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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 February 2025 (27.02.2025)

WIPO | PCT

(10) International Publication Number
WO 2025/041222 A1

(51) International Patent Classification:
G06T 7/00 (2017.01)

(21) International Application Number:
PCT/JP2023/029987

(22) International Filing Date:
21 August 2023 (21.08.2023)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **NEC CORPORATION** [JP/JP]; 7-1, Shiba 5-chome, Minato-ku, Tokyo, 1088001 (JP).

(72) Inventors: **RODRIGUES Royston**; c/o NEC Corporation, 7-1, Shiba 5-chome, Minato-ku, Tokyo, 1088001 (JP). **TANI Masahiro**; c/o NEC Corporation, 7-1, Shiba 5-chome, Minato-ku, Tokyo, 1088001 (JP).

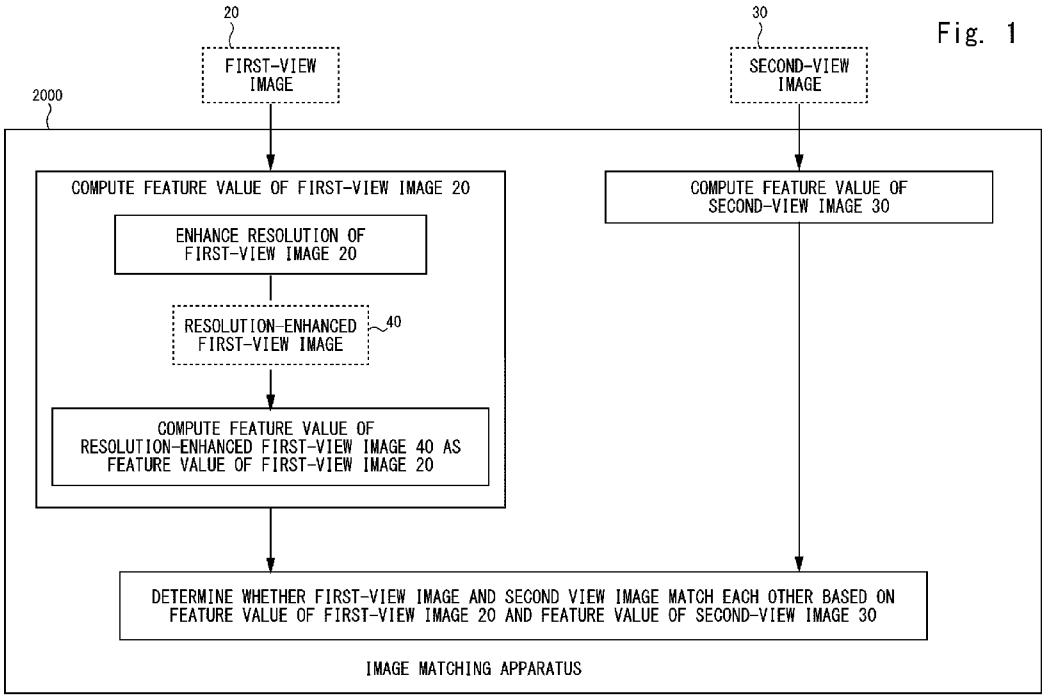
(74) Agent: **IEIRI Takeshi**; HIBIKI IP Law Firm, Urban Center Yokohama West 5th Floor, 3-33-8, Tsuruya-cho, Kanagawa-ku, Yokohama-shi, Kanagawa, 2210835 (JP).

(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, CV,

(54) Title: IMAGE MATCHING APPARATUS, IMAGE MATCHING METHOD, TRAINING APPARATUS, TRAINING METHOD, AND NON-TRANSITORY COMPUTER-READABLE MEDIUM

Fig. 1



(57) Abstract: An image matching apparatus is configured to: acquire a first-view image and a second-view image; compute a feature value of the first-view image; compute a feature value of the second-view image; and determine whether the first-view image and the second-view image match each other based on the feature value of the first-view image and the feature value of the second-view image. The computation of the feature value of the first-view image includes: enhancing a resolution of the first-view image to generate a resolution-enhanced first-view image; and computing a feature value of the resolution-enhanced first-view image as the feature value of the first-view image. The first-view image and the second-view image are respectively a ground-view image and an aerial-view image, or the first-view image and the second-view image are respectively an aerial-view image and a ground-view image.

WO 2025/041222 A1

[Continued on next page]

GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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))*

Description

Title of Invention: IMAGE MATCHING APPARATUS, IMAGE MATCHING METHOD, TRAINING APPARATUS, TRAINING METHOD, AND NON-TRANSITORY COMPUTER-READABLE MEDIUM

Technical Field

[0001] The present disclosure generally relates to an image matching apparatus, an image matching method, a training apparatus, a training method, and a non-transitory computer-readable medium.

Background Art

[0002] A computer system that performs ground-to-aerial cross-view matching (matching between a ground-view image and an aerial-view image) has been developed. For example, NPL1 discloses a system comprising a set of CNNs (Convolutional Neural Networks) to match a ground-view image against an aerial-view image. Specifically, one of the CNNs acquires a set of a ground-view image and orientation maps that indicate orientations (azimuth and altitude) for each location captured on the ground-view image, and extracts features therefrom. The other one acquires a set of an aerial-view image and orientation maps that indicate orientations (azimuth and range) for each location captured on the aerial-view image, and extracts features therefrom. Then, the system determines whether the ground-view image matches the aerial-view image based on the extracted features.

Citation List

Non Patent Literature

[0003] NPL1: Liu Liu and Hongdong Li, "Lending Orientation to Neural Networks for Cross-view Geo-localization", [online], March 29, 2019, [retrieved on 2021-09-24], retrieved from <arXiv, https://arxiv.org/pdf/1903.12351>

Summary of Invention

Technical Problem

[0004] NPL1 does not mention resolution of images. An objective of the present disclosure is to provide a novel technique to determine whether or not a ground-view image and an aerial-view image match each other.

Solution to Problem

[0005] The present disclosure provides an image matching apparatus comprising at least one memory that is configured to store instructions and at least one processor that is configured to execute the instructions to: acquire a first-view image and a second-view

image; compute a feature value of the first-view image; compute a feature value of the second-view image; and determine whether the first-view image and the second-view image match each other based on the feature value of the first-view image and the feature value of the second-view image.

The computation of the feature value of the first-view image includes: enhancing a resolution of the first-view image to generate a resolution-enhanced first-view image; and computing a feature value of the resolution-enhanced first-view image as the feature value of the first-view image.

The first-view image and the second-view image are respectively a ground-view image and an aerial-view image, or the first-view image and the second-view image are respectively an aerial-view image and a ground-view image.

[0006] The present disclosure further provides an image matching method that is performed by a computer. The image matching method comprises: acquiring a first-view image and a second-view image; computing a feature value of the first-view image; computing a feature value of the second-view image; and determining whether the first-view image and the second-view image match each other based on the feature value of the first-view image and the feature value of the second-view image.

The computation of the feature value of the first-view image includes: enhancing a resolution of the first-view image to generate a resolution-enhanced first-view image; and computing a feature value of the resolution-enhanced first-view image as the feature value of the first-view image.

The first-view image and the second-view image are respectively a ground-view image and an aerial-view image, or the first-view image and the second-view image are respectively an aerial-view image and a ground-view image.

[0007] The present disclosure further provides a non-transitory computer-readable medium storing a program that cause a computer to execute: acquiring a first-view image and a second-view image; computing a feature value of the first-view image; computing a feature value of the second-view image; and determining whether the first-view image and the second-view image match each other based on the feature value of the first-view image and the feature value of the second-view image.

The computation of the feature value of the first-view image includes: enhancing a resolution of the first-view image to generate a resolution-enhanced first-view image; and computing a feature value of the resolution-enhanced first-view image as the feature value of the first-view image.

The first-view image and the second-view image are respectively a ground-view image and an aerial-view image, or the first-view image and the second-view image are respectively an aerial-view image and a ground-view image.

[0008] The present disclosure further provides a training apparatus comprising at least one

memory that is configured to store instructions and at least one processor that is configured to execute the instructions to: acquire a target image; and perform training of one or more models.

The training of the one or more models includes: reducing a resolution of the acquired target image to generate a resolution-reduced target image; inputting the resolution-reduced target image into a resolution enhancing model to generate a resolution-reenhanced target image; computing a loss using the acquired target image and the resolution-reenhanced target image; and updating the resolution enhancing model based on the loss.

The target image is a ground-view image or an aerial-view image.

[0009] The present disclosure further provides a training method that is performed by a computer. The training method comprises: acquiring a target image; and performing training of one or more models.

The training of the one or more models includes: reducing a resolution of the acquired target image to generate a resolution-reduced target image; inputting the resolution-reduced target image into a resolution enhancing model to generate a resolution-reenhanced target image; computing a loss using the acquired target image and the resolution-reenhanced target image; and updating the resolution enhancing model based on the loss.

The target image is a ground-view image or an aerial-view image.

[0010] The present disclosure further provides a non-transitory computer-readable medium storing a program that causes a computer to execute: acquiring a target image; and performing training of one or more models.

The training of the one or more models includes: reducing a resolution of the acquired target image to generate a resolution-reduced target image; inputting the resolution-reduced target image into a resolution enhancing model to generate a resolution-reenhanced target image; computing a loss using the acquired target image and the resolution-reenhanced target image; and updating the resolution enhancing model based on the loss.

The target image is a ground-view image or an aerial-view image.

[0011] According to the present disclosure, it is possible to provide a novel technique to determine whether a ground-view image and an aerial-view image match each other.

Brief Description of Drawings

[0012] [Fig.1]Fig. 1 illustrates an overview of an image matching apparatus.

[Fig.2]Fig. 2 illustrates an example of the ground-view image and the aerial-view image.

[Fig.3]Fig. 3 illustrates an overview of the image matching apparatus that handles the

second-view image with the resolution of the first level.

[Fig.4]Fig. 4 is a block diagram showing an example of the functional configuration of the image matching apparatus

[Fig.5]Fig. 5 is a block diagram illustrating an example of the hardware configuration of a computer 1000 realizing the image matching apparatus.

[Fig.6]Fig. 6 is a flowchart illustrating an example flow of processing performed by the image matching apparatus.

[Fig.7]Fig. 7 illustrates a geo-localization system that includes the image matching apparatus.

[Fig.8]Fig. 8 illustrates an example structure of the first feature extracting unit.

[Fig.9]Fig. 9 illustrates a first example of a structure of the second feature extracting unit.

[Fig.10]Fig. 10 illustrates a second example of a structure of the second feature extracting unit.

[Fig.11]Fig. 11 is a diagram illustrating an example of functional configuration of the training apparatus.

[Fig.12]Fig. 12 is a flowchart illustrating an example flow of processing performed by the training apparatus.

[Fig.13]Fig. 13 illustrates a first example way of training the first feature extracting unit.

[Fig.14]Fig. 14 illustrates an image matching apparatus that handles the first-view image with the resolution of the second level and the second-view image with the resolution of the second level.

[Fig.15]Fig. 15 illustrates an additional training performed on the feature extracting model.

[Fig.16]Fig. 16 illustrates an example way of training the resolution enhancing model separately from the feature extracting model.

[Fig.17]Fig. 17 illustrates a first example way of training the second feature extracting unit.

[Fig.18]Fig. 18 illustrates an additional training performed on the feature extracting model.

[Fig.19]Fig. 19 illustrates an example way of training the resolution enhancing model separately from the feature extracting model.

Description of Embodiments

[0013] Example embodiments according to the present disclosure will be described hereinafter with reference to the drawings. The same numeral signs are assigned to the same elements throughout the drawings, and redundant explanations are omitted as

necessary. In addition, predetermined information (e.g., a predetermined value or a predetermined threshold) is stored in advance in a storage device to which a computer using that information has access unless otherwise described.

[0014] FIRST EXAMPLE EMBODIMENT

<Overview>

Fig. 1 illustrates an overview of an image matching apparatus 2000. The image matching apparatus 2000 functions as a discriminator that performs matching between a ground-view image and an aerial-view image (so-called ground-to-aerial cross-view matching). Fig. 2 illustrates an example of the ground-view image 10 and the aerial-view image 15.

[0015] The ground-view image 10 is a digital image that includes a ground view of a place, e.g., an RGB or gray-scale image of ground scenery. For example, the ground-view image is generated by a ground camera that is held by a pedestrian or installed in a car. The ground-view image may be panoramic (having 360-degree field of view), or may have limited (less than 360-degree) field of view.

[0016] The aerial-view image 15 is a digital image that includes a top view of a place, e.g., an RGB or gray-scale image of aerial scenery. For example, the aerial-view image is generated by an aerial camera installed in a drone, an air plane, or a satellite.

[0017] As images to be compared with each other, the image matching apparatus 2000 acquires a first-view image 20 and a second-view image 30. One of them is a ground-view image 10 and the other one of them is an aerial-view image 15. In other words, when the image matching apparatus 2000 is configured to acquire a ground-view image 10 as the first-view image 20, the image matching apparatus 2000 is configured to acquire an aerial-view image 15 as the second-view image 30. On the other hand, when the image matching apparatus 2000 is configured to acquire an aerial-view image 15 as the first-view image 20, the image matching apparatus 2000 is configured to acquire a ground-view image 10 as the second-view image 30.

[0018] The image matching apparatus 2000 is used under a situation where the first-view image 20 has a resolution of a first level while the second-view image 30 has a resolution of a first level or a second level. The first level of resolution may include only a specific image resolution or may include image resolutions within a specific rage. The same applies to the second level of resolution.

[0019] The first level and the second level of resolution are defined so that the resolution of the second level is higher than the resolution of the first level. Suppose that the resolution is quantified so that the higher a resolution of an image is, the smaller a value representing the resolution of the image is. An example of this case is a case where a unit of "cm/pixel" is used to represent the resolution of the image.

[0020] In this case, if the first level and the second level of resolution are respectively

defined by specific real numbers r1 and r2, the numbers r1 and r2 satisfy a condition of "r1>r2". For example, the first level of resolution is defined as "50cm/pixel" while the second level of resolution is defined as "25cm/pixel". If the first level and the second level of resolution are respectively defined by rages R1 and R2, the ranges R1 and R2 satisfy a condition of "the supremum of the rage R2 is less than the infimum of the rage R1". For example, the first level of resolution is defined as a rage of (45[cm/pixel], 55[cm/pixel]) while the second level of resolution is defined as a range of (20[cm/pixel], 30[cm/pixel]).

[0021] It is noted that the levels of the resolution of first-view image and the levels of the resolution of second-view image may be defined separately from each other. It means that the first level of the resolution of first-view image is not necessarily equivalent to the first level of the resolution of second-view image. Similarly, the second level of the resolution of first-view image is not necessarily equivalent to the second level of the resolution of second-view image.

[0022] Since the resolution of the first-view image 20 is of the first level, the image matching apparatus 2000 performs resolution enhancement on the first-view image 20 to increase the level of the resolution of the first-view image 20 to the second level. A first-view image that is obtained as a result of the resolution enhancement is called "resolution-enhanced first-view image 40". The image matching apparatus 2000 computes a feature value of the resolution-enhanced first-view image 40 as the feature value of the first-view image 20.

[0023] Regarding the second-view image 30, the image matching apparatus 2000 does not perform resolution enhancement on the second-view image 30 when the resolution of the second-view image 30 is of the second level. In this case, the image matching apparatus 2000 computes the feature value of the second-view image 30 directly therefrom. Fig. 1 illustrates this case.

[0024] On the other hand, when the resolution of the second-view image 30 is of the first level, the image matching apparatus 2000 performs resolution enhancement on the second-view image 30 to increase the level of the resolution of the second-view image 30 to the second level. Fig. 3 illustrates an overview of the image matching apparatus 2000 that handles the second-view image 30 with the resolution of the first level. A second-view image that is obtained as a result of the resolution enhancement is called "resolution-enhanced second-view image 50". The image matching apparatus 2000 computes a feature value of the resolution-enhanced second-view image 50 as the feature value of the second-view image 30.

[0025] <Example of Advantageous Effect>

As described above, the image matching apparatus 2000 acquires the first-view image 20 and the second-view image 30, and determine whether the first-view image

20 and the second-view image 30 match each other by comparing their feature values. To compute the feature value of the first-view image 20, the first-view image 20 is converted into the resolution-enhanced first-view image 40 by performing resolution enhancement on the first-view image 20. The feature value of the resolution-enhanced first-view image 40 is used as the feature value of the first-view image 20.

[0026] Since the image matching apparatus 2000 has a mechanism of enhancing the resolution of the first-view image 20 before computing the feature value thereof, the image matching apparatus 2000 can handle first-view images whose resolution is not high enough to determine whether the first-view image matches the second-view image 30. Furthermore, in the case where the image matching apparatus 2000 has a mechanism of enhancing the resolution of the second-view image 30 before computing the feature value thereof, the image matching apparatus 2000 can handle second-view images whose resolution is not sufficiently high.

[0027] The image matching apparatus 2000 is useful in various situations. One of such the situations is a situation in which it is difficult to make it sure to acquire a first-view image with high resolution. Suppose that the first-view image 20 is a ground-view image provided by a user. The user may use her/his smartphone to take a picture around her/him, and that picture may be provided as the first-view image 20.

[0028] In this case, the resolution of the first-view image 20 depends on the performance of the smartphone. Thus, there may be some smartphones that provide the first-view image with low resolution to the image matching apparatus 2000. The image matching apparatus 2000 can determine whether the first-view image 20 and the second-view image 30 match each other even in this case.

