US 20250259426A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2025/0259426 A1**

IWAMOTO (43) **Pub. Date: Aug. 14, 2025**

(54) **IMAGE PROCESSING APPARATUS AND CONTROL METHOD THEREOF, IMAGE CAPTURING APPARATUS, AND STORAGE MEDIUM**

(71) Applicant: **CANON KABUSHIKI KAISHA**, Tokyo (JP)

(72) Inventor: **Yasuhiko IWAMOTO**, Tokyo (JP)

(21) Appl. No.: **19/195,824**

(22) Filed: **May 1, 2025**

Related U.S. Application Data

(63) Continuation of application No. PCT/JP2023/035350, filed on Sep. 28, 2023.

(30) **Foreign Application Priority Data**

Nov. 25, 2022 (JP) 2022-188698

Publication Classification

(51) **Int. Cl.**

G06V 10/774	(2022.01)
G06V 10/40	(2022.01)
G06V 10/764	(2022.01)
G06V 10/77	(2022.01)

(52) **U.S. Cl.**

CPC ***G06V 10/774*** (2022.01); ***G06V 10/40*** (2022.01); ***G06V 10/764*** (2022.01); ***G06V 10/7715*** (2022.01)

(57) **ABSTRACT**

This image processing device: acquires a dictionary obtained by machine learning, and a plurality of learning images used for the machine learning of the dictionary; selects, from among the acquired plurality of learning images, one or more learning images to be used as information indicating characteristics of the dictionary; and generates linking information that link the selected one or more learning images with the dictionary.

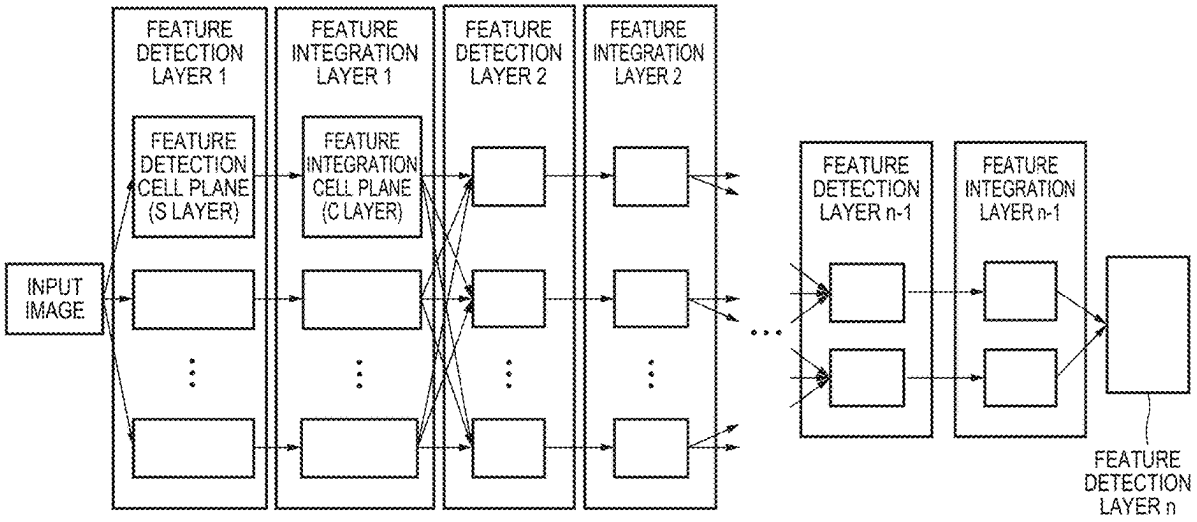


FIG. 1A

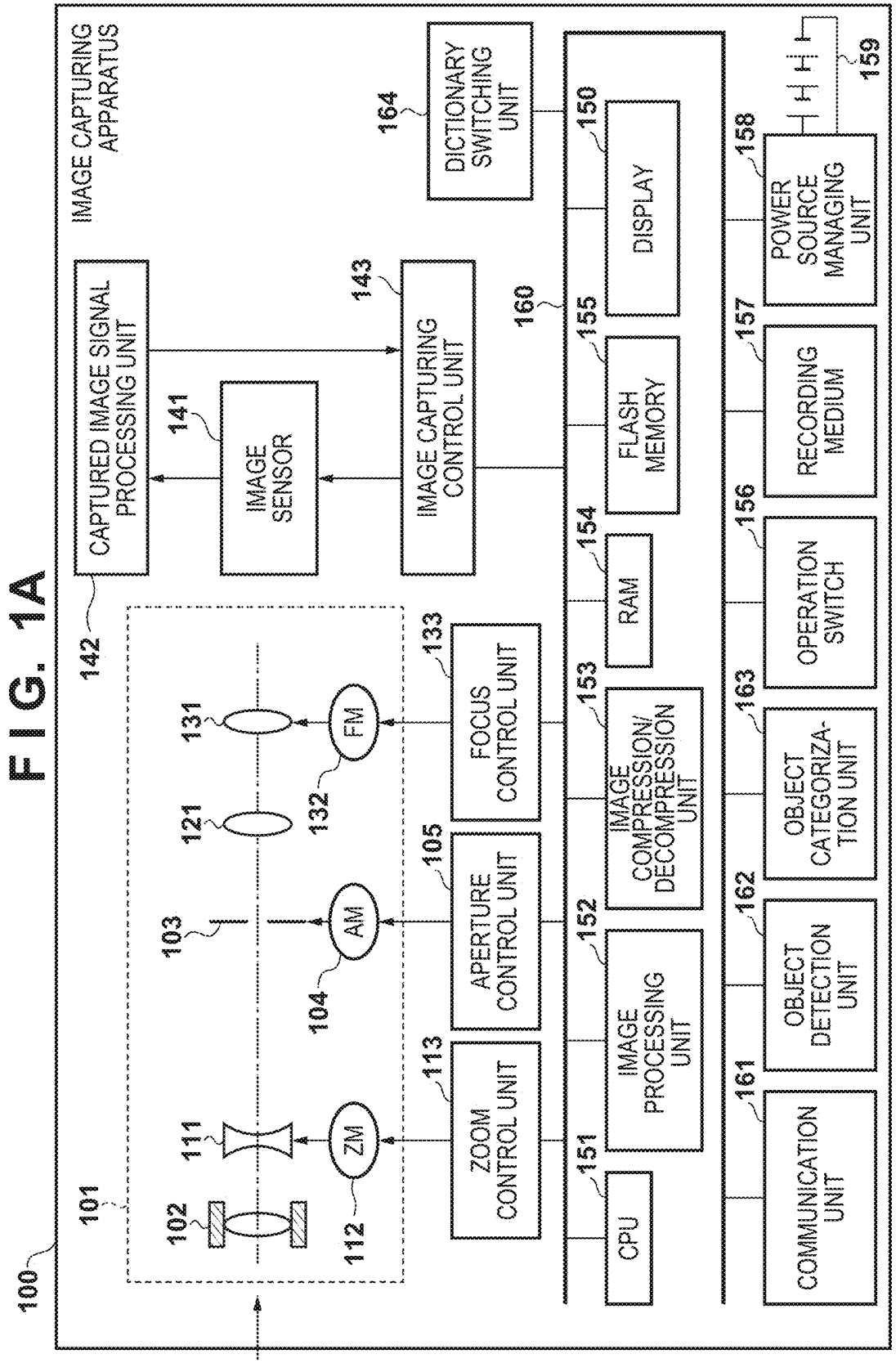

100
IMAGE CAPTURING APPARATUS
101
102
111
112
ZM
103
104
AM
121
131
132
FM
142
CAPTURED IMAGE SIGNAL PROCESSING UNIT
141
IMAGE SENSOR
143
IMAGE CAPTURING CONTROL UNIT
160
164
DICTIONARY SWITCHING UNIT
113
ZOOM CONTROL UNIT
105
APERTURE CONTROL UNIT
133
FOCUS CONTROL UNIT
151
CPU
152
IMAGE PROCESSING UNIT
153
IMAGE COMPRESSION/ DECOMPRESSION UNIT
154
RAM
155
FLASH MEMORY
150
DISPLAY
161
COMMUNICATION UNIT
162
OBJECT DETECTION UNIT
163
OBJECT CATEGORIZATION UNIT
156
OPERATION SWITCH
157
RECORDING MEDIUM
158
POWER SOURCE MANAGING UNIT
159

FIG. 1B

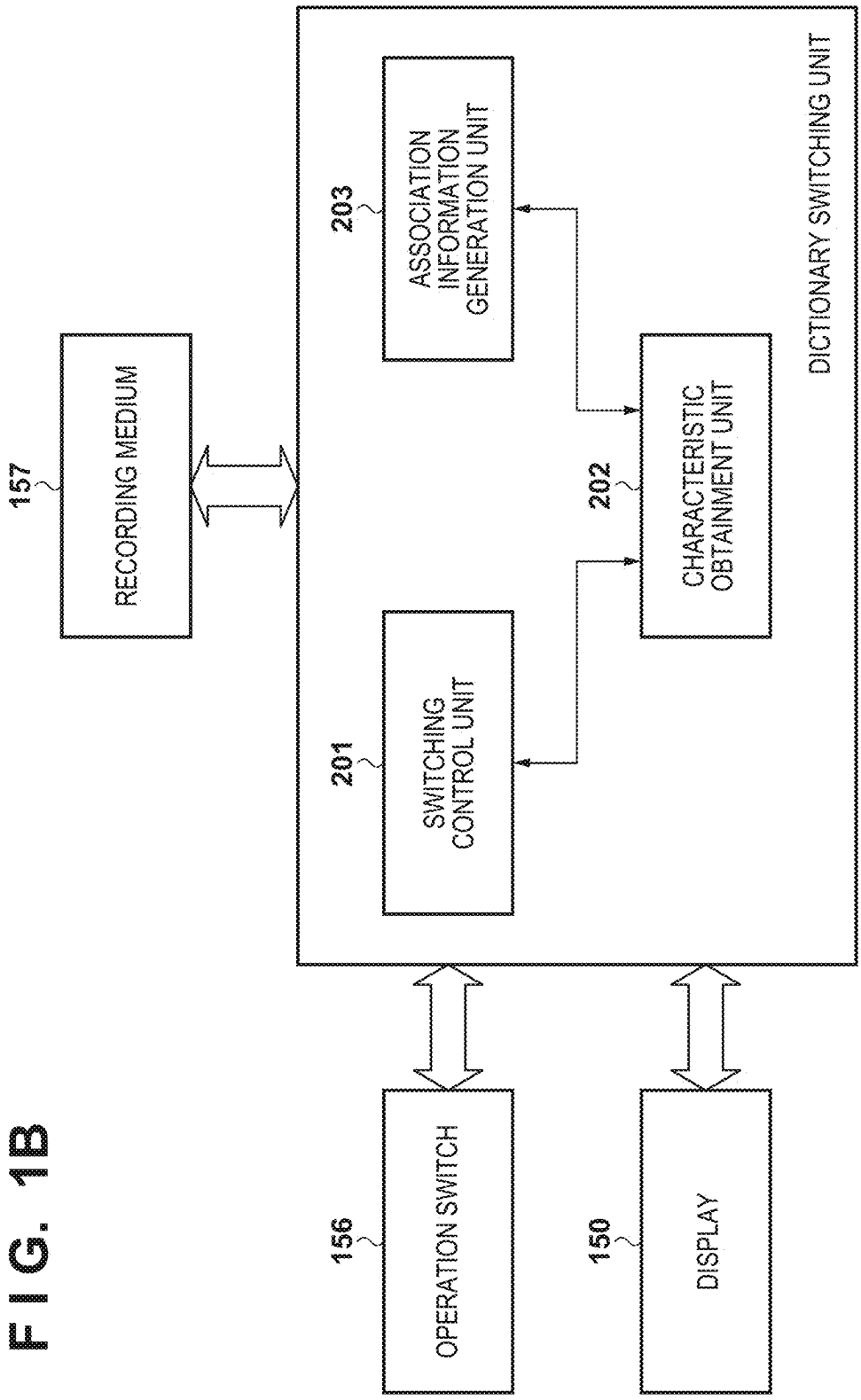

DICTIONARY SWITCHING UNIT
ASSOCIATION INFORMATION GENERATION UNIT
203
CHARACTERISTIC OBTAINMENT UNIT
202
SWITCHING CONTROL UNIT
201
RECORDING MEDIUM
157
OPERATION SWITCH
156
DISPLAY
150

