comprises:
generating a base normal map from depth values in the three-dimensional image;
generating a detailed normal from luminance values in the three-dimensional image;
and

transforming the detailed normal map onto the base normal map to generate a combined normal map.

18. The non-transitory computer-readable storage medium of claim 10, wherein the plurality of input images having varying levels of overlap and orientation changes.
19. A system comprising:
 - a processor; and
 - a non-transitory computer-readable storage medium storing instruction for generating a three-dimensional image, the instructions when executed by a processor causing the processor to perform steps including:
 - receiving a plurality of input images of a scene taken from different vantage points;
 - processing the plurality of input images to generate a sparse reconstruction representation of the scene, the sparse reconstruction representation including a sparse point cloud specifying locations of a plurality of points that correspond to three-dimensional locations of surfaces of objects in the scene;
 - based in part on the sparse reconstruction representation, processing the plurality of input images to generate respective dense reconstruction representations of each of the plurality of input images, the respective dense reconstruction representations each including respective depth images for the plurality of input images, the depth images including color and depth information;
 - projecting front surfaces of the depth images using a forward depth test to generate a plurality of front-warped images;
 - projecting back surfaces of the depth images using an inverted depth test to generate a plurality of back-warped images;
 - stitching the front-warped images and the back-warped images to generate a two-layer panorama having a front surface panorama and a back surface panorama;
 - fusing the front surface panorama and the back surface panorama in the two-layer panorama to generate the three-dimensional image comprising a multi-layered geometric mesh suitable for rendering the scene in the three-dimensional space.

20. The system of claim 19, wherein stitching the front-warped images and the back-warped images to generate a two-layer panorama comprises:
- stitching a depth panorama using depth values from the front-warped images;
 - stitching the front surface panorama using color values from the front-warped images and stitched depth values from the depth panorama;
 - stitching the back surface panorama using color values from the back-warped images and the stitched depth values from the depth panorama; and
 - combining the front surface panorama and the back surface panorama into the two-layer panorama.
21. A method comprising:
- receiving a plurality of input images of a scene taken from different vantage points;
 - processing the plurality of input images to generate a sparse reconstruction representation of the scene, the sparse reconstruction representation including a sparse point cloud specifying locations of a plurality of points that correspond to three-dimensional locations of surfaces of objects in the scene;
 - based in part on the sparse reconstruction representation, processing the plurality of input images to generate respective dense reconstruction representations of each of the plurality of input images, the respective dense reconstruction representations each including respective depth images for the plurality of input images, the depth images including color and depth information;
 - projecting front surfaces of the depth images using a forward depth test to generate a plurality of front-warped images;
 - projecting back surfaces of the depth images using an inverted depth test to generate a plurality of back-warped images;
 - stitching the front-warped images and the back-warped images to generate a two-layer panorama having a front surface panorama and a back surface panorama; and
 - fusing the front surface panorama and the back surface panorama in the two-layer panorama to generate the three-dimensional image comprising a multi-layered geometric mesh suitable for rendering the scene in the three-dimensional space.
22. The method of claim 21, wherein processing the plurality of input images to generate the sparse reconstruction representation comprises:
- applying a surface-from-motion algorithm to the plurality of input images.

23. The method of claim 21 or 22, wherein processing the plurality of input images to generate the respective dense reconstruction representations comprises:
- generating a near envelope prior that assigns a cost to estimated depth values in front of a near envelope; and
 - applying a multi-view stereo processing algorithm to estimate the depth values based on a cost function including the near envelope prior;
- optionally, wherein generating the near envelope prior comprises:
- identifying anchor pixels in the plurality of input images that have high confidence depth estimates;
 - propagating the depth estimates of the anchor pixels to other pixels in the plurality of input images to generate approximate depth maps;
 - filtering the approximate depth maps to determine a near envelope; and
 - generating the near envelope prior based on the depth estimates and the near envelope.
24. The method of any of claims 21 to 23, wherein stitching the front-warped images and the back-warped images to generate a two-layer panorama comprises:
- stitching a depth panorama using depth values from the front-warped images;
 - stitching the front surface panorama using color values from the front-warped images and stitched depth values from the depth panorama;
 - stitching the back surface panorama using color values from the back-warped images and the stitched depth values from the depth panorama; and
 - combining the front surface panorama and the back surface panorama into the two-layer panorama.
25. The method of any of claims 21 to 24, wherein fusing the front surface panorama and the back surface panorama comprises:
- removing background pixels from the back surface panorama that match corresponding foreground pixels in the front surface panorama;
 - storing connections between neighboring pixels meeting a threshold similarity in depth and color information; and
 - hallucinating color and depth information in missing pixel locations.
26. The method of any of claims 21 to 25, further comprising:
- generating a normal map for the multi-layered geometric mesh, the normal map estimating for each pixel, an angle normal to a surface depicted by the pixel;
- optionally wherein generating the normal map comprises:

generating a base normal map from depth values in the three-dimensional image;
generating a detailed normal from luminance values in the three-dimensional image;
and
transforming the detailed normal map onto the base normal map to generate a
combined normal map.

27. The method of any of claims 21 to 26, wherein the plurality of input images having varying levels of overlap and orientation changes.
28. A non-transitory computer-readable storage medium storing instructions, the instructions when executed by a processor causing the processor to perform steps including:
receiving a plurality of input images of a scene taken from different vantage points;
processing the plurality of input images to generate a sparse reconstruction representation of the scene, the sparse reconstruction representation including a sparse point cloud specifying locations of a plurality of points that correspond to three-dimensional locations of surfaces of objects in the scene;
based in part on the sparse reconstruction representation, processing the plurality of input images to generate respective dense reconstruction representations of each of the plurality of input images, the respective dense reconstruction representations each including respective depth images for the plurality of input images, the depth images including color and depth information;
projecting front surfaces of the depth images using a forward depth test to generate a plurality of front-warped images;
projecting back surfaces of the depth images using an inverted depth test to generate a plurality of back-warped images;
stitching the front-warped images and the back-warped images to generate a two-layer panorama having a front surface panorama and a back surface panorama;
fusing the front surface panorama and the back surface panorama in the two-layer panorama to generate the three-dimensional image comprising a multi-layered geometric mesh suitable for rendering the scene in the three-dimensional space.
29. The non-transitory computer-readable storage medium of claim 28, wherein processing the plurality of input images to generate the sparse reconstruction representation comprises:
applying a surface-from-motion algorithm to the plurality of input images.

30. The non-transitory computer-readable storage medium of claim 28 or 29, wherein processing the plurality of input images to generate the respective dense reconstruction representations comprises:
- generating a near envelope prior that assigns a cost to estimated depth values in front of a near envelope; and
 - applying a multi-view stereo processing algorithm to estimate the depth values based on a cost function including the near envelope prior;
- optionally wherein generating the near envelope prior comprises:
- identifying anchor pixels in the plurality of input images that have high confidence depth estimates;
 - propagating the depth estimates of the anchor pixels to other pixels in the plurality of input images to generate approximate depth maps;
 - filtering the approximate depth maps to determine a near envelope; and
 - generating the near envelope prior based on the depth estimates and the near envelope.
31. The non-transitory computer-readable storage medium of any of claims 28 to 30, wherein stitching the front-warped images and the back-warped images to generate a two-layer panorama comprises:
- stitching a depth panorama using depth values from the front-warped images;
 - stitching the front surface panorama using color values from the front-warped images and stitched depth values from the depth panorama;
 - stitching the back surface panorama using color values from the back-warped images and the stitched depth values from the depth panorama; and
 - combining the front surface panorama and the back surface panorama into the two-layer panorama.
32. The non-transitory computer-readable storage medium of any of claims 28 to 31, wherein fusing the front surface panorama and the back surface panorama comprises:
- removing background pixels from the back surface panorama that match corresponding foreground pixels in the front surface panorama;
 - storing connections between neighboring pixels meeting a threshold similarity in depth and color information; and
 - hallucinating color and depth information in missing pixel locations.