[0029] Hereinafter, more detailed explanation of the image matching apparatus 2000 will be described.

[0030] <Example of Functional Configuration>

Fig. 4 is a block diagram showing an example of the functional configuration of the image matching apparatus 2000. The image matching apparatus 2000 includes an acquiring unit 2020, a first feature extracting unit 2040, a second feature extracting unit 2060, and a determining unit 2080. The acquiring unit 2020 acquires the first-view image 20 and the second-view image 30. The first feature extracting unit 2040 computes the feature value of the first-view image 20. Specifically, the first feature extracting unit 2040 enhances the resolution of the first-view image 20 to generate the resolution-enhanced first-view image 40. Then, the first feature extracting unit 2040 computes the feature value of the resolution-enhanced first-view image 40 as the feature value of the first-view image 20.

[0031] The second feature extracting unit 2060 computes the feature value of the second-view image 30. The determining unit 2080 determines whether the first-view image 20

and the second-view image 30 match each other based on the feature value of the first-view image 20 and the feature value of the second-view image 30.

[0032] It is noted that, in a case where the image matching apparatus 2000 is configured to acquire the second-view image with the resolution of the first level, the second feature extracting unit 2060 is configured to enhance the resolution of the second-view image 30 to generate the resolution-enhanced second-view image 50. The second feature extracting unit 2060 is also configured to compute the feature value of the resolution-enhanced second-view image 50 as the second-view image 30.

[0033] <Example of Hardware Configuration>

The image matching apparatus 2000 may be realized by one or more computers. Each of the one or more computers may be a special-purpose computer manufactured for implementing the image matching apparatus 2000, or may be a general-purpose computer like a personal computer (PC), a server machine, or a mobile device.

[0034] The image matching apparatus 2000 may be realized by installing an application in the one or more computers. The application is implemented with a program that causes the one or more computers to function as the image matching apparatus 2000. In other words, the program is an implementation of the functional units of the image matching apparatus 2000.

[0035] Fig. 5 is a block diagram illustrating an example of the hardware configuration of a computer 1000 realizing the image matching apparatus 2000. In Fig. 5, the computer 1000 includes a bus 1020, a processor 1040, a memory 1060, a storage device 1080, an input/output (I/O) interface 1100, and a network interface 1120.

[0036] The bus 1020 is a data transmission channel in order for the processor 1040, the memory 1060, the storage device 1080, and the I/O interface 1100, and the network interface 1120 to mutually transmit and receive data. The processor 1040 is a processer, such as a CPU (Central Processing Unit), GPU (Graphics Processing Unit), FPGA (Field-Programmable Gate Array), or DSP (Digital Signal Processor). The memory 1060 is a primary memory component, such as a RAM (Random Access Memory) or a ROM (Read Only Memory). The storage device 1080 is a secondary memory component, such as a hard disk, an SSD (Solid State Drive), or a memory card. The I/O interface 1100 is an interface between the computer 1000 and peripheral devices, such as a keyboard, mouse, or display device. The network interface 1120 is an interface between the computer 1000 and a network. The network may be a LAN (Local Area Network) or a WAN (Wide Area Network).

[0037] The storage device 1080 may store the program mentioned above. The processor 1040 reads the program from the storage device 1080, and executes the program to realize each functional unit of the image matching apparatus 2000.

[0038] The hardware configuration of the computer 1000 is not restricted to that shown in

Fig. 5. For example, as mentioned-above, the image matching apparatus 2000 may be realized by plural computers. In this case, those computers may be connected with each other through the network.

[0039] <Flow of Processing>

Fig. 6 is a flowchart illustrating an example flow of processing performed by the image matching apparatus 2000. The acquiring unit 2020 acquires the first-view image 20 and the second-view image 30 (S102). The first feature extracting unit 2040 computes the feature value of the first-view image 20 (S104). The second feature extracting unit 2060 computes the feature value of the second-view image 30 (S106). The determining unit 2080 determines whether the first-view image 20 and the second-view image 30 match each other based on the feature value of the first-view image 20 and the feature value of the second-view image 30 (S108).

[0040] It is noted that a flow of processing performed by the image matching apparatus 2000 is not limited to that illustrated by Fig. 6. For example, the computation of the feature value of the first-view image 20 (i.e., Step S104) and the computation of the feature value of the second-view image 30 (i.e., Step S106) may be performed in parallel with each other or in an order opposite to the order illustrated by Fig. 6.

[0041] <Example Application of Image Matching Apparatus 2000>

There are various possible applications of the image matching apparatus 2000. For example, the image matching apparatus 2000 can be used as a part of a system (hereinafter, a geo-localization system) that performs image geo-localization. Image geo-localization is a technique to determine the place at which an input image is captured. The geo-localization system 500 may be implemented by one or more arbitrary computers such as ones depicted by Fig. 5. It is noted that the geo-localization system is merely an example of the application of the image matching apparatus 2000, and the application of the image matching apparatus 2000 is not restricted to being used in the geo-localization system.

[0042] Fig. 7 illustrates a geo-localization system 500 that includes the image matching apparatus 2000. The geo-localization system 500 includes the image matching apparatus 2000 and the location database 600. The location database 600 includes a plurality of aerial-view images to each of which location information is attached. An example of the location information may be GPS (Global Positioning System) coordinates of the place captured on the center of the corresponding aerial-view image.

[0043] The geo-localization system 500 receives a query that includes a ground-view image from a client (e.g., user terminal). Then, the geo-localization system 500 searches the location database 600 for the aerial-view image that matches the ground-view image in the received query, thereby determining the place at which the ground-view image is captured. Specifically, until the aerial-view image that matches the

ground-view image in the query is detected, the geo-localization system 500 repeatedly executes: acquiring one of the aerial-view images from the location database 600; inputting the ground-view image acquired from the query and the aerial-view acquired from the location database 600 into the image matching apparatus 2000; and determining whether the output of the image matching apparatus 2000 indicates that the ground-view image and the aerial-view image match each other. It is noted that, as mentioned above, the image matching apparatus 2000 may be configured to handle the ground-view image and the aerial-view image as the first-view image 20 and the second-view image 30, respectively, or may be configured to handle the ground-view image and the aerial-view image as the second-view image 30 and the first-view image 20, respectively.

[0044] By repeatedly executing the above-mentioned processes, the geo-localization system 500 can find the aerial-view image that includes the place at which the ground-view image is captured. Since the detected aerial-view image is associated with the location information such as the GPS coordinates, the geo-localization system 500 can determine that where the ground-view image is captured is the place that is indicated by the location information associated with the aerial-view image that matches the ground-view image.

[0045] It is noted that the ground-view image and the aerial-view image are used in an opposite way in the geo-localization system 500. In this case, the location database 600 stores a plurality of ground-view images to each of which the location information is attached. The geo-localization system 500 receives a query including an aerial-view image, and searches the location database 600 for the ground-view image that matches the aerial-view image in the query, thereby determining the location of the place that is captured on the aerial-view image.

[0046] <Acquisition of Data: S102>

The acquiring unit 2020 acquires the first-view image 20 and the second-view image 30 (S102). There are various ways to acquire those data. In some implementations, the acquiring unit 2020 may receive the first-view image 20, the second-view image 30, or both sent from another computer. In other implementations, the acquiring unit 2020 may retrieve the first-view image 20, the second-view image 30, or both from a storage device to which the acquiring unit 2020 has access.

[0047] The first-view image 20 and the second-view image 30 may be acquired in the manner same as each other or may be acquired in different manners from each other. For example, the acquiring unit 2020 receives the first-view image 20 from another computer while the acquiring unit 2020 retrieves the second-view image 30 from a storage device, or vice versa.

[0048] <Computation of Feature Value of First-View Image: S104>

The first feature extracting unit 2040 computes the feature value of the first-view image 20 (S104). As described above, resolution enhancement is performed on the first-view image 20 to generate the resolution-enhanced first-view image 40. Then, the first feature extracting unit 2040 computes a feature value of the resolution-enhanced first-view image 40 as the feature value of the first-view image 20.

[0049] Fig. 8 illustrates an example structure of the first feature extracting unit 2040. The first feature extracting unit 2040 may include two machine learning-based models (e.g., neural networks) called "resolution enhancing model 100" and "feature extracting model 110".

[0050] The resolution enhancing model 100 is configured to take an image as input, and output another image whose size is the same as the input image. In addition, the resolution enhancing model 100 is pre-trained to, in response to a first-view image with a resolution of the first level being input thereinto, enhance the resolution of the input first-view image to the second level and thereby generate a first-view image with a resolution of the second level. How to train the resolution enhancing model 100 will be explained later.

[0051] The feature extracting model 110 is configured to take an image as input, and output a value (e.g., vector or tensor) that is computed based on the input image. In addition, the feature extracting model 110 is pre-trained to, in response to a first-view image with a resolution of the second level being input thereinto, compute a feature value of that input image. How to train the first feature extracting model 110 will be explained later.

[0052] The first feature extracting unit 2040 inputs the first-view image 20 acquired by the acquiring unit 2020 into the resolution enhancing model 100. As a result, the resolution enhancing model 100 outputs the resolution-enhanced first-view image 40. Then, the resolution-enhanced first-view image 40 is fed into the feature extracting model 110. As a result, the feature extracting model 110 outputs the feature value of the resolution-enhanced first-view image 40. The feature value of the resolution-enhanced first-view image 40 is handled as the feature value of the first-view image 20 by the determining unit 2080.

[0053] <Computation of Feature Value of Second-View Image: S106>

The second feature extracting unit 2060 computes the feature value of the second-view image 30 (S106). Fig. 9 illustrates a first example of a structure of the second feature extracting unit 2060. In the example illustrated by Fig. 9, it is assumed that the second-view image 30 has a resolution of the second level. Thus, the second feature extracting unit 2060 does not have a function of enhancing the second-view image 30. Specifically, the second feature extracting model 2060 may include a machine learning-based model (e.g., neural network) called "feature extracting model 120".

[0054] The feature extracting model 120 is configured to take an image as input, and output a value (e.g., vector or tensor) that is computed based on the input image. In addition, the feature extracting model 120 is pre-trained to, in response to a second-view image with a resolution of the second level being input thereinto, compute a feature value of that input image. How to train the second feature extracting model 120 will be explained later.

[0055] The second feature extracting unit 2060 inputs the second-view image 30 acquired by the acquiring unit 2020 into the feature extracting model 120, thereby acquiring the feature value of the second-view image 30 that is output by the feature extracting model 120.

[0056] When the second-view image 30 has a resolution of the first level, the second feature extraction unit 2060 further includes a function of enhancing a resolution of the second-view image 30 to the second level to generate the resolution-enhanced second-view image 50. Fig. 10 illustrates a second example of a structure of the second feature extracting unit 2060. In the example illustrated by Fig. 10, it is assumed that the second-view image 30 has a resolution of the first level. The second feature extracting unit 2060 includes a resolution enhancing model 130 as well as the feature extracting model 120.

[0057] The resolution enhancing model 130 is configured to take an image as input, and output another image whose size is the same as the input image. In addition, the resolution enhancing model 130 is pre-trained to, in response to a second-view image with a resolution of the first level being input thereinto, enhance the resolution of the input second-view image to the second level and thereby generate a second-view image with a resolution of the second level. How to train the resolution enhancing model 130 will be explained later.

[0058] In the case where the image matching apparatus 2000 is configured to handle the second-view image 30 with the resolution of the first level, the second feature extracting unit 2060 inputs the second-view image 30 acquired by the acquiring unit 2020 into the resolution enhancing model 130. As a result, the resolution enhancing model 130 outputs the resolution-enhanced second-view image 50. Then, the resolution-enhanced second-view image 50 is fed into the feature extracting model 120. As a result, the feature extracting model 120 outputs the feature value of the resolution-enhanced second-view image 50. The feature value of the resolution-enhanced second-view image 50 is handled as the feature value of the second-view image 30 by the determining unit 2080.

[0059] <Matching: S108>

The determining unit 2080 determines whether the first-view image 20 and the second-view image 30 match each other (S108). Specifically, the determining unit

2080 performs the determination by comparing the feature value of the first-view image 20 and the feature value of the second-view image 30.

[0060] The determining unit 2080 may compute a similarity score, which represents a degree of similarity between the feature value of the first-view image 20 and the feature value of the second-view image 30. There are various metrics to quantify a degree of similarity between feature values, and any one of them can be used to compute the similarity score. For example, the similarity score may be computed as one of various types of distance (e.g., L2 distance), correlation, cosine similarity, or neural network (NN) based similarity between the feature value of the first-view image 20 and the feature value of the second-view image 30. The NN based similarity is the degree of similarity computed by a neural network that is trained to compute the degree of similarity between two input data (in this disclosure, the feature value of the first-view image 20 and the feature value of the second-view image 30).

[0061] The determining unit 2080 determines whether the first-view image 20 and the second-view image 30 match each other based on the similarity score computed for them. Conceptually, the higher the degree of similarity between the feature value of the first-view image 20 and the feature value of the second-view image 30 is, the higher the possibility of that the first-view image 20 and the second-view image 30 match each other. Therefore, for example, the determining unit 2080 determines whether the similarity score is equal to or larger than a predefined threshold. If the similarity score is equal to or larger than the predefined threshold, the determining unit 2080 determines that the first-view image 20 and the second-view image 30 match each other. On the other hand, if the similarity score is less than a predefined threshold, the determining unit 2080 determines that the first-view image 20 and the second-view image 30 do not match each other.

[0062] It is noted that, in the case mentioned above, the similarity score is assumed to become larger as the degree of similarity between the feature values becomes higher. Thus, if a metric (e.g., distance) with which a value computed for the compared feature values becomes less as the degree of similarity between the compared feature values becomes higher is used, the similarity score may be defined as a reciprocal of the value computed for the compared feature values.

[0063] In another example, in the case where the similarity score becomes less as the degree of similarity between the compared feature values becomes higher, the determining unit 2080 may determine whether the similarity score is equal to or less than a predefined threshold. If the similarity score is equal to or less than the predefined threshold, the determining unit 2080 determines that the first-view image 20 and the second-view image 30 match each other. On the other hand, if the similarity score is larger than the predefined threshold, the determining unit 2080 determines that the

first-view image 20 and the second-view image 30 do not match each other.

[0064] <Output from Image Matching Apparatus>

The image matching apparatus 2000 may output information (hereinafter, output information) related to a result of the determination. For example, the output information may indicate whether the first-view image 20 and the second-view image 30 match each other. In addition, as explained with referring to Fig. 7, the output information may further include the location information that indicates the location at which the queried image is captured. The queried image is either the first-view image 20 or the second-view image 30. In other words, the queried image is either the ground-view image or the aerial-view image.

[0065] There are various ways to output the output information. For example, the image matching apparatus 2000 may put the output information into a storage device. In another example, the image matching apparatus 2000 may output the output information to a display device so that the display device displays the contents of the output information. In another example, the image matching apparatus 2000 may output the output information to another computer, such as one included in the geo-localization system 500 shown in Fig. 7.

[0066] <Training of Models>

As described above, machine learning-based models may be used to compute the feature value of the first-view image 20 and the feature value of the second-view image 30. Those models are trained in advance of being used by the image matching apparatus. Hereinafter, an apparatus that performs training of those models are called "training apparatus".

[0067] Fig. 11 is a diagram illustrating an example of functional configuration of the training apparatus 3000. The training apparatus 3000 includes an acquiring unit 3020, a computing unit 3040, and an updating unit 3060. The acquiring unit 3020 acquires a training data. The computing unit 3020 applies the training data to a model to be trained, and computes a loss based on data output by the model. The updating unit 3060 updates the model based on the loss.

[0068] It is noted that models can be updated by updating their trainable parameter based on the loss. When the model is a neural network, trainable parameters may include weights assigned to edges and biases.

[0069] The training apparatus 3000 may have hardware configuration similar to the hardware configuration of the image matching apparatus 2000. For example, like the hardware configuration of the image matching apparatus 2000, the hardware configuration of the training apparatus 3000 may be illustrated by Fig. 5. However, the storage device of the training apparatus 3000 includes a program that implements the functions of the training apparatus 3000.

[0070] Fig. 12 is a flowchart illustrating an example flow of processing performed by the training apparatus 3000. The acquiring unit 3020 acquires the training data (S202). The computing unit 3040 inputs the training data to the model to be trained (S204). The computing unit 3040 computes a loss based on data output by the model (S206). The updating unit 3060 updates the model based on the loss (S208).

[0071] It is noted that the processing illustrated by Fig. 12 is repeatedly performed until the model is sufficiently trained.

[0072] Hereinafter, example ways of training the models will be described. Specifically, example ways of training the models in the first feature extracting unit 2040 will be described first. Then, example ways of training the models in the second feature extracting unit 2060 will be described.

[0073] <<Training of First Feature Extracting unit 2040>>

As depicted by Fig. 8, the first feature extracting unit 2040 may include the resolution enhancing model 100 and the feature extracting model 110. Fig. 13 illustrates a first example way of training the first feature extracting unit 2040. First, the acquiring unit 3020 acquires a first-view image 150 as a training data. The first-view image 150 is a first-view image with the resolution of the second level.

[0074] Then, the computing unit 3040 preforms resolution reduction on the first-view image 150 to generate a resolution-reduced first-view image 160, which is a first-view image with the resolution of the first level. For example, the computing unit 3040 performs down-sampling on the first-view image 150 and resizes the obtained image to the size same as the first-view image 150. By doing so, the computing unit 3040 can obtain, as the resolution-reduced first-view image 160, a first-view image whose size is the same as the first-view image 150 and whose resolution is lower than the first-view image 150.

