

[DESCRIPTION]

[Title of Invention]

TRAINING APPARATUS, TRAINING METHOD, AND NON-TRANSITORY COMPUTER-
5 READABLE MEDIUM

[Technical Field]

[0001]

The present disclosure generally relates to a training apparatus, a training method, and a non-transitory computer-readable medium.

10 [Background Art]

[0002]

A computer system that performs cross-view matching 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
15 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
20 the extracted features.

[Citation List]

[Non Patent Literature]

[0003]

NPL1: Liu Liu and Hongdong Li, "Lending Orientation to Neural Networks for Cross-view Geo-
25 localization", [online], March 29, 2019, [retrieved on 2023-12-12], retrieved from <arXiv, <https://arxiv.org/pdf/1903.12351.pdf>>

[Summary of Invention]

[Technical Problem]

[0004]

30 The system disclosed by NPL1 handles only ground-view images and aerial-view images for cross-view matching. An objective of the present disclosure is to provide a novel technique to handle three or more types of data for cross-view matching.

[Solution to Problem]

[0005]

The present disclosure 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 first training data and a second training data, the first training data including a first-type representation data and a second-type representation data that represent a same place as each other in different types of representation from each other, the second training data including a first-type representation data and a third-type representation data that represent a same place as each other in different types of representation from each other; train a second feature extracting model based on a feature value of the first-type representation data of the first training data computed by a pre-trained first feature extracting model and a feature value of the second-type representation data computed by the second feature extracting model; and train a third feature extracting model based on a feature value of the first-type representation data of the second training data computed by the pre-trained first feature extracting model and a feature value of the third-type representation data computed by the third feature extracting model.

15 [0006]

The present disclosure further provides a training method that is performed by a computer. The training method comprises: acquiring a first training data and a second training data, the first training data including a first-type representation data and a second-type representation data that represent a same place as each other in different types of representation from each other, the second training data including a first-type representation data and a third-type representation data that represent a same place as each other in different types of representation from each other; training a second feature extracting model based on a feature value of the first-type representation data of the first training data computed by a pre-trained first feature extracting model and a feature value of the second-type representation data computed by the second feature extracting model; and training a third feature extracting model based on a feature value of the first-type representation data of the second training data computed by the pre-trained first feature extracting model and a feature value of the third-type representation data computed by the third feature extracting model.

[0007]

The present disclosure further provides a non-transitory computer-readable medium storing a program that cause a computer to execute: acquiring a first training data and a second training data, the first training data including a first-type representation data and a second-type representation data that represent a same place as each other in different types of representation from each other, the second training data including a first-type representation data and a third-type

representation data that represent a same place as each other in different types of representation from each other; training a second feature extracting model based on a feature value of the first-type representation data of the first training data computed by a pre-trained first feature extracting model and a feature value of the second-type representation data computed by the second feature
 5 extracting model; and training a third feature extracting model based on a feature value of the first-type representation data of the second training data computed by the pre-trained first feature extracting model and a feature value of the third-type representation data computed by the third feature extracting model.

[0008]

- 10 According to the present disclosure, it is possible to provide a novel technique to provide a novel technique to handle three or more types of data for cross-view matching.

[Brief Description of Drawings]

[0009]

[Fig. 1]

- 15 Fig. 1 illustrates an overview of a training apparatus.

[Fig. 2]

Fig. 2 illustrates an example operation of the image matching apparatus.

[Fig. 3]

- Fig. 3 illustrates how the feature extracting models work before the training of the second feature
 20 extracting model and the third feature extracting model.

[Fig. 4]

Fig. 4 illustrates how the feature extracting models work after the training of the second feature extracting model and the third feature extracting model.

[Fig. 5]

- 25 Fig. 5 is a block diagram showing an example of the functional configuration of the training apparatus.

[Fig. 6]

Fig. 6 is a block diagram illustrating an example of the hardware configuration of a computer realizing the image matching apparatus.

- 30 [Fig. 7]

Fig. 7 illustrates an example of the ground-view image, the aerial-view image, and the map image.

[Fig. 8]

Fig. 8 illustrates an example of the 3D shape data.

[Fig. 9]

Fig. 9 illustrates examples of keyword matrix data.

[Fig. 10]

Fig. 10 illustrates a geo-localization system that includes the image matching apparatus.

5 [Fig. 11]

Fig. 11 is a flowchart illustrating an example flow of processes performed by the training apparatus.

[Description of Embodiments]

[0010]

Example embodiments according to the present disclosure will be described hereinafter
 10 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.

15 [0011]

FIRST EXAMPLE EMBODIMENT

<Overview>

Fig. 1 illustrates an overview of a training apparatus 2000. The training apparatus 2000 performs training for an image matching apparatus 300 that functions as a discriminator that
 20 performs matching between two-types of representation data. The representation data is data that represents a specific place in the real world: e.g., a ground-view image on which a ground view of a specific place is captured. The image matching apparatus 300 should determine that two representation data match each other when the two representation data represent the same place as each other.

25 [0012]

The types of representation data may include a ground-view image, an aerial-view image, a map image, a three-dimensional (3D) shape data, a keyword matrix data, etc. The detailed explanations about each type of representation data will be described later.

[0013]

30 The image matching apparatus 300 is configured to handle at least three types of representation data. Specifically, the image matching apparatus 300 includes a feature extracting model for each one of the types of representation data that the image matching apparatus 300 handles. Suppose that the image matching apparatus 300 is configured to handle ground-view

images, aerial-view images, and map images. In this case, the image matching apparatus 300 has a feature extracting model to compute a feature value of a ground-view image, a feature extracting model to compute a feature value of an aerial-view image, and a feature extracting model to compute a feature value of a map image.

5 [0014]

To be capable of extracting the feature value from at least three types of representation data, the image matching apparatus 300 includes at least three type of feature extracting models: a first feature extracting model 10 computes a feature value of a first-type representation data; a second feature extracting model 20 that computes a feature value of a second-type representation data; and a third feature extracting model 30 that computes a feature value of a third-type representation data. The first feature extracting model 10 is a pre-trained machine learning-based model, whereas the second feature extracting model 20 and the third feature extracting model 30 are machine learning-based models to be trained by the training apparatus 2000 using the first feature extracting model 10.