33. The non-transitory computer-readable storage medium of any of claims 28 to 32, wherein the instructions when executed by processor further cause the processor to perform steps including:
- generating a normal map for the multi-layered geometric mesh, the normal map
 - estimating for each pixel, an angle normal to a surface depicted by the pixel;
- optionally wherein generating the normal map comprises:
- generating a base normal map from depth values in the three-dimensional image;
 - generating a detailed normal from luminance values in the three-dimensional image;
 - and
 - transforming the detailed normal map onto the base normal map to generate a combined normal map.
34. The non-transitory computer-readable storage medium of any of claims 28 to 33, wherein the plurality of input images having varying levels of overlap and orientation changes.
35. A system comprising:
- a processor; and
 - a non-transitory computer-readable storage medium storing instruction for generating a three-dimensional image, the instructions when executed by a processor causing the processor to perform a method according to any of claims 21 to 27, in particular steps including:
 - receiving a plurality of input images of a scene taken from different vantage points;
 - processing the plurality of input images to generate a sparse reconstruction representation of the scene, the sparse reconstruction representation including a sparse point cloud specifying locations of a plurality of points that correspond to three-dimensional locations of surfaces of objects in the scene;
 - based in part on the sparse reconstruction representation, processing the plurality of input images to generate respective dense reconstruction representations of each of the plurality of input images, the respective dense reconstruction representations each including respective depth images for the plurality of input images, the depth images including color and depth information;
 - projecting front surfaces of the depth images using a forward depth test to generate a plurality of front-warped images;

projecting back surfaces of the depth images using an inverted depth test to generate a plurality of back-warped images;

stitching the front-warped images and the back-warped images to generate a two-layer panorama having a front surface panorama and a back surface panorama;

fusing the front surface panorama and the back surface panorama in the two-layer panorama to generate the three-dimensional image comprising a multi-layered geometric mesh suitable for rendering the scene in the three-dimensional space;

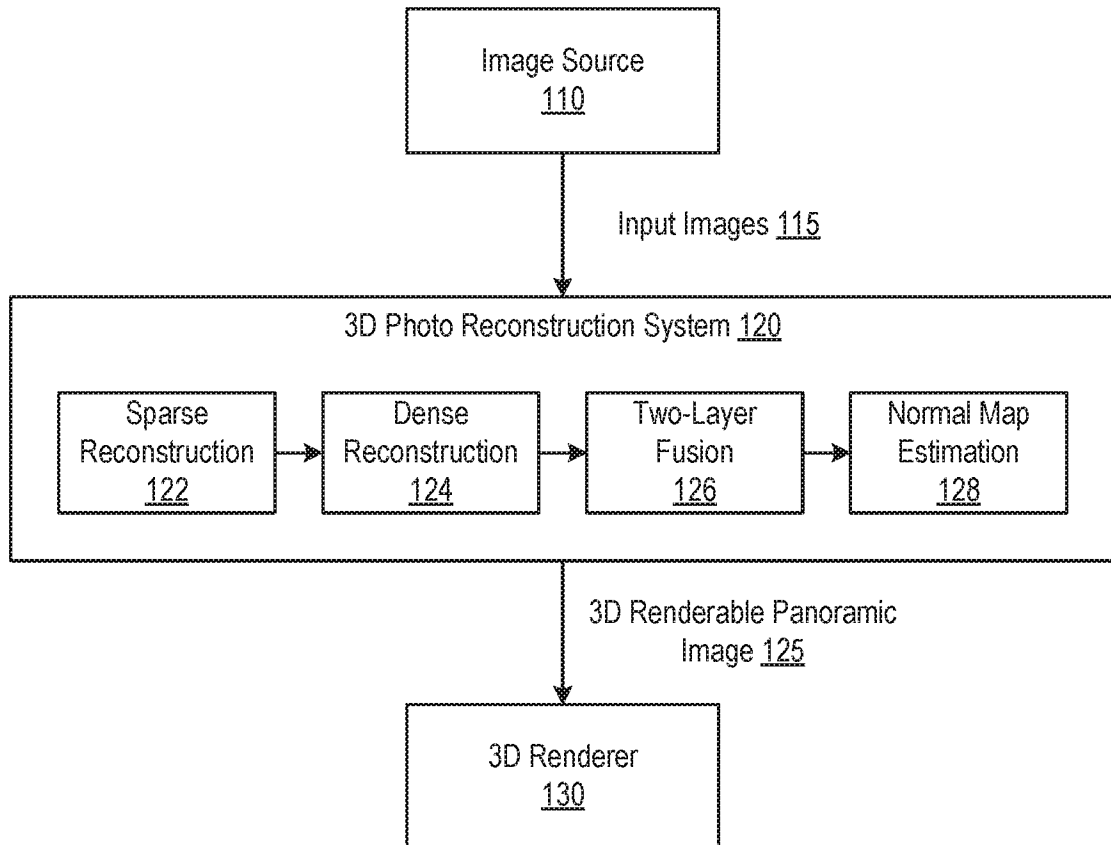
optionally wherein stitching the front-warped images and the back-warped images to generate a two-layer panorama comprises:

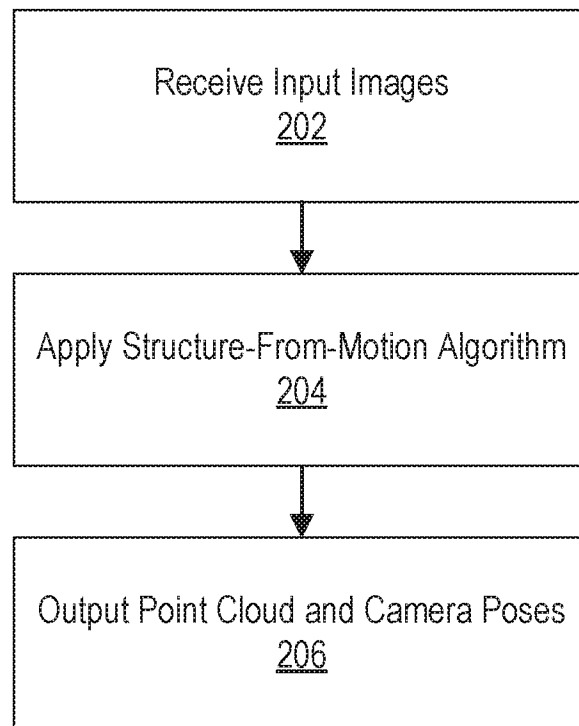
stitching a depth panorama using depth values from the front-warped images;

stitching the front surface panorama using color values from the front-warped images and stitched depth values from the depth panorama;

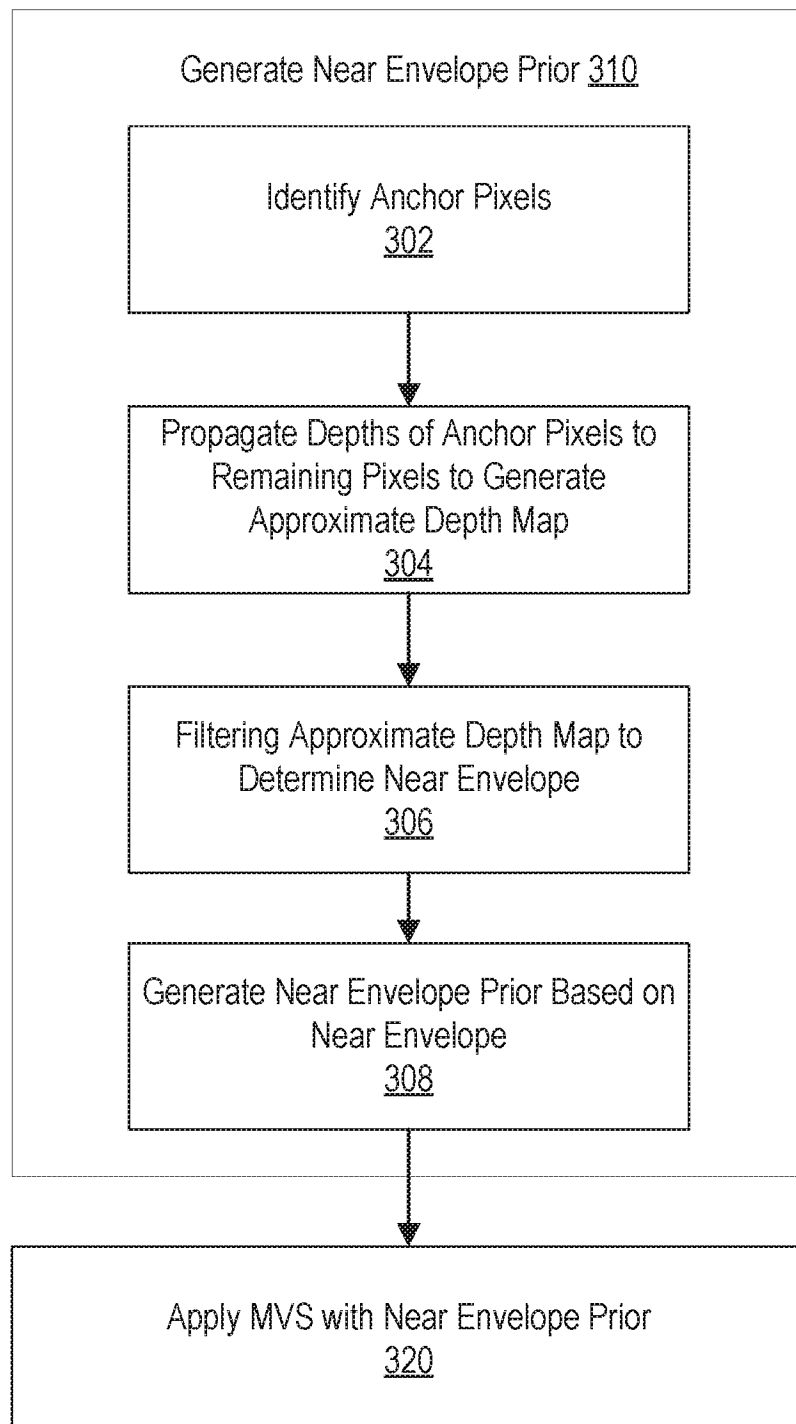
stitching the back surface panorama using color values from the back-warped images and the stitched depth values from the depth panorama; and