[0075] The computing unit 3040 inputs the resolution-reduced first-view image 160 into the resolution enhancing model 100, thereby obtaining a resolution-reenhanced first-view image 170. The resolution-reenhanced first-view image 170 is supposed to be equivalent to the first-view image 150 when the resolution enhancing model 100 is already trained sufficiently.

[0076] Then, the resolution-reenhanced first-view image 170 is fed into the feature extracting model 110. As a result, a feature value of the resolution-reenhanced first-view image 170 is output by the feature extracting model 110.

[0077] The computing unit 3040 also computes a feature value of the first-view image 150. Specifically, the computing unit 3040 inputs the first-view image 150 into a pre-trained feature extracting model 180, which is a machine learning-based model that is trained in advance to compute a feature value of a first-view image with the resolution of the second level.

[0078] The feature value of the resolution-reenhanced first-view image 170 and the feature value of the first-view image 150 are supposed to be equivalent to each other when the resolution enhancing model 100 and the feature extracting model 110 are already trained sufficiently. Thus, the training apparatus 3000 computes a loss that represents a degree of difference between the feature value of the first-view image 150 and the feature value of the resolution-reenhanced first-view image 170.

[0079] The updating unit 3060 uses the computed loss to update the resolution enhancing model 100 and the feature extracting model 110. Specifically, the updating unit 3060 updates trainable parameters of the resolution enhancing model 100 and trainable parameters of the feature extracting model 110 based on the loss.

[0080] According to the training apparatus 3000 that performs the training depicted by Fig. 13, it is possible to train the resolution enhancing model 100 so that the resolution enhancing model 100 can accurately enhance the resolution of a first-view image from the first level to the second level. In addition, it is possible to train the feature extracting model 110 so that the feature extracting model 110 can accurately extract the feature value of a first-view image with the resolution of the second level.

[0081] As mentioned above, the pre-trained feature extracting model 180 is trained in advance. For example, the pre-trained feature extracting model 180 is trained to be used a part of another image matching apparatus, which is configured to take a first-view image with the resolution of the second level and a second-view image with the resolution of the second level, and to determine whether the acquired first-view image and the acquired second-view image match each other.

[0082] Fig. 14 illustrates an image matching apparatus 400 that handles the first-view image with the resolution of the second level and the second-view image with the resolution of the second level. The image matching apparatus 400 acquires a first-view image 410 and a second-view image 420, whose resolutions are of the second level. The image matching apparatus 400 has a feature extracting model 430 and a feature extracting model 440. The feature extracting model 430 takes the first-view image 410 as input, and computes a feature value of the first-view image 410. The feature extracting model 440 takes the second-view image 420 as input, and computes a feature value of the second-view image 420. The image matching apparatus 400 determines whether the first-view image 410 and the second-view image 420 match each other by comparing their feature values.

[0083] What the feature extracting model 430 performs is to compute a feature value of the first-view image whose resolution is of the second level, and this is the same as what the pre-trained feature extracting model 180 performs. Thus, the feature extracting model 430 may be employed in the training apparatus 3000 to be used as the pre-trained feature extracting model 180.

[0084] In some embodiments, after finishing the training of the first feature extracting unit 2040 depicted by Fig. 13, the training apparatus 3000 may further perform additional training on the feature extracting model 110. Fig. 15 illustrates an additional training performed on the feature extracting model 110. The acquiring unit 3020 acquires a first-view image 200 and a second-view image 210 as training data. The first-view image 200 is a first-view image with the resolution of the first level while the second-view image 210 is a second-view image with the resolution of the second level.

[0085] The computing unit 3040 inputs the first-view image 200 into the resolution enhancing model 100 that has been trained in the way depicted by Fig. 13, thereby obtaining a resolution-enhanced first-view image 220. Then, the resolution-enhanced first-view image 220 is fed into the feature extracting model 110 that has been trained in the way depicted by Fig. 13. As a result, a feature value of the resolution-enhanced first-view image 220 is obtained.

[0086] Also, the computing unit 3040 inputs the second-view image 210 into a pre-trained feature extracting model 230, thereby obtaining a feature value of the second-view image 210. It is noted that the feature extracting model 440 depicted by Fig. 14 can be used as the pre-trained feature extracting model 230.

[0087] The computing unit 3040 computes a loss that represents a degree of difference between the feature value of the first-view image 200 and the feature value of the second-view image 210, and update the feature extracting model 110 based on the loss. It is noted that the feature extracting model 110 may be trained so that the loss becomes smaller when the first-view image 200 and the second-view image 210 are supposed to match each other. On the other hand, the feature extracting model 110 may be trained so that the loss becomes larger when the first-view image 200 and the second-view image 210 are supposed not to match each other.

[0088] More specifically, the training apparatus 3000 may use, as training data, a set of the first-view image 200, a positive example of the second-view image 210, and a negative example of the second-view image 210. The positive example of the second-view image 210 is a second-view image 210 that is supposed to match the first-view image 200. On the other hand, the negative example of the second-view image 210 is a second-view image 210 that is supposed not to match the first-view image 200. In this case, the training apparatus 3000 may compute a triplet loss based on the feature value of the first-view image 200, the feature value of the positive example of the second-view image 210, and the feature value of the negative example of the second-view image 210, and update the feature extracting model 110 based on the triplet loss.

[0089] According to the training depicted by Fig. 15, it is possible to train the feature extracting model 110 so that the feature extracting model 110 can accurately compute the feature value of the first-view image from the viewpoint of matching between the first-

view image and the second-view image.

[0090] In some embodiments, the resolution enhancing model 100 may be trained separately from the feature extracting model 110. Fig. 16 illustrates an example way of training the resolution enhancing model 100 separately from the feature extracting model 110.

[0091] The acquiring unit 3020 acquires a first-view image 240 as a training data. The first-view image 240 is a first-view image with the resolution of the second level. The computing unit 3040 preforms resolution reduction on the first-view image 240 to generate a resolution-reduced first-view image 250, which is a first-view image with the resolution of the first level. The computing unit 3040 inputs the resolution-reduced first-view image 250 into the resolution enhancing model 100, thereby obtaining a resolution-reenhanced first-view image 260.

[0092] The resolution-reenhanced first-view image 260 is supposed to be equivalent to the first-view image 240 when the resolution enhancing model 100 is already trained sufficiently. Thus, the computing unit 3040 computes a loss that represents a degree of difference between the first-view image 240 and the resolution-reenhanced first-view image 260, and updates the resolution enhancing model 100 based on the loss.

[0093] <<Training of Second Feature Extracting Unit 2060>>

As illustrated by Fig. 9, when the resolution of the second-view image 30 is of the second level, the second feature extracting unit 2060 may include the feature extracting model 120 and not include the resolution enhancing model 130. In this case, the feature extracting model 440 depicted by Fig. 14 may be used as the feature extracting model 120.

[0094] On the other hand, when the resolution of the second-view image 30 is of the first level, the second feature extracting unit 2060 may include the resolution enhancing model 130 and the feature extracting model 120. In this case, the second extracting model 2060 may be trained in the same manner as the manner of training the first feature extracting unit 2040.

[0095] Fig. 17 illustrates a first example way of training the second feature extracting unit 2060. The acquiring unit 3020 acquires a second-view image 270 as a training data. The second-view image 270 is a second-view image with the resolution of the second level.

[0096] The computing unit 3040 preforms resolution reduction on the second-view image 270 to generate a resolution-reduced second-view image 280, which is a second-view image with the resolution of the first level. The computing unit 3040 inputs the resolution-reduced second-view image 280 into the resolution enhancing model 130, thereby obtaining a resolution-reenhanced second-view image 290.

[0097] The resolution-reenhanced second-view image 290 is fed into the feature ex-

tracting model 120. As a result, a feature value of the resolution-reenhanced second-view image 290 is output by the feature extracting model 120.

[0098] The computing unit 3040 also computes a feature value of the second-view image 270. Specifically, the computing unit 3040 inputs the second-view image 270 into a pre-trained feature extracting model 300, which is a machine learning-based model that is trained in advance to compute a feature value of a second-view image with the resolution of the second level. The feature extracting model 340 depicted by Fig. 14 may be used as the pre-trained feature extracting model 300.

[0099] The feature value of the resolution-reenhanced first-view image 170 and the feature value of the first-view image 150 are supposed to be equivalent to each other when the resolution enhancing model 130 and the feature extracting model 120 are already trained sufficiently. Thus, the computing unit 3040 computes a loss that represents a degree of difference between the feature value of the second-view image 270 and the feature value of the resolution-reenhanced second-view image 290.

[0100] The updating unit 3060 uses the computed loss to update the resolution enhancing model 130 and the feature extracting model 120. Specifically, the updating unit 3060 updates trainable parameters of the resolution enhancing model 130 and trainable parameters of the feature extracting model 120 based on the loss.

[0101] According to the training apparatus 3000 that performs the training depicted by Fig. 17, it is possible to train the resolution enhancing model 130 so that the resolution enhancing model 130 can accurately enhance the resolution of a second-view image from the first level to the second level. In addition, it is possible to train the feature extracting model 120 so that the feature extracting model 120 can accurately extract the feature value of a second-view image with the resolution of the second level.

[0102] In some embodiments, after finishing the training of the second feature extracting unit 2060 depicted by Fig. 17, the training apparatus 3000 may further perform additional training on the feature extracting model 120. Fig. 18 illustrates an additional training performed on the feature extracting model 120. The acquiring unit 3020 acquires a first-view image 310 and a second-view image 320 as training data. The first-view image 310 is a first-view image with the resolution of the second level while the second-view image 320 is a second-view image with the resolution of the first level.

[0103] The computing unit 3040 inputs the second-view image 320 into the resolution enhancing model 130 that has been trained in the way depicted by Fig. 17, thereby obtaining a resolution-enhanced second-view image 330. Then, the resolution-enhanced second-view image 330 is fed into the feature extracting model 120 that has been trained in the way depicted by Fig. 17. As a result, a feature value of the resolution-enhanced second-view image 330 is obtained.

[0104] Also, the computing unit 3040 inputs the first-view image 310 into a pre-trained feature extracting model 340, thereby obtaining a feature value of the first-view image 310. It is noted that the feature extracting model 430 depicted by Fig. 14 can be used as the pre-trained feature extracting model 340.

[0105] The computing unit 3040 computes a loss that represents a degree of difference between the feature value of the first-view image 310 and the feature value of the second-view image 320, and update the feature extracting model 120 based on the loss. It is noted that the feature extracting model 120 may be trained so that the loss becomes smaller when the first-view image 310 and the second-view image 320 are supposed to match each other. On the other hand, the feature extracting model 120 may be trained so that the loss becomes larger when the first-view image 310 and the second-view image 320 are supposed not to match each other.

[0106] More specifically, the training apparatus 3000 may use, as training data, a set of the first-view image 310, a positive example of the second-view image 320, and a negative example of the second-view image 320. The positive example of the second-view image 320 is a second-view image 320 that is supposed to match the first-view image 310. On the other hand, the negative example of the second-view image 320 is a second-view image 320 that is supposed not to match the first-view image 310. In this case, the training apparatus 3000 may compute a triplet loss based on the feature value of the first-view image 310, the feature value of the positive example of the second-view image 320, and the feature value of the negative example of the second-view image 320, and update the feature extracting model 120 based on the triplet loss.

[0107] According to the training depicted by Fig. 18, it is possible to train the feature extracting model 120 so that the feature extracting model 120 can accurately compute the feature value of the second-view image from the viewpoint of matching between the first-view image and the second-view image.

[0108] In some embodiments, the resolution enhancing model 130 may be trained separately from the feature extracting model 120. Fig. 19 illustrates an example way of training the resolution enhancing model 130 separately from the feature extracting model 120.

[0109] The acquiring unit 3020 acquires a second-view image 350 as a training data. The second-view image 350 is a second-view image with the resolution of the second level. The computing unit 3040 preforms resolution reduction on the second-view image 350 to generate a resolution-reduced second-view image 360, which is a second-view image with the resolution of the first level. The computing unit 3040 inputs the resolution-reduced second-view image 360 into the resolution enhancing model 130, thereby obtaining a resolution-reenhanced second-view image 370.

[0110] The resolution-reenhanced second-view image 370 is supposed to be equivalent to

the second-view image 350 when the resolution enhancing model 130 is already trained sufficiently. Thus, the computing unit 3040 computes a loss that represents a degree of difference between the second-view image 350 and the resolution-reenhanced second-view image 370, and updates the resolution enhancing model 130 based on the loss.

[0111] The program can be stored and provided to a computer using any type of non-transitory computer readable media. Non-transitory computer readable media include any type of tangible storage media. Examples of non-transitory computer readable media include magnetic storage media (such as floppy disks, magnetic tapes, hard disk drives, etc.), optical magnetic storage media (e.g., magneto-optical disks), CD-ROM (compact disc read only memory), CD-R (compact disc recordable), CD-R/W (compact disc rewritable), and semiconductor memories (such as mask ROM, PROM (programmable ROM), EPROM (erasable PROM), flash ROM, RAM (random access memory), etc.). The program may be provided to a computer using any type of transitory computer readable media. Examples of transitory computer readable media include electric signals, optical signals, and electromagnetic waves. Transitory computer readable media can provide the program to a computer via a wired communication line (e.g., electric wires, and optical fibers) or a wireless communication line.

[0112] While the present disclosure has been particularly shown and described with reference to example embodiments thereof, the present disclosure is not limited to these example embodiments. It will be understood by those of ordinary skill in the art that various changes in form and details may be made therein without departing from the spirit and scope of the present disclosure as defined by the claims. And each embodiment can be appropriately combined with at least one of embodiments.

[0113] Each of the drawings or figures is merely an example to illustrate one or more example embodiments. Each figure may not be associated with only one particular example embodiment, but may be associated with one or more other example embodiments. As those of ordinary skill in the art will understand, various features or steps described with reference to any one of the figures can be combined with features or steps illustrated in one or more other figures, for example, to produce example embodiments that are not explicitly illustrated or described. Not all of the features or steps illustrated in any one of the figures to describe an example embodiment are necessarily essential, and some features or steps may be omitted. The order of the steps described in any of the figures may be changed as appropriate.

[0114] The whole or part of the example embodiments disclosed above can be described as, but not limited to, the following supplementary notes.

<Supplementary notes>

(Supplementary Note 1)

An image matching apparatus comprising:

at least one memory that is configured to store instructions; and

at least one processor that is configured to execute the instructions to:

acquire a first-view image and a second-view image;

compute a feature value of the first-view image;

compute a feature value of the second-view image; and

determine whether the first-view image and the second-view image match each other based on the feature value of the first-view image and the feature value of the second-view image,

wherein the computation of the feature value of the first-view image includes:

enhancing a resolution of the first-view image to generate a resolution-enhanced first-view image; and

computing a feature value of the resolution-enhanced first-view image as the feature value of the first-view image,

wherein the first-view image and the second-view image are respectively a ground-view image and an aerial-view image, or the first-view image and the second-view image are respectively an aerial-view image and a ground-view image.

(Supplementary Note 2)

An image matching apparatus according to supplementary note 1,

wherein the computation of the feature value of the second-view image includes:

enhancing a resolution of the second-view image to generate a resolution-enhanced second-view image; and

computing a feature value of the resolution-enhanced second-view image as the feature value of the second-view image.

(Supplementary Note 3)

An image matching method performed by a computer, comprising:

acquiring a first-view image and a second-view image;

computing a feature value of the first-view image;

computing a feature value of the second-view image; and

determining whether the first-view image and the second-view image match each other based on the feature value of the first-view image and the feature value of the second-view image,

wherein the computation of the feature value of the first-view image includes:

enhancing a resolution of the first-view image to generate a resolution-enhanced first-view image; and

computing a feature value of the resolution-enhanced first-view image as the feature value of the first-view image,

wherein the first-view image and the second-view image are respectively a ground-

view image and an aerial-view image, or the first-view image and the second-view image are respectively an aerial-view image and a ground-view image.

(Supplementary Note 4)

The image matching method according to supplementary note 3,

wherein the computation of the feature value of the second-view image includes:

enhancing a resolution of the second-view image to generate a resolution-enhanced second-view image; and

computing a feature value of the resolution-enhanced second-view image as the feature value of the second-view image.

(Supplementary Note 5)

A non-transitory computer-readable medium storing a program that cause a computer to execute:

acquiring a first-view image and a second-view image;

computing a feature value of the first-view image;

computing a feature value of the second-view image; and

determining whether the first-view image and the second-view image match each other based on the feature value of the first-view image and the feature value of the second-view image,

wherein the computation of the feature value of the first-view image includes:

enhancing a resolution of the first-view image to generate a resolution-enhanced first-view image; and

computing a feature value of the resolution-enhanced first-view image as the feature value of the first-view image,

wherein the first-view image and the second-view image are respectively a ground-view image and an aerial-view image, or the first-view image and the second-view image are respectively an aerial-view image and a ground-view image.

(Supplementary Note 6)

The medium according to supplementary note 5,

wherein the computation of the feature value of the second-view image includes:

enhancing a resolution of the second-view image to generate a resolution-enhanced second-view image; and

computing a feature value of the resolution-enhanced second-view image as the feature value of the second-view image.

(Supplementary Note 7)

A training apparatus comprising:

at least one memory that is configured to store instructions; and

at least one processor that is configured to execute the instructions to:

acquire a target image; and

perform training of one or more models,

wherein the training of the one or more models includes:

reducing a resolution of the acquired target image to generate a resolution-reduced target image;

inputting the resolution-reduced target image into a resolution enhancing model to generate a resolution-reenhanced target image;

computing a loss using the acquired target image and the resolution-reenhanced target image; and

updating the resolution enhancing model based on the loss,

wherein the target image is a ground-view image or an aerial-view image.

