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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 June 2021 (10.06.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/112406 A1

(51) International Patent Classification:
G06T 19/00 (2011.01) *G06F 3/041* (2006.01)
G06T 15/00 (2006.01) *G06T 7/50* (2017.01)
G06T 19/20 (2011.01) *G06N 3/08* (2006.01)

(21) International Application Number:
PCT/KR2020/014721

(22) International Filing Date:
27 October 2020 (27.10.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10-2019-0159394 03 December 2019 (03.12.2019) KR

(71) Applicant: **SAMSUNG ELECTRONICS CO., LTD.** [KR/KR]; 129, Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16677 (KR).

(72) Inventors: **KIM, Yongsung**; 129, Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16677 (KR). **BAN, Daehyun**; 129, Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16677 (KR). **LEE, Dongwan**; 129, Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16677 (KR). **LEE, Hongpyo**; 129, Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16677 (KR). **ZHANG, Lei**; 129, Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16677 (KR).

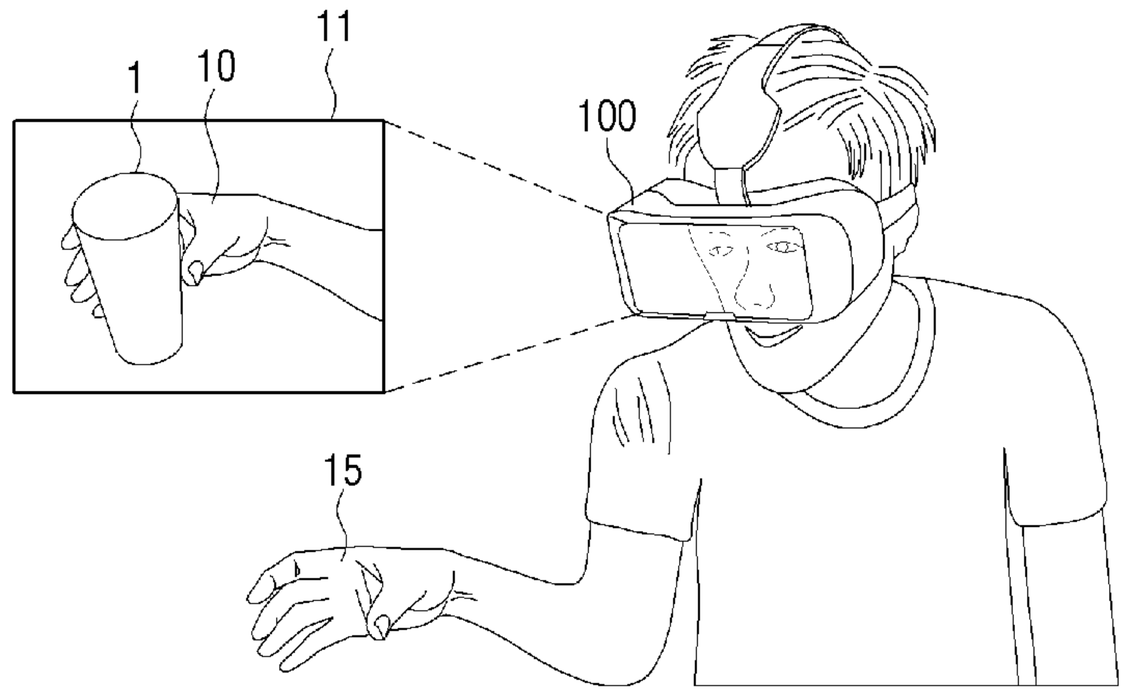
(74) Agent: **KIM, Tae-hun** et al.; 9th Floor, Shinduk Bldg., 343, Gangnam-daero, Seocho-gu, Seoul 06626 (KR).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

(54) Title: ELECTRONIC APPARATUS AND METHOD FOR CONTROLLING THEREOF



(57) Abstract: An electronic apparatus is provided. The electronic apparatus includes a display, a camera configured to capture a rear of the electronic apparatus facing a front of the electronic apparatus in which the display displays an image, and a processor configured to render a virtual object based on the image captured by the camera, based on a user body being detected from the captured image, estimate a plurality of joint coordinates with respect to the detected user body using a pre-trained learning model, generate an augmented reality image using the estimated plurality of joint coordinates, the rendered virtual object, and the captured image, and control the display to display the generated augmented reality image, wherein the processor is configured to identify whether the user body touches the virtual object based on the plurality of estimated joint coordinates, and change a transmittance of the virtual object based on the touch being identified.

WO 2021/112406 A1

Description

Title of Invention: ELECTRONIC APPARATUS AND METHOD FOR CONTROLLING THEREOF

Technical Field

[1] The disclosure relates to an electronic apparatus and method for controlling thereof. More particularly, the disclosure relates to an electronic apparatus that renders a virtual object on an image captured by a camera, and displays an AR image using the rendered virtual image and the captured image, and a method for controlling thereof.

Background Art

[2] Augmented reality (AR) technology is a technology that superimposes a 3D virtual image on a real image or background and displays it as a single image. The AR technology is being used in various ways in everyday life, such as not only video games but also smartphones, a head-up display (HUD) on windshield of vehicle, or the like.

[3] However, in the case of the AR technology, an image is output by superimposing a virtual object on an image received by a camera, but there has been a problem in that even when a user's hand is closer to the camera than a virtual object, the virtual object is formed on the user's hand so that the user hand is seen to be farther away from the camera than the virtual object.

[4] In addition, the AR technology has a problem in that a plurality of cameras must capture a user and a space from various viewpoints for interaction between the user and a virtual object. Also, a high-performance equipment was required to process an image captured by the plurality of cameras in real time.

[5] The above information is presented as background information only to assist with an understanding of the disclosure. No determination has been made, and no assertion is made, as to whether any of the above might be applicable as prior art with regard to the disclosure.

Disclosure of Invention

Technical Problem

[6] Aspects of the disclosure are to address at least the above-mentioned problems and/or advantages and to provide at least the advantages described below. Accordingly, aspects of the disclosure is to provide an electronic apparatus configured to render a virtual object on an image captured by one camera, and display an augmented reality image by using the rendered virtual object and the captured image, and a method of controlling thereof.

Solution to Problem

[7] Additional aspects will be set forth in part in the description which follows and, in part, will be apparent from the description, or may be learned by practice of the presented embodiments.

[8] In accordance with an aspect of the disclosure, an electronic apparatus is provided. The apparatus includes a display, a camera configured to capture a rear of the electronic apparatus facing a front of the electronic apparatus in which the display displays an image, and a processor configured to render a virtual object based on the image captured by the camera, based on a user body being detected from the captured image, estimate a plurality of joint coordinates with respect to the detected user body using a pre-trained learning model, generate an augmented reality image using the estimated plurality of joint coordinates, the rendered virtual object, and the captured image, and control the display to display the generated augmented reality image, wherein the processor is configured to identify whether the user body touches the virtual object based on the estimated plurality of joint coordinates, and change a transmittance of the virtual object based on the touch being identified.

[9] The processor may be configured to estimate a plurality of joint coordinates corresponding to a finger joint and a palm using the pre-trained learning model based on the detected user body being identified to be a hand.

[10] The processor may be configured to render a virtual hand object and the virtual object based on the estimated plurality of joint coordinates.

[11] The processor may be configured to change a transmittance of one area of the virtual object corresponding to the touch.

[12] The processor may be configured to change a transmittance of the user body and transparently display the user body based on the touch being identified.

[13] The processor may be configured to receive depth data of the captured image from the camera, and generate the augmented reality image using the received depth data.

[14] The pre-trained learning model may be configured to be trained through a plurality of learning data including hand images by using a convolutional neural network (CNN).

[15] The plurality of learning data may be configured to include a first data in which a 3D coordinate is matched to at least one area of the hand image, and a second data in which the 3D coordinate is not matched to the hand image, and the pre-trained learning model is configured be trained by updating a weight value of the CNN based on the first data and the second data.

[16] In accordance with another aspect of the disclosure, a method of controlling an electronic apparatus is provided. The method includes capturing a rear of the electronic apparatus facing a front of the electronic apparatus in which a display displays an image, rendering a virtual object based on a captured image, based on a user body being detected from the captured image, estimating a plurality of joint coordinates with

respect to the detected user body using a pre-trained learning model, generating an augmented reality image using the estimated plurality of joint coordinates, the rendered virtual object, and the captured image, displaying the generated augmented reality image, identifying whether the user body touches the virtual object based on the estimated plurality of joint coordinates, and changing a transmittance of the virtual object based on the touch being identified.

[17] The estimating may include estimating a plurality of joint coordinates corresponding to a finger joint and a palm using the pre-trained learning model based on the detected user body being identified to be a hand.

[18] The rendering may include rendering a virtual hand object and the virtual object based on the estimated plurality of joint coordinates.

[19] The changing may include changing a transmittance of one area of the virtual object corresponding to the touch.

[20] The method may further include changing a transmittance of the user body and transparently displaying the user body based on the touch being identified.

[21] The generating may include receiving depth data of the captured image from the camera, and generating the augmented reality image using the received depth data.

[22] The pre-trained learning model may be configured to be trained through a plurality of learning data including hand images by using a convolutional neural network (CNN).

[23] The plurality of learning data may be configured to include a first data in which a 3D coordinate is matched to at least one area of the hand image, and a second data in which the 3D coordinate is not matched to the hand image, and wherein the pre-trained learning model is configured be trained by updating a weight value of the CNN based on the first data and the second data.