15 [0015]

The image matching apparatus 300 acquires two types of representation data to be compared, extracts a feature value from each of the acquired representation data, and determines whether the acquired representation data match each other based on the feature value extracted therefrom. Fig. 2 illustrates an example operation of the image matching apparatus 300. The image matching apparatus 300 acquires an i-th-type representation data and a j-th type representation data. When the image matching apparatus 300 is configured to handle three types of representation data, i and j are 1, 2, or 3 and $i \neq j$.

[0016]

It is noted that the image matching apparatus 300 selects the feature extracting models to be used based on the types of representation data input thereinto. For example, when the first-type representation data and the third-type representation data are input into the image matching apparatus 300, the first feature extracting model 10 and the third feature extracting model 30 are selected. In this case, i is 1 and j is 3 in Fig. 2

[0017]

30 The image matching apparatus 300 inputs the i-th-type representation data into an i-th feature extracting model to compute the feature value of the i-th-type representation data while the image matching apparatus 300 inputs the j-th-type representation data into a j-th feature extracting model to compute the feature value of the j-th-type representation data. The image matching

apparatus 300 compares the feature value of the i-th-type representation data and the feature value of the j-th-type representation data to determine whether the i-th-type representation data the j-th-type representation data match each other.

[0018]

5 To determine whether the i-th-type representation data the j-th-type representation data match each other, the image matching apparatus 300 may compute a degree of similarity between the feature value of the i-th-type representation data and the feature value of the j-th type representation data. In some implementations, the degree of similarity between two feature values is represented by a distance between the two feature values.

10 [0019]

When the degree of similarity between the feature value of the i-th-type representation data and the feature value of the j-th type representation data is substantially high (e.g., larger than or equal to a predefined threshold), the image matching apparatus 300 determines that the i-th-type representation data the j-th-type representation data match each other. On the other hand, 15 when the degree of similarity between the feature value of the i-th-type representation data and the feature value of the j-th type representation data is not substantially high (e.g., smaller than the predefined threshold), the image matching apparatus 300 determines that the i-th-type representation data the j-th-type representation data do not match each other.

[0020]

20 To train the second feature extracting model 20 and the third feature extracting model 30, the training apparatus 2000 acquires two pieces of training data (training data 40-1 and training data 40-2). The training data 40-1 includes a first-type representation data 50-1 and a second-type representation data 60 while the training data 40-2 includes a first-type representation data 50-2 and a third-type representation data 70. The first-type representation data 50-1 and the second 25 type-representation data 60 represent the same place as each other in different formats from each other (in different types of representation from each other). Suppose that the first-type representation data 50-1 is a ground-view image on which a ground-view of a specific facility F1 is captured. In this case, the second-type representation data is, for example, an aerial-view image on which an aerial-view of the facility F1 is captured.

30 [0021]

The first-type representation data 50-2 and the third type-representation data 70 represent the same place as each other in different formats from each other (in different types of representation from each other). Suppose that the first-type representation data 50-2 is a ground-

view image on which a ground-view of a specific facility F2 is captured. In this case, the third-type representation data is, for example, a map image that visually represents map information around the facility F2. It is noted that the first-type representation data 50-1 and the first-type representation data 50-2 may be the same data as each other or different data from each other.

5 [0022]

The training apparatus 2000 uses the training data 40-1 to train the second feature extracting model 20 while using the training data 40-2 to train the third feature extracting model 30. Specifically, the training apparatus 2000 inputs the first-type representation data 50-1 into the first feature extracting model 10 to extract the feature value of the first-type representation data 10 50-1. In addition, the training apparatus 2000 inputs the second-type representation data 60 into the second feature extracting model 20 to extract the feature value of the second-type representation data 60. Then, the training apparatus 2000 computes a loss for the feature value of the first-type representation data 50-1 and the feature value of the second-type representation data 60, and updates trainable parameters of the second feature extracting model 20 to train the second 15 feature extracting model 20.

[0023]

Similarly, the training apparatus 2000 inputs the first-type representation data 50-2 into the first feature extracting model 10 to extract the feature value of the first-type representation data 50-2. In addition, the training apparatus 2000 inputs the third-type representation data 70 into the 20 third feature extracting model 30 to extract the feature value of the third-type representation data 70. Then, the training apparatus 2000 computes a loss for the feature value of the first-type representation data 50-2 and the feature value of the third-type representation data 70, and updates trainable parameters of the third feature extracting model 30 to train the third feature extracting model 30.

25 [0024]

<Example of Advantageous Effect>

According to the training apparatus 2000, the second feature extracting model 20 is trained based on the loss computed for the feature value of the first-type representation data 50-1 extracted by the first feature extracting model 10 and the feature value of the second-type representation data 60 extracted by the second feature extracting model 20. Thus, the image matching apparatus 300 becomes capable of performing matching between the first-type representation data 50 and the second-type representation data 60. In addition, the third feature extracting model 30 is trained based on the loss computed for the feature value of the first-type

representation data 50-2 extracted by the first feature extracting model 10 and the feature value of the third-type representation data 70 extracted by the third feature extracting model 30. Thus, the image matching apparatus 300 becomes capable of performing matching between the first-type representation data 50 and the third-type representation data 70.

5 [0025]

Furthermore, the image matching apparatus 300 becomes capable of performing matching between the second-type representation data 60 and the third-type representation data 70, even without training the second feature extracting model 20 and the third feature extracting model 30 based on a loss computed for a feature value of the second-type representation data 60 and the
10 feature value of the third-type representation data 70. Thus, the training performed by the training apparatus 2000 is advantageous in that there is no need to prepare pairs of the second-type representation data 60 and the third-type representation data 70.

[0026]

The reason why the image matching apparatus 300 becomes capable of performing
15 matching between the second-type representation data 60 and the third-type representation data 70 will be explained with referring to Fig. 3 and 4. Fig. 3 illustrates how the feature extracting models work before the training of the second feature extracting model 20 and the third feature extracting model 30. Before the training, the first feature extracting model 10, the second feature extracting model 20, and the third feature extracting model 30 map the representation data to different feature
20 spaces from each other.