(Supplementary Note 8)

The training apparatus according to supplementary note 7,

wherein the reduction of the resolution of the acquired target image includes:

down-sampling the acquired target image; and

increase a size of an image that is acquired by the down-sampling to a size of the acquired target image, thereby generating the resolution-reduced target image.

(Supplementary Note 9)

The training apparatus comprising according to supplementary note 7 or 8,

wherein the loss is computed so as to represent a degree of difference between the acquired target image and the resolution-reenhanced target image.

(Supplementary Note 10)

The training apparatus comprising according to supplementary note 7 or 8,

wherein the training of the one or more models further includes:

inputting the resolution-reenhanced target image into a feature extracting model to compute a feature value of the resolution-reenhanced target image;

computing a feature value of the acquired target image;

computing the loss that represents a degree of difference between the feature value of the acquired target image and the feature value of the resolution-reenhanced target image; and

updating the resolution-reenhanced target image and the feature extracting model based on the loss.

(Supplementary Note 11)

A training method performed by a computer, comprising:

acquiring a target image; and

performing training of one or more models,

wherein the training of the one or more models includes:

reducing a resolution of the acquired target image to generate a resolution-reduced target image;

inputting the resolution-reduced target image into a resolution enhancing model to generate a resolution-reenhanced target image;

computing a loss using the acquired target image and the resolution-reenhanced target image; and

updating the resolution enhancing model based on the loss,

wherein the target image is a ground-view image or an aerial-view image.

(Supplementary Note 12)

The training method according to supplementary note 11,

wherein the reduction of the resolution of the acquired target image includes:

down-sampling the acquired target image; and

increase a size of an image that is acquired by the down-sampling to a size of the acquired target image, thereby generating the resolution-reduced target image.

(Supplementary Note 13)

The training method comprising according to supplementary note 11 or 12,

wherein the loss is computed so as to represent a degree of difference between the acquired target image and the resolution-reenhanced target image.

(Supplementary Note 14)

The training method comprising according to supplementary note 11 or 12,

wherein the training of the one or more models further includes:

inputting the resolution-reenhanced target image into a feature extracting model to compute a feature value of the resolution-reenhanced target image;

computing a feature value of the acquired target image;

computing the loss that represents a degree of difference between the feature value of the acquired target image and the feature value of the resolution-reenhanced target image; and

updating the resolution-reenhanced target image and the feature extracting model based on the loss.

(Supplementary Note 15)

A non-transitory computer-readable medium storing a program that causes a computer to execute:

acquiring a target image; and

performing training of one or more models,

wherein the training of the one or more models includes:

reducing a resolution of the acquired target image to generate a resolution-reduced target image;

inputting the resolution-reduced target image into a resolution enhancing model to generate a resolution-reenhanced target image;

computing a loss using the acquired target image and the resolution-reenhanced

target image; and

updating the resolution enhancing model based on the loss,

wherein the target image is a ground-view image or an aerial-view image.

(Supplementary Note 16)

The medium according to supplementary note 15,

wherein the reduction of the resolution of the acquired target image includes:

down-sampling the acquired target image; and

increase a size of an image that is acquired by the down-sampling to a size of the acquired target image, thereby generating the resolution-reduced target image.

(Supplementary Note 17)

The medium according to supplementary note 15 or 16,

wherein the loss is computed so as to represent a degree of difference between the acquired target image and the resolution-reenhanced target image.

(Supplementary Note 18)

The medium according to supplementary note 15 or 16,

wherein the training of the one or more models further includes:

inputting the resolution-reenhanced target image into a feature extracting model to compute a feature value of the resolution-reenhanced target image;

computing a feature value of the acquired target image;

computing the loss that represents a degree of difference between the feature value of the acquired target image and the feature value of the resolution-reenhanced target image; and

updating the resolution-reenhanced target image and the feature extracting model based on the loss.

[0115] Some or all of elements specified in any of Supplementary Notes may be applied to various types of hardware, software, and recording means for recording software, systems, and methods.

Reference Signs List

[0116] 10 ground-view image

15 aerial-view image

20 first-view image

30 second-view image

40 resolution-enhanced first-view image

50 resolution-enhanced second-view image

100 resolution enhancing model

110 feature extracting model

120 feature extracting model

130 resolution-enhancing model
150 first-view image
160 resolution-reduced first-view image
170 resolution-reenhanced first-view image
180 feature extracting model
200 first-view image
210 second-view image
220 resolution-enhanced first-view image
230 feature extracting model
240 first-view image
250 resolution-reduced first-view image
260 resolution-reenhanced first-view image
270 second-view image
280 resolution-reduced second-view image
290 resolution-reenhanced second-view image
300 feature extracting model
310 first-view image
320 second-view image
330 resolution-enhanced second-view image
340 feature extracting model
350 second-view image
360 resolution-reduced second-view image
370 resolution-reenhanced second-view image
400 image matching apparatus
410 first-view image
420 second-view image
430 feature extracting model
440 feature extracting model
500 geo-localization system
600 location database
1000 computer
1020 bus
1040 processor
1060 memory
1080 storage device
1100 input/output interface
1120 network interface
2000 image matching apparatus

2020 acquiring unit
2040 first feature extracting unit
2060 second feature extracting unit

Claims

[Claim 1] An image matching apparatus comprising:

at least one memory that is configured to store instructions; and

at least one processor that is configured to execute the instructions to:

acquire a first-view image and a second-view image;

compute a feature value of the first-view image;

compute a feature value of the second-view image; and

determine whether the first-view image and the second-view image match each other based on the feature value of the first-view image and the feature value of the second-view image,

wherein the computation of the feature value of the first-view image includes:

enhancing a resolution of the first-view image to generate a resolution-enhanced first-view image; and

computing a feature value of the resolution-enhanced first-view image as the feature value of the first-view image,

wherein the first-view image and the second-view image are respectively a ground-view image and an aerial-view image, or the first-view image and the second-view image are respectively an aerial-view image and a ground-view image.

[Claim 2] An image matching apparatus according to claim 1,

wherein the computation of the feature value of the second-view image includes:

enhancing a resolution of the second-view image to generate a resolution-enhanced second-view image; and

computing a feature value of the resolution-enhanced second-view image as the feature value of the second-view image.

[Claim 3] An image matching method performed by a computer, comprising:

acquiring a first-view image and a second-view image;

computing a feature value of the first-view image;

computing a feature value of the second-view image; and

determining whether the first-view image and the second-view image match each other based on the feature value of the first-view image and the feature value of the second-view image,

wherein the computation of the feature value of the first-view image includes:

enhancing a resolution of the first-view image to generate a resolution-enhanced first-view image; and

computing a feature value of the resolution-enhanced first-view image as the feature value of the first-view image,

wherein the first-view image and the second-view image are respectively a ground-view image and an aerial-view image, or the first-view image and the second-view image are respectively an aerial-view image and a ground-view image.

[Claim 4] The image matching method according to claim 3,

wherein the computation of the feature value of the second-view image includes:

enhancing a resolution of the second-view image to generate a resolution-enhanced second-view image; and

computing a feature value of the resolution-enhanced second-view image as the feature value of the second-view image.

[Claim 5] A non-transitory computer-readable medium storing a program that cause a computer to execute:

acquiring a first-view image and a second-view image;

computing a feature value of the first-view image;

computing a feature value of the second-view image; and

determining whether the first-view image and the second-view image match each other based on the feature value of the first-view image and the feature value of the second-view image,

wherein the computation of the feature value of the first-view image includes:

enhancing a resolution of the first-view image to generate a resolution-enhanced first-view image; and

computing a feature value of the resolution-enhanced first-view image as the feature value of the first-view image,

wherein the first-view image and the second-view image are respectively a ground-view image and an aerial-view image, or the first-view image and the second-view image are respectively an aerial-view image and a ground-view image.

[Claim 6] The medium according to claim 5,

wherein the computation of the feature value of the second-view image includes:

enhancing a resolution of the second-view image to generate a resolution-enhanced second-view image; and

computing a feature value of the resolution-enhanced second-view image as the feature value of the second-view image.

[Claim 7] A training apparatus comprising:

at least one memory that is configured to store instructions; and

at least one processor that is configured to execute the instructions to:

acquire a target image; and

perform training of one or more models,

wherein the training of the one or more models includes:

reducing a resolution of the acquired target image to generate a resolution-reduced target image;

inputting the resolution-reduced target image into a resolution enhancing model to generate a resolution-reenhanced target image;

computing a loss using the acquired target image and the resolution-reenhanced target image; and

updating the resolution enhancing model based on the loss,

wherein the target image is a ground-view image or an aerial-view image.

[Claim 8] The training apparatus according to claim 7,

wherein the reduction of the resolution of the acquired target image includes:

down-sampling the acquired target image; and

increase a size of an image that is acquired by the down-sampling to a size of the acquired target image, thereby generating the resolution-reduced target image.

[Claim 9] The training apparatus comprising according to claim 7 or 8,

wherein the loss is computed so as to represent a degree of difference between the acquired target image and the resolution-reenhanced target image.

[Claim 10] The training apparatus comprising according to claim 7 or 8,

wherein the training of the one or more models further includes:

inputting the resolution-reenhanced target image into a feature extracting model to compute a feature value of the resolution-reenhanced target image;

computing a feature value of the acquired target image;

computing the loss that represents a degree of difference between the feature value of the acquired target image and the feature value of the resolution-reenhanced target image; and

updating the resolution-reenhanced target image and the feature extracting model based on the loss.

[Claim 11] A training method performed by a computer, comprising:

acquiring a target image; and

performing training of one or more models,

wherein the training of the one or more models includes:

reducing a resolution of the acquired target image to generate a resolution-reduced target image;

inputting the resolution-reduced target image into a resolution enhancing model to generate a resolution-reenhanced target image;

computing a loss using the acquired target image and the resolution-reenhanced target image; and

updating the resolution enhancing model based on the loss,

wherein the target image is a ground-view image or an aerial-view image.

[Claim 12] The training method according to claim 11,

wherein the reduction of the resolution of the acquired target image includes:

down-sampling the acquired target image; and

increase a size of an image that is acquired by the down-sampling to a size of the acquired target image, thereby generating the resolution-reduced target image.

[Claim 13] The training method comprising according to claim 11 or 12,

wherein the loss is computed so as to represent a degree of difference between the acquired target image and the resolution-reenhanced target image.

[Claim 14] The training method comprising according to claim 11 or 12,

wherein the training of the one or more models further includes:

inputting the resolution-reenhanced target image into a feature extracting model to compute a feature value of the resolution-reenhanced target image;

computing a feature value of the acquired target image;

computing the loss that represents a degree of difference between the feature value of the acquired target image and the feature value of the resolution-reenhanced target image; and

updating the resolution-reenhanced target image and the feature extracting model based on the loss.

[Claim 15] A non-transitory computer-readable medium storing a program that

causes a computer to execute:

acquiring a target image; and

performing training of one or more models,

wherein the training of the one or more models includes:

reducing a resolution of the acquired target image to generate a resolution-reduced target image;

inputting the resolution-reduced target image into a resolution enhancing model to generate a resolution-reenhanced target image;

computing a loss using the acquired target image and the resolution-reenhanced target image; and

updating the resolution enhancing model based on the loss,

wherein the target image is a ground-view image or an aerial-view image.

[Claim 16] The medium according to claim 15,

wherein the reduction of the resolution of the acquired target image includes:

down-sampling the acquired target image; and

increase a size of an image that is acquired by the down-sampling to a size of the acquired target image, thereby generating the resolution-reduced target image.

[Claim 17] The medium according to claim 15 or 16,

wherein the loss is computed so as to represent a degree of difference between the acquired target image and the resolution-reenhanced target image.

[Claim 18] The medium according to claim 15 or 16,

wherein the training of the one or more models further includes:

inputting the resolution-reenhanced target image into a feature extracting model to compute a feature value of the resolution-reenhanced target image;

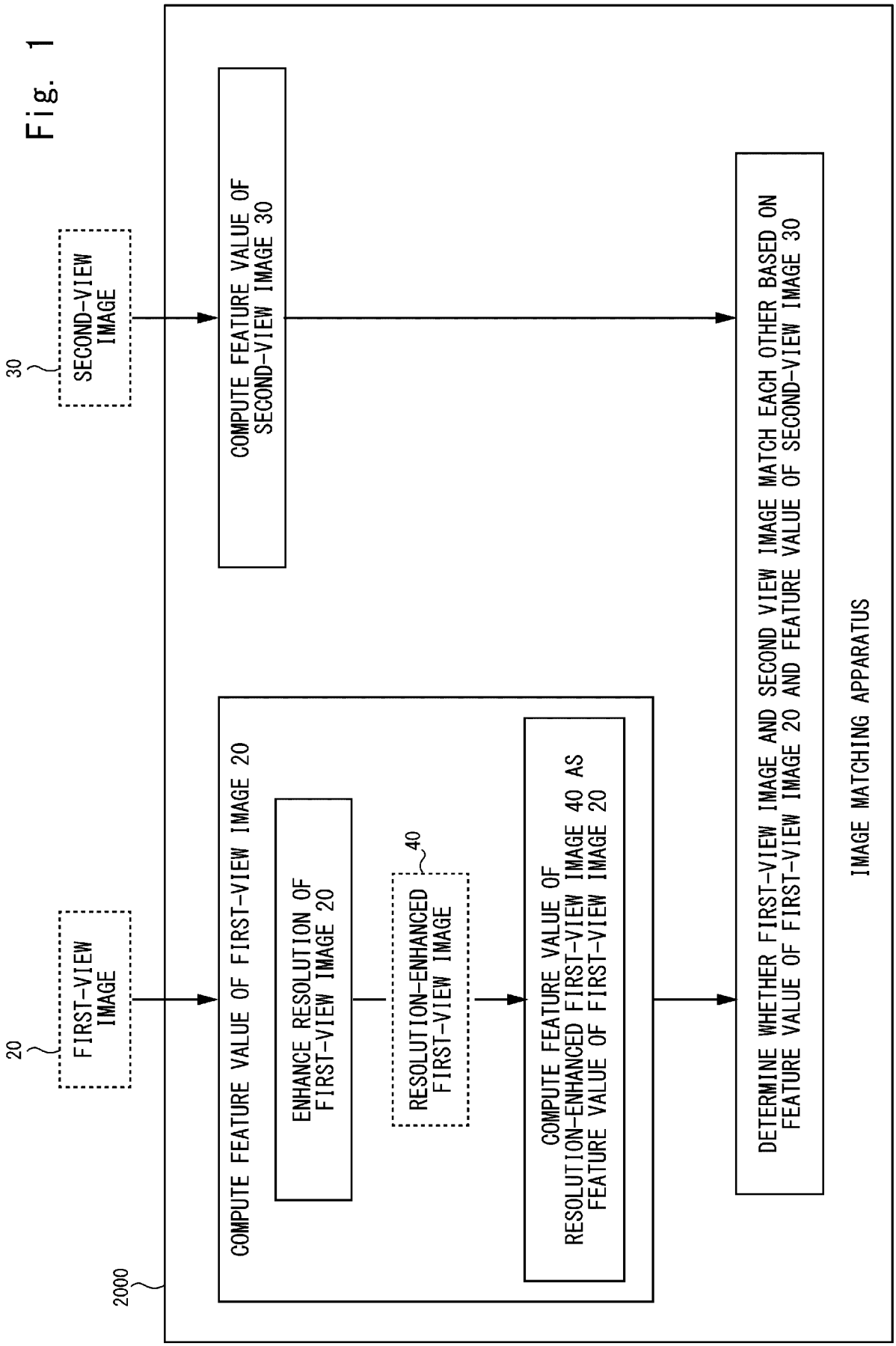
computing a feature value of the acquired target image;

computing the loss that represents a degree of difference between the feature value of the acquired target image and the feature value of the resolution-reenhanced target image; and

updating the resolution-reenhanced target image and the feature extracting model based on the loss.