[24] Other aspects, advantages, and salient features of the disclosure will become apparent to those skilled in the art from the following detailed description, which, taken in conjunction with the annexed drawings, discloses various embodiments of the disclosure.

Brief Description of Drawings

[25] The above and other aspects, features, and advantages of certain embodiments of the disclosure will be more apparent from the following description taken in conjunction with the accompanying drawings, in which:

[26] FIG. 1 is a view illustrating an operation of an electronic apparatus according to an embodiment of the disclosure;

[27] FIG. 2 is a block diagram illustrating a configuration of an electronic apparatus according to an embodiment of the disclosure;

[28] FIG. 3 is a block diagram illustrating a detailed configuration of an electronic apparatus according to an embodiment of the disclosure;

[29] FIG. 4A is a view illustrating estimating process of a plurality of joint coordinates according to an embodiment of the disclosure;

[30] FIG 4B is a view illustrating estimating process of a plurality of joint coordinates according to an embodiment of the disclosure;

[31] FIG. 5 is a view illustrating an AR image displayed by an electronic apparatus according to an embodiment of the disclosure;

[32] FIG. 6 is a view illustrating a process of rendering a virtual hand object on a user body according to an embodiment of the disclosure;

[33] FIG. 7 is a view illustrating an event corresponding to an object touch according to an embodiment of the disclosure;

[34] FIG. 8 is a view illustrating an object touch according to an embodiment of the disclosure; and

[35] FIG. 9 is a flowchart illustrating a method of controlling an electronic apparatus according to an embodiment of the disclosure.

[36] The same reference numerals are used to represent the same elements throughout the drawings.

Mode for the Invention

[37] The following description with reference to the accompanying drawings is provided to assist in a comprehensive understanding of various embodiments of the disclosure as defined by the claims and their equivalents. It includes various specific details to assist in that understanding but these are to be regarded as merely exemplary. Accordingly, those of ordinary skill in the art will recognize that various changes and modifications of the various embodiments described herein can be made without departing from the scope and spirit of the disclosure. In addition, descriptions of well-known functions and constructions may be omitted for clarity and conciseness.

[38] The terms and words used in the following description and claims are not limited to the bibliographical meanings, but, are merely used by the inventor to enable a clear and consistent understanding of the disclosure. Accordingly, it should be apparent to those skilled in the art that the following description of various embodiments of the disclosure is provided for illustration purpose only and not for the purpose of limiting the disclosure as defined by the appended claims and their equivalents.

[39] It is to be understood that the singular forms “a,” “an,” and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to “a component surface” includes reference to one or more of such surfaces.

[40] The terms “have”, “may have”, “include”, and “may include” used in the embodiments of the disclosure indicate the presence of corresponding features (for example, elements such as numerical values, functions, operations, or parts), and do

not preclude the presence of additional features.

[41] In addition, the disclosure describes components necessary for disclosure of each embodiment of the disclosure, and thus is not limited thereto. Accordingly, some components may be changed or omitted, and other components may be added. In addition, they may be distributed and arranged in different independent devices.

[42] Hereinafter, the disclosure will be described in more detail with reference to the drawings.

[43] FIG. 1 is a view illustrating an operation of an electronic apparatus according to an embodiment of the disclosure.

[44] Referring to FIG. 1, a user wears an electronic apparatus 100 and the user interacts with a virtual object 1 using an Augmented reality (AR) image 11 displayed on the electronic apparatus 100.

[45] The electronic apparatus 100 is a device including a camera and a display. As shown in FIG. 1, the electronic apparatus 100 may be implemented in a form of wearable augmented reality (AR) glasses that can be worn by a user. Alternatively, as an embodiment, it may be implemented as at least one of a display apparatus, a smartphone, a laptop PC, a laptop computer, a desktop PC, a server, a camera device, and a wearable device including a communication function.

[46] The electronic apparatus 100 may provide the AR image 11 to the user using a display, and the electronic apparatus 100 may capture a rear of the electronic apparatus 100 facing a front of the electronic apparatus 100 in which the display displays an image by using a camera to move the user's hand 15, thereby interacting the virtual object 1 with the user body 10.

[47] The AR image 11 is an image provided by the electronic apparatus 100 through a display, and may display the user body 10 and the virtual object 1. In addition, the camera included in the electronic apparatus 100 captures a space where the user exists, and the AR image 11 may provide an object and a surrounding environment that are actually exist to the user through the captured image.

[48] A display is disposed on the front of the electronic apparatus 100 to provide the AR image 11 to the user, and a camera is disposed on the rear of the electronic apparatus 100 to capture the user's surroundings and the user body. According to an embodiment of the disclosure, since the electronic apparatus 100 captures the user's surroundings and the user body depending on a direction of the user's gaze, that is, a first person perspective, and provides an image generated based on the perspective, the electronic apparatus 100 may provide a realistic AR image.

[49] According to an embodiment of the disclosure, the electronic apparatus 100 may include a camera. The electronic apparatus 100 may guarantee real-time performance of image processing and may not require a high-performance device by using a single

camera. The electronic apparatus 100 may estimate 3D coordinates of the user body by using a pre-trained learning model even when only a portion of the user body (e.g., hand) is captured. In addition, the electronic apparatus 100 may estimate an exact location and motion of the user body despite using a single camera, and based on this, may provide a service capable of interacting with a virtual object to the user.

[50] FIG. 2 is a block diagram illustrating a configuration of an electronic apparatus according to an embodiment of the disclosure.

[51] Referring to FIG. 2, the electronic apparatus 100 may include a camera 110, a display 120, and a processor 130.

[52] The camera 110 may capture the rear of the electronic apparatus facing the front of the electronic apparatus 100 in which the display 120 displays an image (S210). The camera 110 may capture a space where the user exists and the user body. The camera 110 may be disposed on the rear or side of the electronic apparatus 100 to capture the rear of the electronic apparatus 100. The camera 110 is disposed on the rear or side of the electronic apparatus 100, but the electronic apparatus 100 may be implemented as a wearable glass AR device or a smartphone, etc. as illustrated in FIG. 1. Thus, the camera 110 may capture an image with a variable direction depending on the user's movement and the user's gaze.

[53] In addition, the electronic apparatus 100 may be connected to the processor 130 in a wired or wireless communication method. The image captured by the camera 110 may be provided to the user in real time after a series of processing by the processor 130. In addition, the image captured by the camera 110 may be used as a basis for generating an AR image by the processor 130 described below. The image captured by the camera 110 may be an RGB image including RGB data. Alternatively, according to another embodiment of the disclosure, the camera 110 may be a 3D camera capable of acquiring depth data. The processor 130 may acquire the depth data from the image captured by the camera 110, and use the acquired depth data as a basis for generating an AR image.

[54] The display 120 may be disposed in front of the electronic apparatus 100. In addition, the display 120 may be connected to the processor 130 by wired or wireless, and the display 120 may display various information under the control of the processor 130. In particular, the display 120 may display an AR image generated by the processor 130 (S250). Since the display 120 displays the AR image generated based on the image captured by the camera 110 disposed on the rear of the electronic apparatus 100, the AR image displayed by the display 120 may be a first person perspective image.

[55] In addition, the display 120 may be implemented in a form of a general display such as a Liquid Crystal Display (LCD), a Light Emitting Diode (LED), an Organic Light Emitting Diode (OLED), a Quantum dot Light Emitting Diode (QLED), etc., and

according to another embodiment, the display 120 may also be implemented as a transparent display. Specifically, the display 120 is made of a transparent material, and light outside the electronic apparatus 100 may penetrate the display 120 to reach the user, and the user may observe the user body and external environment by penetrating the display 120. The transparent display may be implemented as a transparent liquid crystal display (LCD) type, a transparent thin-film electroluminescent panel (TFEL) type, a transparent Organic Light Emitting Diode (OLED) type, or the like, and may be implemented in a form of displaying by projecting an image on a transparent screen (e.g., head-up display (HUD)). When the display 120 is implemented as a transparent display, the processor 130 may control the display 120 such that only virtual objects are displayed on the display 120.

[56] The processor 130 may control overall operations and functions of the electronic apparatus 100. In particular, the processor 130 may render a virtual object based on the image captured by the camera 110, estimate a plurality of the user body detected using the pre-trained learning model when a user body is detected in the captured image, and generate an AR image by using estimated joint coordinates, the rendered virtual object, and the captured image, and control the display 120 to display the generated AR image.

[57] The processor 130 may be electrically connected to the camera 110, and may receive data including the image captured by the camera 110 from the camera 110. The processor 130 may render a virtual object based on the image captured by the camera 110 (S220). Specifically, rendering may refer to generating a second image including a virtual object to correspond to a first image captured by the camera 110. In other words, rendering may mean generating a virtual object to correspond to a certain area of the captured image. Since the processor 130 renders a virtual object based on the captured image, the rendered virtual object may include depth information about space.

[58] When the user body is detected from the captured image, the processor 130 may estimate a plurality of joint coordinates for the detected user body using a pre-trained learning model (S230). Specifically, the processor 130 may estimate the plurality of joint coordinates for the user body using RGB data included in the captured image. First, the processor 130 may detect a user body from the captured image, and the processor 130 may extract RGB data including the user body from the captured image. In addition, the processor 130 may estimate motions, shapes, and predicted coordinates of the user body by inputting the extracted RGB data. According to another embodiment, depth data may be included in an image captured by the camera 110 and estimated coordinates may be further estimated using the depth data.

