

(51) International Patent Classification:
G06T 7/00 (2017.01)(21) International Application Number:
PCT/CN2019/091346(22) International Filing Date:
14 June 2019 (14.06.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
PCT/CN2018/083350
17 April 2018 (17.04.2018) CN(71) Applicant: **SHANGHAITECH UNIVERSITY** [CN/CN];
Office of Research Administration, Shanghai Tech University, Room 207, Administration Center, No. 393 Middle Huaxia Road, Pudong New District, Shanghai 201210 (CN).(72) Inventors: **WU, Minye**; Office of Research Administration, Shanghai Tech University, Room 207, Administration Center, No. 393 Middle Huaxia Road, Pudong New District, Shanghai 201210 (CN). **SHI, Zhiru**; Office of Research Administration, Shanghai Tech University, Room 207, Administration Center, No. 393 Middle Huaxia Road, Pudong New District, Shanghai 201210 (CN).(74) Agent: **SHANGHAI SAVVY INTELLECTUAL PROPERTY AGENCY**; Unit 606, Shenergy International Building, 1 Middle Fuxing Road, Huangpu District, Shanghai 200021 (CN).(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, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) Title: LIGHT FIELD SYSTEM OCCLUSION REMOVAL

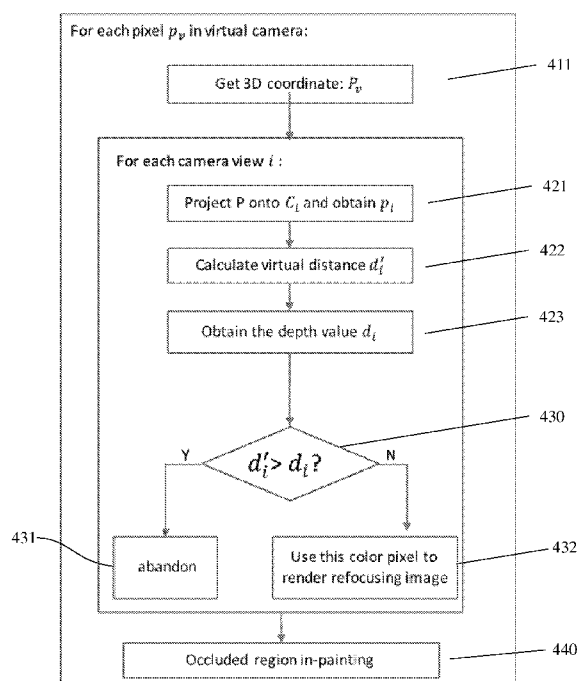


Figure 4

(57) Abstract: A method of image processing for occlusion removal in images and videos captured by light field camera systems. The method comprises: capturing a plurality of camera views using a plurality of cameras; capturing a plurality of depth maps using a plurality of depth sensors; generating a depth map for each camera view; calculating a target view on a focal plane corresponding to a virtual camera; set a weighting function on the pixels on the camera views based on the depth map and a virtual distance; and blending the pixels in accordance with the weighting function to generate a refocused target view.

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))*
- *with information concerning request for restoration of the right of priority in respect of one or more priority claims (Rules 26bis.3 and 48.2(b)(vii))*

LIGHT FIELD SYSTEM OCCLUSION REMOVAL

CROSS REFERENCE TO RELATED APPLICATIONS

The present application is based on and claims the benefit of priority to International Application under Patent Cooperation Treaty, No. PCT/CN2018/083350, filed on April 17, 2018, the entire contents of which are incorporated herein by reference.

TECHNICAL FIELD

The present invention relates to image processing, and more particularly, to light field image processing for removing occlusion.

BACKGROUND

Over the last decades, intelligent video surveillance systems have becoming increasingly important due to the huge number of cameras used for security and surveillance in public areas. Directly monitoring the huge amount of information collected by human is impractical. Hence, reliable automatic processing of the collected video information is in high demand. Facial recognition is one of the advanced technologies that can be used to process such information, as it could be used to spot individual faces within a crowd and prevent crime in real-time. However, crowd occlusion is one of the issues that affects the accuracy of facial recognition.

For instance, in security surveillance, the capability of seeing a subject of interest through occluders can be crucial. However, refocusing technology alone often fails to adequately remove or blur an occluding object for sufficient identification and/or recognition of the occluded object at the full potential. In certain circumstances, this may determine whether a suspected terrorist or a dangerous criminal can be identified or not.