[Fig. 1]

Fig. 1


2000
IMAGE MATCHING APPARATUS
20
FIRST-VIEW IMAGE
30
SECOND-VIEW IMAGE
COMPUTE FEATURE VALUE OF FIRST-VIEW IMAGE 20
ENHANCE RESOLUTION OF FIRST-VIEW IMAGE 20
40
RESOLUTION-ENHANCED FIRST-VIEW IMAGE
COMPUTE FEATURE VALUE OF RESOLUTION-ENHANCED FIRST-VIEW IMAGE 40 AS FEATURE VALUE OF FIRST-VIEW IMAGE 20
COMPUTE FEATURE VALUE OF SECOND-VIEW IMAGE 30
DETERMINE WHETHER FIRST-VIEW IMAGE AND SECOND VIEW IMAGE MATCH EACH OTHER BASED ON FEATURE VALUE OF FIRST-VIEW IMAGE 20 AND FEATURE VALUE OF SECOND-VIEW IMAGE 30

[Fig. 2]

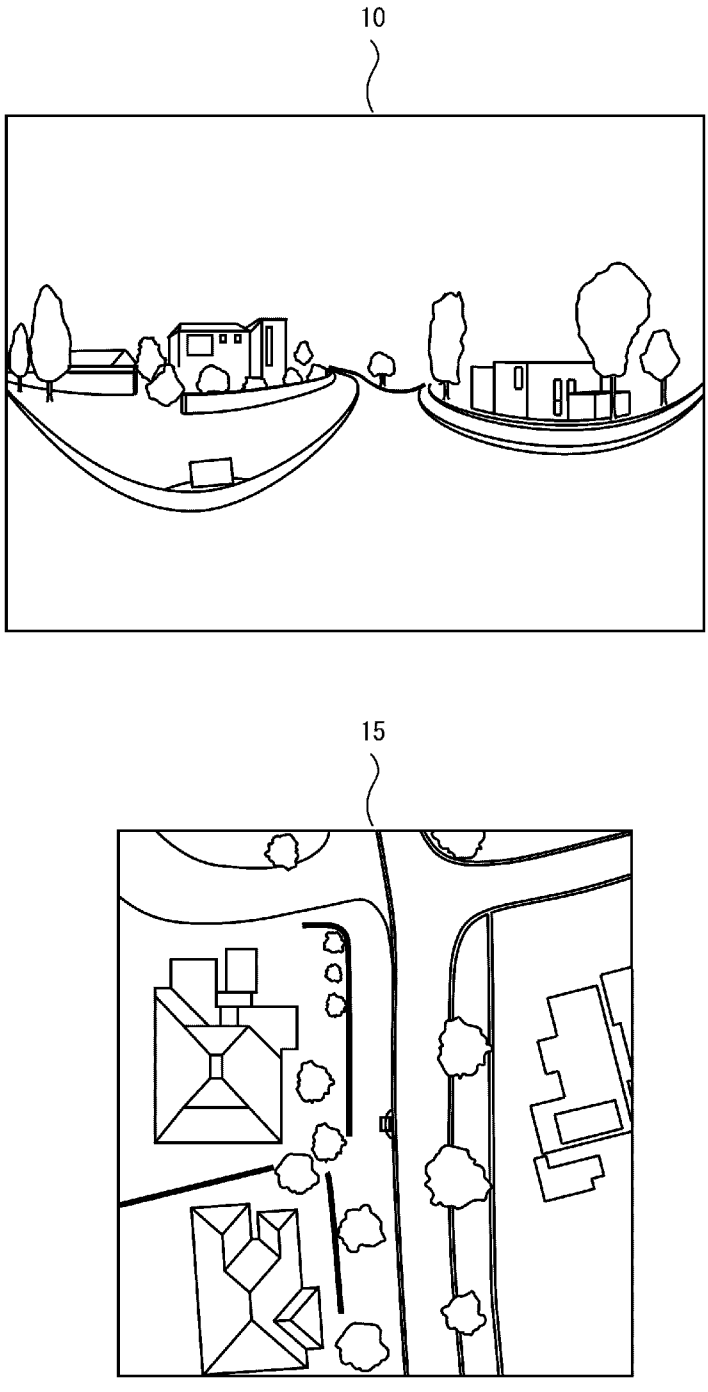


Fig. 2

[Fig. 3]

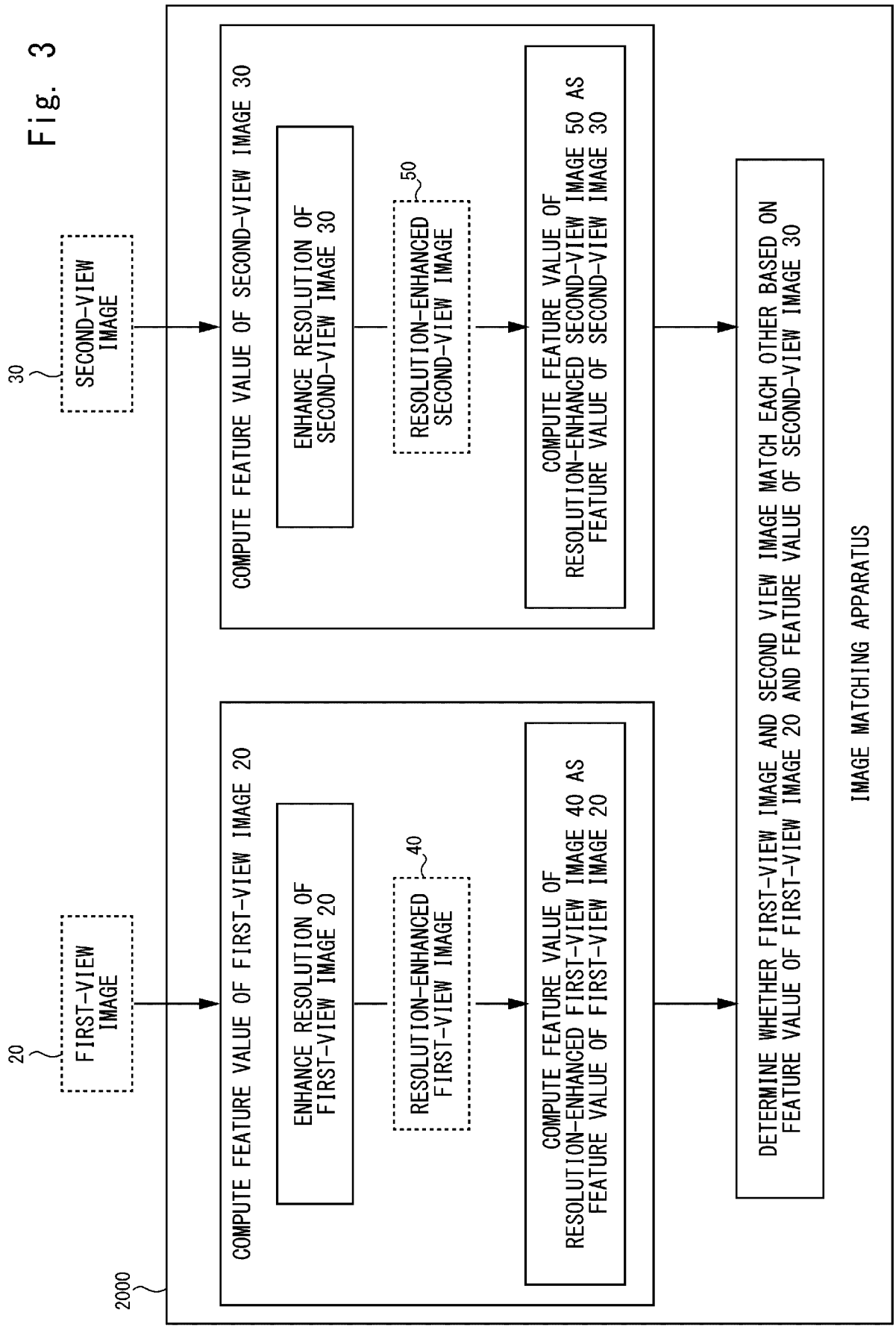


Fig. 3
2000
IMAGE MATCHING APPARATUS
20
FIRST-VIEW IMAGE
COMPUTE FEATURE VALUE OF FIRST-VIEW IMAGE 20
ENHANCE RESOLUTION OF FIRST-VIEW IMAGE 20
40
RESOLUTION-ENHANCED FIRST-VIEW IMAGE
COMPUTE FEATURE VALUE OF RESOLUTION-ENHANCED FIRST-VIEW IMAGE 40 AS FEATURE VALUE OF FIRST-VIEW IMAGE 20
30
SECOND-VIEW IMAGE
COMPUTE FEATURE VALUE OF SECOND-VIEW IMAGE 30
ENHANCE RESOLUTION OF SECOND-VIEW IMAGE 30
50
RESOLUTION-ENHANCED SECOND-VIEW IMAGE
COMPUTE FEATURE VALUE OF RESOLUTION-ENHANCED SECOND-VIEW IMAGE 50 AS FEATURE VALUE OF SECOND-VIEW IMAGE 30
DETERMINE WHETHER FIRST-VIEW IMAGE AND SECOND VIEW IMAGE MATCH EACH OTHER BASED ON FEATURE VALUE OF FIRST-VIEW IMAGE 20 AND FEATURE VALUE OF SECOND-VIEW IMAGE 30

[Fig. 4]

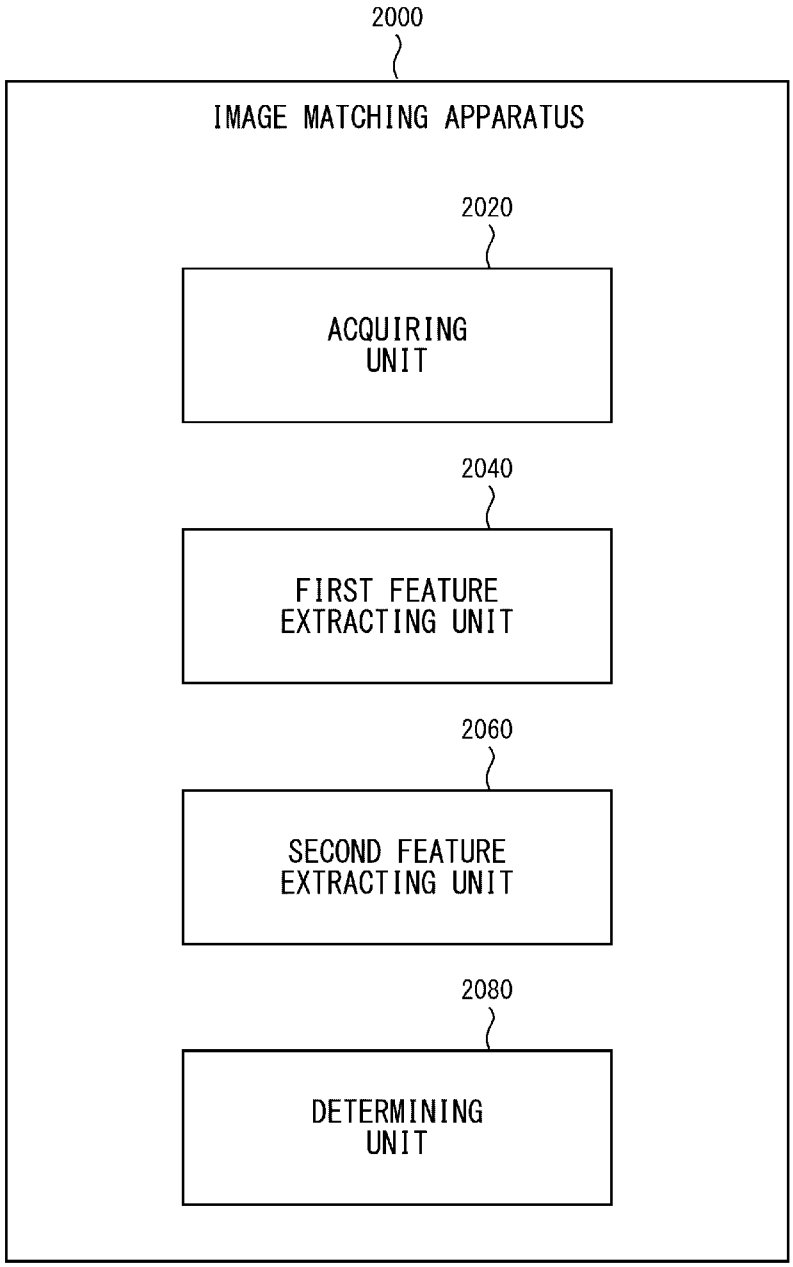

2000
IMAGE MATCHING APPARATUS
2020
ACQUIRING UNIT
2040
FIRST FEATURE EXTRACTING UNIT
2060
SECOND FEATURE EXTRACTING UNIT
2080
DETERMINING UNIT

Fig. 4

[Fig. 5]

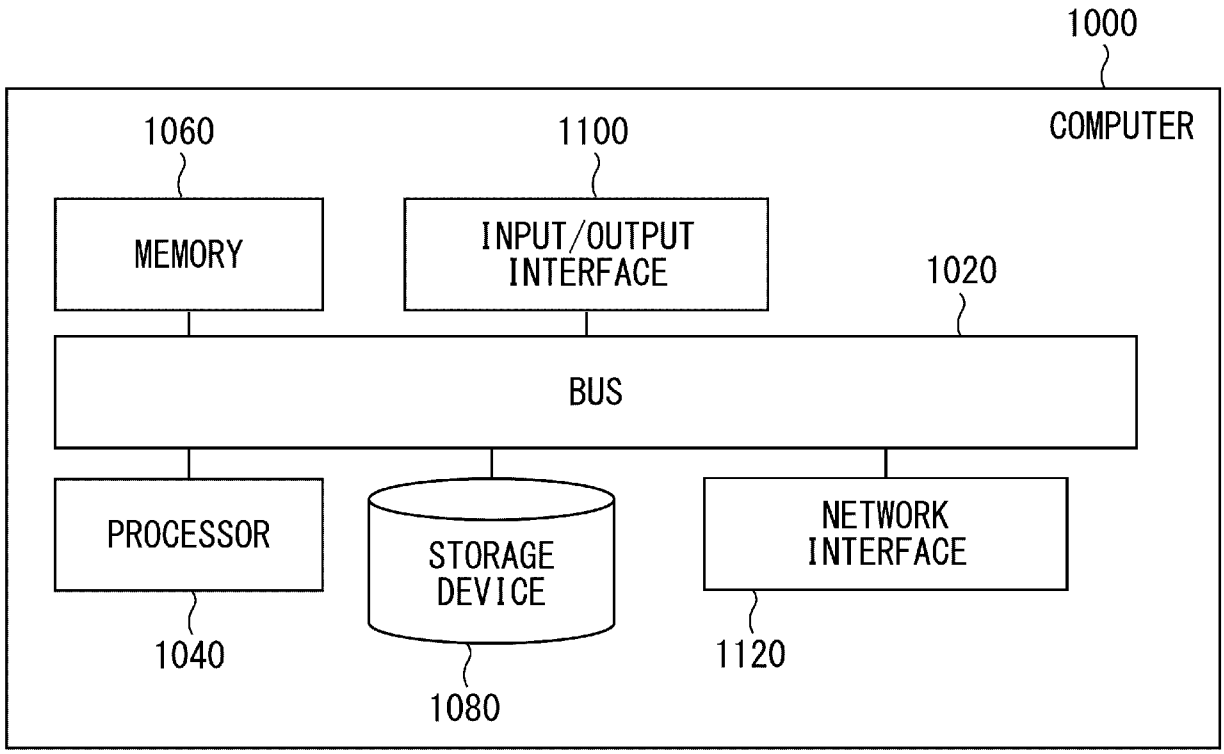

1000
COMPUTER
1060
MEMORY
1100
INPUT/OUTPUT INTERFACE
1020
BUS
PROCESSOR
1040
STORAGE DEVICE
1080
NETWORK INTERFACE
1120

Fig. 5

[Fig. 6]

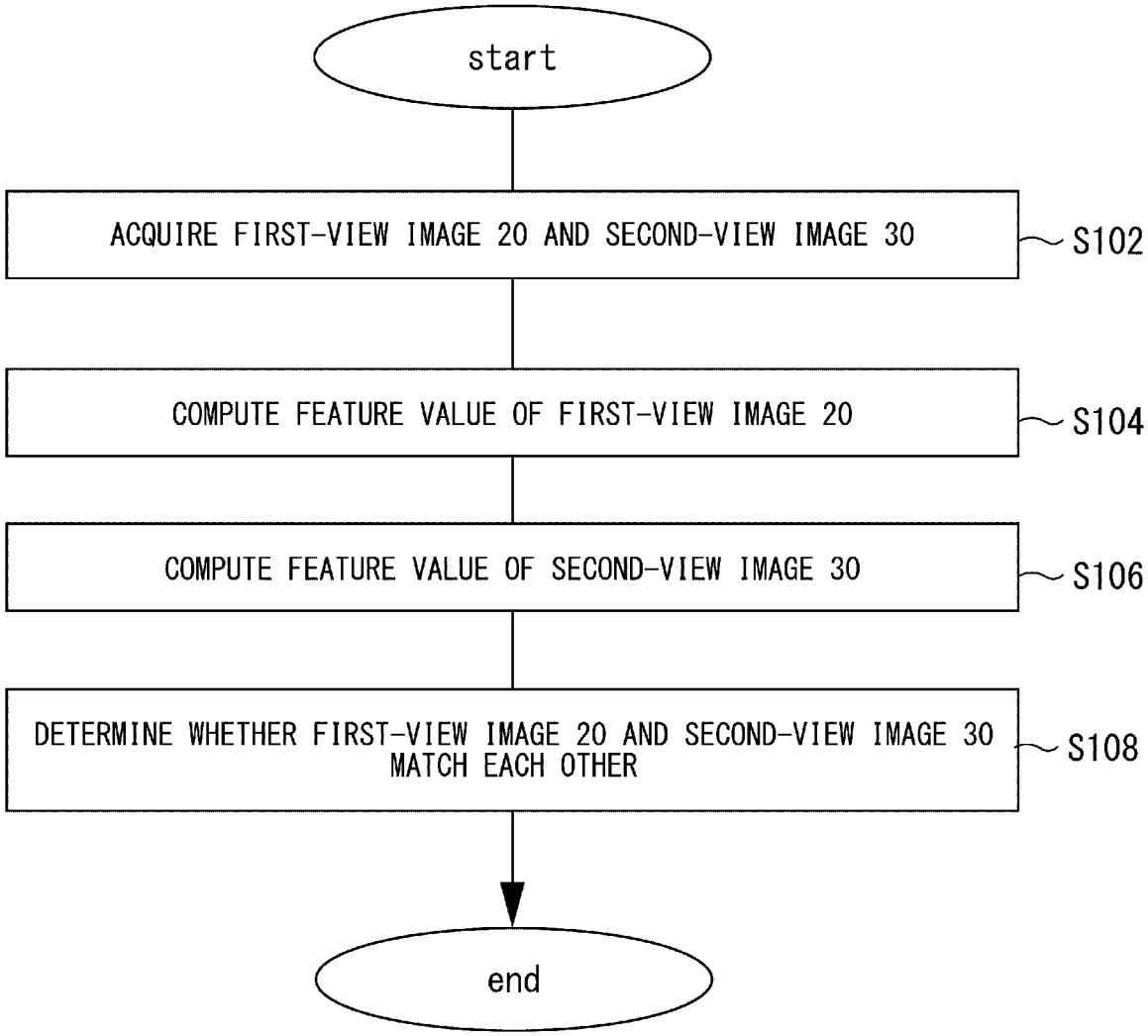

start
ACQUIRE FIRST-VIEW IMAGE 20 AND SECOND-VIEW IMAGE 30
S102
COMPUTE FEATURE VALUE OF FIRST-VIEW IMAGE 20
S104
COMPUTE FEATURE VALUE OF SECOND-VIEW IMAGE 30
S106
DETERMINE WHETHER FIRST-VIEW IMAGE 20 AND SECOND-VIEW IMAGE 30 MATCH EACH OTHER
S108
end