[59] The pre-trained learning model may be a learning model trained through a plurality of learning data including a hand image using a convolutional neural network (CNN).

When the user body is a hand, a method of estimating joint coordinates using a learning model will be described below in detail with reference to FIGS. 4A and 4B.

[60] The processor 130 may generate an AR image using the estimated joint coordinates, the rendered virtual object, and the captured image (S240). The AR image may refer to a third image generated by matching or calibrating the first image captured by the camera and the second image including the virtual object.

[61] The processor 130 may control the display 120 to display the generated AR image. When the AR image is displayed on the display 120, the user may interact with the virtual object through the AR image. Specifically, the processor 130 may check whether the user body touches the virtual object based on the estimated joint co-ordinates, and perform an event corresponding to the object touch when the object touch is detected. The event may mean changing a transmittance of the virtual object. The processor 130 may change an alpha value of the virtual object included in the generated AR image by a unit of pixel.

[62] When a touch of the user body and the virtual object is identified, the processor 130 may change the transmittance of the virtual object (S260). Alternatively, the processor 130 may change only the transmittance of an area of the virtual object corresponding to the touch. In other words, the processor 130 may change the transmittance of the virtual object by changing alpha values of all pixels corresponding to the virtual object or by changing only an alpha value of the pixel in the area of the virtual object.

[63] Also, the processor 130 may identify an object touch based on whether the estimated joint coordinates are positioned at coordinates corresponding to the rendered virtual object. In addition, the processor 130 may track joint coordinates in real time or at a predetermined time interval through an image captured in real time by the camera 110. As described above with reference to FIG. 1, since only the image captured through one camera is used and only the plurality of joint coordinates for the user body are estimated by using RGB data of the captured image, the processor 130 may perform real-time image processing.

[64] FIG. 3 is a block diagram illustrating a detailed configuration of an electronic apparatus according to an embodiment of the disclosure.

[65] Referring to FIG 3, the electronic apparatus 100 may include the camera 110, the display 120, the processor 130, a communication interface 140, a memory 150, and a sensor 160. Meanwhile, since the camera 110, the display 120, and the processor 130 illustrated in FIG. 3 have been described in FIG. 2, redundant description will be omitted.

[66] The communication interface 140 may communicate with an external apparatus (not shown). The communication interface 140 may be connected to an external device through communication via a third device (e.g., a repeater, a hub, an access point, a

server, a gateway, etc.).

[67] In addition, the communication interface 140 may include various communication modules to perform communication with an external device. Specifically, the communication interface 140 may include an NFC module, a wireless communication module, an infrared module, and a broadcast receiving module.

[68] The communication interface 140 may receive information related to the operation of the electronic apparatus 100 from an external device. According to an embodiment, the communication interface 140 may receive a learning model previously learned from an external server and device using the communication interface 140, and control the communication interface 140 to estimate coordinates of the user body using an external high-performance server and device. Further, the communication interface 140 may be used to update information stored in the electronic apparatus 100.

[69] The memory 150, for example, may store a command or data regarding at least one of the other elements of the electronic apparatus 100. The memory 150 may be implemented as a non-volatile memory, a volatile memory, a flash memory, a hard disk drive (HDD) or a solid state drive (SDD). The memory 150 may be accessed by the processor 130, and perform readout, recording, correction, deletion, update, and the like, on data by the processor 130. According to an embodiment of the disclosure, the term of the storage may include the memory 150, read-only memory (ROM) (not illustrated) and random access memory (RAM) (not illustrated) within the processor 130, and a memory card (not illustrated) attached to the electronic apparatus 100 (e.g., micro secure digital (SD) card or memory stick). Also, the memory 150 may store a program, data, and the like for constituting various types of screens that will be displayed in the display area of the display 120.

[70] In addition, the memory 150 may store data for displaying the AR image. Specifically, the memory 150 may store an image captured by the camera 110 and a second image including a virtual object generated by the processor 130. Also, the memory 150 may store the AR image generated based on the captured image and the rendered virtual object. Also, the memory 150 may store the plurality of joint coordinates of the user body estimated by the processor 130 in real time.

[71] The sensor 160 may detect an object. Specifically, the sensor 160 may sense an object by sensing physical changes such as heat, light, temperature, pressure, sound, or the like. Also, the sensor 160 may output coordinate information about the sensed object. Specifically, the sensor 160 may acquire 3D point information of the sensed object or output coordinate information based on a distance.

[72] For example, the sensor 160 may be a lidar sensor, a radar sensor, an infrared sensor, an ultrasonic sensor, a radio frequency (RF) sensor, a depth sensor, and a distance measurement sensor. The sensor 160 is a type of an active sensor and may transmit a

specific signal to measure a time of flight (ToF). The ToF is a flight time distance measurement method, and may be a method of measuring a distance by measuring a time difference between a reference time point at which a pulse is fired and a detection time point of a pulse reflected back from a measurement object.

[73] FIGS. 4A and 4B are views illustrating estimating process of a plurality of joint coordinates according to an embodiment of the disclosure.

[74] Referring to FIG. 4A, a user body 40, a plurality of joint coordinates 41 to 46 corresponding to a finger joint and a palm are illustrated.

[75] Referring to FIG. 4B, a hand image to which a plurality of joint coordinates 41 to 46 of FIG. 4A are connected when the hand is illustrated in a plurality of configurations.

[76] If the user body detected from an image captured by the camera 110 is a hand, the electronic apparatus 100 may use a pre-trained learning model as a method of estimating a plurality of joint coordinates corresponding to a finger joint and a palm.

[77] The learning model may be trained through a plurality of learning data or training data including a human hand image using a convolutional neural network (CNN). The learning data may be learned based on data in which 3D coordinates are input to at least one region of the hand image and data in which the 3D coordinates are not matched to the hand image. First, data input with 3D coordinates in at least one region of the hand image may be learned using a learning model. Subsequently, output data may be obtained from the learning model using data in which the 3D coordinates are not matched to the hand image, and a loss function or error between the output data and the data in which 3D coordinates are input in at least one region of the hand image may be calculated. The learning model may be trained through a process of updating a weight value of the CNN using the calculated loss function or error.

[78] Referring again to FIG. 4A, the learning data or training data may be a hand image in which 3D coordinates are assigned to a plurality of joints including a fingertip and a finger joint.

[79] Referring again to FIG. 4B, the learning model may be formed by machine learning or may be learned through deep learning after inputting 21 coordinates, respectively, as shown in FIG. 4A in each of the hand images including various hand shapes and size. In other words, the hand image illustrated in FIG. 4B may be learning data or training data for learning the learning model.

[80] FIG. 4A shows a user body 40 composed of 20 points corresponding to each finger and one point corresponding to the palm, a total of 21 points, are illustrated. The 21 points included in the user body 40 may be the basis for estimating a plurality of joint coordinates for the user body.

[81] For example, when the learning model is trained using experimental data or learning data that includes more than 21 points included in the user body 40 of FIG. 4A, the

electronic apparatus 100 may more accurately grasp the user body joint coordinates and locations included in the captured image. However, the number of points included in the user body 40 of FIG. 4A is only an embodiment, and the number of points set according to the performance of the electronic apparatus 100 may be appropriately selected in the range of 5 to 40 during implementation.

[82] As shown in FIG. 4B, the plurality of hand images may be learning data or training data for learning the learning model. The electronic apparatus 100 may estimate the joint coordinates of the user by learning the hand image illustrated in FIG. 4B. In particular, the electronic apparatus 100 may estimate coordinates of hidden finger with only a part of the captured image even when a part of the hand, for example, fingers are hidden by another object or subject.

[83] For example, it may be assumed that the learning model learns the data of FIG. 4B and a finger image similar to the data of FIG. 4B is given. In this case, according to the prior art, RGB and depth data for the fingers hidden by another object or subject may not be obtained, and accurate coordinates for the hidden fingers may not be obtained with only given data. However, the electronic apparatus 100 according to the disclosure may estimate coordinates using a pre-trained learning model for fingers hidden by other objects or subjects. However, the above example is an example for convenience of description, and data related to all movements and postures of the hand are not required to estimate the coordinates of the hand according to an embodiment of the disclosure.

[84] FIG. 5 is a view illustrating an AR image displayed by an electronic apparatus according to an embodiment of the disclosure.

[85] Referring to FIG. 5, a first image 51 including a user body 50, a second image 52 including a virtual object 5, and a third image 53 including a user body 50 and a virtual object 5 are illustrated.

[86] The first image 51 may be an image captured by a camera included in the electronic apparatus 100. Specifically, the rear of the electronic apparatus 100 may by captured by the camera included in the electronic apparatus 100, and the captured image may include the user body 50. The user body 50 captured by the camera is illustrated in the first image 51.

[87] In addition, the second image 52 may be an image generated by the virtual object 5 rendered by the electronic apparatus 100 based on the captured image so as to correspond to the first image captured by the camera included in the electronic apparatus 100. Since the virtual object 5 rendered by the electronic apparatus 100 is generated based on the captured image, depth information on space may be included.

[88] The third image 53 may be an image generated by the first image 51 and the second image 52 being matched and calibrated by the electronic apparatus 100.

[89] Specifically, the electronic apparatus 100 may extract RGB data including the user body from the first image captured by the camera included in the electronic apparatus 100. Then, the electronic apparatus 100 may estimate motions, shapes, and predicted coordinates of the user body by inputting the extracted RGB data. Alternatively, the electronic apparatus 100 may include depth data in the captured image, and estimate a plurality of joint coordinates for the user body by using the depth data.