Therefore, there are needs in the field of light field image rendering for

improved methods and systems to create more effective see-through effects.

SUMMARY OF THE INVENTION

To address issues in the prior art, embodiments of the present invention provide a method of processing light field images for occlusion removal in images and videos captured by light field camera systems.

In accordance with an embodiment of the present invention, a method of image processing is provided, the method comprising: capturing a plurality of camera views using a plurality of cameras; generating a depth map for each camera view; calculating a target view on a focal plane corresponding to a virtual camera; projecting a pixel on the target view to the plurality of camera views to obtain a plurality of projected pixels; setting a weighting of a projected pixel to 0 if the depth value of the projected pixel on the first camera view is less than a virtual distance between the pixel on the target view and the corresponding camera, and to 1 if otherwise; and blending the projected pixels in accordance with the weightings to generate a refocused target view.

Preferably, the pixel on the target view comprises a color value or a grayscale value, and blending the projected pixel comprises calculating the color value or a grayscale value of the pixel on the target view by weighted averaging a color value or a grayscale value of the projected pixels on the plurality of camera views.

Preferably, the method further comprises capturing a plurality of depth maps using a plurality of depth sensors.

Preferably, generating a depth map for a camera view comprising warping a depth image captured by a depth sensor in accordance with intrinsic matrices and extrinsic matrices of the depth sensor and camera.

In accordance with an embodiment of the present invention, warping the depth image generated by a depth sensor according to a warping algorithm comprises forwarding warping the depth image generated by the depth sensor to

obtain an intermediate depth image; filling holes in the intermediate depth image; and backward warping the intermediate depth image to obtain the depth map for the camera view.

In accordance with an embodiment of the present invention, the plurality of cameras are configured to capture a plurality of videos, and a refocused target video corresponding to the virtual camera is generated in real time.

Preferably, the plurality of video comprise 30 frames per second.

BRIEF DESCRIPTION OF THE DRAWINGS

To better illustrate the technical features of the embodiments of the present invention, various embodiments of the present invention will be briefly described in conjunction with the accompanying drawings.

Figure 1 is an exemplary schematic diagram of a light field camera system with depth sensors in accordance with an embodiment of the present invention.

Figure 2 is an exemplary schematic diagram for calibrating the camera system in accordance with another embodiment of the present invention.

Figure 3 is an exemplary schematic diagram showing the geometry indicators for the refocusing algorithm in accordance with yet another embodiment of the present invention.

Figure 4 an exemplary schematic diagram showing the occlusion removal algorithm in accordance with yet another embodiment of the present invention and two prior art techniques.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In accordance with embodiments of the present invention, a light field camera array system with wide baseline is introduced to remove occlusion in real-time. Figure 1 is an exemplary schematic diagram of a light field camera system with depth sensors in accordance with an embodiment of the present invention. As shown in Figure 1, the system contains several color image sensors and depth sensors. All of these sensors have a synchronous mechanism.

The depth information can be generated by various depth sensors, such as Kinect and Realsense sensors. Alternatively, the 3D information can also be produced by a light field color image system by conducting stereo matching. Depth information is used for occlusion detection and occluded region in-painting. A 3D point cloud on the focal plane is generated.

Compared to traditional pin-hole camera system, photographs from multiple angles of light field provide extra information to in-painting the occluded regions. The light field camera system allows us to see “around” the occluder by filling in unseen areas of the photograph. With the real-time light field render, this system can provide reliable occlusion-free footage for face recognition analysis. The occlusion removal mechanism is consisted in the following steps.

1. Camera System Calibration

First of all, we need to calibrate our camera system. We use traditional calibration methods to calibrate cameras and obtain their intrinsic and extrinsic matrices. Figure 2 is an exemplary schematic diagram for calibrating the camera system in accordance with another embodiment of the present invention.

As shown in Figure 2, in step 201, a checkerboard pattern is randomly moved in front of the camera system. In step 202, the system captures sample images. In step 203, the intrinsic and distortion parameters of each camera are estimated. In step 204, the relative pose between each camera and the checkerboard is estimated. In step 205, the absolute pose of each camera is initialized. Finally, in step 206, all parameters are globally optimized to minimize re-projection error.

2. Depth maps and color images capture

When the system is calibrated, we can start processing pipeline for each frame we captured. All color images and depth maps are captured synchronously,

which ensures that all image from different sensors are captured at the same time.