Fig. 6

[Fig. 7]

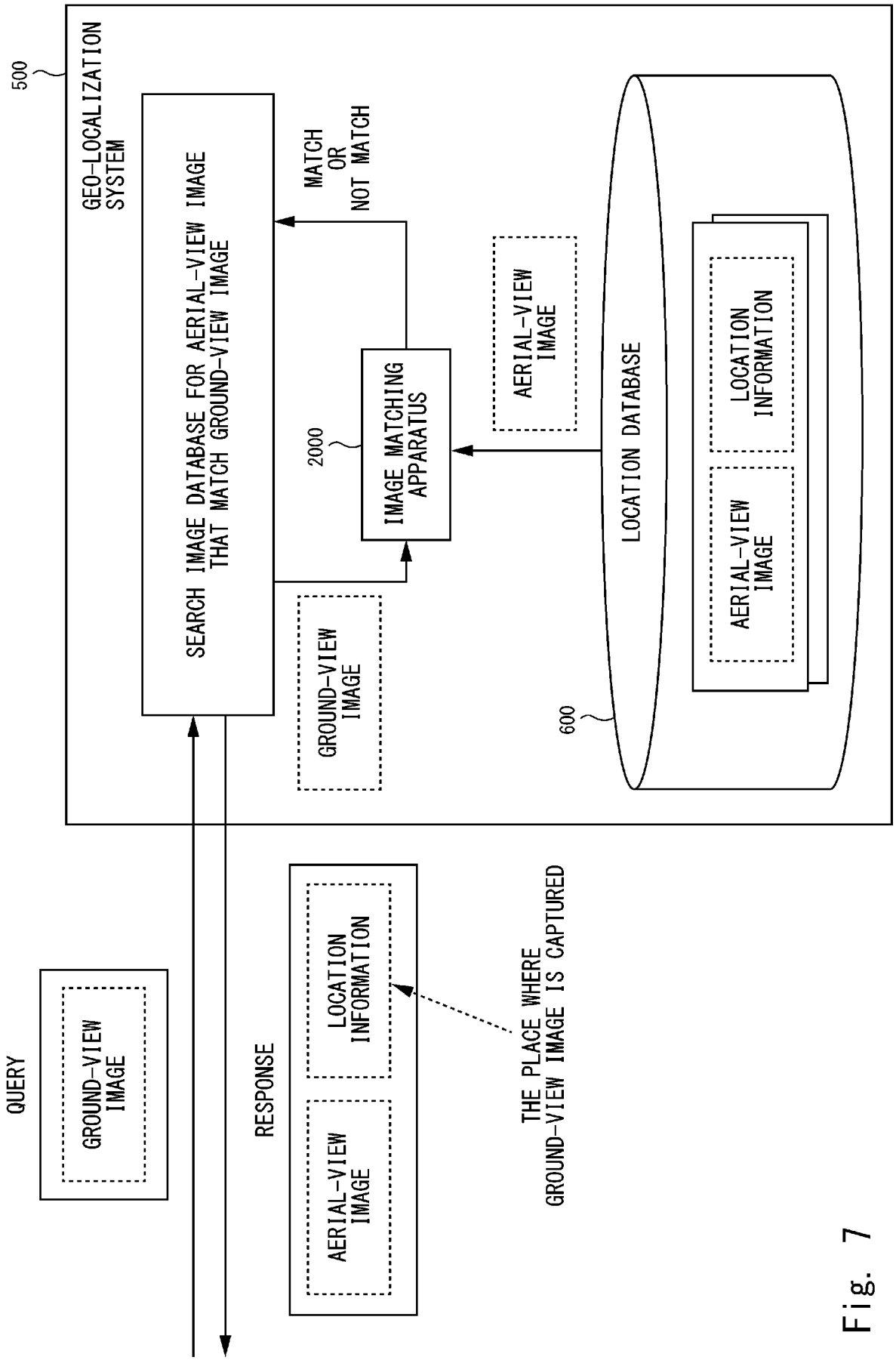

500
GEO-LOCALIZATION SYSTEM
SEARCH IMAGE DATABASE FOR AERIAL-VIEW IMAGE THAT MATCH GROUND-VIEW IMAGE
MATCH OR NOT MATCH
2000
IMAGE MATCHING APPARATUS
GROUND-VIEW IMAGE
AERIAL-VIEW IMAGE
LOCATION DATABASE
600
LOCATION INFORMATION
AERIAL-VIEW IMAGE
QUERY
GROUND-VIEW IMAGE
RESPONSE
LOCATION INFORMATION
AERIAL-VIEW IMAGE
THE PLACE WHERE GROUND-VIEW IMAGE IS CAPTURED

Fig. 7

[Fig. 8]

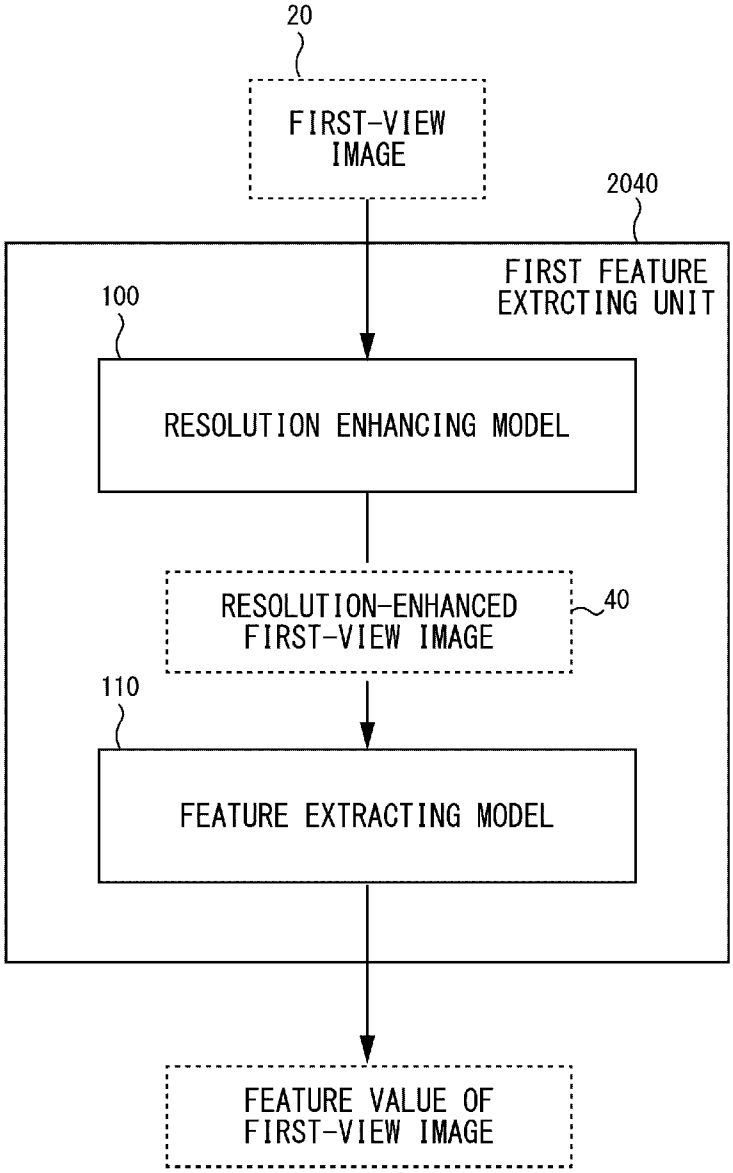

20
FIRST-VIEW IMAGE
2040
FIRST FEATURE EXTRCTING UNIT
100
RESOLUTION ENHANCING MODEL
RESOLUTION-ENHANCED FIRST-VIEW IMAGE
40
110
FEATURE EXTRACTING MODEL
FEATURE VALUE OF FIRST-VIEW IMAGE

Fig. 8

[Fig. 9]

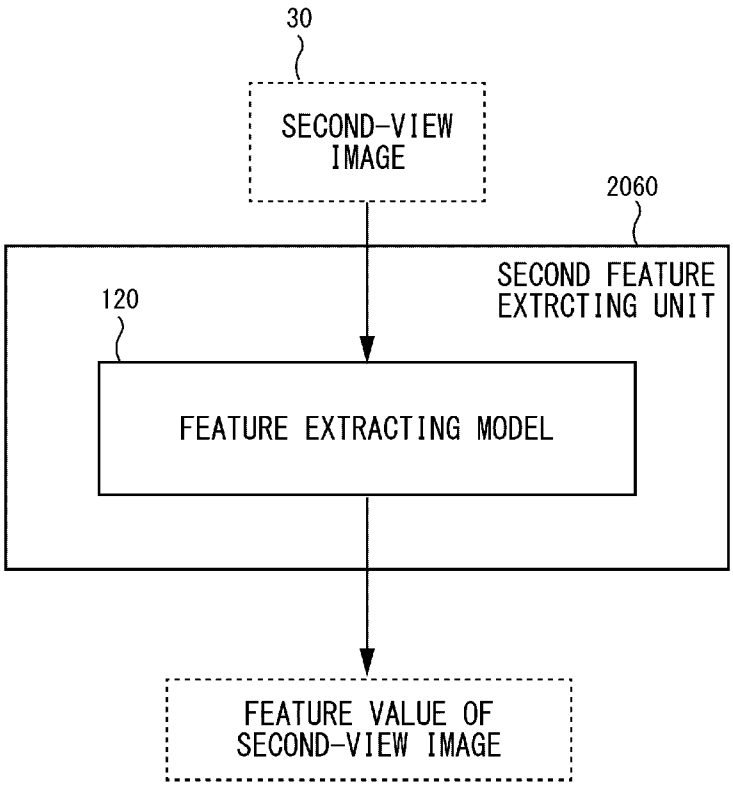

30
SECOND-VIEW IMAGE
2060
SECOND FEATURE EXTRCTING UNIT
120
FEATURE EXTRACTING MODEL
FEATURE VALUE OF SECOND-VIEW IMAGE

Fig. 9

[Fig. 10]

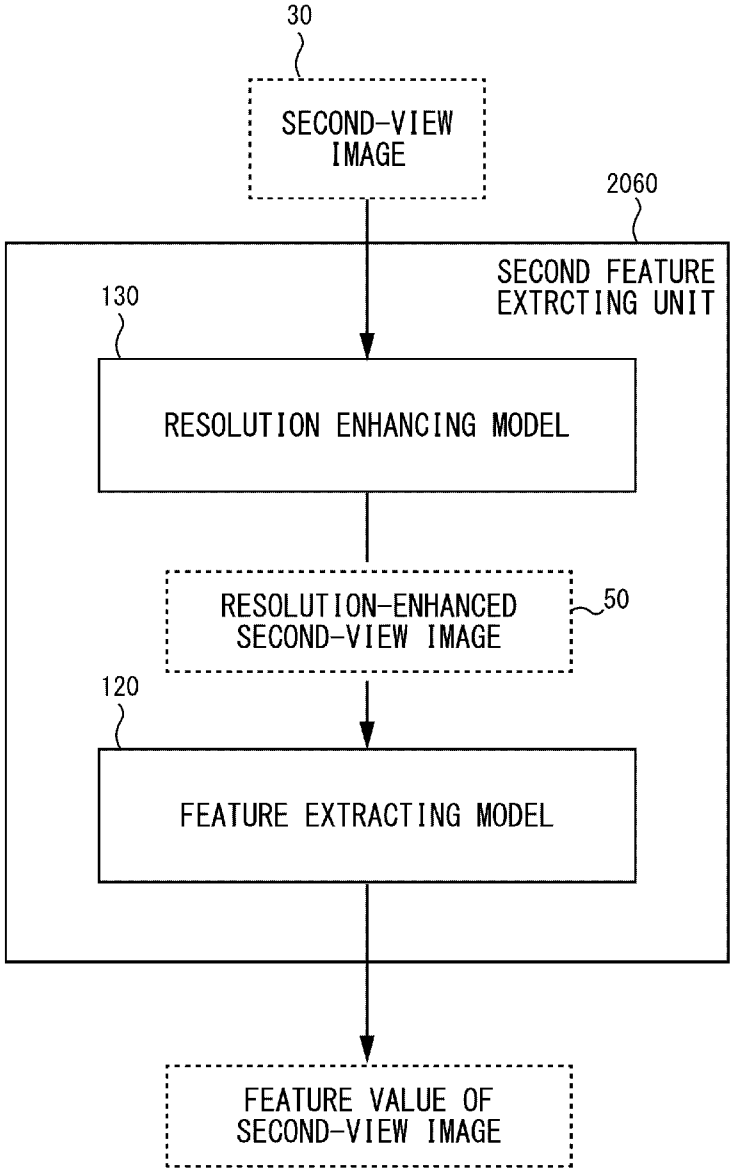

30
SECOND-VIEW IMAGE
2060
SECOND FEATURE EXTRCTING UNIT
130
RESOLUTION ENHANCING MODEL
RESOLUTION-ENHANCED SECOND-VIEW IMAGE
50
120
FEATURE EXTRACTING MODEL
FEATURE VALUE OF SECOND-VIEW IMAGE

Fig. 10

[Fig. 11]

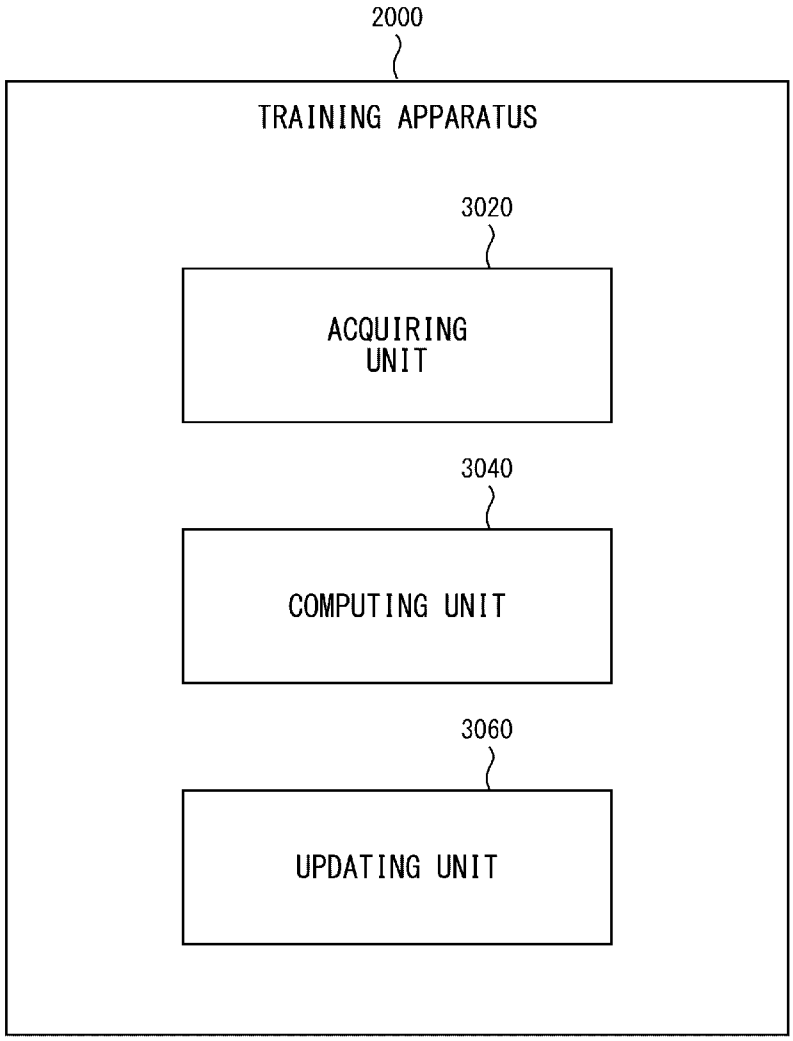

2000
TRAINING APPARATUS
3020
ACQUIRING UNIT
3040
COMPUTING UNIT
3060
UPDATING UNIT

Fig. 11

[Fig. 12]