[90] For example, the electronic apparatus 100 may detect a hand of the user body 50 included in the first image 51, and estimate coordinates of the user body 50 detected by using the pre-trained learning model. Alternatively, the electronic apparatus 100 may estimate the coordinates of the user body 50 by additionally using the depth data of the user body 50 from the captured image.

[91] The electronic apparatus 100 may match or calibrate the first image 51 and the second image 52 by using the estimated joint coordinates of the user body 50 and the depth information of the virtual object 5. Since the electronic apparatus 100 may generate an augmented reality image using the estimated coordinates of the user body and the coordinates of the virtual object, the user may be able to interact with the virtual object displayed through the display of the electronic apparatus 100.

[92] FIG. 6 is a view illustrating a process of rendering a virtual hand object on a user body according to an embodiment of the disclosure.

[93] Referring to FIG. 6, an augmented reality image 61 including a virtual hand object 60 is illustrated. As described above, the electronic apparatus 100 may estimate a plurality of joint coordinates for the user body, and render a virtual hand object based on the estimated plurality of joint coordinates.

[94] Specifically, the electronic apparatus 100 may detect the user's hand from the captured image and estimate the plurality of joint coordinates for the detected user's hand. The electronic apparatus 100 may render a virtual hand object at coordinates corresponding to the estimated plurality of joint coordinates, and the electronic apparatus 100 may output a virtual hand object instead of the user body. The electronic apparatus 100 may track movements of the user's hand in real time, and the electronic apparatus 100 superimposes a virtual hand object on the user's hand, so that the user may move the virtual hand object 60 in the augmented reality image 61 just like moving the user's hand, and interact with the virtual object.

[95] When the electronic apparatus 100 detects an object touch, at least one of a user body or a virtual object may be transparently displayed. An embodiment in which the electronic apparatus 100 transparently displays a virtual object or a user body will be described with reference to FIGS. 7 and 9.

[96] FIG. 7 is a view illustrating an event corresponding to an object touch according to an embodiment of the disclosure.

[97] Referring to FIG. 7, an augmented reality image 71 including a virtual object 7 and a user body 70 is illustrated. The electronic apparatus 100 may identify whether the user body 70 touches the virtual object 7 based on the estimated joint coordinates, and perform an event corresponding to the object touch when the object touch is detected. The event may refer to changing a transmittance of the virtual object 7 or a color of the displayed virtual object 7. Meanwhile, the electronic apparatus 100 may change an alpha value of the virtual object 7 included in the generated augmented reality image 71 by a unit of pixel. When a touch of the user body 70 and the virtual object 7 is identified, the electronic apparatus 100 may change the transmittance of the virtual object 7. Alternatively, the electronic apparatus 100 may change only the transmittance of an area of the virtual object 7 corresponding to the touch. In other words, the electronic apparatus 100 may change the transmittance of the virtual object by changing the alpha values of all pixels corresponding to the virtual object 7 or by changing only the alpha values of pixels in one region of the virtual object 7.

[98] FIG. 8 is a view illustrating an object touch according to an embodiment of the disclosure.

[99] Referring to FIG. 8, an augmented reality image 81 including a virtual object 8 and user bodies 80a and 80b is illustrated. The electronic apparatus 100 may identify whether the user body 80a and 80b touch the virtual object 8 based on the estimated joint coordinates, and perform an event corresponding to the object touch when the object touch is detected. The event may refer to changing a transmittance of the rendered virtual hand object 80b.

[100] Meanwhile, the electronic apparatus 100 may change an alpha value of the virtual hand object 80b included in the generated augmented reality image by a unit of pixel. The electronic apparatus 100 may change a transmittance of the virtual hand object 80b when a touch of the user bodies 80a and 80b and the virtual object 8 is identified. Alternatively, the electronic apparatus 100 may change only the transmittance of one area of the virtual hand object 80b corresponding to the touch.

[101] Specifically, referring to FIG. 8, the virtual object 8 may be rendered by the electronic apparatus 100 and occupy a certain area in space (e.g., an area of a certain size of a cuboid). In addition, the electronic apparatus 100 may grasp 3D coordinates of the virtual object 8 and also estimate the 3D coordinates of the user bodies 80a and 80b. In other words, the electronic apparatus 100 may identify whether to touch the object by comparing the 3D coordinates of the virtual object 8 and the user's bodies 80a and 80b. In addition, the electronic apparatus 100 may identify whether a part of the user's bodies 80a and 80b touches the virtual object 8. The electronic apparatus 100 may change a transmittance of a part area (the virtual hand object 80b) of the user body determined to have generated object touch with the virtual object 8, or reverse the

color of the rendered virtual hand object 80b.

[102] FIG. 9 is a flowchart illustrating a method of controlling an electronic apparatus according to an embodiment of the disclosure.

[103] Referring to FIG. 9, the electronic apparatus 100 may include a camera, a display, and a processor. The electronic apparatus 100 may capture a rear of the electronic apparatus 100 facing a front of which the display displays an image (S910). The front and the rear are described for convenience of description, and may be first and second surfaces.

[104] The electronic apparatus 100 may render a virtual object based on the captured image (S920). Specifically, when an image captured by the camera is referred to as a first image or a first layer, the electronic apparatus 100 may generate a second image or a second layer including a virtual object based on the first image or the first layer.

[105] When the user body is detected from the captured image, the electronic apparatus 100 may estimate a plurality of joint coordinates for the detected user body using a pre-trained learning model (S930). Specifically, when the user body detected from the captured image is a hand, the electronic apparatus 100 may estimate a plurality of joint coordinates corresponding to the finger joint and the palm by using the pre-trained learning model. The pre-trained learning model may be learned through a plurality of learning data including hand images.

[106] The electronic apparatus 100 may generate an augmented reality image using the estimated plurality of joint coordinates, the rendered virtual object, and the captured image (S940). The electronic apparatus 100 may display the generated augmented reality image (S950). The electronic apparatus 100 may identify whether the user body touches the virtual object in the displayed augmented reality image, and perform an event corresponding to the object touch when the object touch is detected. In particular, the electronic apparatus 100 may identify whether the user body touches the virtual object based on the estimated joint coordinates, and change a transmittance of the virtual object when the touch is confirmed (S960). The electronic apparatus 100 may change an alpha value of the virtual object included in the generated augmented reality image by a unit of pixel, and the electronic apparatus 100 may identify an object touch based on whether the estimated joint coordinates are located at coordinates cor-responding to the rendered virtual object. Also, the electronic apparatus 100 may change only the transmittance of an area of the virtual object corresponding to the touch. In other words, the electronic apparatus 100 may change the transmittance of the virtual object by changing the alpha value of all pixels corresponding to the virtual object or by changing only the pixel alpha value of the area of the virtual object.

[107] The term “module” as used herein includes units made up of hardware, software, or firmware, and may be used interchangeably with terms such as logic, logic blocks,

components, or circuits. A “module” may be an integrally constructed component or a minimum unit or part thereof that performs one or more functions. For example, the module may be composed of an application-specific integrated circuit (ASIC).

[108] The various example embodiments described above may be implemented as an S/W program including an instruction stored on machine-readable (e.g., computer-readable) storage media. The machine is an apparatus which is capable of calling a stored instruction from the storage medium and operating according to the called instruction, and may include an electronic apparatus (e.g., an electronic apparatus A) according to the above-described example embodiments. When the instruction is executed by a processor, the processor may perform a function corresponding to the instruction directly or using other components under the control of the processor. The command may include a code generated or executed by a compiler or an interpreter. A machine-readable storage medium may be provided in the form of a non-transitory storage medium. Herein, the term “non-transitory” only denotes that a storage medium does not include a signal but is tangible, and does not distinguish the case where a data is semi-permanently stored in a storage medium from the case where a data is temporarily stored in a storage medium.

[109] The respective components (e.g., module or program) according to the various example embodiments may include a single entity or a plurality of entities, and some of the corresponding sub-components described above may be omitted, or another sub-component may be further added to the various example embodiments. Alternatively or additionally, some components (e.g., module or program) may be combined to form a single entity which performs the same or similar functions as the corresponding elements before being combined. Operations performed by a module, a program module, or other component, according to various embodiments, may be sequential, parallel, or both, executed iteratively or heuristically, or at least some operations may be performed in a different order, omitted, or other operations may be added.

Claims

[Claim 1] An electronic apparatus comprising:
a display;
a camera configured to capture a rear of the electronic apparatus facing a front of the electronic apparatus in which the display displays an image; and
a processor configured to:
render a virtual object based on the image captured by the camera,
based on a user body being detected from the captured image, estimate a plurality of joint coordinates with respect to the detected user body using a pre-trained learning model,
generate an augmented reality image using the estimated plurality of joint coordinates, the rendered virtual object, and the captured image, and
control the display to display the generated augmented reality image,
wherein the processor is further configured to:
identify whether the user body touches the rendered virtual object based on the estimated plurality of joint coordinates, and
change a transmittance of the rendered virtual object based on the touch being identified.

[Claim 2] The apparatus of claim 1, wherein the processor is further configured to estimate the plurality of joint coordinates corresponding to a finger joint and a palm using the pre-trained learning model based on the detected user body being identified to be a hand.

[Claim 3] The apparatus of claim 2, wherein the processor is further configured to render a virtual hand object and the rendered virtual object based on the estimated plurality of joint coordinates.