3. Depth projections

Next, we need to prepare depth maps for each viewpoint, which generate depth maps for all color images.

Depth images for each viewpoint can be obtained by warping the depth images from depth sensors according to the corresponding transformation between RGB cameras and depth sensors. The algorithm consists of three major components: a forward warping, a backward warping, and a hole filling process.

We do depth projection for each depth sensor. Here we demonstrate the warping algorithm for a single depth sensor. Let $d(x, y)$ be the depth image obtained from the depth sensor, $d_i(x_i, y_i)$ be the expected depth image of i -th viewpoint, K_k and K_i be the intrinsic matrices and C_k and C_i be the extrinsic matrices of depth sensor and i -th RGB camera respectively. We can formulate this relation as:

$$\begin{pmatrix} \omega_i x_i \\ \omega_i y_i \\ \omega_i d_i \\ \omega_i \end{pmatrix} = K_i C_i (K_k C_k)^{-1} \begin{pmatrix} \omega x \\ \omega y \\ \omega d \\ \omega \end{pmatrix} = T_i \begin{pmatrix} x \\ y \\ d \\ 1 \end{pmatrix} \quad (1)$$

By removing the third column and row of T_i , we obtain the forward warping equation given as:

$$\begin{pmatrix} \omega_i x_i \\ \omega_i y_i \\ \omega_i \end{pmatrix} = H_i \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} + d \vec{a} \quad (2)$$

We can rewrite this equation as:

$$\begin{pmatrix} \omega_i x_i \\ \omega_i y_i \\ \omega_i \end{pmatrix} = H_i \begin{pmatrix} \omega_i^* x_i^* \\ \omega_i^* y_i^* \\ \omega_i^* \end{pmatrix} \quad (3)$$

$$\text{with } \begin{pmatrix} \omega_i^* x_i^* \\ \omega_i^* y_i^* \\ \omega_i^* \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} + d \cdot H_i^{-1} \vec{a}. \quad (4)$$

Given the above formulations, we propose the following warping algorithm:

- a. Forward warp the depth image $\mathbf{d}$ to obtain an intermediate depth image according to the equation (4);
- b. Fill holes in the intermediate depth image; and
- c. Backward warp the intermediate depth image to obtain $\mathbf{d}_i$ according to equation (3) (the depth image of the desired viewpoint)

Figure 3 is an exemplary schematic diagram showing the geometry indicators for the refocusing algorithm in accordance with yet another embodiment of the present invention.

4. Occlusion removal

Figure 4 an exemplary schematic diagram showing the occlusion removal algorithm in accordance with yet another embodiment of the present invention and two prior art techniques. For each pixel in virtual camera, the process shown in Figure 4 is performed to remove occlusion.

4.1 Calculate 3D point on the focal plane

Like traditional light field refocusing algorithm, we also have a focal plane which represents the estimated geometry of scene. Focal plane is located in virtual camera's coordinate system, and has a distance away from virtual camera.

4.1.1 Get 3D coordinate

For each pixel of target view, we assume a planar proxy and apply ray tracing technique to calculate its 3D coordinate. The focal plane is perpendicular to the view direction of the target camera and the distance between them is z . Given a pixel $p_v = [u, v]^T$ in the target view, we calculate its 3D coordinate $P_v = [x_v, y_v, z_v]^T$ as

$$\begin{pmatrix} x_v \\ y_v \\ z_v \end{pmatrix} = z K_v^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad (5)$$

where K_v is the intrinsic of the target camera.

4.2 Project 3D point to viewpoint

4.2.1 Project P onto C_i and obtain p_i

We then project the 3D point P_v to all the captured views, from which we generate all the candidates for later pixel blending. Let $P_i = [x_i, y_i, z_i]^T$ and $p_i = [u_i, v_i]^T$ be the corresponding 3D coordinate and 2D pixel coordinate in the i -th viewpoint,

$$\begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} = C_i C_v^{-1} \begin{pmatrix} x_v \\ y_v \\ z_v \end{pmatrix} \quad (6)$$

$$p_i = \begin{pmatrix} u_i \\ v_i \end{pmatrix} = \pi(K_i P_i) \quad (7)$$

where $\pi(\cdot)$ is the normalization operator to map the 3D homogeneous coordinate to the 2D pixel coordinate.