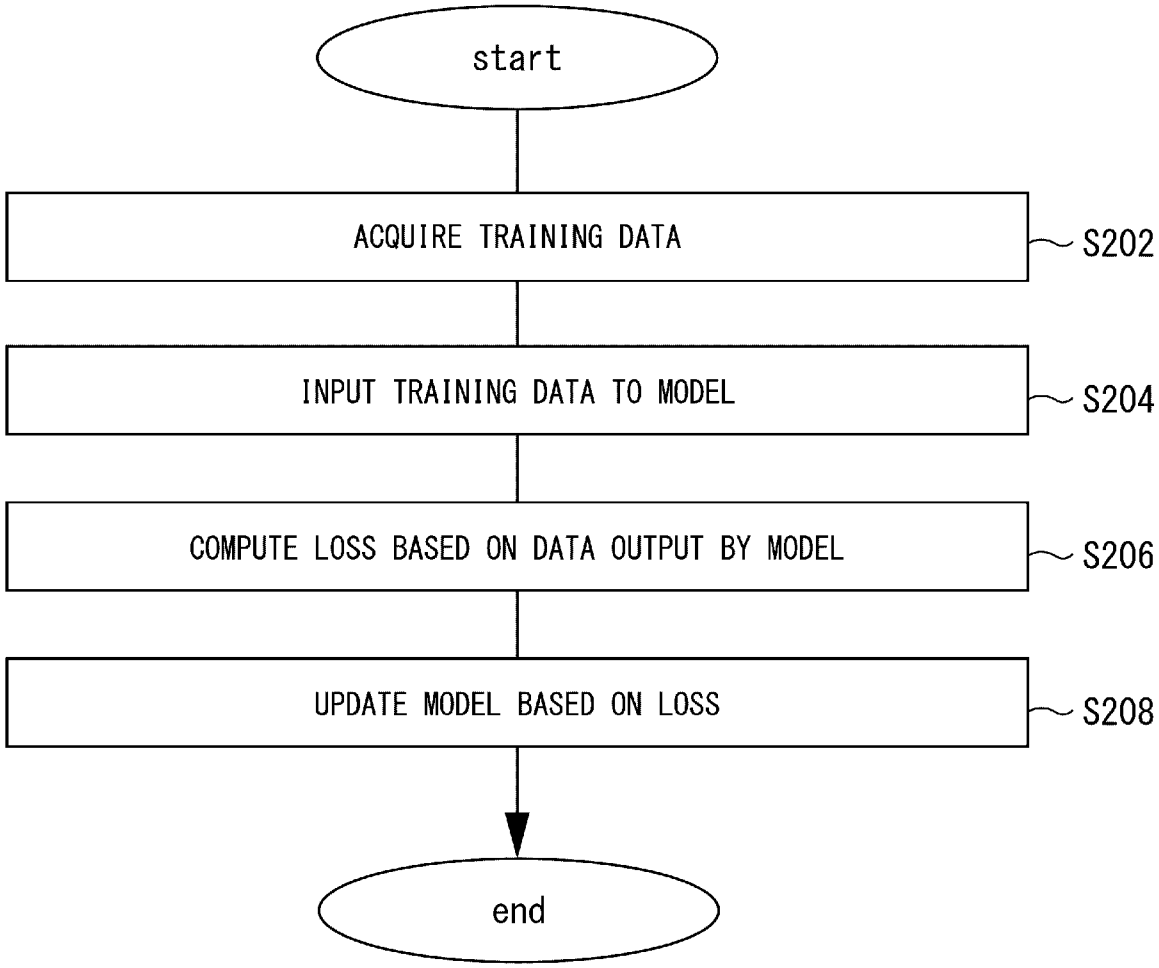

start
ACQUIRE TRAINING DATA
S202
INPUT TRAINING DATA TO MODEL
S204
COMPUTE LOSS BASED ON DATA OUTPUT BY MODEL
S206
UPDATE MODEL BASED ON LOSS
S208
end

Fig. 12

[Fig. 13]

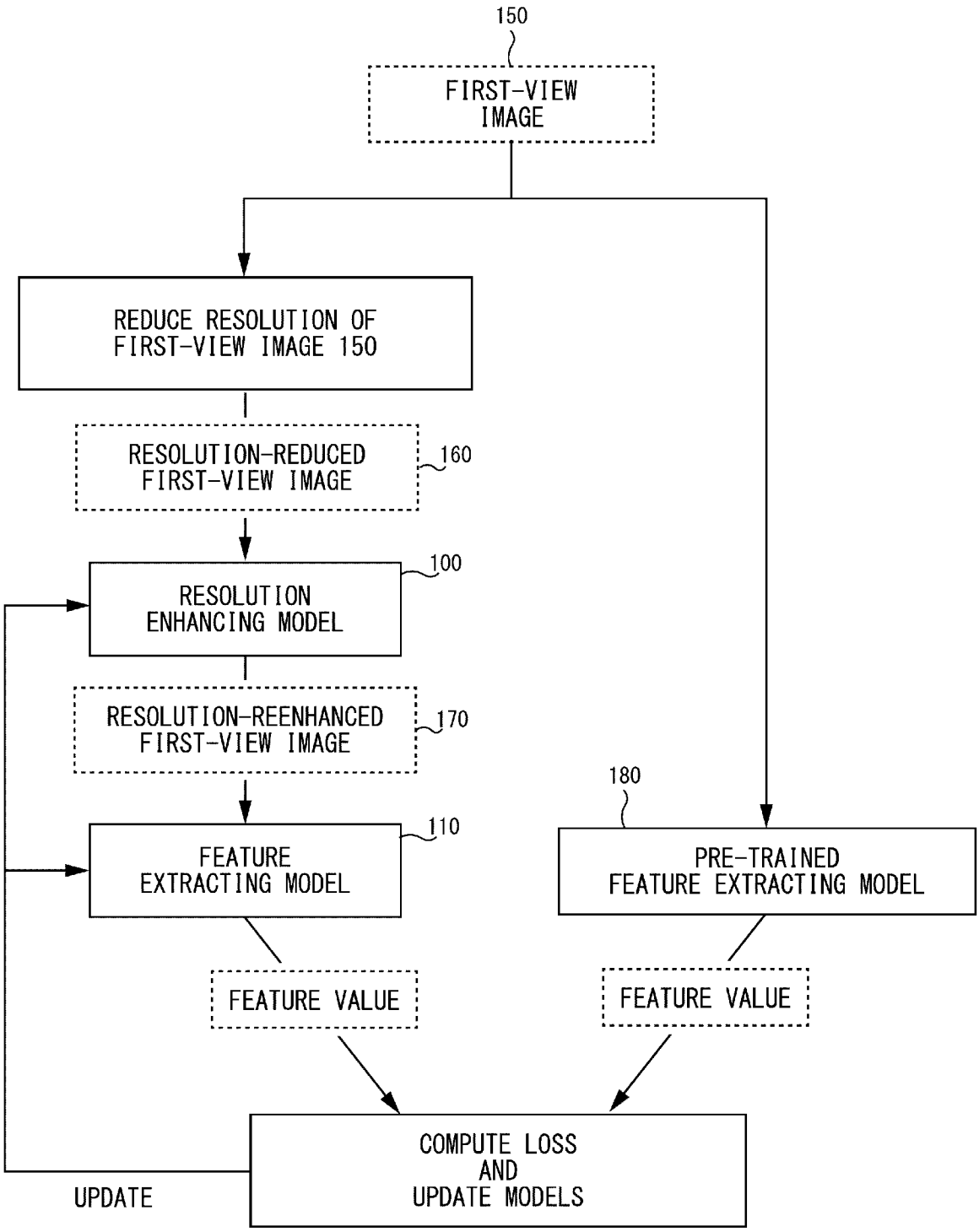

150
FIRST-VIEW IMAGE
REDUCE RESOLUTION OF FIRST-VIEW IMAGE 150
RESOLUTION-REDUCED FIRST-VIEW IMAGE
160
100
RESOLUTION ENHANCING MODEL
RESOLUTION-REENHANCED FIRST-VIEW IMAGE
170
180
110
FEATURE EXTRACTING MODEL
PRE-TRAINED FEATURE EXTRACTING MODEL
FEATURE VALUE
FEATURE VALUE
COMPUTE LOSS AND UPDATE MODELS
UPDATE

Fig. 13

[Fig. 14]

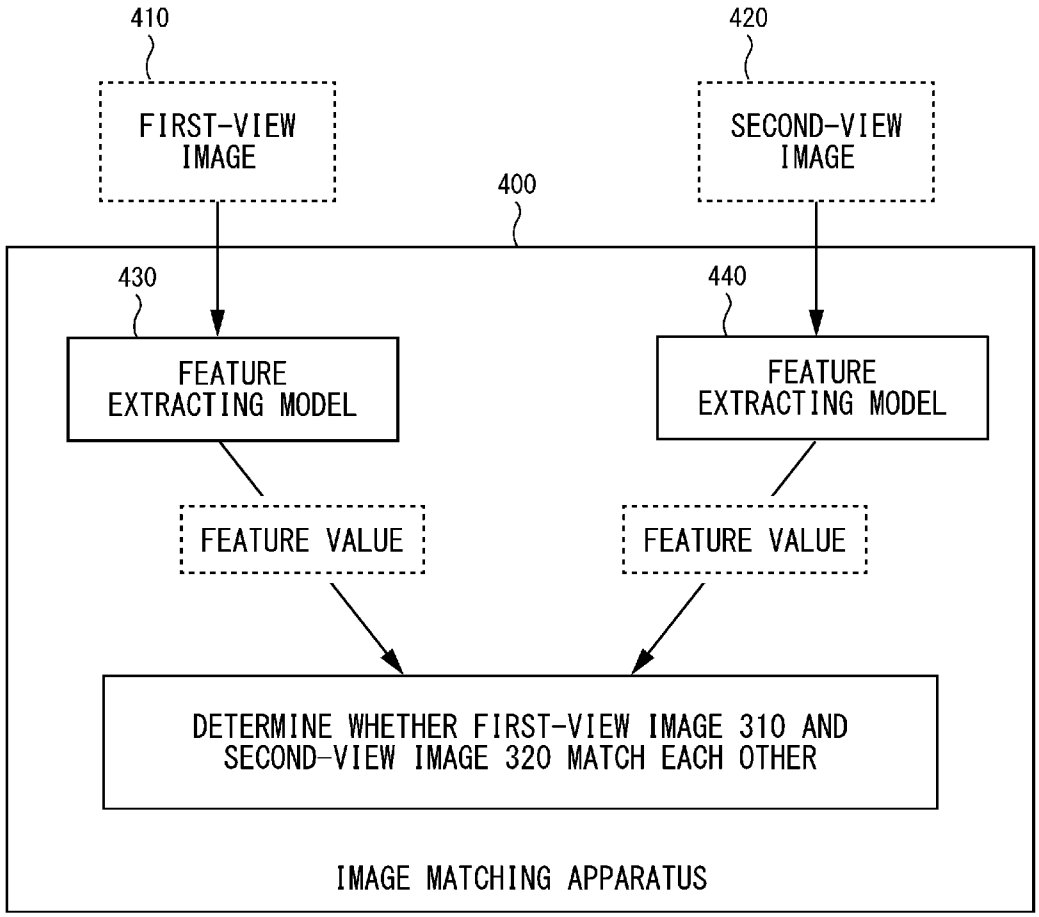

410
FIRST-VIEW IMAGE
420
SECOND-VIEW IMAGE
400
430
FEATURE EXTRACTING MODEL
440
FEATURE EXTRACTING MODEL
FEATURE VALUE
FEATURE VALUE
DETERMINE WHETHER FIRST-VIEW IMAGE 310 AND SECOND-VIEW IMAGE 320 MATCH EACH OTHER
IMAGE MATCHING APPARATUS

Fig. 14

[Fig. 15]

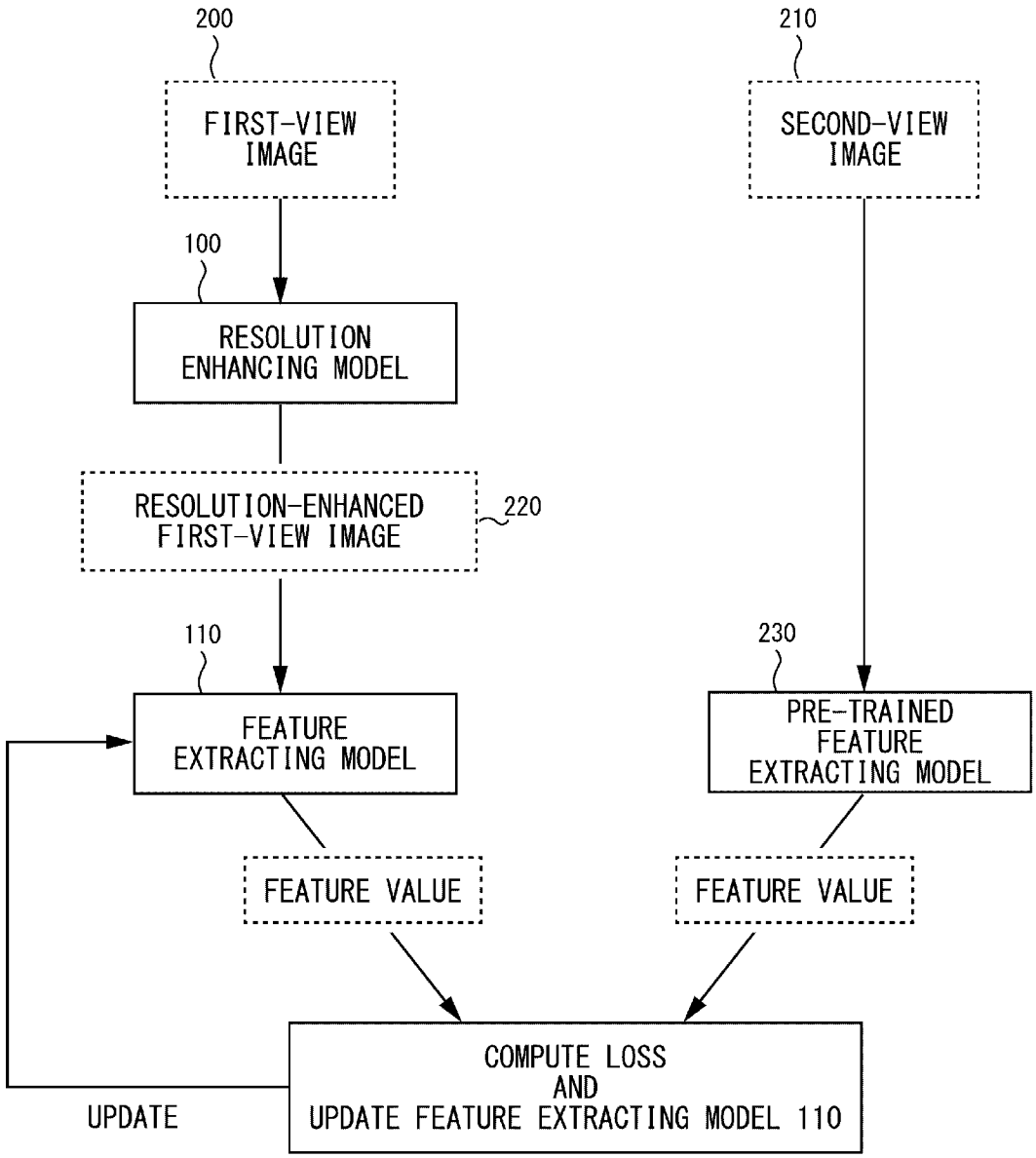

200
FIRST-VIEW IMAGE
210
SECOND-VIEW IMAGE
100
RESOLUTION ENHANCING MODEL
RESOLUTION-ENHANCED FIRST-VIEW IMAGE
220
110
FEATURE EXTRACTING MODEL
230
PRE-TRAINED FEATURE EXTRACTING MODEL
FEATURE VALUE
FEATURE VALUE
COMPUTE LOSS AND UPDATE FEATURE EXTRACTING MODEL 110
UPDATE

Fig. 15

[Fig. 16]