[0027]

Fig. 4 illustrates how the feature extracting models work after the training of the second feature extracting model 20 and the third feature extracting model 30. Since the second feature extracting model 20 is trained based on the loss computed for the feature value of the first-type representation data 50-1 and the feature value of the second-type representation data 60, and since
25 the first feature extracting model 10 is pretrained, the second feature extracting model 20 becomes mapping the second-type representation data to the feature space to which the first feature extracting model 10 maps the first-type representation data 50. Specifically, the second feature extracting model 20 is trained to compute the feature value of the second-type representation data
30 60 having higher similarity to the feature value of the first-type representation data 50-1 when the first-type representation data 50-1 and the second-type representation data 60 are to be matched each other (i.e., they include the same place as each other). In addition, the second feature extracting model 20 is trained to compute the feature value of the second-type representation data

60 having lower similarity to the feature value of the first-type representation data 50-1 when the first-type representation data 50-1 and the second-type representation data 60 are not to be matched each other (i.e., they include different places from each other).

[0028]

5 Similarly, since the third feature extracting model 30 is trained based on the loss computed for the feature value of the first-type representation data 50-2 and the feature value of the third-type representation data 70, and since the first feature extracting model 10 is pretrained, the third feature extracting model 30 becomes mapping the third-type representation data 70 to the feature space to which the first feature extracting model 10 maps the first-type representation data 50.

10 Specifically, the third feature extracting model 30 is trained to compute the feature value of the third-type representation data 70 having higher similarity to the feature value of the first-type representation data 50-2 when the first-type representation data 50-2 and the third-type representation data 70 are to be matched each other (i.e., they include the same place as each other). In addition, the third feature extracting model 30 is trained to compute the feature value of the

15 third-type representation data 70 having lower similarity to the feature value of the first-type representation data 50-2 when the first-type representation data 50-2 and the third-type representation data 70 are not to be matched each other (i.e., they include different places from each other).

[0029]

20 As a result of the training, the first feature extracting model 10, the second feature extracting model 20, the third feature extracting model 30 are configured to map the representation data to the same feature space as each other as illustrated in Fig. 4. Thus, the image matching apparatus 300 also can take a pair of the second-type representation data 60 and the third-type representation data 70 to determine whether they match each other.

25 [0030]

Hereinafter, a more detailed explanation of the image matching apparatus 300 will be described.

[0031]

<Example of Functional Configuration>

30 Fig. 5 is a block diagram showing an example of the functional configuration of the training apparatus 2000. The training apparatus 2000 includes an acquiring unit 2020, a feature extracting unit 2040, and an updating unit 2060. The acquiring unit 2020 acquires the training data 40-1 and 40-2. The training data 40-1 includes the first-type representation data 50-1 and the

second-type representation data 60 while the training data 40-2 includes the first-type representation data 50-2 and the third-type representation data 70.

[0032]

The feature extracting unit 2040 inputs the first-type representation data 50-1 into the first feature extracting model 10 to compute the feature value of the first-type representation data 50-1. The feature extracting unit 2040 inputs the first-type representation data 50-2 into the first feature extracting model 10 to compute the feature value of the first-type representation data 50-2. The feature extracting unit 2040 inputs the second-type representation data 60 into the second feature extracting model 20 to compute the feature value of the second-type representation data 60. The feature extracting unit 2040 inputs the third-type representation data 70 into the third feature extracting model 30 to compute the feature value of the third-type representation data 70.

[0033]

The updating unit 2060 computes a loss for the feature value of the first-type representation data 50-1 and the feature value of the second-type representation data 60, and updates trainable parameters of the second feature extracting model 20 based on the computed loss. The updating unit 2060 computes a loss for the feature value of the first-type representation data 50-2 and the feature value of the third-type representation data 70, and updates trainable parameters of the third feature extracting model 30 based on the computed loss.

[0034]

20 <Example of Hardware Configuration>

The training 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 training apparatus 2000, or may be a general-purpose computer like a personal computer (PC), a server machine, or a mobile device.

25 [0035]

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

30 [0036]

Fig. 6 is a block diagram illustrating an example of the hardware configuration of a computer 1000 realizing the training apparatus 2000. In Fig. 6, 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.

[0037]

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 processor, 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).

[0038]

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 training apparatus 2000.

[0039]

The hardware configuration of the computer 1000 is not restricted to that shown in Fig. 6. For example, as mentioned-above, the training apparatus 2000 may be realized by plural computers. In this case, those computers may be connected with each other through the network.

[0040]

<Example of Representation data>

As described above, there may be various types of representation data that the training apparatus 2000 and the image matching apparatus 300 handle: e.g., ground-view image, aerial-view image, map image, 3D shape data, and keyword matrix data. Hereinafter, each type of representation data is explained in detail.

[0041]

The ground-view image is a digital image (e.g., RGB image or grayscale image) that includes a ground view of a place. The ground-view image is generated by a ground camera that may be held by a pedestrian or installed on a vehicle. The ground-view image may be panoramic (having 360-degree field of view), or may have limited (less than 360-degree) field of view.

[0042]

The aerial-view image is a digital image (e.g., RGB image or grayscale image) that includes an aerial (top, in other words) view of a place. For example, the aerial-view image is generated by an aerial camera that may be installed on a drone, an airplane, or a satellite.

[0043]

The map image is a digital image (e.g., RGB image or grayscale image) that visually describe map information of a place. In some implementations, the map image may include geometric shapes that represent various landmarks, such as buildings, roads, railways, rivers, lakes, mountains, etc. The map image may also show additional information, such as the name of each landmark, an address assigned to each section, etc.

[0044]

Fig. 7 illustrates an example of the ground-view image, the aerial-view image, and the map image. In Fig. 7, the ground-view image 100, the aerial-view image 110, and the map image 120 show the same place as each other in different forms.

[0045]

The 3D shape data is 3D data that represents a three-dimensional shape of a place, such as a point cloud, polygon mesh data, etc. Specifically, the 3D shape data of a place may represent 3D shapes of objects that exist in the place: e.g., 3D shapes of buildings and natural elements in the place. The natural elements may include a ground, a river, a lake, etc. Fig. 8 illustrates an example of the 3D shape data. The 3D shape data 130 shown in Fig. 8 represents 3D shapes of buildings and a 3D shape of the ground around the buildings.