[Claim 4] The apparatus of claim 1, wherein the processor is further configured to change a transmittance of one area of the rendered virtual object corresponding to the touch.

[Claim 5] The apparatus of claim 1, wherein the processor is further configured to change a transmittance of the user body and transparently display the user body, based on the touch being identified.

[Claim 6] The apparatus of claim 1, wherein the processor is further configured to:
receive depth data of the captured image from the camera, and
generate the augmented reality image using the received depth data.

[Claim 7] The apparatus of claim 1, wherein the pre-trained learning model is configured to be trained through a plurality of learning data comprising hand images by using a convolutional neural network (CNN).

[Claim 8] The apparatus of claim 7, wherein the plurality of learning data comprises first data in which a 3D coordinate is matched to at least one area of a hand image, and second data in which the 3D coordinate is not matched to the hand image, and
wherein the pre-trained learning model is configured be trained by updating a weight value of the CNN based on the first data and the second data.

[Claim 9] A method of controlling an electronic apparatus comprising:
capturing a rear of the electronic apparatus facing a front of the electronic apparatus in which a display displays an image;
rendering a virtual object based on a captured image;
based on a user body being detected from the captured image, estimating a plurality of joint coordinates with respect to the detected user body using a pre-trained learning model;
generating an augmented reality image using the estimated plurality of joint coordinates, the rendered virtual object, and the captured image;
displaying the generated augmented reality image;
identifying whether the user body touches the rendered virtual object based on the estimated plurality of joint coordinates; and
changing a transmittance of the rendered virtual object based on the touch being identified.

[Claim 10] The method of claim 9, wherein the estimating comprises estimating a plurality of joint coordinates corresponding to a finger joint and a palm using the pre-trained learning model based on the detected user body being identified to be a hand.

[Claim 11] The method of claim 10, wherein the rendering comprises rendering a virtual hand object and the rendered virtual object based on the estimated plurality of joint coordinates.

[Claim 12] The method of claim 9, wherein the changing comprises changing a transmittance of one area of the rendered virtual object corresponding to the touch.

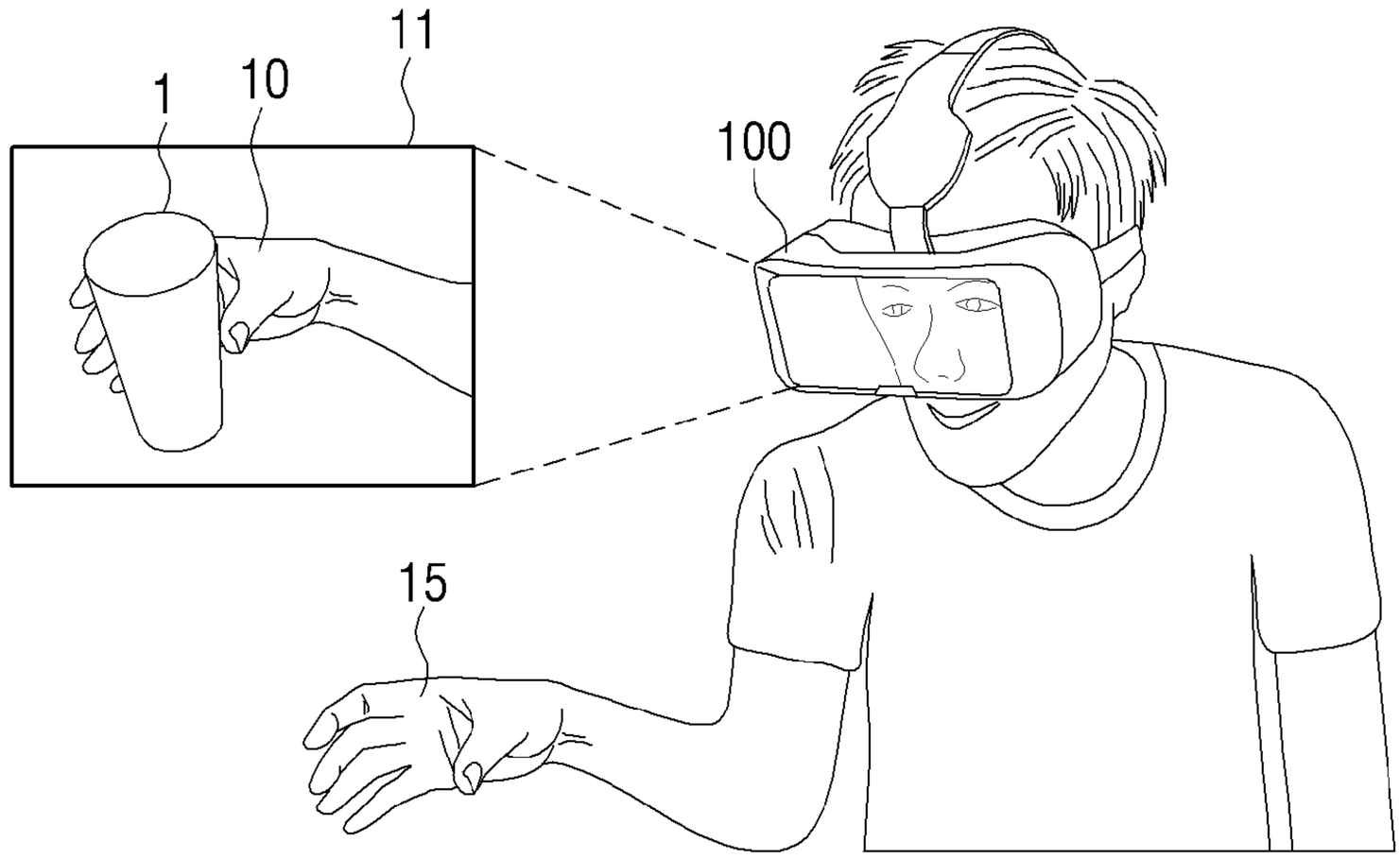
[Claim 13] The method of claim 9, further comprising:
changing a transmittance of the user body and transparently displaying the user body based on the touch being identified.

[Claim 14] The method of claim 9, wherein the generating comprises receiving

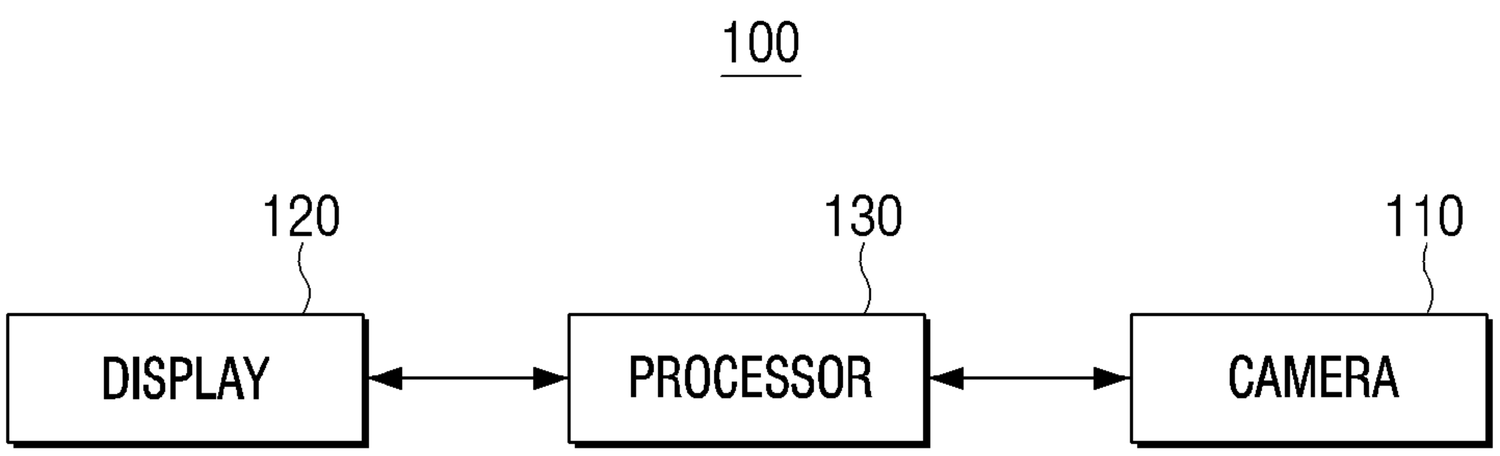
depth data of the captured image from a camera, and generating the augmented reality image using the received depth data.

[Claim 15] The method of claim 9, wherein the pre-trained learning model is configured to be trained through a plurality of learning data comprising hand images by using a convolutional neural network (CNN).