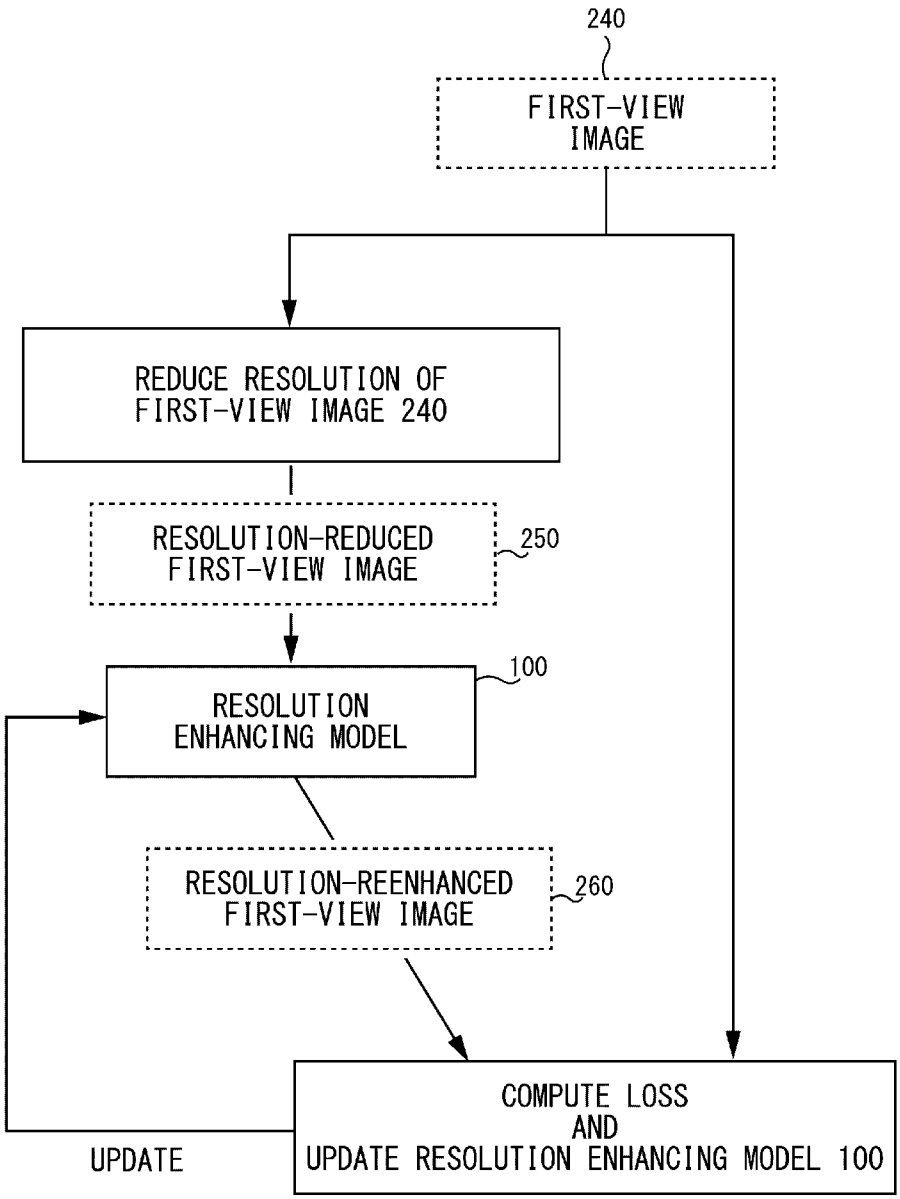

240
FIRST-VIEW IMAGE
REDUCE RESOLUTION OF FIRST-VIEW IMAGE 240
RESOLUTION-REDUCED FIRST-VIEW IMAGE
250
RESOLUTION ENHANCING MODEL
100
RESOLUTION-REENHANCED FIRST-VIEW IMAGE
260
COMPUTE LOSS AND UPDATE RESOLUTION ENHANCING MODEL 100
UPDATE

Fig. 16

[Fig. 17]

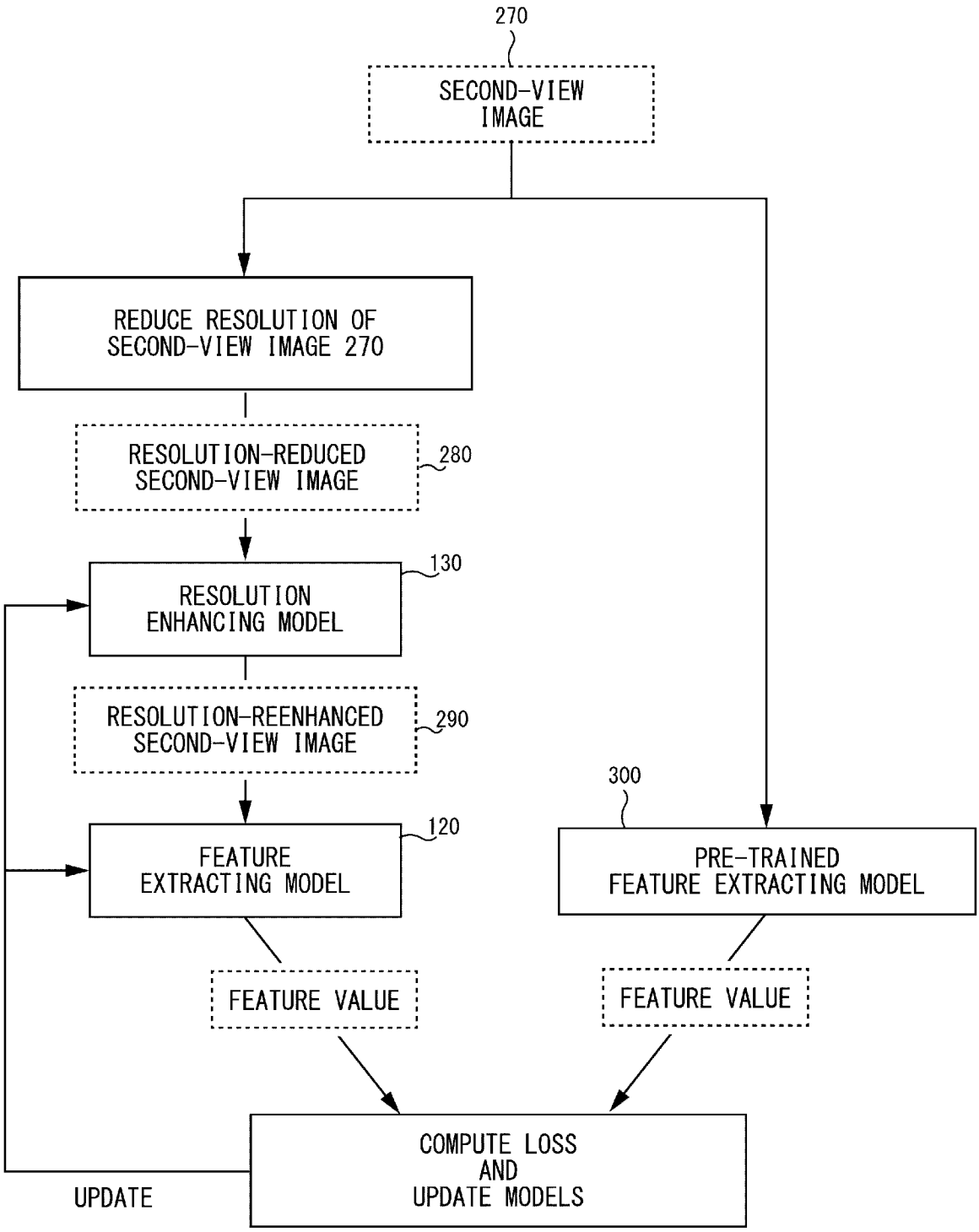

270
SECOND-VIEW IMAGE
REDUCE RESOLUTION OF SECOND-VIEW IMAGE 270
RESOLUTION-REDUCED SECOND-VIEW IMAGE
280
130
RESOLUTION ENHANCING MODEL
RESOLUTION-REENHANCED SECOND-VIEW IMAGE
290
300
120
FEATURE EXTRACTING MODEL
PRE-TRAINED FEATURE EXTRACTING MODEL
FEATURE VALUE
FEATURE VALUE
COMPUTE LOSS AND UPDATE MODELS
UPDATE

Fig. 17

[Fig. 18]

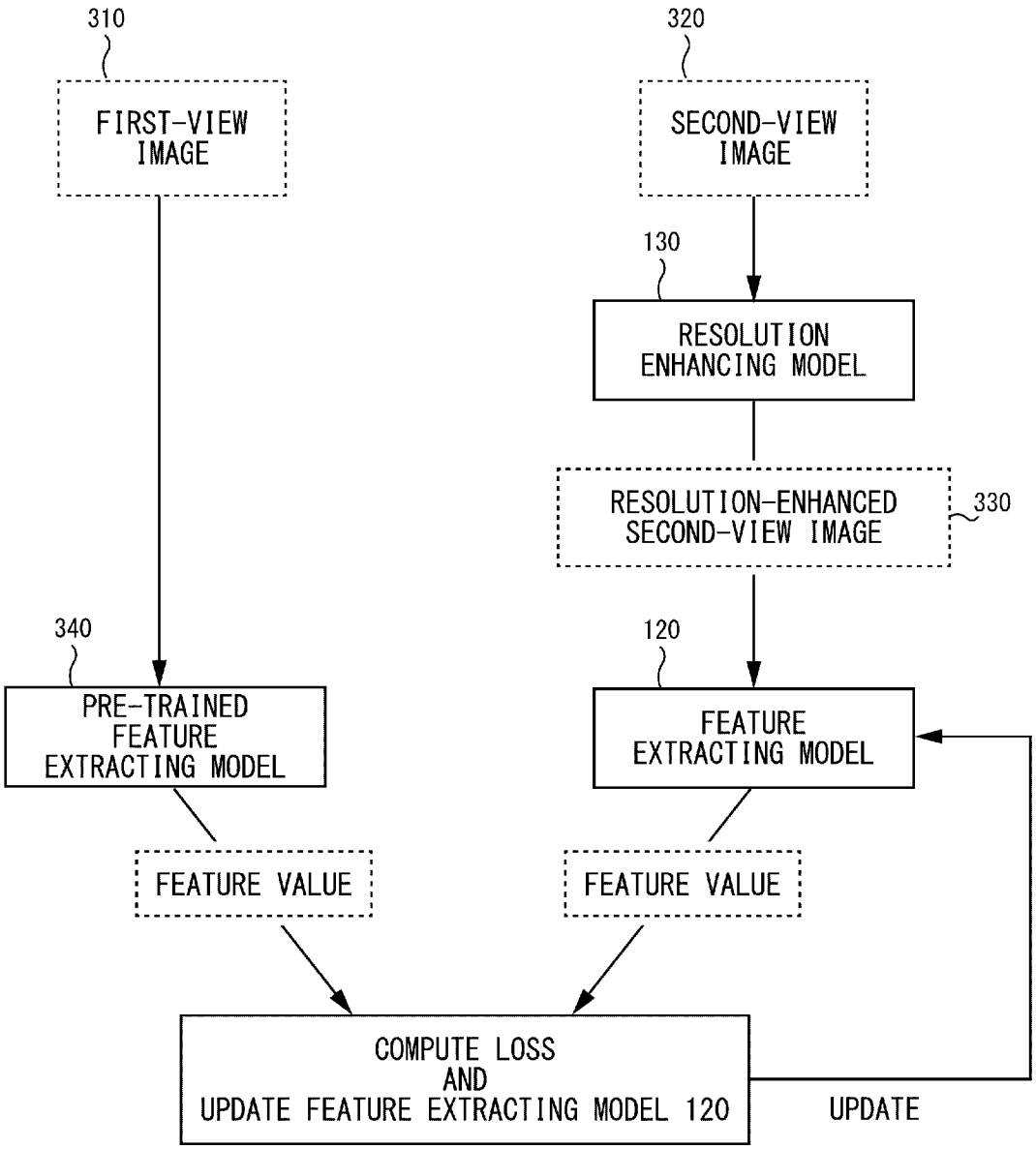

310
FIRST-VIEW IMAGE
320
SECOND-VIEW IMAGE
130
RESOLUTION ENHANCING MODEL
RESOLUTION-ENHANCED SECOND-VIEW IMAGE
330
340
PRE-TRAINED FEATURE EXTRACTING MODEL
120
FEATURE EXTRACTING MODEL
FEATURE VALUE
FEATURE VALUE
COMPUTE LOSS AND UPDATE FEATURE EXTRACTING MODEL 120
UPDATE

Fig. 18

[Fig. 19]

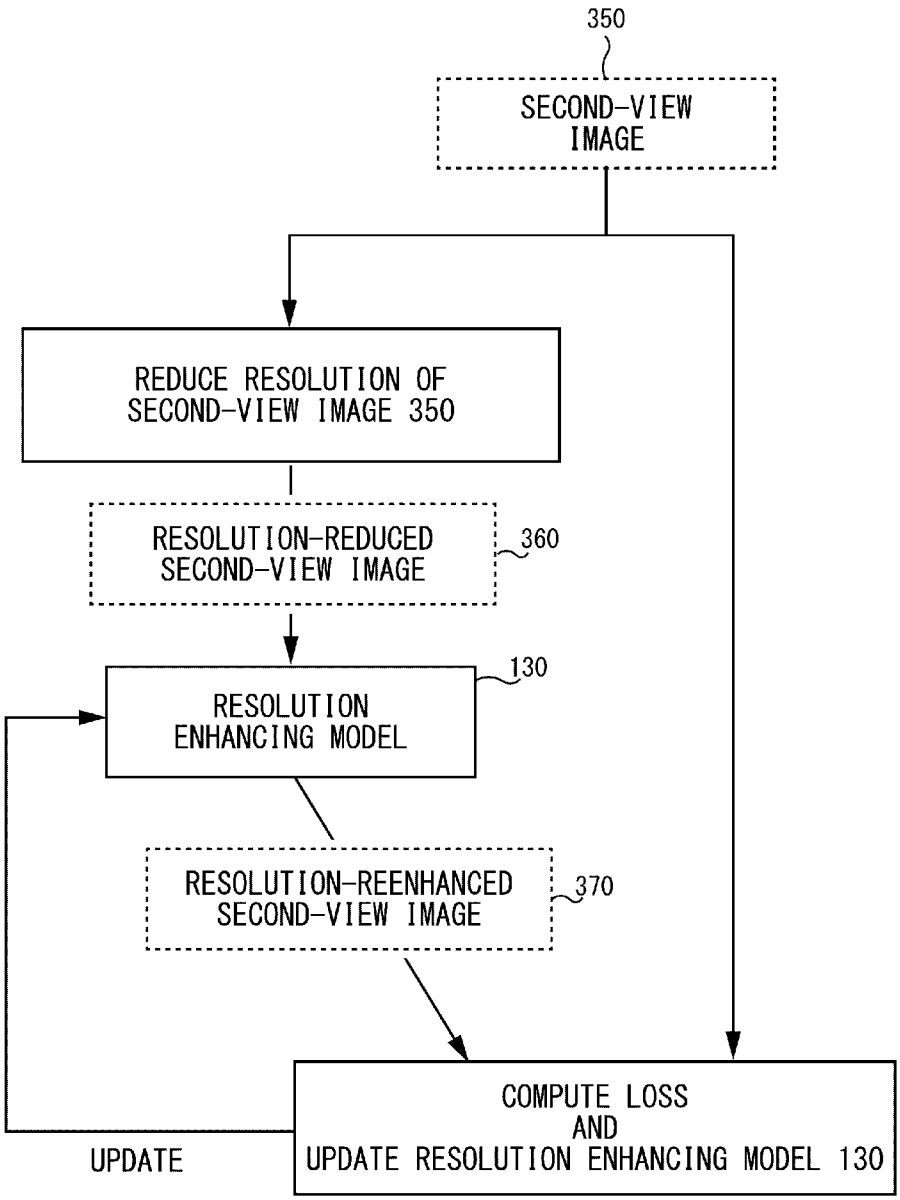

350
SECOND-VIEW IMAGE
REDUCE RESOLUTION OF SECOND-VIEW IMAGE 350
RESOLUTION-REDUCED SECOND-VIEW IMAGE
360
RESOLUTION ENHANCING MODEL
130
RESOLUTION-REENHANCED SECOND-VIEW IMAGE
370
COMPUTE LOSS AND UPDATE RESOLUTION ENHANCING MODEL 130
UPDATE

Fig. 19

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2023/029987

A. CLASSIFICATION OF SUBJECT MATTER

G06T 7/00(2017.01)i
FI: G06T7/00 300F

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)

G06T7/00

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

Published examined utility model applications of Japan 1922-1996
Published unexamined utility model applications of Japan 1971-2023
Registered utility model specifications of Japan 1996-2023
Published registered utility model applications of Japan 1994-2023

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2022/044104 A1 (NEC CORPORATION) 03 March 2022 (2022-03-03) paragraphs [0012]- [0015], [0028]-[0032]	1-18
Y	JP 2005-215883 A (SONY CORPORATION) 11 August 2005 (2005-08-11) paragraphs [0038], [0046]-[0049], [0080]	1-6
Y	WO 2019/230665 A1 (NIPPON TELEGRAPH AND TELEPHONE CORPORATION) 27 May 2019 (2019-05-27) paragraphs [0034], [0040], [0042], [0046], [0059]-[0062]	7-18
Y	WO 2022/243671 A1 (CALIPSA LIMITED) 17 May 2022 (2022-05-17) p. 8, lines 21-32	8,12,16
Y	WO 2016/019484 A1 (TANG, Xiaoou) 11 February 2016 (2016-02-11) paragraphs [0057]-[0058]	8,12,16
Y	CN 114663965 B (LABORATORY IN JIANJIANG RIVER) 21 October 2022 (2022-10-21) paragraphs [0012],[0031]-[0040]	10,14,18

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

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
27 September 2023	**24 October 2023**
Name and mailing address of the ISA/JP **Japan Patent Office** **3-4-3, Kasumigaseki, Chiyoda-ku, Tokyo** **100-8915, Japan**	Authorized officer **ICHIJI, Kazuyuki 5H 2571** Telephone No. **+81-3-3581-1101 Ext. 3545**

INTERNATIONAL SEARCH REPORT

International application No.
PCT/JP2023/029987

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2021/256091 A1 (SONY GROUP CORPORATION) 23 December 2021 (2021-12-23) paragraphs [0028]-[0037], [0079]-[0083], [0090]-[0091]	10,14,18

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/JP2023/029987

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
WO 2022/044104 A1	03 March 2022	(Family: none)	
JP 2005-215883 A	11 August 2005	US 2005/0180636 A1 paragraphs [0038], [0046]-[0049], [0080]	
WO 2019/230665 A1	27 May 2019	JP 2019-211912 A paragraphs [0034], [0040], [0042], [0046], [0059]-[0062]	
WO 2022/243671 A1	17 May 2022	(Family: none)	
WO 2016/019484 A1	11 February 2016	CN 106796716 A	
CN 114663965 B	21 October 2022	(Family: none)	
WO 2021/256091 A1	23 December 2021	EP 4167183 A1 paragraphs [0028]-[0037], [0079]-[0083], [0090]-[0091]	