FIG. 2

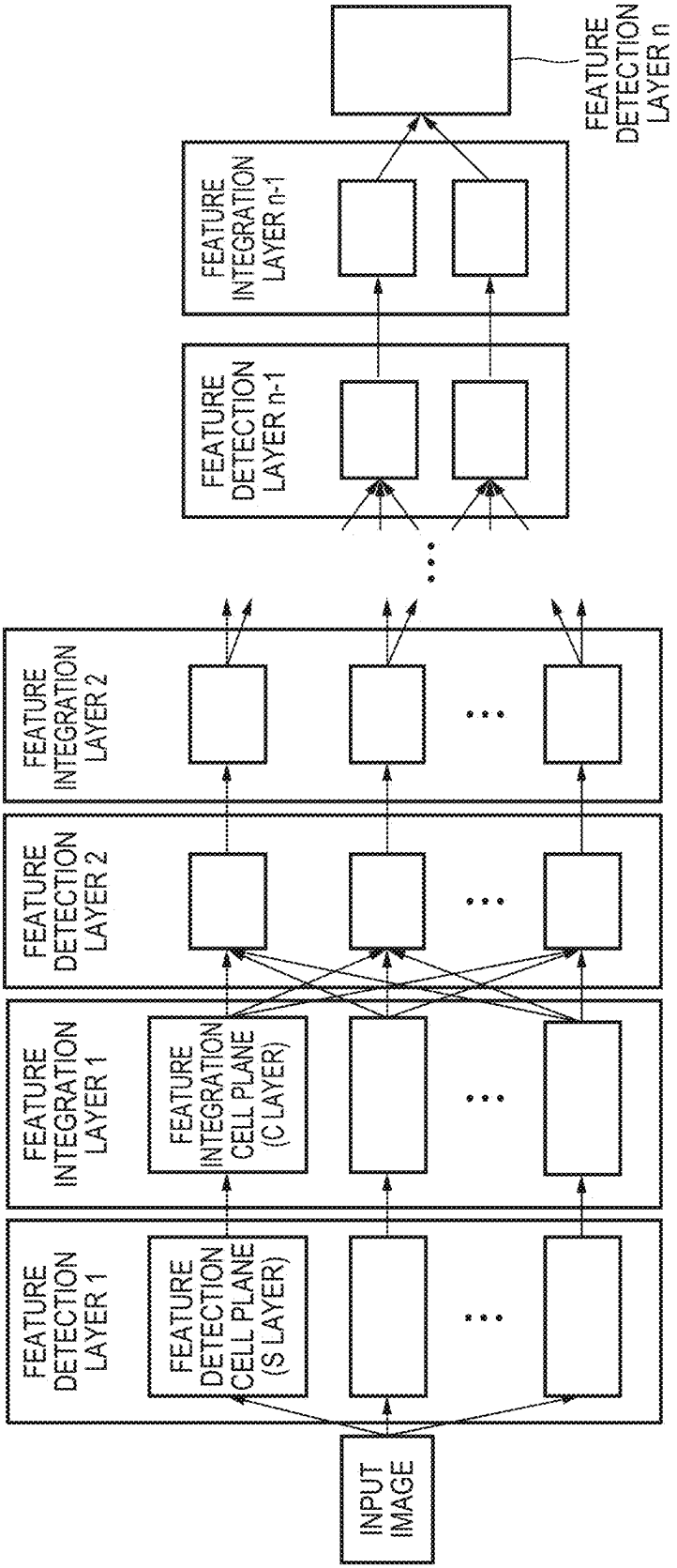

INPUT IMAGE
FEATURE DETECTION LAYER 1
FEATURE DETECTION CELL PLANE (S LAYER)
FEATURE INTEGRATION LAYER 1
FEATURE INTEGRATION CELL PLANE (C LAYER)
FEATURE DETECTION LAYER 2
FEATURE INTEGRATION LAYER 2
FEATURE DETECTION LAYER n-1
FEATURE INTEGRATION LAYER n-1
FEATURE DETECTION LAYER n

FIG. 3

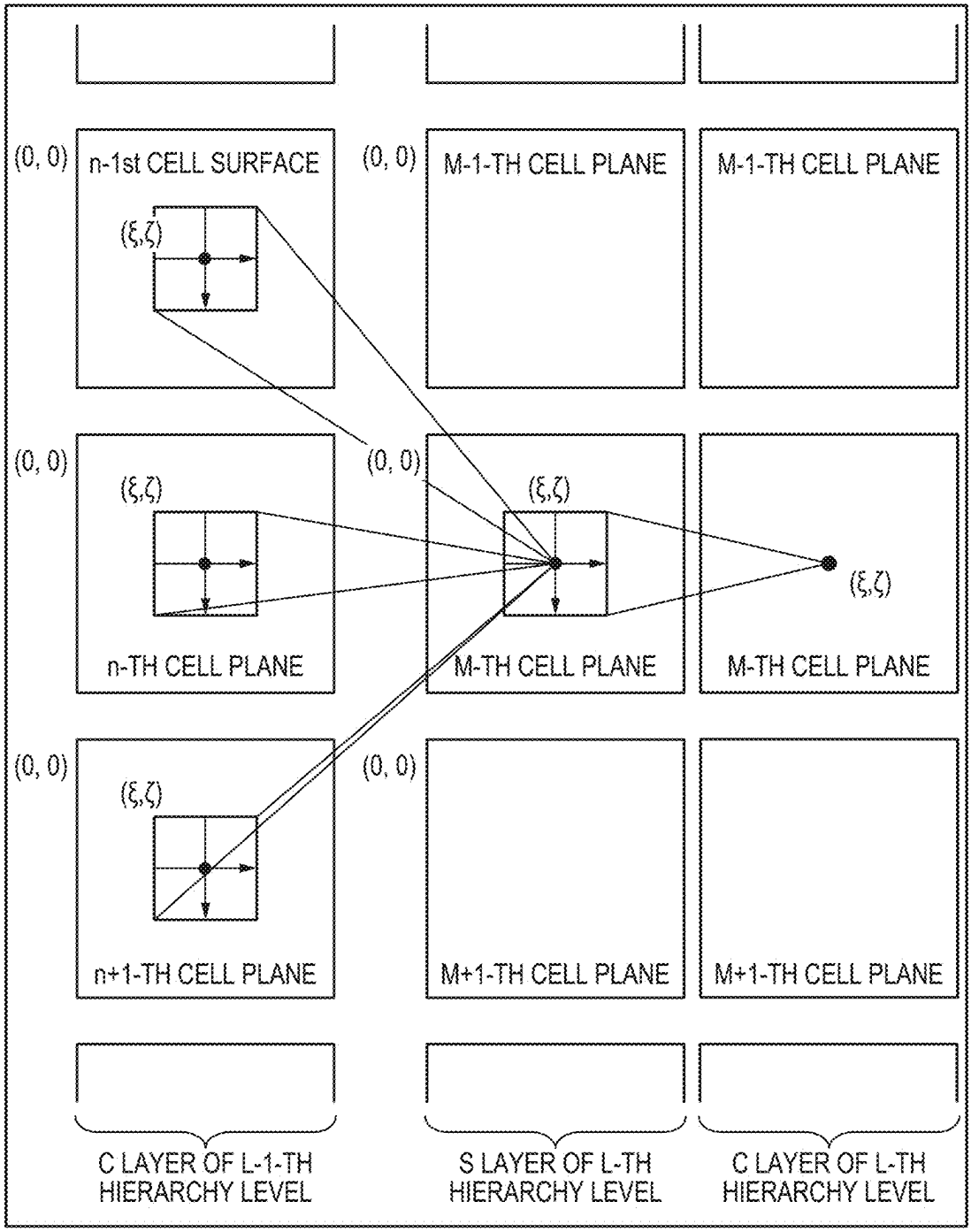

(0, 0)
n-1st CELL SURFACE
(ξ,ζ)
(0, 0)
M-1-TH CELL PLANE
M-1-TH CELL PLANE
(0, 0)
(ξ,ζ)
n-TH CELL PLANE
(0, 0)
(ξ,ζ)
M-TH CELL PLANE
(ξ,ζ)
M-TH CELL PLANE
(0, 0)
(ξ,ζ)
n+1-TH CELL PLANE
(0, 0)
M+1-TH CELL PLANE
M+1-TH CELL PLANE
C LAYER OF L-1-TH HIERARCHY LEVEL
S LAYER OF L-TH HIERARCHY LEVEL
C LAYER OF L-TH HIERARCHY LEVEL