combining the front surface panorama and the back surface panorama into the two-layer panorama.

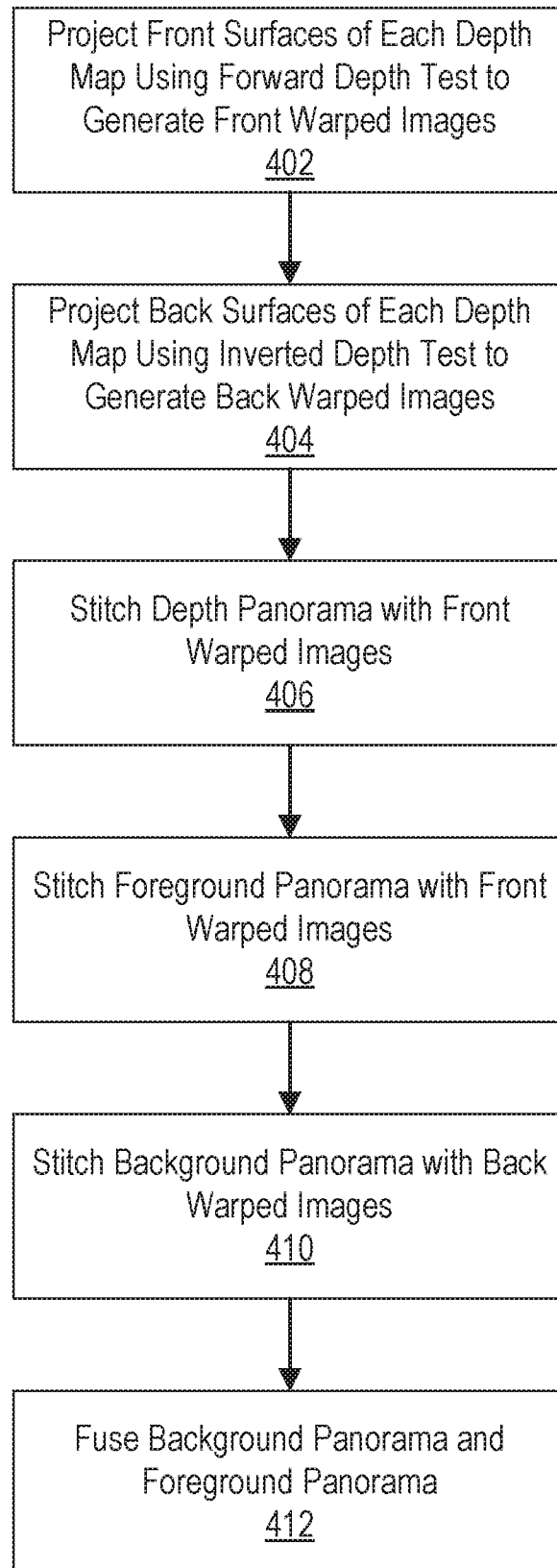
**Fig. 1**

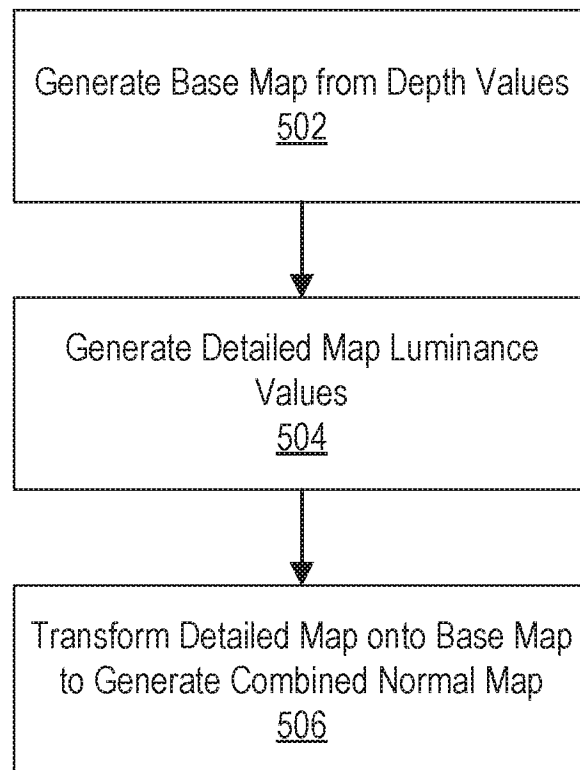
**Fig. 2**

3/5

**Fig. 3**

4/5

**Fig. 4**

**Fig. 5**

A. CLASSIFICATION OF SUBJECT MATTER**G06T 19/00(2011.01)i, G06T 7/20(2006.01)i, G06T 7/50(2017.01)i**

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)

G06T 19/00; G06K 9/36; H04N 5/232; H04N 13/00; G09G 5/377; G06K 9/00; H04N 5/265; G06T 7/20; G06T 7/50

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & keywords: a plurality of input images, front-warped image, back-warped image, stitching, fusing

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2002-0061131 A1 (HARPREET SINGH SAWHNEY et al.) 23 May 2002 See paragraphs [0063]-[0070], [0120]; and figures 2, 10.	1-35
A	US 2008-0152200 A1 (GERARD MEDIONI et al.) 26 June 2008 See paragraphs [0019]-[0028]; claim 1; and figure 1.	1-35
A	US 2016-0037082 A1 (KANG-HUAI WANG et al.) 04 February 2016 See paragraphs [0043]-[0055]; and figures 2-4B.	1-35
A	US 2014-0184640 A1 (GURURAJ GOPAL PUTRAYA et al.) 03 July 2014 See paragraphs [0051]-[0057], [0069]-[0073]; and figures 3, 6.	1-35
A	US 2014-0118482 A1 (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY) 01 May 2014 See paragraphs [0044]-[0094]; and figures 2-7.	1-35



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

19 October 2017 (19.10.2017)

Date of mailing of the international search report