4.2.2 Calculate virtual distance d'_i

By taking the z coordinate from P_i , we obtain the z -depth d'_i between the 3D point and the i -th camera.

$$d'_i = z_i \quad (8)$$

4.2.3 Obtain the depth value d_i

At the same time, we retrieve the warped depth value $d_i(u_i, v_i)$ and the color $c_i(u_i, v_i)$, which can be referred to section 3.

4.3 Occlusion detection

We compare these two depth values. If $d_i(u_i, v_i) < d'_i$, we become confident that the expected object in the target view is occluded in the i -th viewpoint, therefore we abandon the pixel color from i -th viewpoint by setting the

weighting m_i to zero. Otherwise we accept the pixel color by setting the weighting m_i to one.

4.4 Occluded region in-painting

The final pixel blending calculates the pixel color for p_v

$$c(u, v) = \frac{\sum_i m_i c_i(u_i, v_i)}{\sum_i m_i}$$

with this pixel selecting scheme, the refocused target image is guaranteed to avoid any obstructions in front of the focal plane.

After one frame is processed, we repeat these procedures above and obtain frames (video) with occlusions removed.

Claim

1. A method of image processing, comprising:

capturing a plurality of camera views using a plurality of cameras;
generating a depth map for each camera view;
calculating a target view on a focal plane corresponding to a virtual

camera;

projecting a pixel on the target view to the plurality of camera
views to obtain a plurality of projected pixels;

setting a weighting of a projected pixel to 0 if the depth value of the
projected pixel on the first camera view is less than a virtual distance between
the pixel on the target view and the corresponding camera, and to 1 if otherwise;
and

blending the projected pixels in accordance with the weightings to
generate a refocused target view.

2. The method of claim 1, wherein the pixel on the target view comprises a
color value or a grayscale value, and blending the projected pixel comprises
calculating the color value or a grayscale value of the pixel on the target view by
weighted averaging a color value or a grayscale value of the projected pixels on
the plurality of camera views.

3. The method of claim 1, further comprising:

capturing a plurality of depth maps using a plurality of depth sensors.

4. The method of claim 3, wherein generating a depth map for a camera view
comprises warping a depth image captured by a depth sensor in accordance with
intrinsic matrices and extrinsic matrices of the depth sensor and camera.

5. The method of claim 4, wherein warping the depth image generated by a depth sensor according to a warping algorithm comprises:

forwarding warping the depth image generated by the depth sensor to obtain an intermediate depth image;

filling holes in the intermediate depth image; and

backward warping the intermediate depth image to obtain the depth map for the camera view.

6. The method of claim 5, further comprising forwarding warping the depth image to obtain an intermediate depth image in accordance with

$$\begin{pmatrix} \omega_i^* x_i^* \\ \omega_i^* y_i^* \\ \omega_i^* \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} + d \cdot H_i^{-1} \vec{a}.$$

wherein $\begin{pmatrix} \omega_i^* x_i^* \\ \omega_i^* y_i^* \\ \omega_i^* \end{pmatrix}$ is the intermediate depth image, and $\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$ is the depth map generated by the depth sensor.

7. The method of claim 6, further comprising obtaining H_i by removing the third row of T_i , wherein $T_i = K_i C_i (K_k C_k)^{-1}$, K_k and K_i are intrinsic matrices of the depth sensor and the camera, and C_k and C_i are extrinsic matrices of the depth sensor and the camera.

8. The method of claim 6, further comprising backward warping the intermediate depth image to obtain the depth map for the camera view in accordance with

$$\begin{pmatrix} \omega_i x_i \\ \omega_i y_i \\ \omega_i \end{pmatrix} = H_i \begin{pmatrix} \omega_i^* x_i^* \\ \omega_i^* y_i^* \\ \omega_i^* \end{pmatrix}$$

wherein $\begin{pmatrix} \omega_i^* x_i^* \\ \omega_i^* y_i^* \\ \omega_i^* \end{pmatrix}$ is the intermediate depth image, and $\begin{pmatrix} \omega_i x_i \\ \omega_i y_i \\ \omega_i \end{pmatrix}$ is the depth map for the camera view.

9. The method of claim 4, wherein the plurality of cameras are configured to capture a plurality of videos, and a refocused target video corresponding to the virtual camera is generated in real time.