FIG. 4

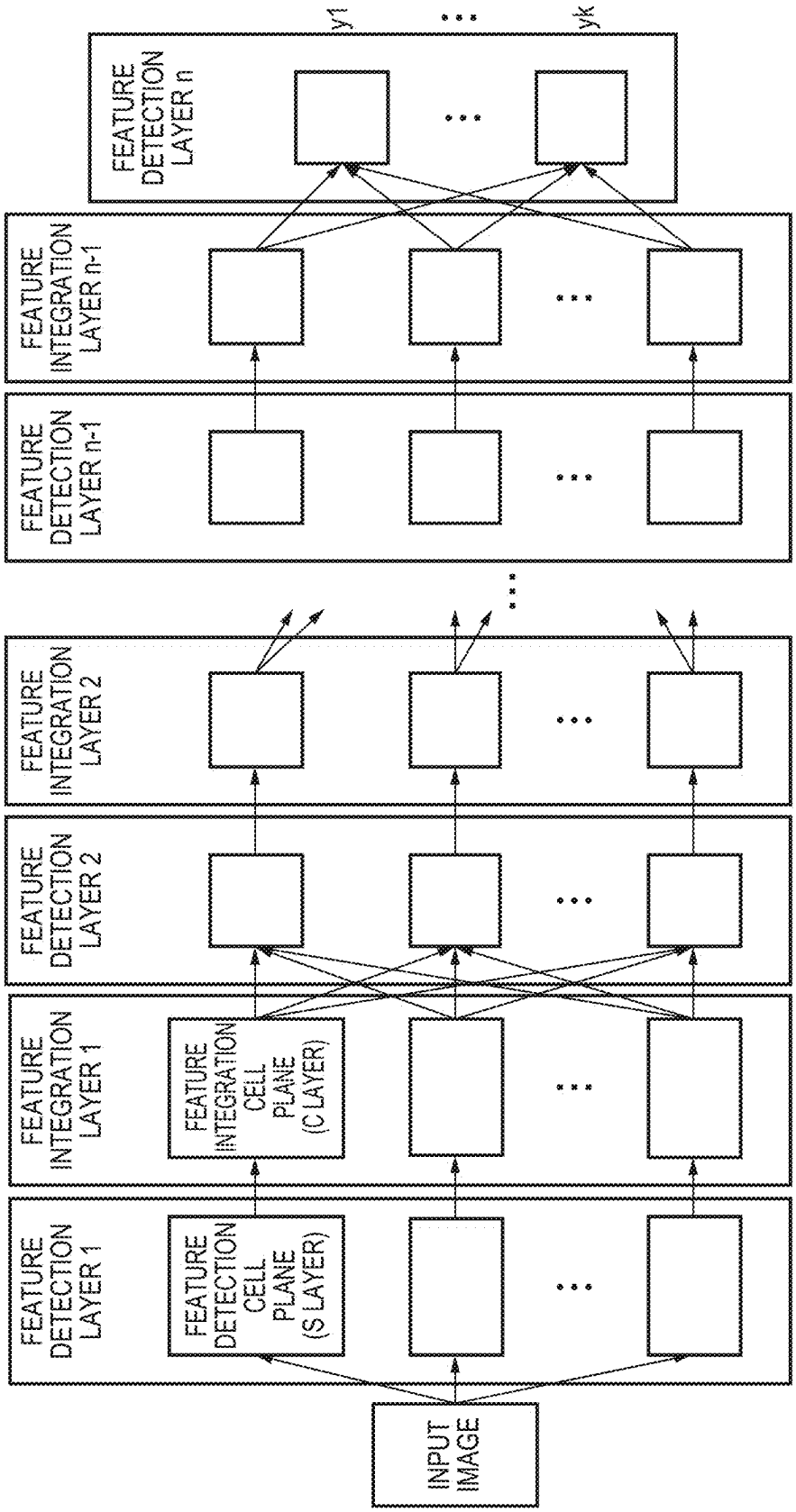

INPUT IMAGE
FEATURE DETECTION LAYER 1
FEATURE DETECTION CELL PLANE (S LAYER)
FEATURE INTEGRATION LAYER 1
FEATURE INTEGRATION CELL PLANE (C LAYER)
FEATURE DETECTION LAYER 2
FEATURE INTEGRATION LAYER 2
FEATURE DETECTION LAYER n-1
FEATURE INTEGRATION LAYER n-1
FEATURE DETECTION LAYER n
y1
yk

FIG. 5

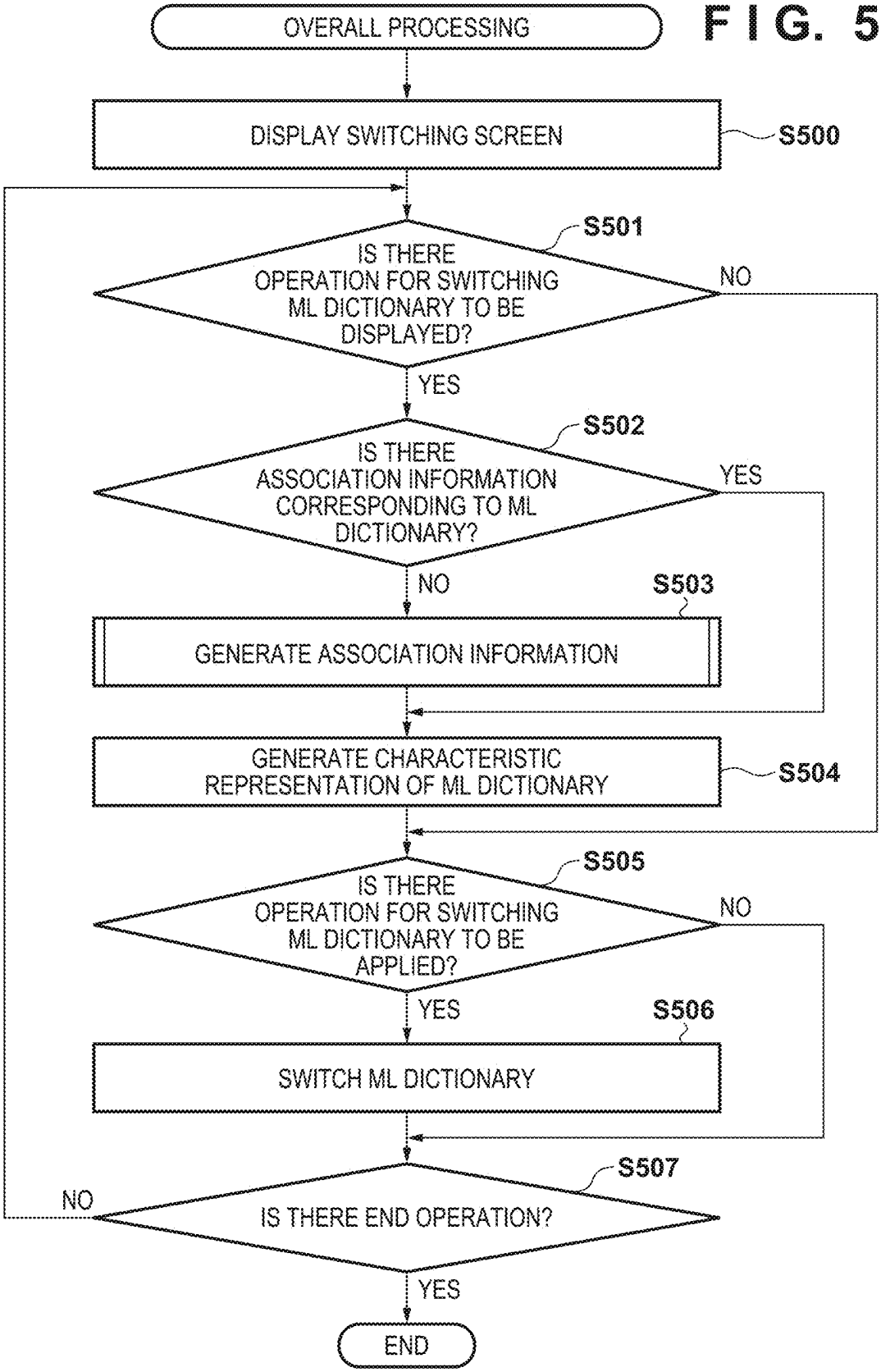
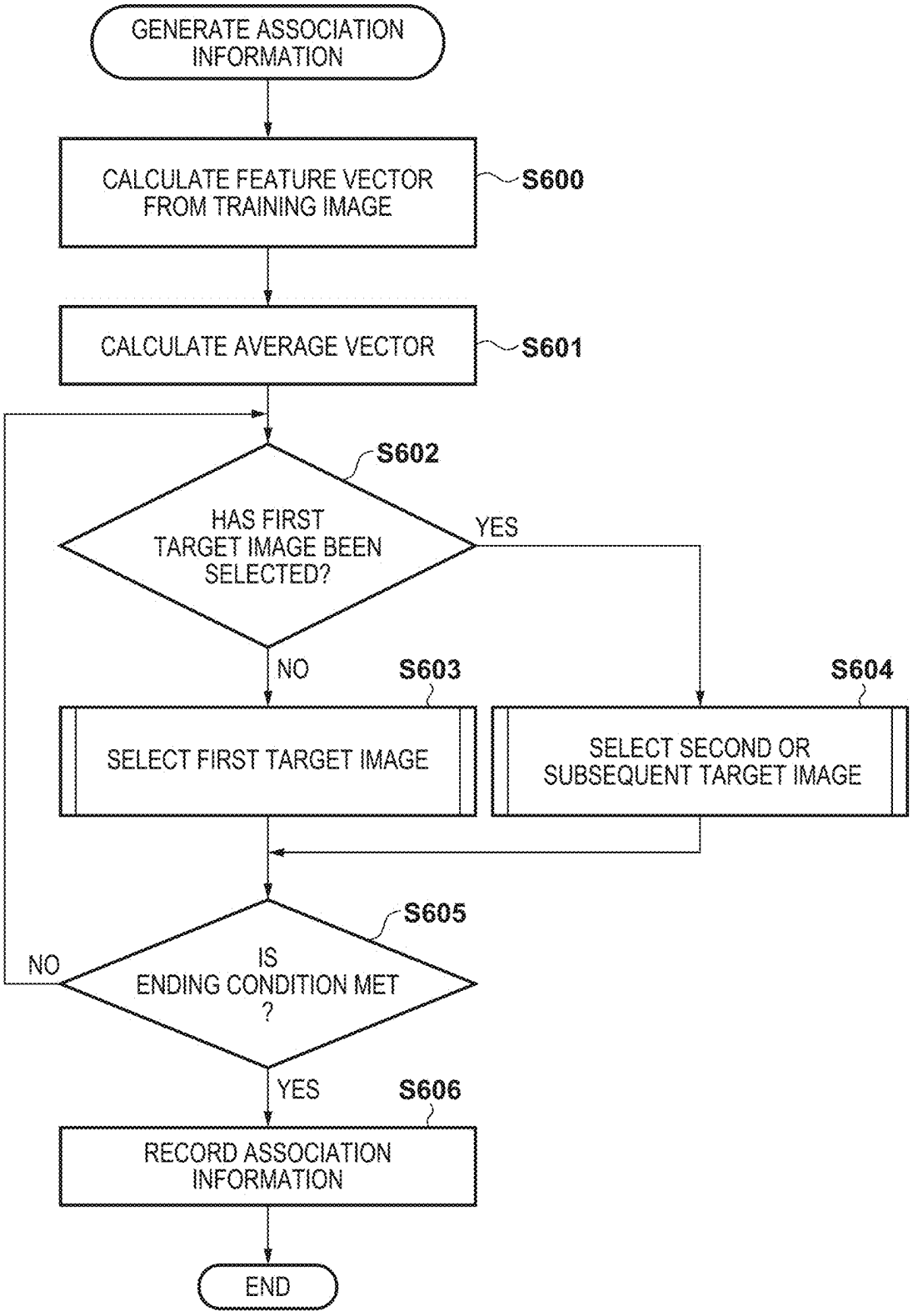

OVERALL PROCESSING
DISPLAY SWITCHING SCREEN
S500
S501
IS THERE OPERATION FOR SWITCHING ML DICTIONARY TO BE DISPLAYED?
NO
YES
S502
IS THERE ASSOCIATION INFORMATION CORRESPONDING TO ML DICTIONARY?
YES
NO
S503
GENERATE ASSOCIATION INFORMATION
GENERATE CHARACTERISTIC REPRESENTATION OF ML DICTIONARY
S504
S505
IS THERE OPERATION FOR SWITCHING ML DICTIONARY TO BE APPLIED?
NO
YES
S506
SWITCH ML DICTIONARY
S507
NO
IS THERE END OPERATION?
YES
END

FIG. 6


GENERATE ASSOCIATION INFORMATION
CALCULATE FEATURE VECTOR FROM TRAINING IMAGE
S600
CALCULATE AVERAGE VECTOR
S601
S602
HAS FIRST TARGET IMAGE BEEN SELECTED?
YES
NO
S603
SELECT FIRST TARGET IMAGE
S604
SELECT SECOND OR SUBSEQUENT TARGET IMAGE
S605
IS ENDING CONDITION MET ?
NO
YES
S606
RECORD ASSOCIATION INFORMATION
END