19 October 2017 (19.10.2017)

Name and mailing address of the ISA/KR

International Application Division

Korean Intellectual Property Office

189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea



Facsimile No. +82-42-481-8578

Authorized officer

LEE, Dal Kyong

Telephone No. +82-42-481-8440



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2017/031839

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2002-0061131 A1	23/05/2002	EP 1337961 A1 JP 2004-525437 A US 7085409 B2 WO 02-45001 A1	27/08/2003 19/08/2004 01/08/2006 06/06/2002
US 2008-0152200 A1	26/06/2008	CN 101395613 A EP 1982292 A2 JP 2009-525543 A KR 10-2008-0108430 A US 2007-0183653 A1 US 2008-0152213 A1 US 7856125 B2 US 8126261 B2 WO 2008-013575 A2 WO 2008-013575 A3	25/03/2009 22/10/2008 09/07/2009 15/12/2008 09/08/2007 26/06/2008 21/12/2010 28/02/2012 31/01/2008 19/06/2008
US 2016-0037082 A1	04/02/2016	CN 105308621 A EP 3005232 A2 JP 2016-519968 A TW 201509363 A TW I532460 B US 2015-0221116 A1 US 2017-0046825 A1 US 9324172 B2 US 9672620 B2 WO 2014-193670 A2 WO 2014-193670 A3 WO 2014-193670 A4 WO 2016-160862 A1 WO 2017-030747 A1	03/02/2016 13/04/2016 11/07/2016 16/03/2015 11/05/2016 06/08/2015 16/02/2017 26/04/2016 06/06/2017 04/12/2014 29/01/2015 19/03/2015 06/10/2016 23/02/2017
US 2014-0184640 A1	03/07/2014	EP 2715449 A1 EP 2715449 B1 ES 2607214 T3 WO 2012-164148 A1	09/04/2014 09/11/2016 29/03/2017 06/12/2012
US 2014-0118482 A1	01/05/2014	KR 10-1370718 B1	06/03/2014