[0046]

Keyword matrix data is a matrix data that shows a spatial distribution of objects in a place. The keyword matrix data may show the spatial distribution of objects from a ground-level perspective or from an aerial perspective. The keyword matrix data may be generated by performing an analysis, such as semantic segmentation, on a ground-view image or an aerial-view image. In some implementations, each cell of the keyword matrix indicates a class identifier of an object that is captured on a region of the image corresponding to that cell.

[0047]

Fig. 9 illustrates examples of keyword matrix data. The keyword matrix data 140-1 shows a spatial distribution of objects in the place captured on the ground-view image 100 while the keyword matrix data 140-2 shows a spatial distribution of objects in the place captured on the aerial-view image 110. In reference to the keyword matrix data 140-1, there are five types of objects that are assigned class identifiers: road (1), building (2), tree (3), ground (4), and sky (5).

In reference to the keyword matrix data 140-2, there are four types of objects that are assigned class identifiers: road (1), building (2), tree (3), and ground (4).

[0048]

<Example Application of Image matching apparatus 300>

5 There are various possible applications of the image matching apparatus 300. For example, the image matching apparatus 300 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 200 may be implemented by one or more arbitrary computers such as ones depicted by Fig. 6. It is noted
10 that the geo-localization system is merely an example of the application of the image matching apparatus 300, and the application of the image matching apparatus 300 is not restricted to being used in the geo-localization system.

[0049]

Fig. 10 illustrates a geo-localization system 200 that includes the image matching
15 apparatus 300. The geo-localization system 200 includes the image matching apparatus 300 and the location database 400. The location database 400 includes a plurality of aerial-view images to each of which location information is attached and a plurality of map images to each of which location information is attached. An example of the location information attached to an aerial-view image may be GPS (Global Positioning System) coordinates of the place captured on the
20 center of the corresponding aerial-view image. Similarly, an example of the location information attached to a map image may be GPS coordinates of the place that is located at the center of the corresponding map image.

[0050]

In the example shown in Fig. 10, the first-type representation data, the second-type
25 location data, and the third-type location data are the ground-view image, the aerial-view image, and the map image, respectively. The first feature extracting model 10 is configured to take a ground-view image as input and computes the feature value of the ground-view image input thereinto. The second feature extracting model 20 is configured to take an aerial-view image as input and computes the feature value of the aerial-view image input thereinto. The third feature
30 extracting model 30 is configured to take a map image as input and computes the feature value of the map image input thereinto.

[0051]

The image matching apparatus 300 receives a query that includes a ground-view image

from a client (e.g., user terminal). Then, the image matching apparatus 300 searches the location database 400 for the aerial-view image or the map image that matches the ground-view image in the received query, thereby determining the place at which the ground-view image is captured. For example, the image matching apparatus 300 first tries to find the aerial-view image that matches the ground-view image in the query. Specifically, the image matching apparatus 300 inputs the ground-view image into the first extracting model thereby obtaining the feature value of the ground-view image. Then, the image matching apparatus 300 repeatedly executes: acquiring one of the aerial-view images from the location database 400; inputting the acquired aerial-view image into the second feature extracting model to compute the feature value thereof; computing a degree of similarity between the feature value of the ground-view image and the feature value of the aerial-view image; and determines whether the ground-view image matches the aerial-view image based on the computed degree of similarity.

[0052]

When it is determined that the ground-view image matches the aerial-view image, the image matching apparatus 300 sends a response to the client. The response includes the aerial-view image that is determined to match the ground-view image and the location information corresponding to that aerial-view image. Since this location information indicates the location of the place captured on the aerial-view image that matches the ground-view image, this location information can be used as data that indicates the location of the place at which the ground-view image is captured.

[0053]

When no aerial-view image is determined to match the ground-view image, the image matching apparatus 300 tries to find the map image that matches the ground-view image in the query. Specifically, the image matching apparatus 300 repeatedly executes: acquiring one of the map images from the location database 400; inputting the acquired map image into the third feature extracting model to compute the feature value thereof; computing a degree of similarity between the feature value of the ground-view image and the feature value of the map image; and determines whether the ground-view image matches the map image based on the computed degree of similarity.

[0054]

When it is determined that ground-view image matches the map image, the image matching apparatus 300 sends a response to the client. The response includes the map image that is determined to match the ground-view image and the location information corresponding to that

map image. When no map image is determined to match the ground-view image, the image matching apparatus 300 may send the client a response to notify the failure of finding the place where the ground-view image is captured.

[0055]

5 Since the aerial-view images include more detailed information than map images, the aerial-view images enable the image matching apparatus 300 to perform cross-view matching more accurately than the map images do. However, the aerial-view images offer narrower coverage of areas compared to the map images due to some reasons, such as regulations of governments. This is a reason why the image matching apparatus 300 uses the aerial-view images
10 first, then uses the map images after no aerial-view image is determined to match the ground-view image.

[0056]

It is noted that three or more types of representation data can be prepared in the location database 400. Suppose that the location database also includes pairs of 3D shape data and location
15 information. In this case, for example, 3D shape data, aerial-view images, and map images are used by the image matching apparatus 300 in this order since they include more information in this order. However, the order of the usage of those representation data can be determined taking various factors into consideration, and is not limited to the order mentioned above.

[0057]

20 When the image matching apparatus 300 handles four types of representation data, the image matching apparatus 300 also includes a fourth feature extracting model that takes fourth-type representation data as input and compute a feature value of the fourth-type representation data. The training apparatus 2000 acquires a third training data that includes a first-type representation data and a fourth-type representation data and performs training on the fourth feature extracting
25 model in the same manner as training the second feature extracting model and the third feature extracting model.

[0058]

<Flow of Process>

Fig. 11 is a flowchart illustrating an example flow of processes performed by the training
30 apparatus 2000. The acquiring unit 2020 acquires the training data 40-1 and the training data 40-2 (S102).