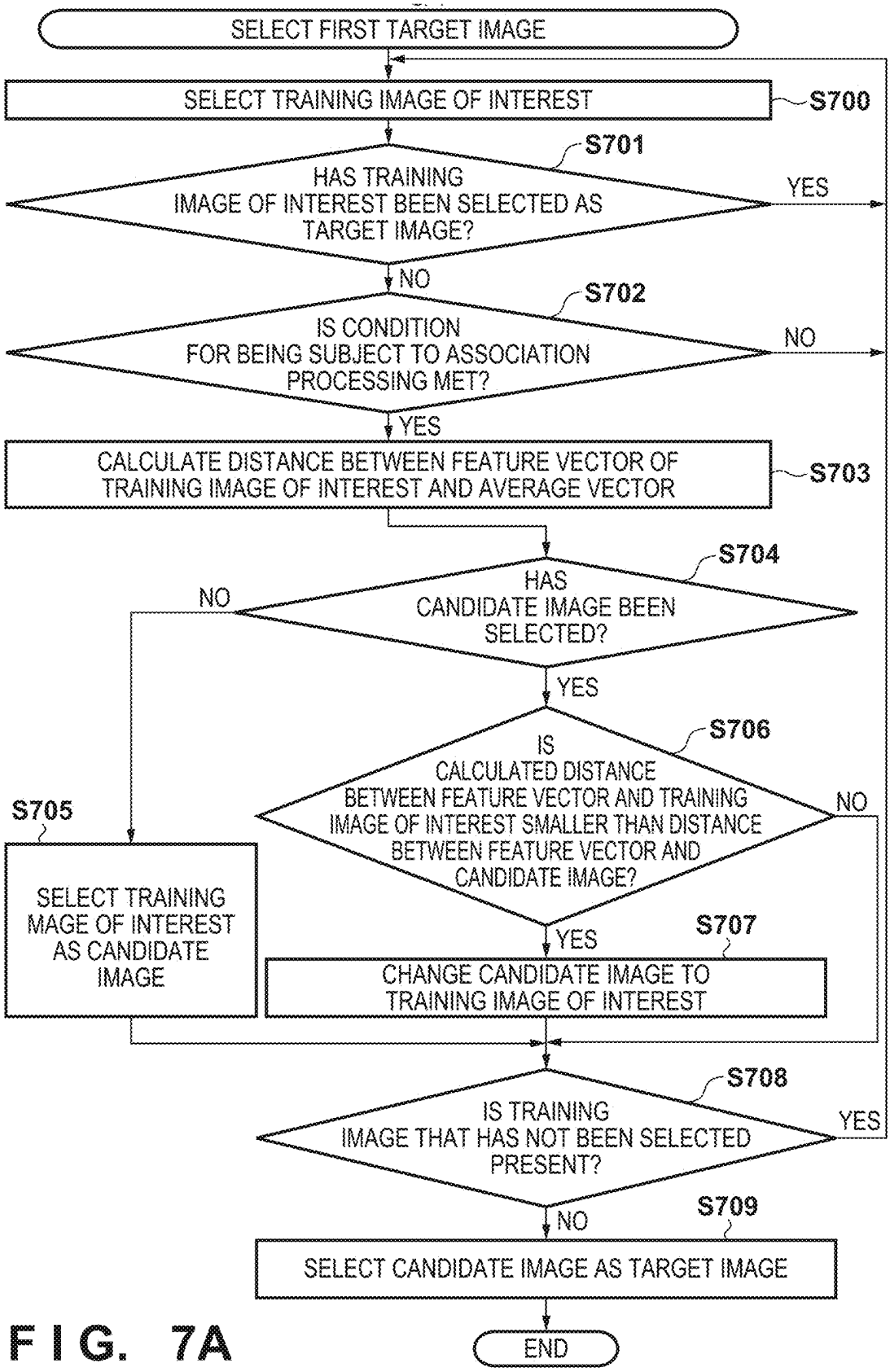

SELECT FIRST TARGET IMAGE
SELECT TRAINING IMAGE OF INTEREST
S700
S701
HAS TRAINING IMAGE OF INTEREST BEEN SELECTED AS TARGET IMAGE?
YES
NO
S702
IS CONDITION FOR BEING SUBJECT TO ASSOCIATION PROCESSING MET?
NO
YES
CALCULATE DISTANCE BETWEEN FEATURE VECTOR OF TRAINING IMAGE OF INTEREST AND AVERAGE VECTOR
S703
S704
HAS CANDIDATE IMAGE BEEN SELECTED?
NO
YES
S706
IS CALCULATED DISTANCE BETWEEN FEATURE VECTOR AND TRAINING IMAGE OF INTEREST SMALLER THAN DISTANCE BETWEEN FEATURE VECTOR AND CANDIDATE IMAGE?
NO
S705
SELECT TRAINING MAGE OF INTEREST AS CANDIDATE IMAGE
YES
S707
CHANGE CANDIDATE IMAGE TO TRAINING IMAGE OF INTEREST
S708
IS TRAINING IMAGE THAT HAS NOT BEEN SELECTED PRESENT?
YES
NO
S709
SELECT CANDIDATE IMAGE AS TARGET IMAGE
END

FIG. 7A

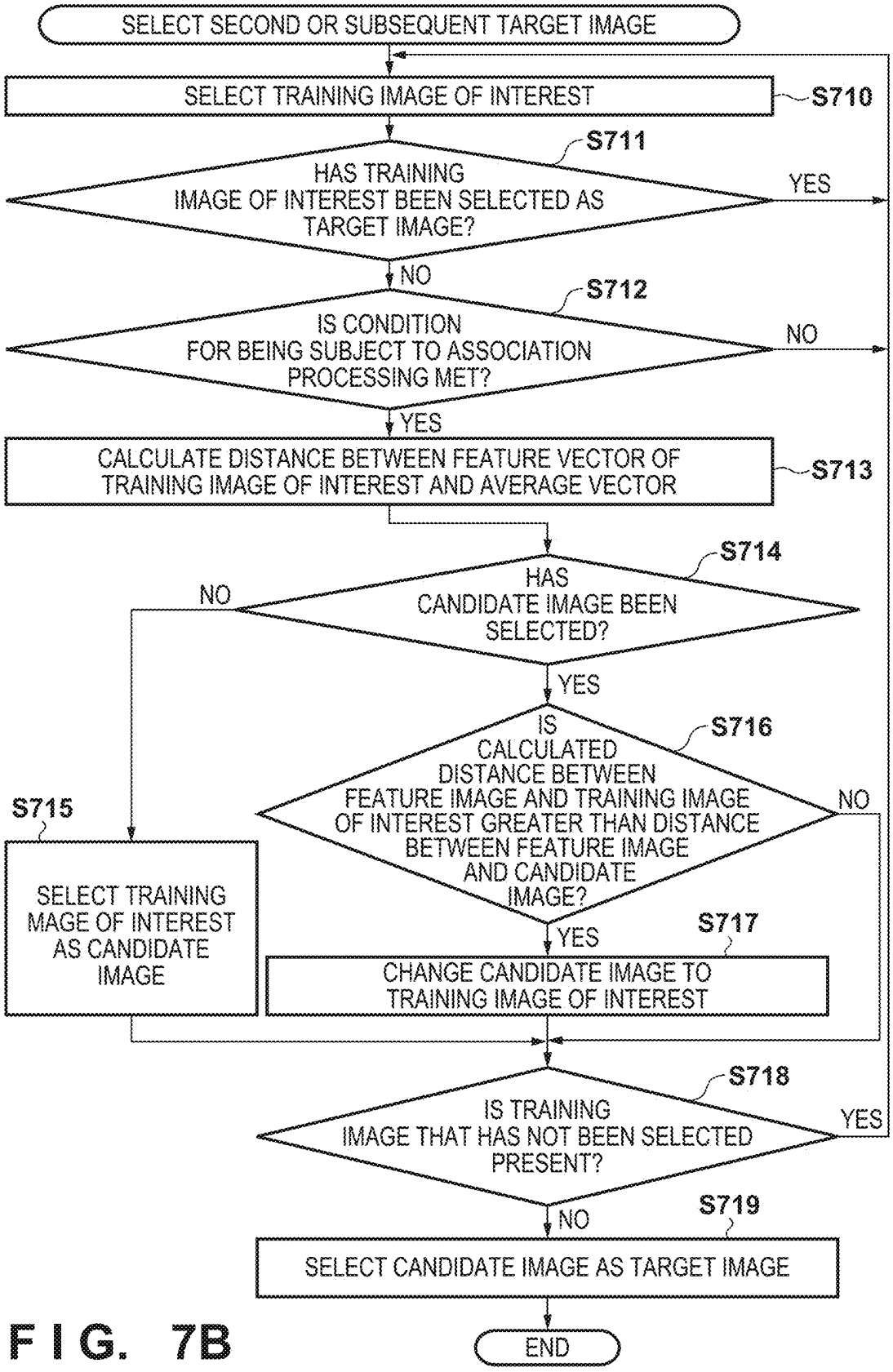

SELECT SECOND OR SUBSEQUENT TARGET IMAGE
SELECT TRAINING IMAGE OF INTEREST
S710
S711
HAS TRAINING IMAGE OF INTEREST BEEN SELECTED AS TARGET IMAGE?
YES
NO
S712
IS CONDITION FOR BEING SUBJECT TO ASSOCIATION PROCESSING MET?
NO
YES
CALCULATE DISTANCE BETWEEN FEATURE VECTOR OF TRAINING IMAGE OF INTEREST AND AVERAGE VECTOR
S713
S714
HAS CANDIDATE IMAGE BEEN SELECTED?
NO
YES
S716
IS CALCULATED DISTANCE BETWEEN FEATURE IMAGE AND TRAINING IMAGE OF INTEREST GREATER THAN DISTANCE BETWEEN FEATURE IMAGE AND CANDIDATE IMAGE?
NO
YES
S715
SELECT TRAINING MAGE OF INTEREST AS CANDIDATE IMAGE
S717
CHANGE CANDIDATE IMAGE TO TRAINING IMAGE OF INTEREST
S718
IS TRAINING IMAGE THAT HAS NOT BEEN SELECTED PRESENT?
YES
NO
S719
SELECT CANDIDATE IMAGE AS TARGET IMAGE
END