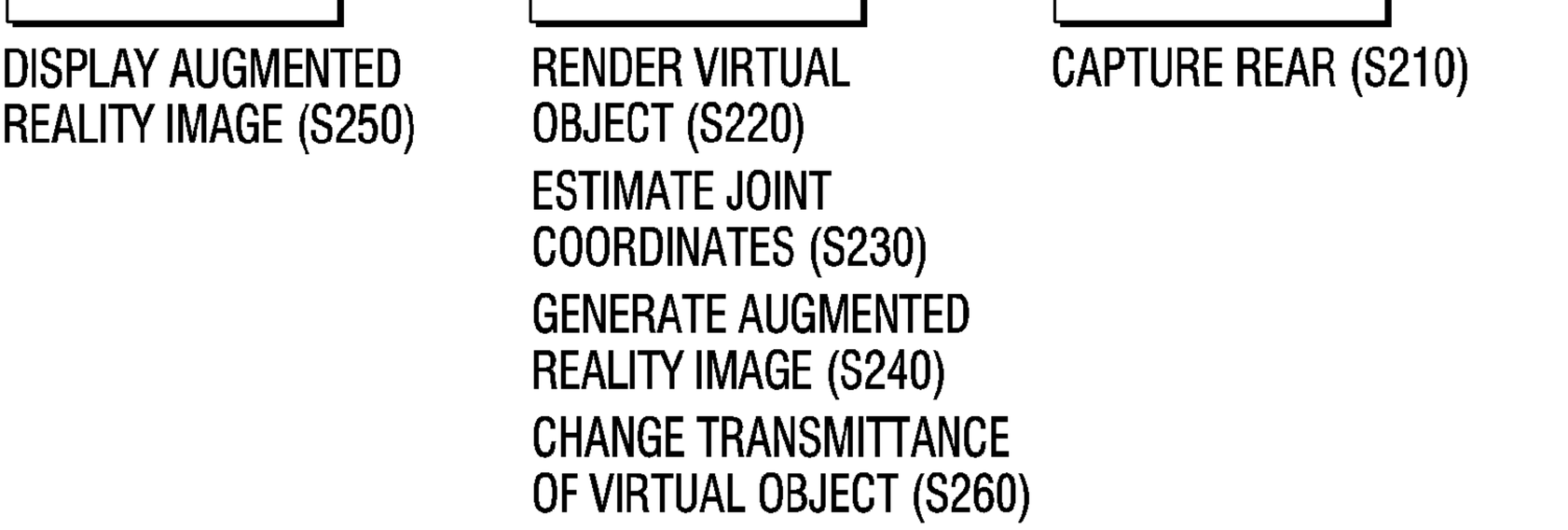
[Fig. 1]



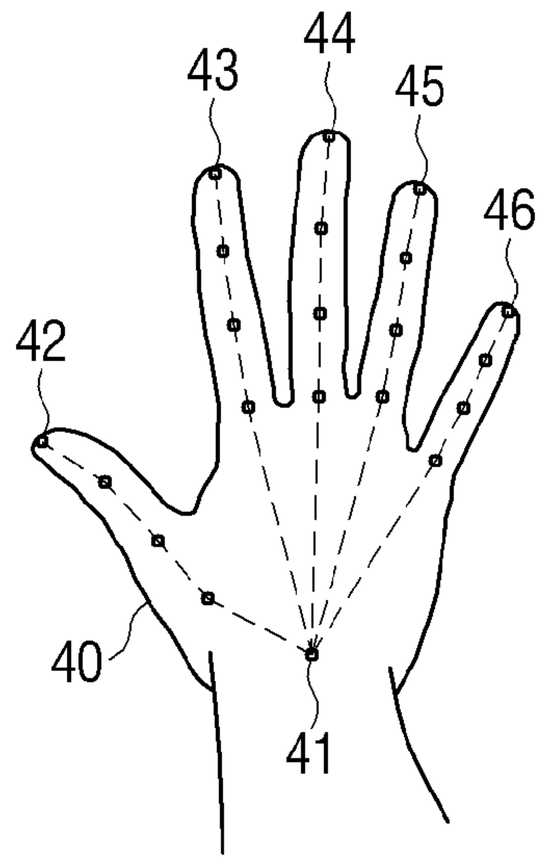
[Fig. 2]



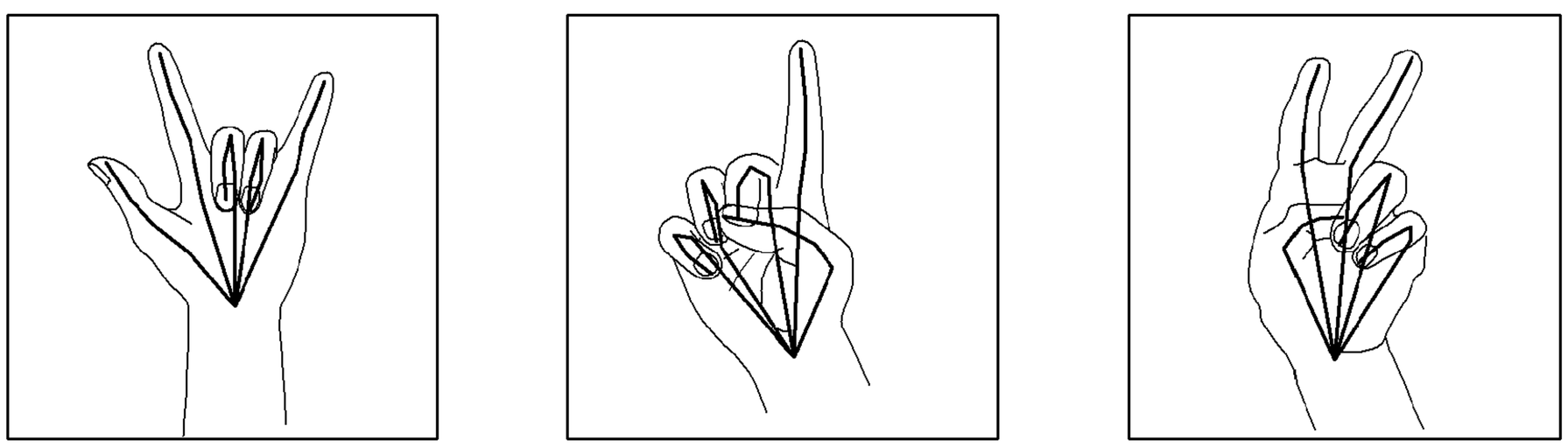
[Fig. 3]



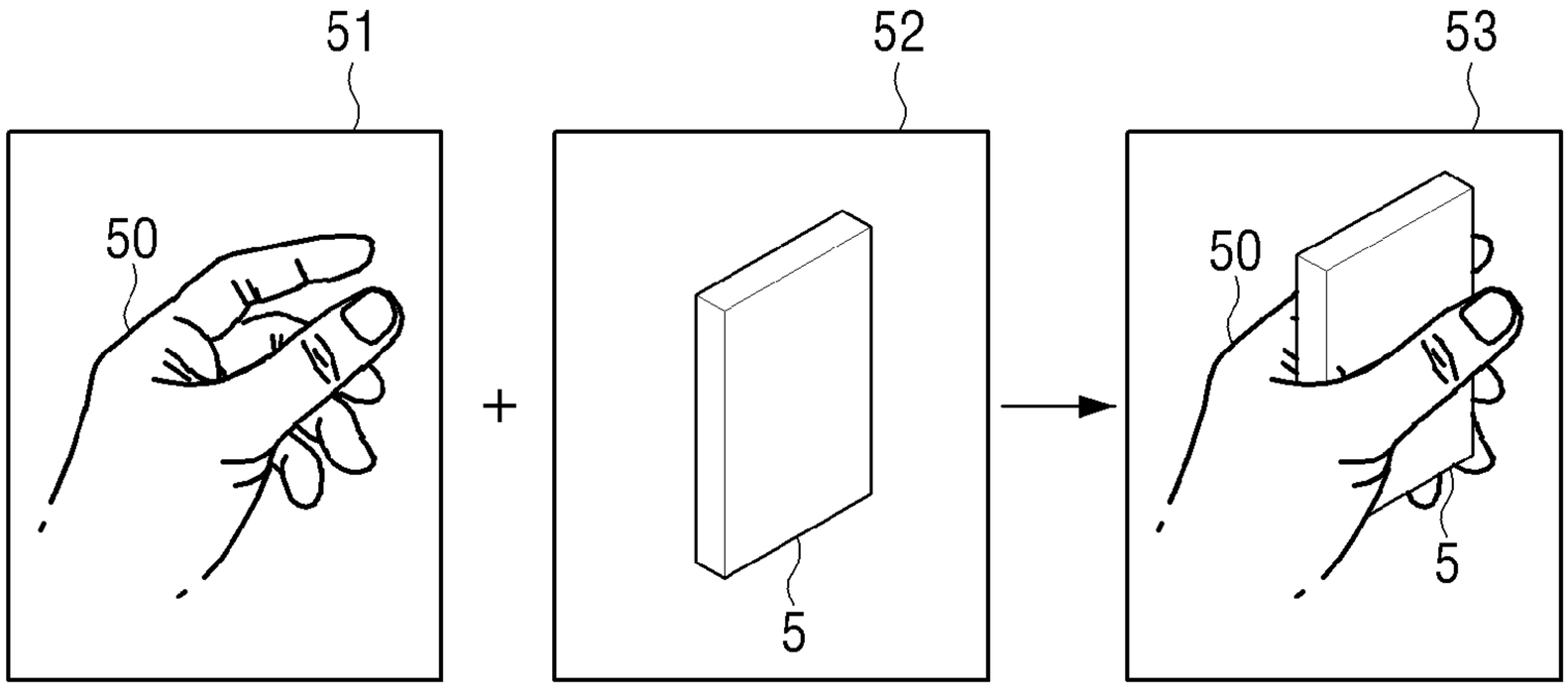
[Fig. 4A]