[0059]

The training apparatus 2000 performs a training of the second feature extracting model

20 using the training data 40-1 (S104 to S110). Specifically, the feature extracting unit 2040 inputs the first-type representation data 50-1 into the first feature extracting model 10 to compute the feature value of the first-type representation data 50-1 (S104). The feature extracting unit 2040 inputs the second-type representation data 60 into the second feature extracting model 20 to compute the feature value of the second-type representation data 60 (S106). The updating unit 2060 computes a loss for the feature value of the first-type representation data 50-1 and the feature value of the second-type representation data 60 (S108). The updating unit 2060 updates the trainable parameters of the second feature extracting model 20 based on the loss computed for the feature value of the first-type representation data 50-1 and the feature value of the second-type representation data 60 (S110).

[0060]

The training apparatus 2000 also performs a training of the third feature extracting model 30 using the training data 40-2 (S112 to S118). Specifically, the feature extracting unit 2040 inputs the first-type representation data 50-2 into the first feature extracting model 10 to compute the feature value of the first-type representation data 50-2 (S112). The feature extracting unit 2040 inputs the third-type representation data 70 into the third feature extracting model 30 to compute the feature value of the third-type representation data 70 (S114). The updating unit 2060 computes a loss for the feature value of the first-type representation data 50-2 and the feature value of the third-type representation data 70 (S116). The updating unit 2060 updates the trainable parameters of the third feature extracting model 30 based on the loss computed for the feature value of the first-type representation data 50-2 and the feature value of the third-type representation data 70 (S118).

[0061]

It is noted that a flow of processes performed by the training apparatus 2000 is not limited to that illustrated in Fig. 6. For example, the training of the second feature extracting model 20 and the training of the third feature extracting model 30 may be performed in the opposite order from that shown in Fig. 6 or may be performed in parallel with each other.

[0062]

<Acquisition of Training Data: S102>

The acquiring unit 2020 acquires the training data 40-1 and 40-2 (S102). There may be various ways to acquire training data to be used. For example, the training data 40-1 and 40-2 are stored in advance in a storage unit to which the training apparatus 2000 has access. In this case, the acquiring unit 2020 acquires the training data 40-1 and 40-2 from the storage unit.

[0063]

For another example, the training data 40-1 and 40-2 are sent from another apparatus (e.g., an apparatus that generated the training data 40-1 and 40-2) to the training apparatus 2000. In this case, the acquiring unit 2020 acquires the training data 40-1 and 40-2 by receiving them.

5 [0064]

As described above, the training data 40-1 includes the first-type representation data 50-1 and the second-type representation data 60 that show the same place as each other. This means that the second-type representation data 60 is a positive example for the training of the second feature extracting model 20. However, not only positive examples but also negative examples
10 may be used to train the second feature extracting model 20. The negative example for the training of the second feature extracting model 20 is the second-type representation data 60 that represents a different place from the place represented by the first-type representation data 50-1.

[0065]

In the case where negative examples are also used for the training of the second feature
15 extracting model 20, the training data 40-1 may further include a label that indicates whether the second-type representation data 60 therein is a positive example or a negative example.

[0066]

In some implementations, the training data 40-1 may include both a positive example of the second-type representation data 60 and a negative example of the second-type representation
20 data 60. In this case, for example, the updating unit 2060 computes a triplet loss for the first-type representation data 50-1, the positive example of the second-type representation data 60, and the negative example of the second-type representation data 60 to update the second feature extracting model 20.

[0067]

25 The same may apply to the training data 40-2. Specifically, not only positive examples of the third-type representation data but also negative examples of the third-type representation data may be used to train the third feature extracting model 30. The negative example for the training of the third feature extracting model 30 is the third-type representation data 70 that represents a different place from the place represented by the first-type representation data 50-2.
30 In this case, the training data 40-2 may further include a label that indicates whether the third-type representation data 70 therein is a positive example or a negative example.

[0068]

In some implementations, the training data 40-2 may include both a positive example of

the third-type representation data 70 and a negative example of the third-type representation data 70. In this case, for example, the updating unit 2060 computes a triplet loss for the first-type representation data 50-2, the positive example of the third-type representation data 70, and the negative example of the third-type representation data 70 to update the third feature extracting model 30.

[0069]

<Computation of Feature Value of First-Type Representation data: S104, S112>

The feature extracting unit 2040 inputs the first-type representation data 50-1 into the first feature extracting model 10 to compute the feature value of the first-type representation data 50-1 (S104). In addition, the feature extracting unit 2040 inputs the first-type representation data 50-2 into the first feature extracting model 10 to compute the feature value of the first-type representation data 50-2 (S112). The first feature extracting model 10 may be implemented as a machine learning-based model, such as a neural network

[0070]

The first feature extracting model 10 is configured to take a first-type representation data as input, perform computations on the input data, and output a value (e.g., vector or tensor) that represents the feature value of the input data. In response to the first-type representation data 50-1 being input into the first feature extracting model 10 by the feature extracting unit 2040, the first feature extracting model 10 performs computations on the input data and thus outputs the feature value of the first-type representation data 50-1 input thereinto. Similarly, in response to the first-type representation data 50-2 being input into the first feature extracting model 10 by the feature extracting unit 2040, the first feature extracting model 10 performs computations on the input data and thus outputs the feature value of the first-type representation data 50-2 input thereinto.

[0071]

As mentioned above, the first feature extracting model 10 is pre-trained to extract the feature value from the first-type representation data 50. The first feature extracting model 10 may be pre-trained in an arbitrary manner.

[0072]

<Computation of Feature Value of Second-Type Representation data: S106>

The feature extracting unit 2040 inputs the second-type representation data 60 into the second feature extracting model 20 to compute the feature value of the second-type representation data 60 (S106). The second feature extracting model 20 may be implemented as a machine learning-based model, such as a neural network.

[0073]

The second feature extracting model 20 is configured to take a second-type representation data as input, perform computations on the input data, and output a value (e.g., vector or tensor) that is supposed to represent the feature value of the input data. In response to the second-type representation data 60 being input into the second feature extracting model 20 by the feature extracting unit 2040, the second feature extracting model 20 performs computations on the input data and thus outputs the feature value of the second-type representation data 60 input therein.