FIG. 7B

FIG. 8A

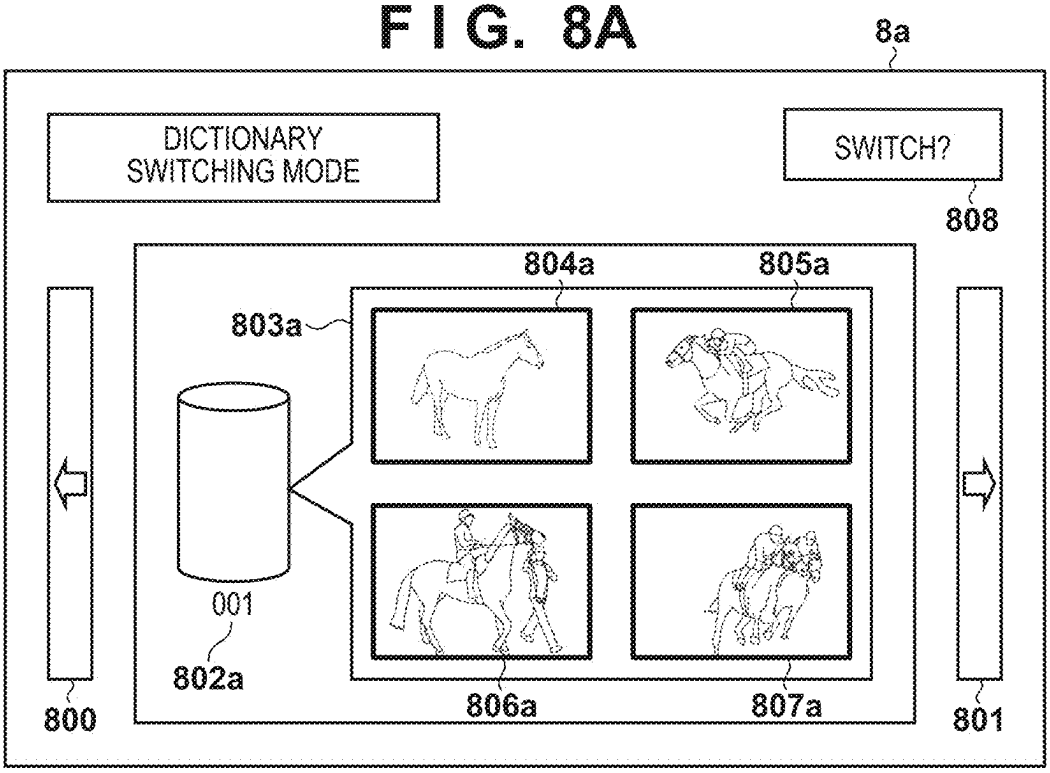
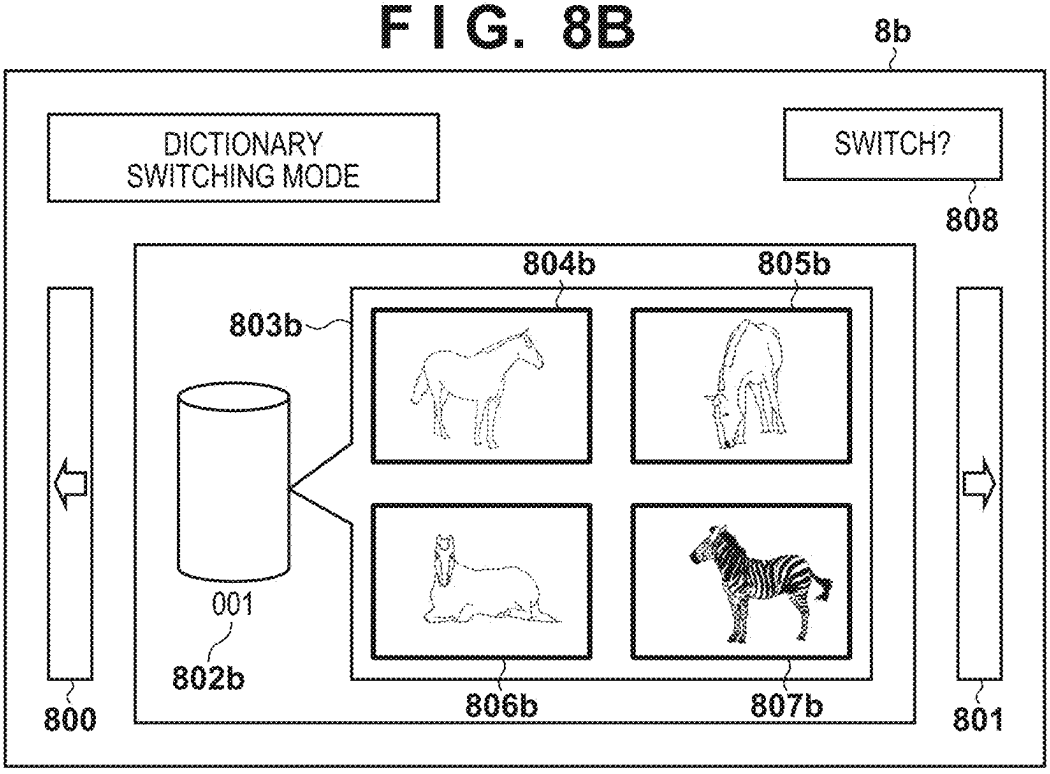

8a
DICTIONARY SWITCHING MODE
SWITCH?
808
804a
805a
803a
001
802a
806a
807a
800
801

FIG. 8B


8b
DICTIONARY SWITCHING MODE
SWITCH?
808
804b
805b
803b
001
802b
806b
807b
800
801

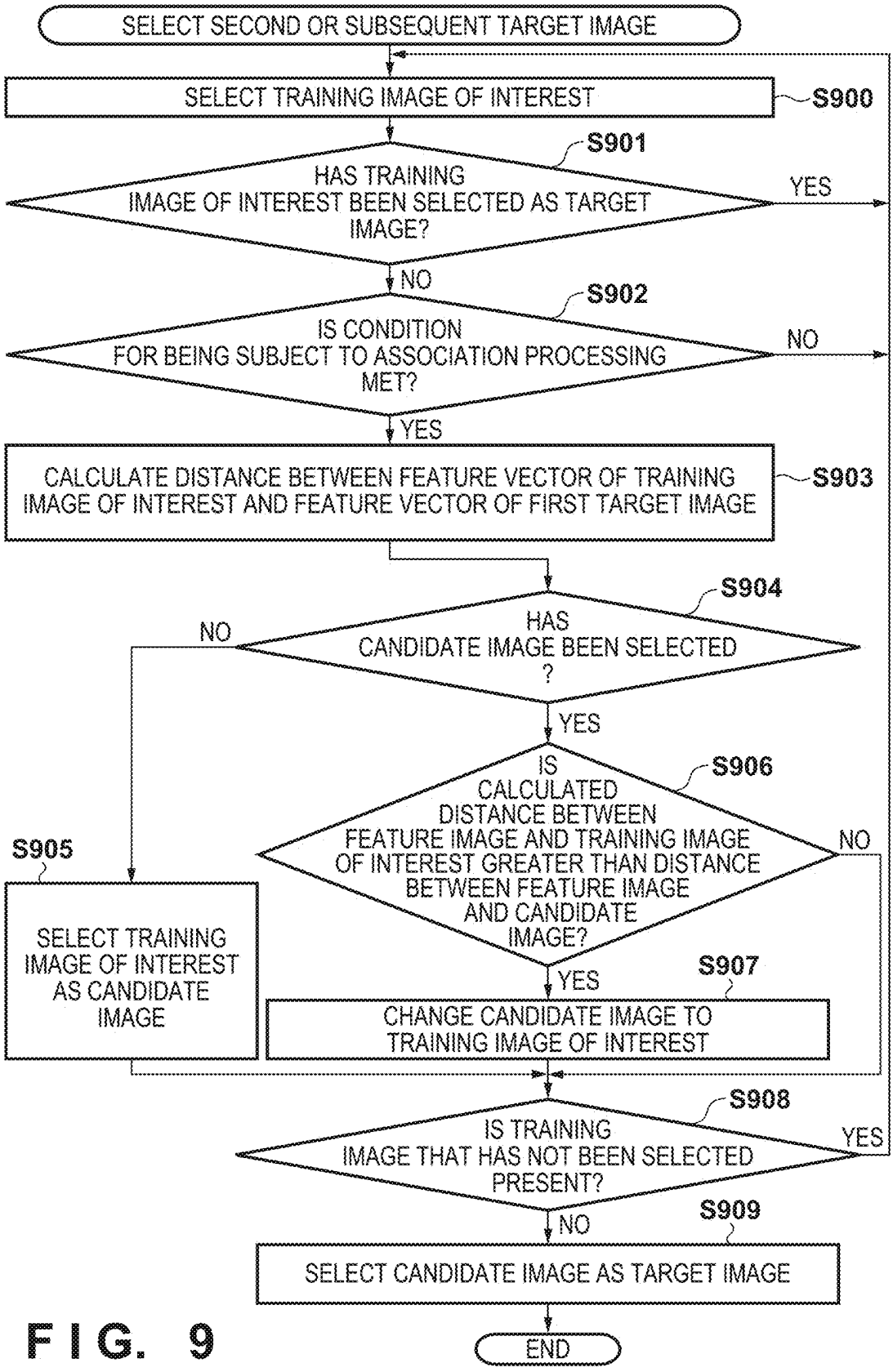

SELECT SECOND OR SUBSEQUENT TARGET IMAGE
SELECT TRAINING IMAGE OF INTEREST
S900
S901
HAS TRAINING IMAGE OF INTEREST BEEN SELECTED AS TARGET IMAGE?
YES
NO
S902
IS CONDITION FOR BEING SUBJECT TO ASSOCIATION PROCESSING MET?
NO
YES
CALCULATE DISTANCE BETWEEN FEATURE VECTOR OF TRAINING IMAGE OF INTEREST AND FEATURE VECTOR OF FIRST TARGET IMAGE
S903
S904
HAS CANDIDATE IMAGE BEEN SELECTED ?
NO
YES
S906
IS CALCULATED DISTANCE BETWEEN FEATURE IMAGE AND TRAINING IMAGE OF INTEREST GREATER THAN DISTANCE BETWEEN FEATURE IMAGE AND CANDIDATE IMAGE?
NO
S905
SELECT TRAINING IMAGE OF INTEREST AS CANDIDATE IMAGE
YES
S907
CHANGE CANDIDATE IMAGE TO TRAINING IMAGE OF INTEREST
S908
IS TRAINING IMAGE THAT HAS NOT BEEN SELECTED PRESENT?
YES
NO
S909
SELECT CANDIDATE IMAGE AS TARGET IMAGE
END