10. The method of claim 9, wherein the plurality of video comprise 30 frames per second.

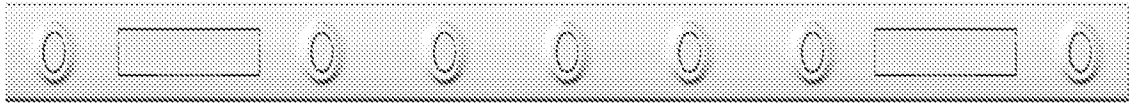


Figure 1

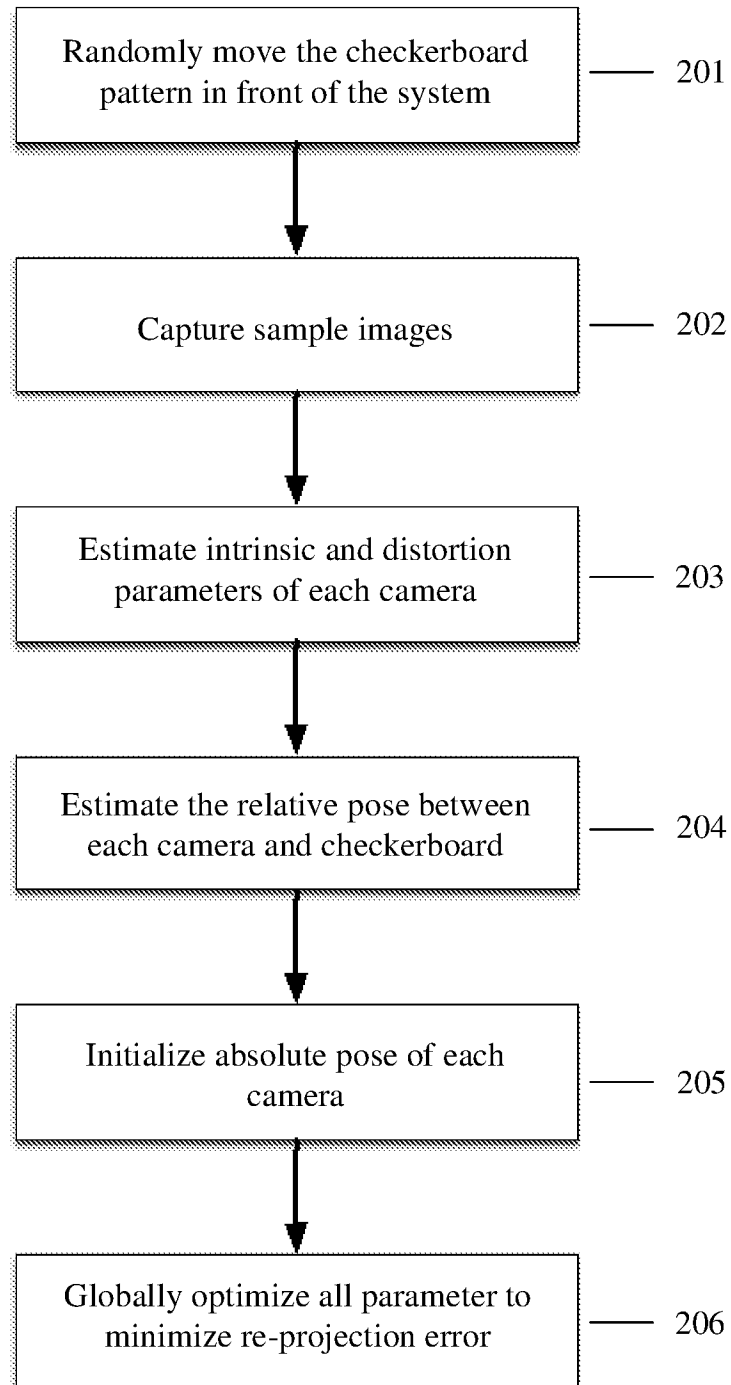


Figure 2

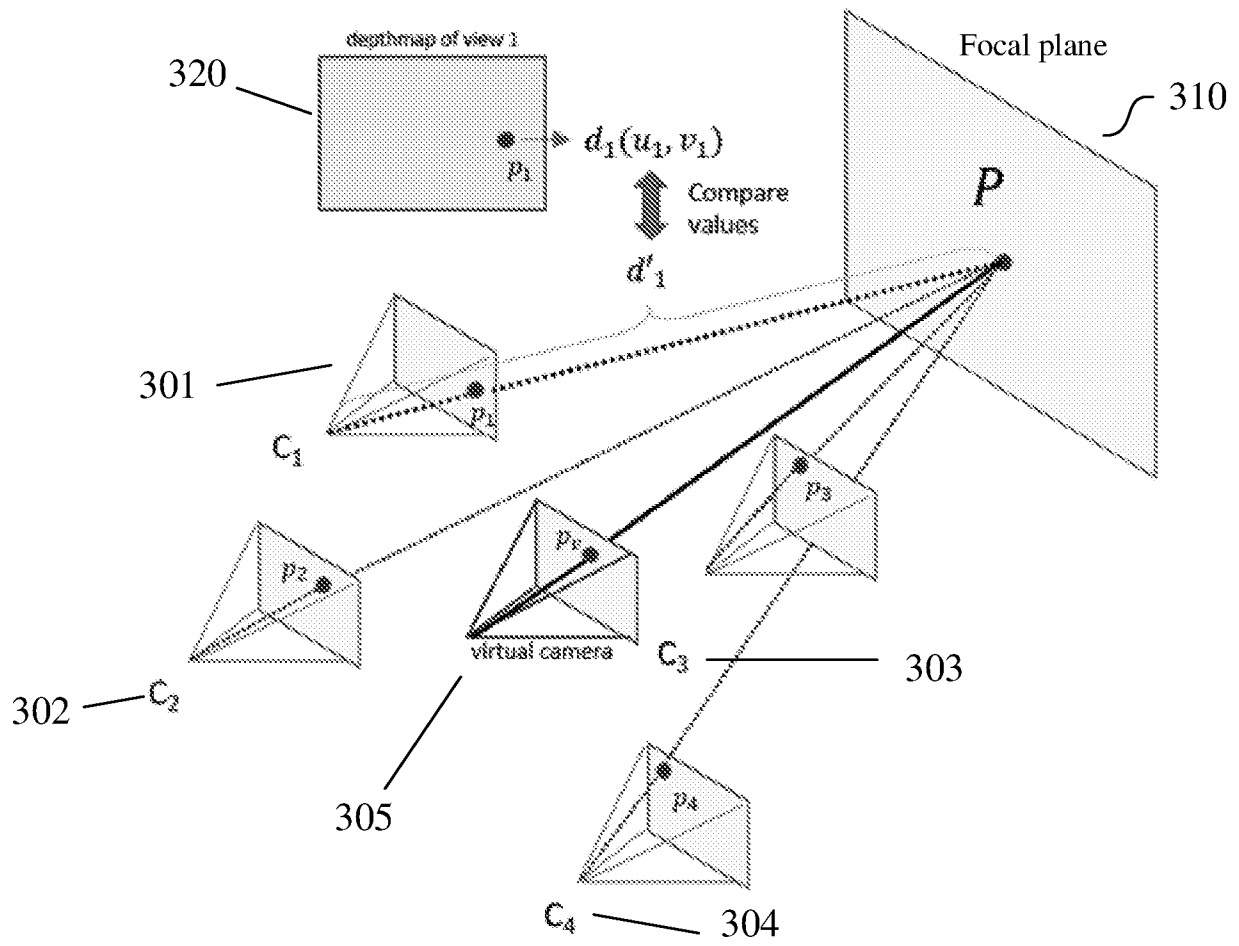


Figure 3

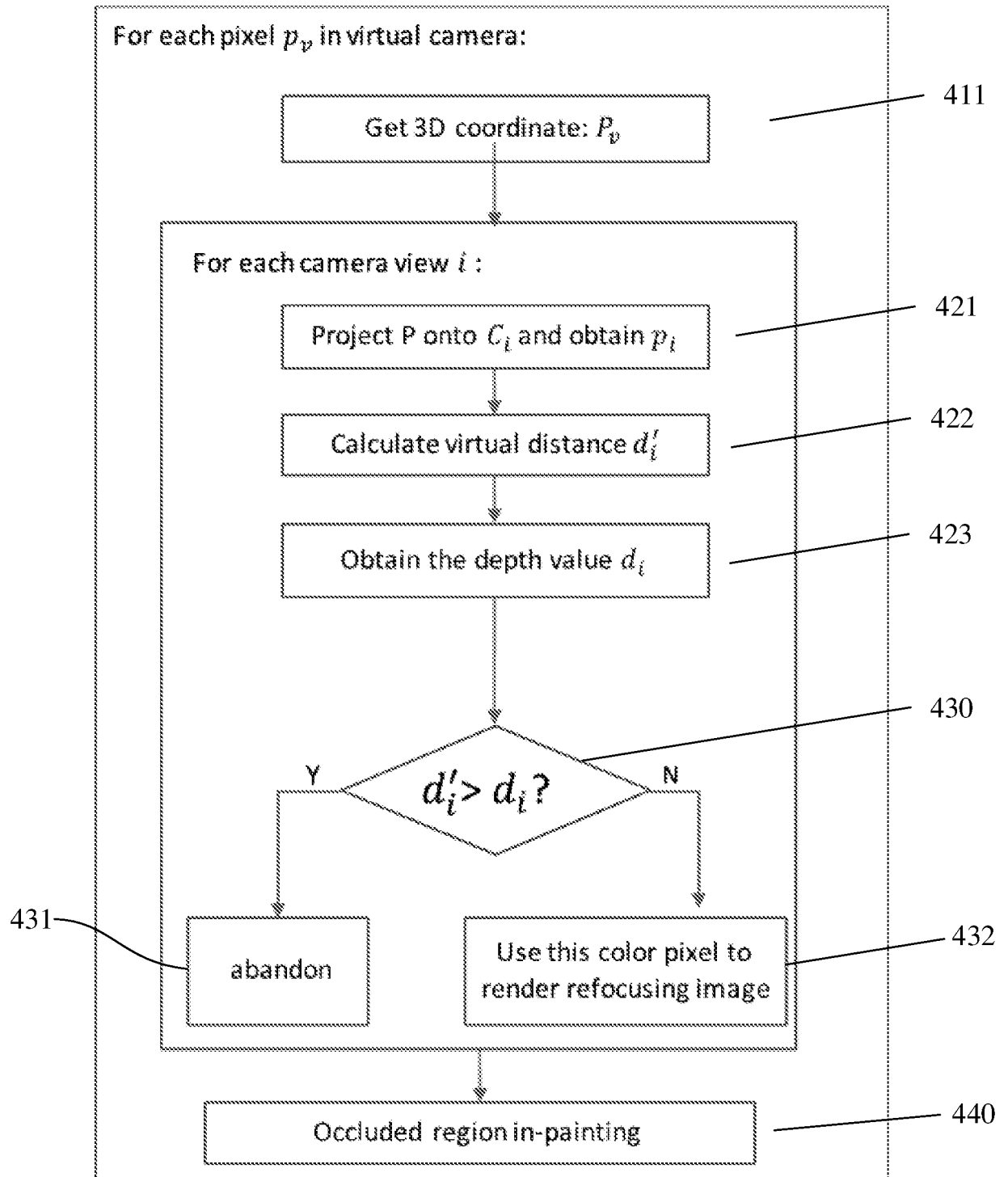


Figure 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/091346

A. CLASSIFICATION OF SUBJECT MATTER

G06T 7/00(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; H04N

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

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