[0074]

<Computation of Feature Value of Third-Type Representation data: S114>

10 The feature extracting unit 2040 inputs the third-type representation data 70 into the third feature extracting model 30 to compute the feature value of the third-type representation data 70 (S114). The third feature extracting model 30 may be implemented as a machine learning-based model, such as a neural network.

[0075]

15 The third feature extracting model 30 is configured to take a third-type representation data as input, perform computations on the input data, and output a value (e.g., vector or tensor) that is supposed to represent the feature value of the input data. In response to the third-type representation data 70 being input into the third feature extracting model 30 by the feature extracting unit 2040, the third feature extracting model 30 performs computations on the input data and thus outputs the feature value of the third-type representation data 70 input therein.

[0076]

<Computation of Loss: S108, S116>

25 The updating unit 2060 computes a loss for the feature value of the first-type representation data 50-1 and the second-type representation data 60 (S108). This loss may be computed by applying the feature value of the first-type representation data 50-1 and the feature value of the second-type representation data 60 to a pre-defined loss function. There are various types of loss functions, and one of them can be employed as the loss function to compute the loss for the feature value of the first-type representation data 50-1 and the second-type representation data 60.

30 [0077]

As described above, the training data 40-1 may include both the positive example of the second-type representation data 60 and the negative example of the second-type representation data 60 in some implementations. In this case, the updating unit 2060 may apply the feature value

of the first-type representation data 50-1, the feature value of the positive example of the second-type representation data 60, and the feature value of the negative example of the second-type representation data 60 to a pre-defined triplet loss function to compute a triplet loss for those three feature values.

5 [0078]

The same may apply to the computation of the loss for the feature value of the first-type representation data 50-1 and the third-type representation data 70 (S116). Specifically, this loss may be computed by applying the feature value of the first-type representation data 50-2 and the feature value of the third-type representation data 70 to a pre-defined loss function. One of various
10 types of loss function may be employed as the loss function to compute the loss for the feature value of the first-type representation data 50-2 and the third-type representation data 70. When the training data 40-2 includes both the positive example of the third-type representation data and the negative example of the third-type representation data, the updating unit 2060 may apply the feature value of the first-type representation data 50-2, the feature value of the positive example
15 of the third-type representation data 70, and the feature value of the negative example of the third-type representation data 70 to a pre-defined triplet loss function to compute a triplet loss for those three feature values.

[0079]

<Update of Models: S110, S118>

20 The updating unit 2060 updates the trainable parameters of the second feature extracting model 20 to train the second feature extracting model (S110). When the second feature extracting model 20 is implemented as a neural network, the trainable parameters of the second feature extracting model 20 include biases and weights assigned to connections between nodes.

[0080]

25 The second feature extracting model 20 may be updated based on individual losses or based on set of losses. In the former case, every time the loss is computed for a single piece of the training data 40-1, the updating unit 2060 updates the trainable parameters of the second feature extracting model 20 based on that loss. In the latter case, the updating unit 2060 computes the loss for each one of two or more pieces of the training data 40-1, computes a statistical value (e.g., an
30 average) of those losses, and updates the trainable parameters of the second feature extracting model 20 based on the statistical value of those losses.

[0081]

The updating unit 2060 also updates the trainable parameters of the third feature

extracting model 30 to train the third feature extracting model (S118). The third feature extracting model 30 may be updated in the same manner as the second feature extracting model 20. Specifically, the third feature extracting model 30 may be updated based on individual losses or based on set of losses. In the former case, for example, every time the loss is computed for a single
 5 piece of the training data 40-2, the updating unit 2060 updates the trainable parameters of the third feature extracting model 30 based on that loss. In the latter case, the updating unit 2060 computes the loss for each one of two or more pieces of the training data 40-2, computes a statistical value (e.g., an average) of those losses, and updates the trainable parameters of the third feature extracting model 30 based on the statistical value of those losses.

10 [0082]

<Output from Training Apparatus>

The training apparatus 2000 may output information (hereinafter, output information) related to a result of the training of the second feature extracting model 20 and the third feature extracting model 30. The output information may include the trainable parameters of the second
 15 feature extracting model 20 and those of the third feature extracting model 30, which are determined as a result of the training of the second feature extracting model 20 and the training of the third feature extracting model 30. The output information may also include other information related to the second feature extracting model 20 and the third feature extracting model: e.g., hyperparameters of the second feature extracting model 20 and the third feature extracting model,
 20 and programs of the second feature extracting model 20 and the third feature extracting model 30.

[0083]

There are various ways to output the output information. For example, the training apparatus 2000 may put the output information into a storage device. For another example, the training apparatus 2000 may output the output information to a display device so that the display
 25 device displays the contents of the output information. For another example, the training apparatus 2000 may output the output information to another computer, such as the image matching apparatus 300.

[0084]

The program includes instructions (or software codes) that, when loaded into a computer,
 30 cause the computer to perform one or more of the functions described in the embodiments. The program may be stored in a non-transitory computer readable medium or a tangible storage medium. By way of example, and not a limitation, non-transitory computer readable media or tangible storage media can include a random-access memory (RAM), a read-only memory (ROM),

a flash memory, a solid-state drive (SSD) or other types of memory technologies, a CD-ROM, a digital versatile disc (DVD), a Blu-ray disc or other types of optical disc storage, and magnetic cassettes, magnetic tape, magnetic disk storage or other types of magnetic storage devices. The program may be transmitted on a transitory computer readable medium or a communication
5 medium. By way of example, and not a limitation, transitory computer readable media or communication media can include electrical, optical, acoustical, or other forms of propagated signals.

[0085]

Although the present disclosure is explained above with reference to example
10 embodiments, the present disclosure is not limited to the above-described example embodiments. Various modifications that can be understood by those skilled in the art can be made to the configuration and details of the present disclosure within the scope of the invention.