FIG. 9

FIG. 10

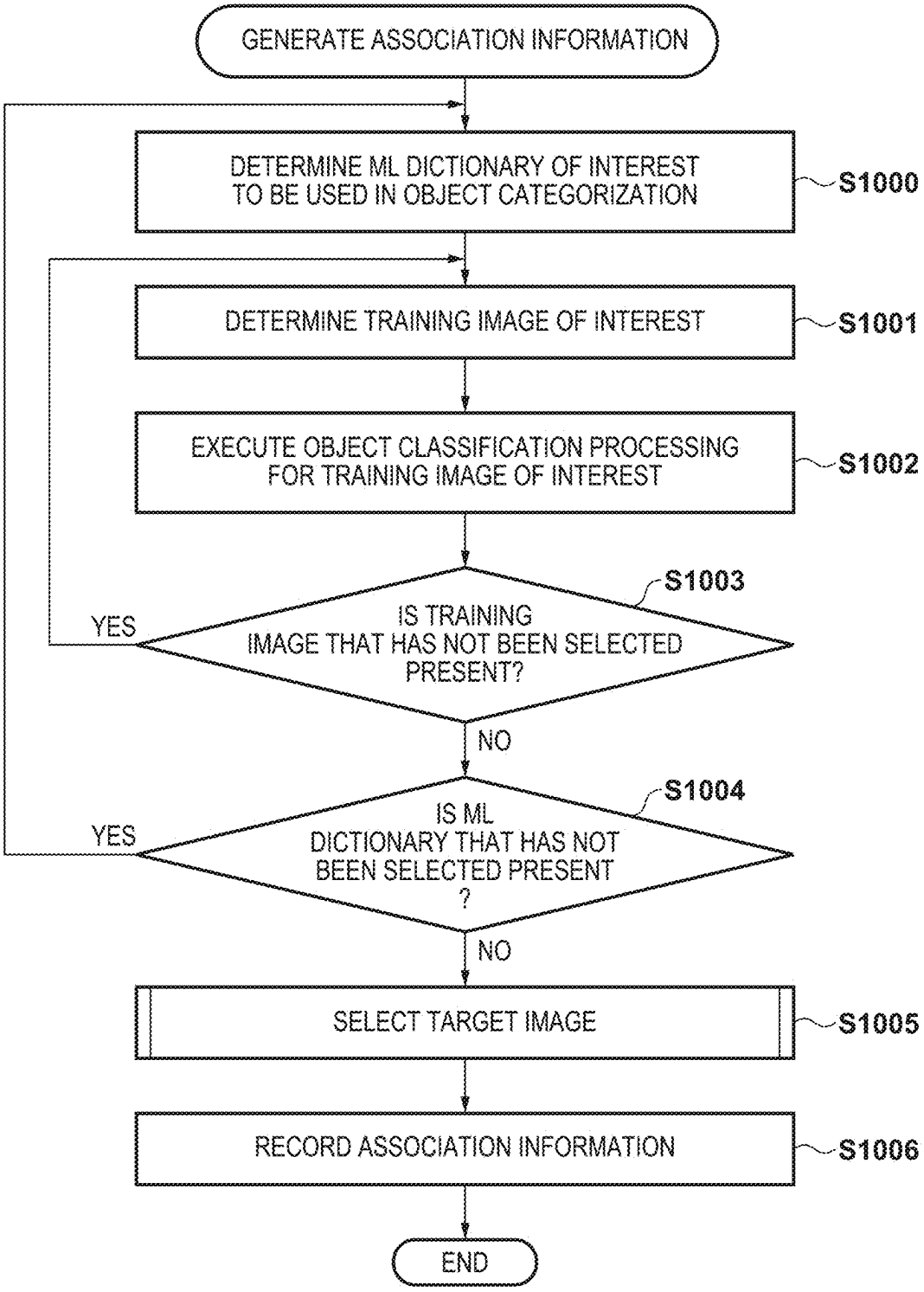

GENERATE ASSOCIATION INFORMATION
DETERMINE ML DICTIONARY OF INTEREST TO BE USED IN OBJECT CATEGORIZATION
S1000
DETERMINE TRAINING IMAGE OF INTEREST
S1001
EXECUTE OBJECT CLASSIFICATION PROCESSING FOR TRAINING IMAGE OF INTEREST
S1002
S1003
IS TRAINING IMAGE THAT HAS NOT BEEN SELECTED PRESENT?
YES
NO
S1004
IS ML DICTIONARY THAT HAS NOT BEEN SELECTED PRESENT ?
YES
NO
SELECT TARGET IMAGE
S1005
RECORD ASSOCIATION INFORMATION
S1006
END

FIG. 11

CATEGORY	Class1	Class2	Class3	Class4
TYPE	HORSE:49000	ZEBRA:500	DONKEY:400	OTHER:100
HAIR COLOR	BAY HAIR:10000	BLACK HAIR:10000	CHESTNUT HAIR:10000	OTHER COLOR HAIR:20000
POSTURE	STANDING ON FOUR LEGS:45000	STANDING ON TWO LEGS:4000	PRONE:900	SLEEPING:100
MASK	YES:40000	NO:10000		
SADDLE	YES:40000	NO:10000		

FIG. 12

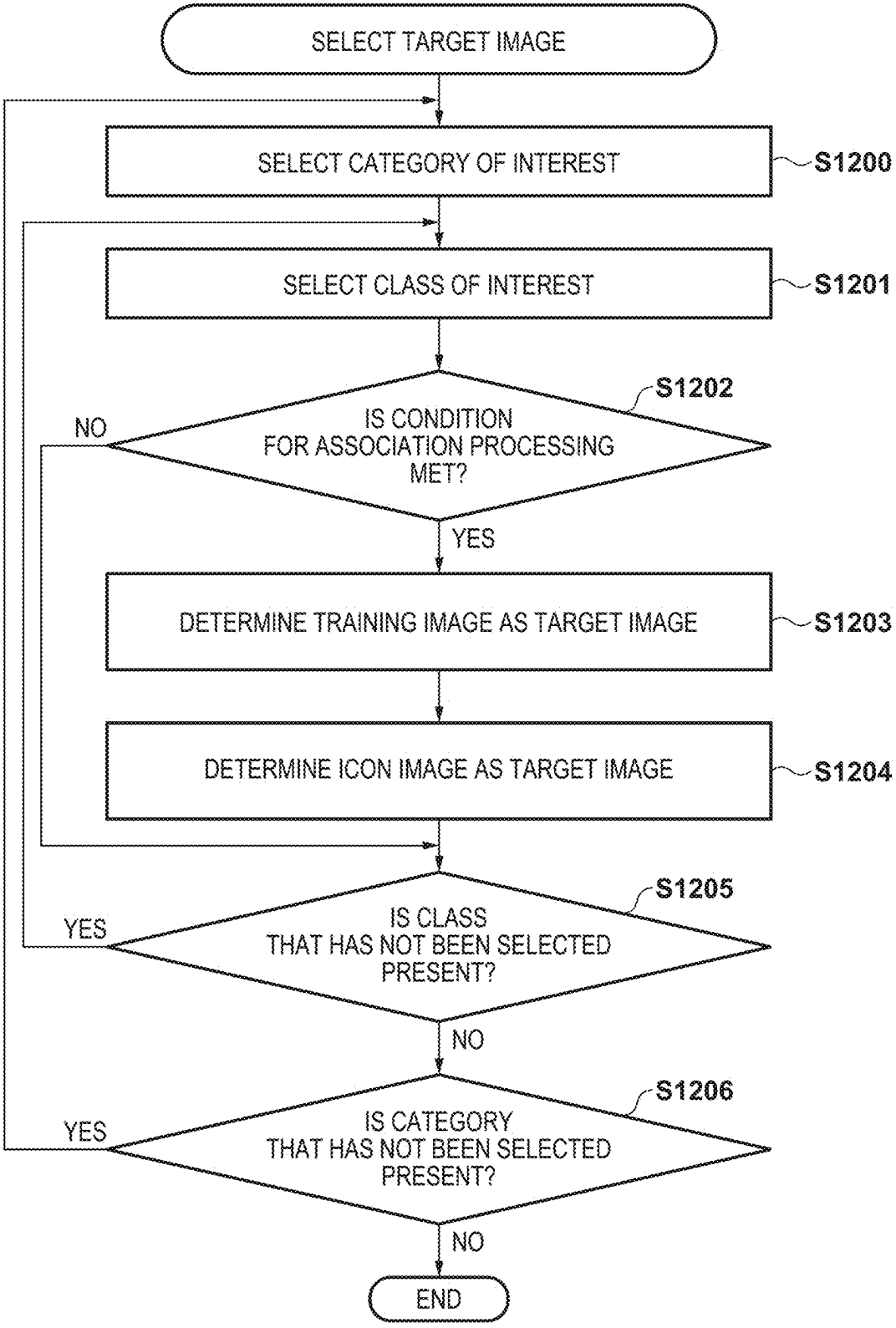

SELECT TARGET IMAGE
SELECT CATEGORY OF INTEREST
S1200
SELECT CLASS OF INTEREST
S1201
S1202
IS CONDITION FOR ASSOCIATION PROCESSING MET?
NO
YES
DETERMINE TRAINING IMAGE AS TARGET IMAGE
S1203
DETERMINE ICON IMAGE AS TARGET IMAGE
S1204
S1205
IS CLASS THAT HAS NOT BEEN SELECTED PRESENT?
YES
NO
S1206
IS CATEGORY THAT HAS NOT BEEN SELECTED PRESENT?
YES
NO
END