CNPAT,CNKI,WPLEPODOC: SHANGHAITECH UNIVERSITY, camera, depth, projected pixel, weight, virtual, distance, view, blend, compose, synthesize.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 105719317 A (SHAANXI NORMAL UNIVERSITY) 29 June 2016 (2016-06-29) description, paragraphs [0046]-[0088]	1-10
A	CN 103237155 A (BEIJING UNIVERSITY OF TECHNOLOGY) 07 August 2013 (2013-08-07) the whole document	1-10
A	CN 104661013 A (NINGBO UNIVERSITY) 27 May 2015 (2015-05-27) the whole document	1-10
A	US 2012162193 A1 (SONY CORPORATION) 28 June 2012 (2012-06-28) the whole document	1-10
A	US 2014292748 A1 (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 02 October 2014 (2014-10-02) the whole document	1-10

☐ 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

02 September 2019

Date of mailing of the international search report

12 September 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088
China

Authorized officer

LI,Hui

Facsimile No. (86-10)62019451

Telephone No. 86-10-53961720

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/091346

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	105719317	A	29 June 2016	None			
CN	103237155	A	07 August 2013	None			
CN	104661013	A	27 May 2015	None			
US	2012162193	A1	28 June 2012	CN	102572465	A	11 July 2012
				EP	2469865	A2	27 June 2012
US	2014292748	A1	02 October 2014	KR	20140119999	A	13 October 2014