[0086]

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

<Supplementary notes>

(Supplementary Note 1)

A training apparatus comprising:

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

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

acquire a first training data and a second training data, the first training data including a first-type representation data and a second-type representation data that represent a same place as each other in different types of representation from each other, the second training data including a first-type representation data and a third-type representation data that represent a same place as
25 each other in different types of representation from each other;

train a second feature extracting model based on a feature value of the first-type representation data of the first training data computed by a pre-trained first feature extracting model and a feature value of the second-type representation data computed by the second feature extracting model; and

30 train a third feature extracting model based on a feature value of the first-type representation data of the second training data computed by the pre-trained first feature extracting model and a feature value of the third-type representation data computed by the third feature extracting model.

(Supplementary Note 2)

The training apparatus according to claim 1,

wherein the training of the second feature extracting model includes:

inputting the first-type representation data of the first training data into the pre-trained first feature extracting model to compute the feature value of the first-type representation data of the first training data;

inputting the second-type representation data into the second feature extracting model to compute the feature value of the second-type representation data;

computing a loss based on the feature value of the first-type representation data of the first training data and the feature value of the second-type representation data; and

updating trainable parameters of the second feature extracting model based on that loss, and

wherein the training of the third feature extracting model includes:

inputting the first-type representation data of the second training data into the pre-trained first feature extracting model to compute the feature value of the first-type representation data of the second training data;

inputting the third-type representation data into the third feature extracting model to compute the feature value of the third-type representation data;

computing a loss based on the feature value of the first-type representation data of the second training data and the feature value of the third-type representation data; and

updating trainable parameters of the third feature extracting model based on that loss.

(Supplementary Note 3)

The training apparatus according to claim 1,

wherein each of the first-type representation data, the second-type representation data, and the third-type representation data belongs to a distinct type among ground-view image, aerial-view image, map image, three-dimensional shape data, and keyword matrix data.

(Supplementary Note 4)

The training apparatus according to claim 1,

wherein the first-type representation data of the first training data and the first-type representation data of the second training data are same data as each other.

(Supplementary Note 5)

A training method, performed by a computer, comprising:

acquiring a first training data and a second training data, the first training data including a first-type representation data and a second-type representation data that represent a same place as each other in different types of representation from each other, the second training data including a first-type representation data and a third-type representation data that represent a same place as each other in different types of representation from each other;

training a second feature extracting model based on a feature value of the first-type representation data of the first training data computed by a pre-trained first feature extracting model and a feature value of the second-type representation data computed by the second feature extracting model; and

training a third feature extracting model based on a feature value of the first-type representation data of the second training data computed by the pre-trained first feature extracting model and a feature value of the third-type representation data computed by the third feature extracting model.

(Supplementary Note 6)

The training method according to claim 5,

wherein the training of the second feature extracting model includes:

inputting the first-type representation data of the first training data into the pre-trained first feature extracting model to compute the feature value of the first-type representation data of the first training data;

inputting the second-type representation data into the second feature extracting model to compute the feature value of the second-type representation data;

computing a loss based on the feature value of the first-type representation data of the first training data and the feature value of the second-type representation data; and

updating trainable parameters of the second feature extracting model based on that loss, and

wherein the training of the third feature extracting model includes:

inputting the first-type representation data of the second training data into the pre-trained first feature extracting model to compute the feature value of the first-type representation data of the second training data;

inputting the third-type representation data into the third feature extracting model to compute the feature value of the third-type representation data;

computing a loss based on the feature value of the first-type representation data of the second training data and the feature value of the third-type representation data; and

updating trainable parameters of the third feature extracting model based on that loss.

(Supplementary Note 7)

The training method according to claim 5,

5 wherein each of the first-type representation data, the second-type representation data, and the third-type representation data belongs to a distinct type among ground-view image, aerial-view image, map image, three-dimensional shape data, and keyword matrix data.

(Supplementary Note 8)

The training method according to claim 5,

10 wherein the first-type representation data of the first training data and the first-type representation data of the second training data are same data as each other.

(Supplementary Note 9)

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

15 acquiring a first training data and a second training data, the first training data including a first-type representation data and a second-type representation data that represent a same place as each other in different types of representation from each other, the second training data including a first-type representation data and a third-type representation data that represent a same place as each other in different types of representation from each other;

20 training a second feature extracting model based on a feature value of the first-type representation data of the first training data computed by a pre-trained first feature extracting model and a feature value of the second-type representation data computed by the second feature extracting model; and

25 training a third feature extracting model based on a feature value of the first-type representation data of the second training data computed by the pre-trained first feature extracting model and a feature value of the third-type representation data computed by the third feature extracting model.

(Supplementary Note 10)

The medium according to claim 9,

30 wherein the training of the second feature extracting model includes:

inputting the first-type representation data of the first training data into the pre-trained first feature extracting model to compute the feature value of the first-type representation data of the first training data;

inputting the second-type representation data into the second feature extracting model to compute the feature value of the second-type representation data;

computing a loss based on the feature value of the first-type representation data of the first training data and the feature value of the second-type representation data; and

5 updating trainable parameters of the second feature extracting model based on that loss, and

wherein the training of the third feature extracting model includes:

inputting the first-type representation data of the second training data into the pre-trained first feature extracting model to compute the feature value of the first-type
10 representation data of the second training data;

inputting the third-type representation data into the third feature extracting model to compute the feature value of the third-type representation data;

computing a loss based on the feature value of the first-type representation data of the second training data and the feature value of the third-type representation data; and

15 updating trainable parameters of the third feature extracting model based on that loss.

(Supplementary Note 11)

The medium according to claim 9,

wherein each of the first-type representation data, the second-type representation data,
20 and the third-type representation data belongs to a distinct type among ground-view image, aerial-view image, map image, three-dimensional shape data, and keyword matrix data.

(Supplementary Note 12)

The medium according to claim 9,

wherein the first-type representation data of the first training data and the first-type
25 representation data of the second training data are same data as each other.