FIG. 13A

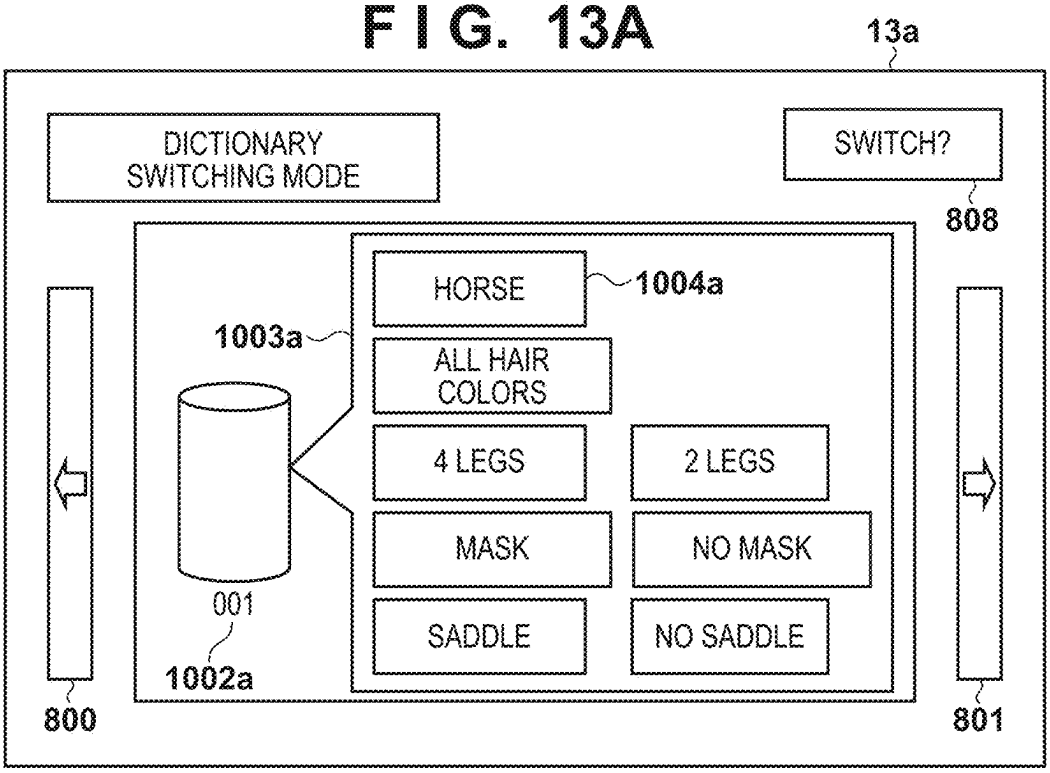


FIG. 13B

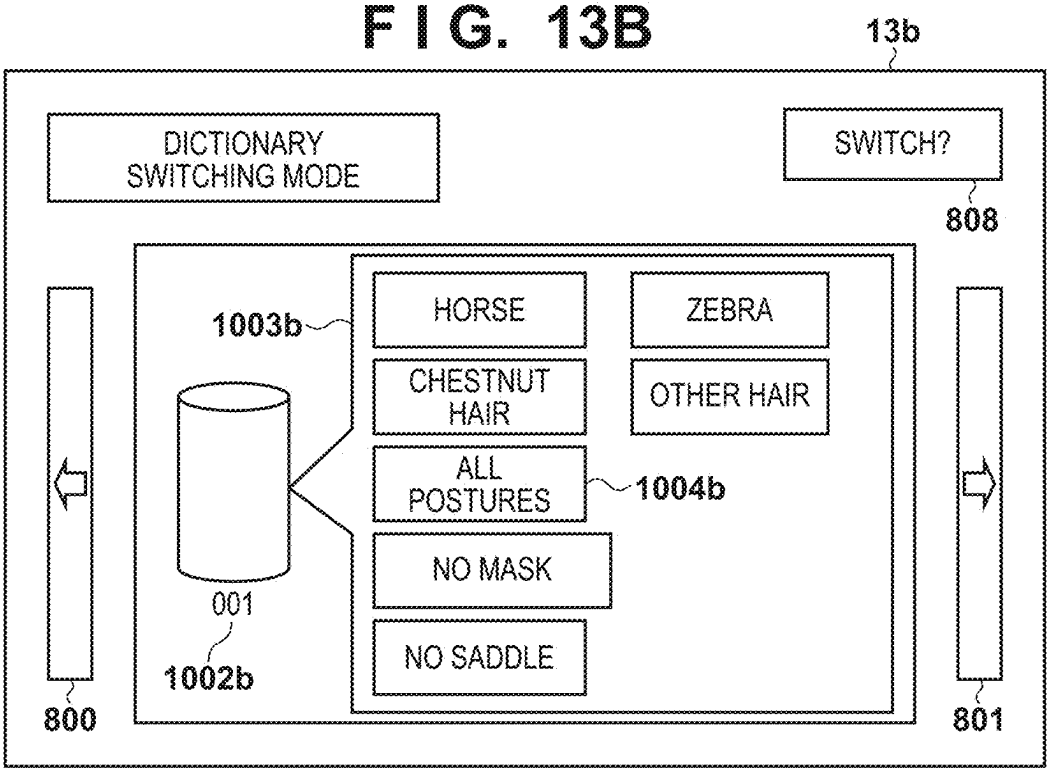


FIG. 14

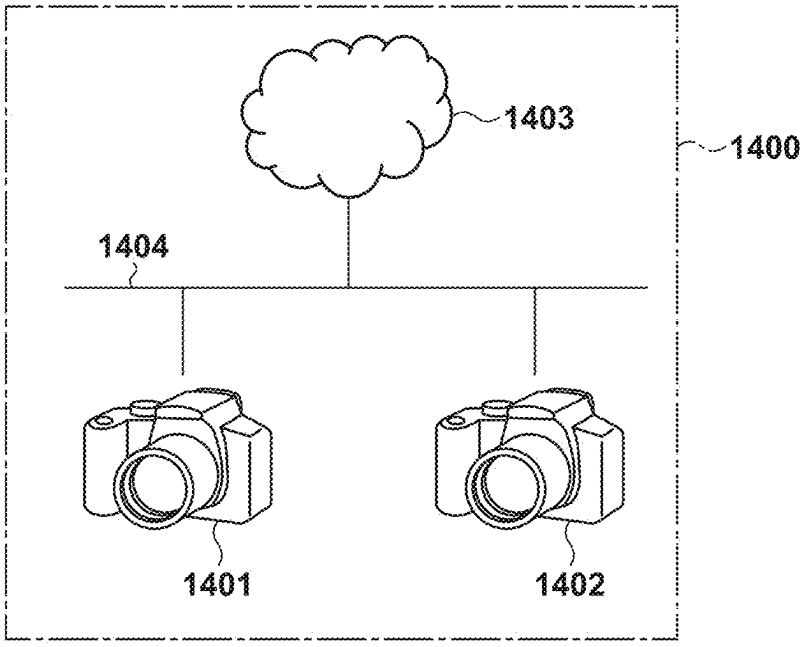

1403
1400
1404
1401
1402

FIG. 15

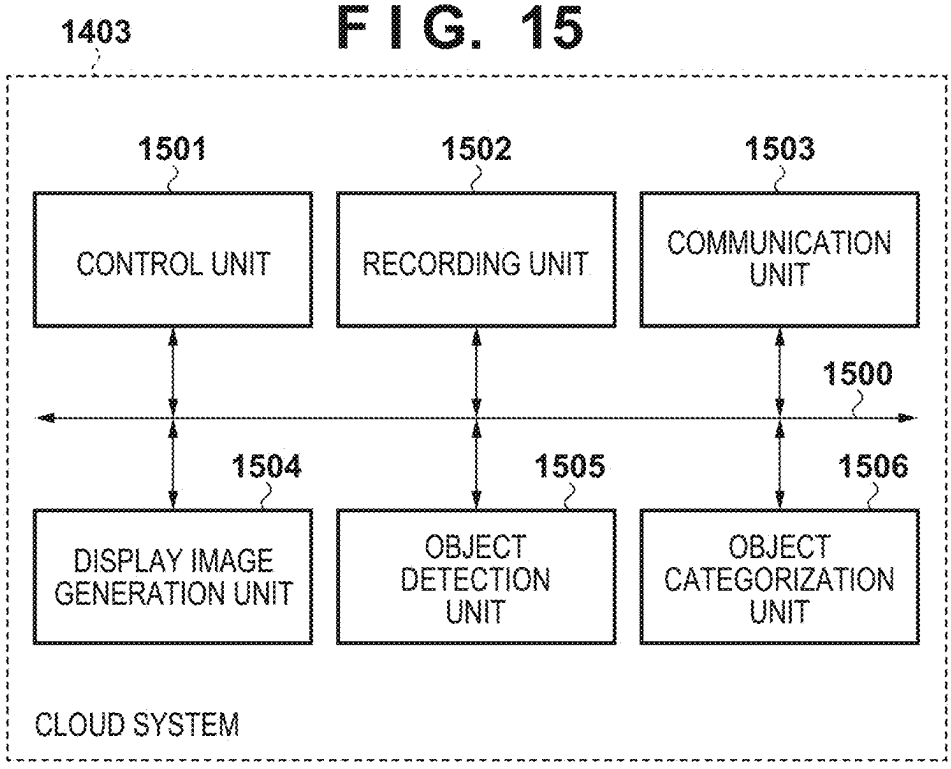

1403
1501
1502
1503
CONTROL UNIT
RECORDING UNIT
COMMUNICATION UNIT
1500
1504
1505
1506
DISPLAY IMAGE GENERATION UNIT
OBJECT DETECTION UNIT
OBJECT CATEGORIZATION UNIT
CLOUD SYSTEM

FIG. 16A

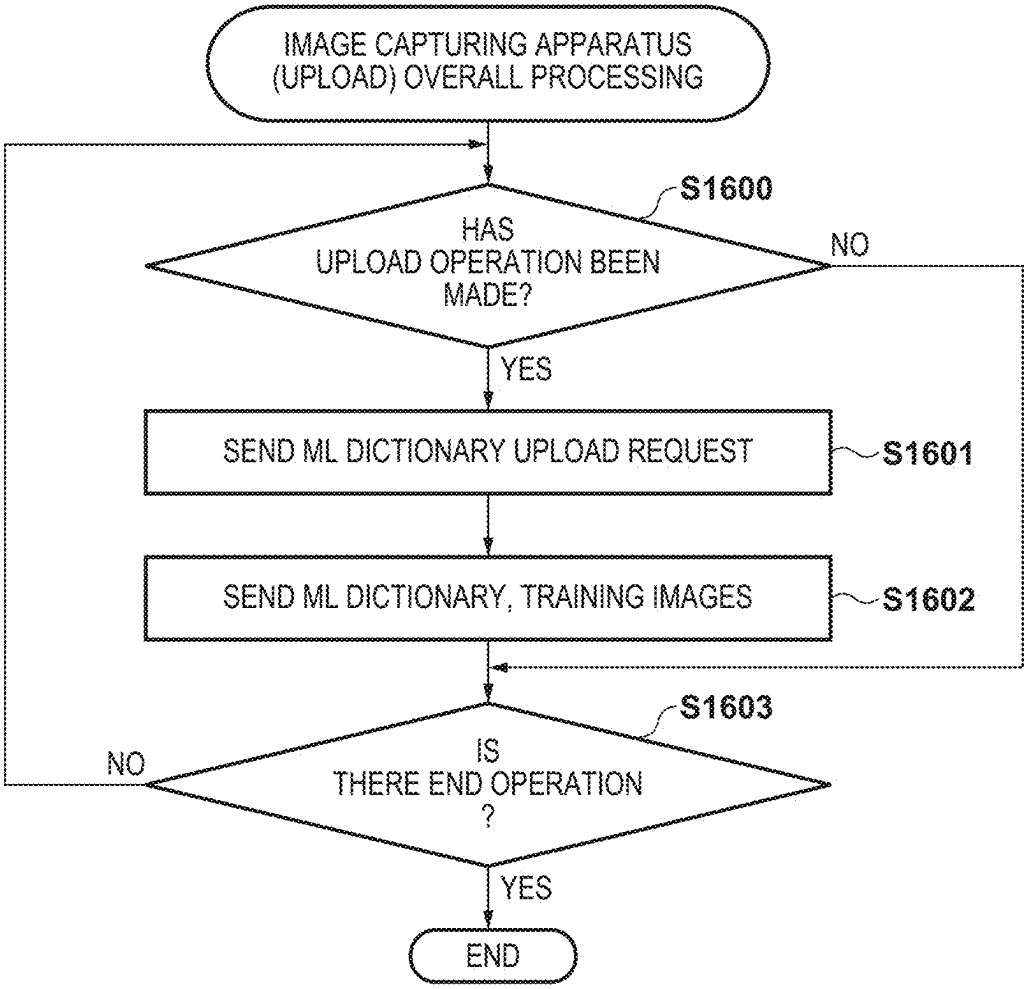
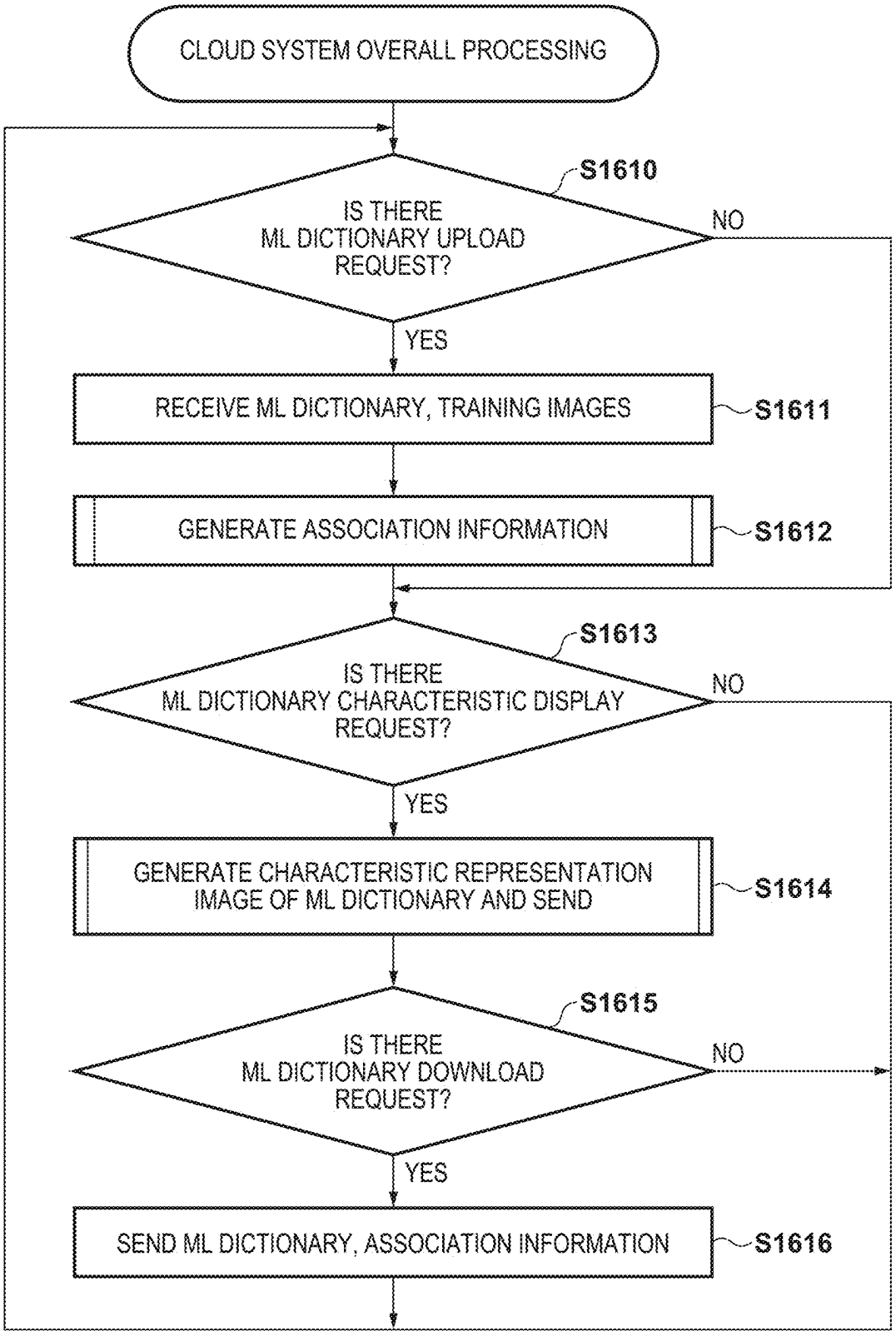