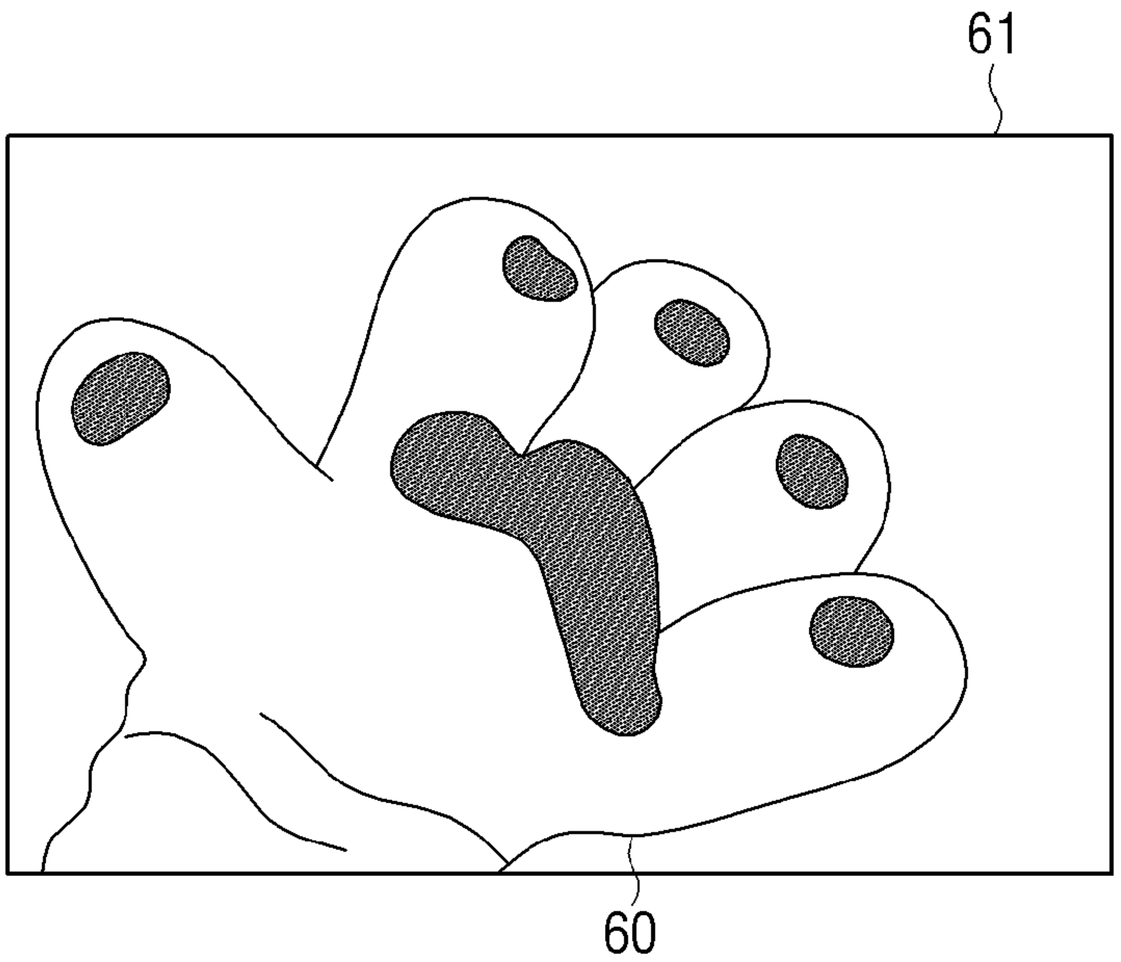
[Fig. 4B]



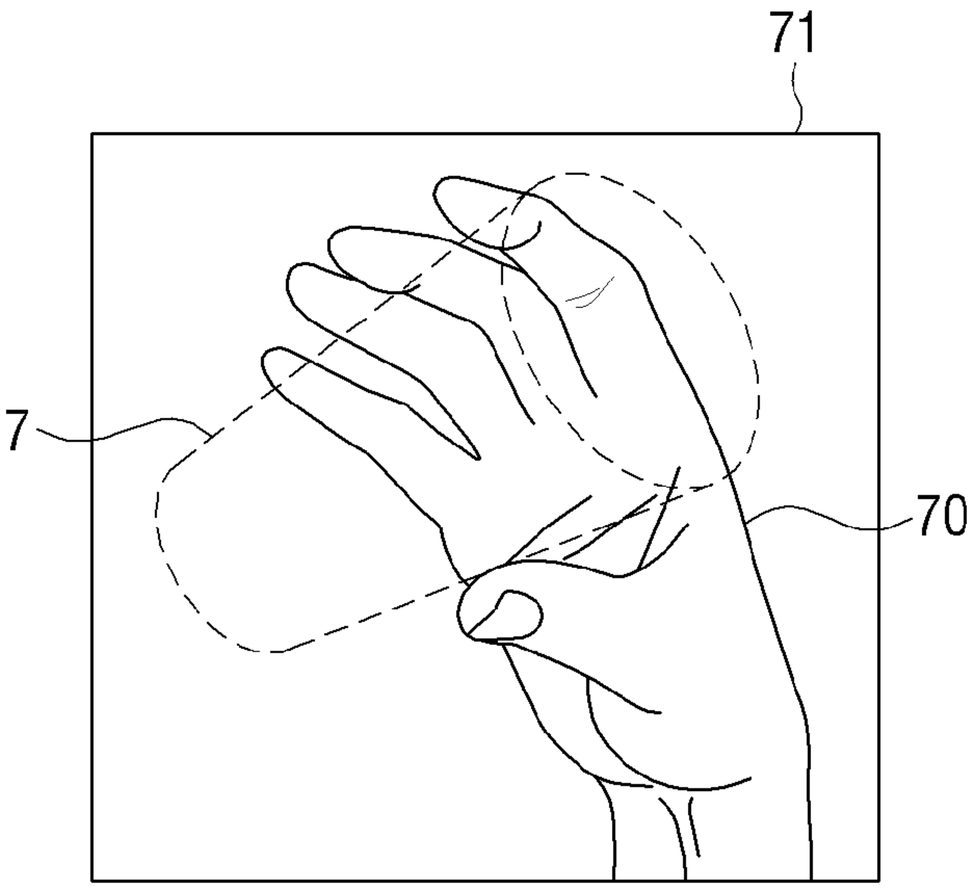
[Fig. 5]



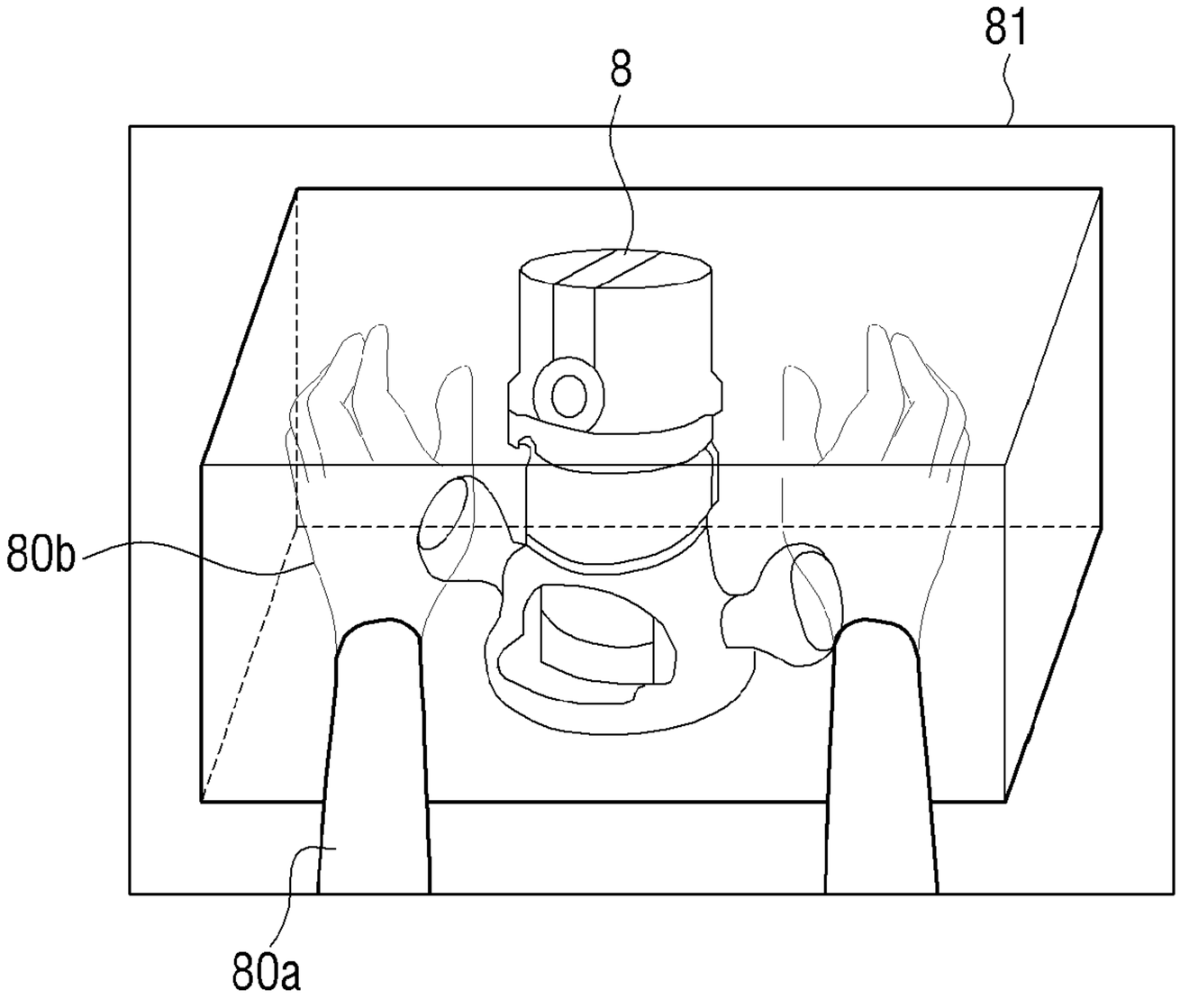
[Fig. 6]