[Reference Signs List]

[0087]

10 first feature extracting model

20 second feature extracting model

30 30 third feature extracting model

40 training data

50 first-type representation data

60 second-type representation data

- 70 third-type representation data
- 100 ground-view image
- 110 aerial-view image
- 120 map image
- 5 130 3D shape data
- 140 keyword matrix data
- 200 geo-localization system
- 300 image matching apparatus
- 400 location database
- 10 1000 computer
- 1020 bus
- 1040 processor
- 1060 memory
- 1080 storage device
- 15 1100 input/output interface
- 1120 network interface
- 2000 image matching apparatus
- 2020 acquiring unit
- 2040 feature extracting unit
- 20 2060 updating unit

[CLAIMS]

[Claim 1]

A training apparatus comprising:

- 5 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 training data and a second training data, the first training data including a first-type representation data and a second-type representation data that represent a same place as each other in different types of representation from each other, the second training data including
10 a first-type representation data and a third-type representation data that represent a same place as each other in different types of representation from each other;

 train a second feature extracting model based on a feature value of the first-type representation data of the first training data computed by a pre-trained first feature extracting model and a feature value of the second-type representation data computed by the second feature
15 extracting model; and

 train a third feature extracting model based on a feature value of the first-type representation data of the second training data computed by the pre-trained first feature extracting model and a feature value of the third-type representation data computed by the third feature
 extracting model.

20

[Claim 2]

The training apparatus according to claim 1,

wherein the training of the second feature extracting model includes:

 inputting the first-type representation data of the first training data into the pre-
25 trained first feature extracting model to compute the feature value of the first-type representation data of the first training data;

 inputting the second-type representation data into the second feature extracting model to compute the feature value of the second-type representation data;

 computing a loss based on the feature value of the first-type representation data
30 of the first training data and the feature value of the second-type representation data; and

 updating trainable parameters of the second feature extracting model based on that loss, and

 wherein the training of the third feature extracting model includes:

inputting the first-type representation data of the second training data into the pre-trained first feature extracting model to compute the feature value of the first-type representation data of the second training data;

inputting the third-type representation data into the third feature extracting model
5 to compute the feature value of the third-type representation data;

computing a loss based on the feature value of the first-type representation data of the second training data and the feature value of the third-type representation data; and

updating trainable parameters of the third feature extracting model based on that loss.

10

[Claim 3]

The training apparatus according to claim 1,

wherein each of the first-type representation data, the second-type representation data, and the third-type representation data belongs to a distinct type among ground-view image, aerial-
15 view image, map image, three-dimensional shape data, and keyword matrix data.

[Claim 4]

The training apparatus according to claim 1,

wherein the first-type representation data of the first training data and the first-type
20 representation data of the second training data are same data as each other.

[Claim 5]

A training method, performed by a computer, comprising:

acquiring a first training data and a second training data, the first training data including
25 a first-type representation data and a second-type representation data that represent a same place as each other in different types of representation from each other, the second training data including a first-type representation data and a third-type representation data that represent a same place as each other in different types of representation from each other;

training a second feature extracting model based on a feature value of the first-type
30 representation data of the first training data computed by a pre-trained first feature extracting model and a feature value of the second-type representation data computed by the second feature extracting model; and

training a third feature extracting model based on a feature value of the first-type

representation data of the second training data computed by the pre-trained first feature extracting model and a feature value of the third-type representation data computed by the third feature extracting model.

5 [Claim 6]

The training method according to claim 5,

wherein the training of the second feature extracting model includes:

inputting the first-type representation data of the first training data into the pre-trained first feature extracting model to compute the feature value of the first-type representation
10 data of the first training data;

inputting the second-type representation data into the second feature extracting model to compute the feature value of the second-type representation data;

computing a loss based on the feature value of the first-type representation data of the first training data and the feature value of the second-type representation data; and

15 updating trainable parameters of the second feature extracting model based on that loss, and

wherein the training of the third feature extracting model includes:

inputting the first-type representation data of the second training data into the pre-trained first feature extracting model to compute the feature value of the first-type
20 representation data of the second training data;

inputting the third-type representation data into the third feature extracting model to compute the feature value of the third-type representation data;

computing a loss based on the feature value of the first-type representation data of the second training data and the feature value of the third-type representation data; and

25 updating trainable parameters of the third feature extracting model based on that loss.

[Claim 7]

The training method according to claim 5,

30 wherein each of the first-type representation data, the second-type representation data, and the third-type representation data belongs to a distinct type among ground-view image, aerial-view image, map image, three-dimensional shape data, and keyword matrix data.

[Claim 8]

The training method according to claim 5,

wherein the first-type representation data of the first training data and the first-type representation data of the second training data are same data as each other.

5

[Claim 9]

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

acquiring a first training data and a second training data, the first training data including
10 a first-type representation data and a second-type representation data that represent a same place as each other in different types of representation from each other, the second training data including a first-type representation data and a third-type representation data that represent a same place as each other in different types of representation from each other;

training a second feature extracting model based on a feature value of the first-type
15 representation data of the first training data computed by a pre-trained first feature extracting model and a feature value of the second-type representation data computed by the second feature extracting model; and

training a third feature extracting model based on a feature value of the first-type
20 representation data of the second training data computed by the pre-trained first feature extracting model and a feature value of the third-type representation data computed by the third feature extracting model.

[Claim 10]

The medium according to claim 9,

25 wherein the training of the second feature extracting model includes:

inputting the first-type representation data of the first training data into the pre-trained first feature extracting model to compute the feature value of the first-type representation data of the first training data;

inputting the second-type representation data into the second feature extracting
30 model to compute the feature value of the second-type representation data;

computing a loss based on the feature value of the first-type representation data of the first training data and the feature value of the second-type representation data; and

updating trainable parameters of the second feature extracting model based on

that loss, and

wherein the training of the third feature extracting model includes:

inputting the first-type representation data of the second training data into the pre-trained first feature extracting model to compute the feature value of the first-type

5 representation data of the second training data;

inputting the third-type representation data into the third feature extracting model to compute the feature value of the third-type representation data;

computing a loss based on the feature value of the first-type representation data of the second training data and the feature value of the third-type representation data; and

10 updating trainable parameters of the third feature extracting model based on that loss.

[Claim 11]

The medium according to claim 9,

15 wherein each of the first-type representation data, the second-type representation data, and the third-type representation data belongs to a distinct type among ground-view image, aerial-view image, map image, three-dimensional shape data, and keyword matrix data.

[Claim 12]

20 The medium according to claim 9,

wherein the first-type representation data of the first training data and the first-type representation data of the second training data are same data as each other.