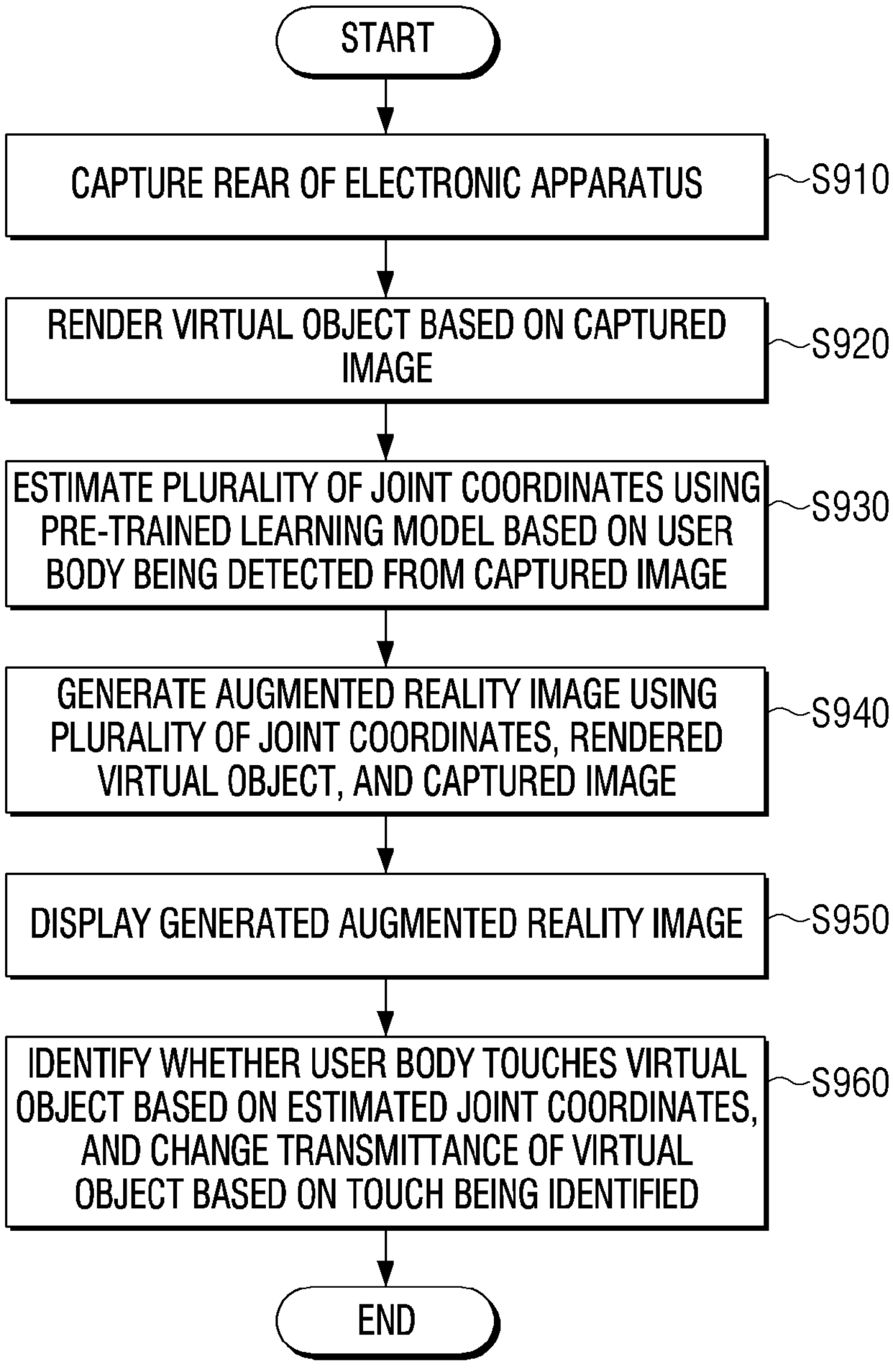
[Fig. 7]



[Fig. 8]



[Fig. 9]


START
CAPTURE REAR OF ELECTRONIC APPARATUS
S910
RENDER VIRTUAL OBJECT BASED ON CAPTURED IMAGE
S920
ESTIMATE PLURALITY OF JOINT COORDINATES USING PRE-TRAINED LEARNING MODEL BASED ON USER BODY BEING DETECTED FROM CAPTURED IMAGE
S930
GENERATE AUGMENTED REALITY IMAGE USING PLURALITY OF JOINT COORDINATES, RENDERED VIRTUAL OBJECT, AND CAPTURED IMAGE
S940
DISPLAY GENERATED AUGMENTED REALITY IMAGE
S950
IDENTIFY WHETHER USER BODY TOUCHES VIRTUAL OBJECT BASED ON ESTIMATED JOINT COORDINATES, AND CHANGE TRANSMITTANCE OF VIRTUAL OBJECT BASED ON TOUCH BEING IDENTIFIED
S960
END

INTERNATIONAL SEARCH REPORT

International application No.
PCT/KR2020/014721

A. CLASSIFICATION OF SUBJECT MATTER

G06T 19/00(2011.01)i; **G06T 15/00**(2006.01)i; **G06T 19/20**(2011.01)i; **G06F 3/041**(2006.01)i; **G06T 7/50**(2017.01)i; **G06N 3/08**(2006.01)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(2011.01); A61B 5/00(2006.01); G06F 3/01(2006.01); G06K 9/62(2006.01); G06N 3/02(2006.01); G06T 11/60(2006.01); G06T 7/73(2017.01); G09G 5/00(2006.01)

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: augmented reality, transparency, overlap, overlay, superposition, touch, hand, finger, Convolutional Neural Network (CNN), joint, coordinates

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2012-0249590 A1 (GIULIANO MACIOCCI et al.) 04 October 2012 (2012-10-04) paragraphs [0111], [0133], [0194], [0206], [0339]; and claims 4, 9, 39	1-15
Y	US 2018-0024641 A1 (USENS, INC.) 25 January 2018 (2018-01-25) paragraphs [0059], [0062]; and claim 10	1-15
A	KR 10-2018-0097949 A (OH CHI MIN) 03 September 2018 (2018-09-03) claims 1-3	1-15
A	WO 2018-071225 A1 (MICROSOFT TECHNOLOGY LICENSING, LLC) 19 April 2018 (2018-04-19) claims 1-12	1-15
A	US 10134166 B2 (AUGMEDICS LTD.) 20 November 2018 (2018-11-20) claims 1-7	1-15

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

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
09 February 2021	**10 February 2021**

Name and mailing address of the ISA/KR	Authorized officer
Korean Intellectual Property Office **189 Cheongsa-ro, Seo-gu, Daejeon** **35208, Republic of Korea**	**YANG, Jeong Rok**
Facsimile No. **+82-42-481-8578**	Telephone No. **+82-42-481-5709**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/KR2020/014721

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
US 2012-0249590 A1	04 October 2012	CN 103460255 A	18 December 2013
		CN 103460256 A	18 December 2013
		CN 103493106 A	01 January 2014
		CN 103562968 A	05 February 2014
		EP 2691935 A1	05 February 2014
		EP 2691936 A1	05 February 2014
		EP 2691937 A1	05 February 2014
		EP 2691938 A1	05 February 2014
		EP 3654146 A1	20 May 2020
		EP 3654147 A1	20 May 2020
		JP 2014-514652 A	19 June 2014
		JP 2014-514653 A	19 June 2014
		JP 2014-515854 A	03 July 2014
		JP 2014-518596 A	31 July 2014
		JP 2015-228256 A	17 December 2015
		KR 10-1591493 B1	03 February 2016
		KR 10-1591579 B1	18 February 2016
		KR 10-1818024 B1	12 January 2018
		KR 10-2013-0136566 A	12 December 2013
		KR 10-2016-0084502 A	13 July 2016
		US 2012-0249416 A1	04 October 2012
		US 2012-0249544 A1	04 October 2012
		US 2012-0249591 A1	04 October 2012
		US 2012-0249741 A1	04 October 2012
		WO 2012-135545 A1	04 October 2012
		WO 2012-135546 A1	04 October 2012
		WO 2012-135547 A1	04 October 2012
		WO 2012-135553 A1	04 October 2012
		WO 2012-135554 A1	04 October 2012
US 2018-0024641 A1	25 January 2018	CN 108369643 A	03 August 2018
		EP 3488324 A1	29 May 2019
		WO 2018-017399 A1	25 January 2018
KR 10-2018-0097949 A	03 September 2018	None	
WO 2018-071225 A1	19 April 2018	CN 109844820 A	04 June 2019
		EP 3526774 A1	21 August 2019
		US 2018-0108325 A1	19 April 2018
US 10134166 B2	20 November 2018	EP 3274985 A1	31 January 2018
		US 2017-0178375 A1	22 June 2017
		US 2018-0182150 A1	28 June 2018
		US 2019-0043238 A1	07 February 2019
		US 2019-0273916 A1	05 September 2019
		US 9928629 B2	27 March 2018
		WO 2016-151506 A1	29 September 2016