IMAGE CAPTURING APPARATUS (UPLOAD) OVERALL PROCESSING
S1600
HAS UPLOAD OPERATION BEEN MADE?
NO
YES
SEND ML DICTIONARY UPLOAD REQUEST
S1601
SEND ML DICTIONARY, TRAINING IMAGES
S1602
S1603
IS THERE END OPERATION ?
NO
YES
END

FIG. 16B


CLOUD SYSTEM OVERALL PROCESSING
S1610
IS THERE ML DICTIONARY UPLOAD REQUEST?
NO
YES
RECEIVE ML DICTIONARY, TRAINING IMAGES
S1611
GENERATE ASSOCIATION INFORMATION
S1612
S1613
IS THERE ML DICTIONARY CHARACTERISTIC DISPLAY REQUEST?
NO
YES
GENERATE CHARACTERISTIC REPRESENTATION IMAGE OF ML DICTIONARY AND SEND
S1614
S1615
IS THERE ML DICTIONARY DOWNLOAD REQUEST?
NO
YES
SEND ML DICTIONARY, ASSOCIATION INFORMATION
S1616

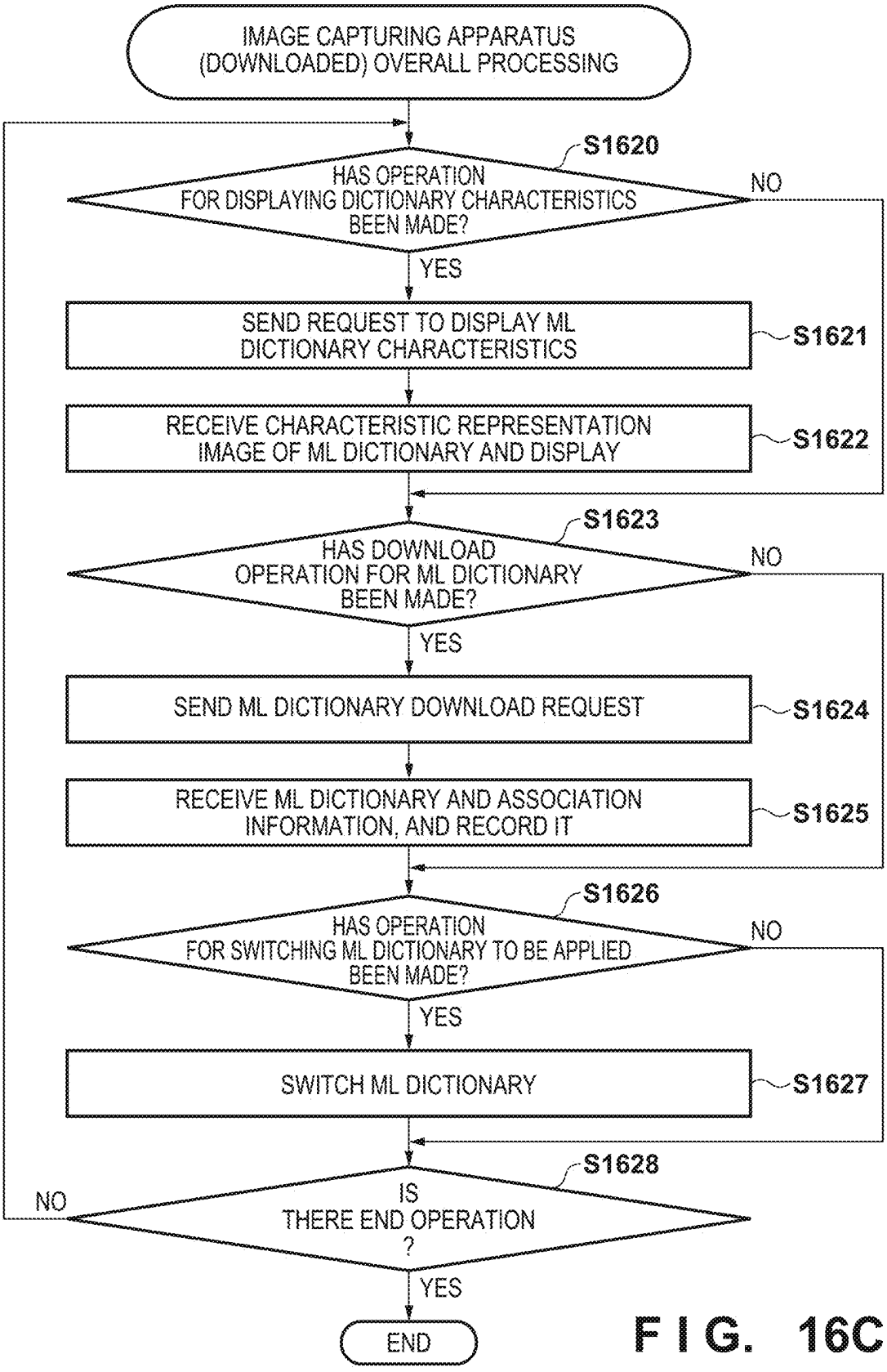


IMAGE CAPTURING APPARATUS (DOWNLOADED) OVERALL PROCESSING
S1620
HAS OPERATION FOR DISPLAYING DICTIONARY CHARACTERISTICS BEEN MADE?
NO
YES
SEND REQUEST TO DISPLAY ML DICTIONARY CHARACTERISTICS
S1621
RECEIVE CHARACTERISTIC REPRESENTATION IMAGE OF ML DICTIONARY AND DISPLAY
S1622
S1623
HAS DOWNLOAD OPERATION FOR ML DICTIONARY BEEN MADE?
NO
YES
SEND ML DICTIONARY DOWNLOAD REQUEST
S1624
RECEIVE ML DICTIONARY AND ASSOCIATION INFORMATION, AND RECORD IT
S1625
S1626
HAS OPERATION FOR SWITCHING ML DICTIONARY TO BE APPLIED BEEN MADE?
NO
YES
SWITCH ML DICTIONARY
S1627
S1628
IS THERE END OPERATION ?
NO
YES
END

FIG. 16C

IMAGE PROCESSING APPARATUS AND CONTROL METHOD THEREOF, IMAGE CAPTURING APPARATUS, AND STORAGE MEDIUM

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a Continuation of International Patent Application No. PCT/JP2023/035350, filed Sep. 28, 2023, which claims the benefit of Japanese Patent Application No. 2022-188698, filed Nov. 25, 2022, both of which are hereby incorporated by reference herein in their entirety.

BACKGROUND

Field of the Technology

[0002] The present disclosure relates to an image processing apparatus and a control method thereof, an image capturing apparatus, and a storage medium.

Description of the Related Art

[0003] Cameras equipped with dictionaries prepared by the manufacturer in advance through machine learning ("ML dictionaries" hereinafter) are being sold. With this type of camera, using the ML dictionary makes it possible to detect various objects, such as people, dogs, horses, and the like, from images that have been shot. As machine learning technology has become more widespread, using a plurality of ML dictionaries in cameras has also been proposed. As a configuration that uses a plurality of ML dictionaries, Patent Literature (PTL) 1 discloses switching among a plurality of ML dictionaries on the basis of past results of detecting continuous images using the ML dictionaries. Additionally, PTL 2 discloses analyzing a detection result from each of a plurality of ML dictionaries, and changing an analysis time for which detection is to be performed using each respective ML dictionary in accordance with the accuracy of each result.

CITATION LIST

Patent Literature

[0004] PTL 1: Japanese Patent Laid-Open No. 2021-132369

[0005] PTL 2: Japanese Patent Laid-Open No. 2022-039667

Non Patent Literature

[0006] NPL 1: S. Haykin, "Neural Networks A Comprehensive Foundation 2nd Edition", Prentice Hall, pp. 156-255, July 1998 (referenced in the embodiments)

[0007] However, these documents do not consider situations where a user selects and uses a preferred ML dictionary from a plurality of ML dictionaries installed in the camera to suit that user's purposes. When a user selects a preferred dictionary from a plurality of ML dictionaries, it is desirable for the user to be able to understand the characteristics of the selected ML dictionary in advance. For example, even among ML dictionaries capable of detecting horses, it is conceivable that a plurality of the ML dictionaries will have different characteristics, such as one ML dictionary being suited to detecting special postures, another ML dictionary being suited to detecting similar species such as zebras, and the like. A method for expressing the performance of an ML dictionary has not yet been proposed, and when a user selects one of a plurality of ML dictionaries to use, it may be difficult for the user to understand the characteristics of the ML dictionary in advance.

SUMMARY

[0008] One aspect of the present disclosure provides a configuration capable of generating information enabling a user to understand the characteristics of a dictionary obtained through machine learning.

[0009] An image processing apparatus according to one aspect of the present disclosure has the following configuration. In other words, the image processing apparatus includes: obtaining means for obtaining a dictionary obtained through machine learning and a plurality of training images used in the machine learning of the dictionary; selecting means for selecting, from the plurality of training images, at least one training image to be used as information expressing a characteristic of the dictionary; and generating means for generating association information that associates the at least one training image selected by the selecting means with the dictionary.

[0010] Features of the present disclosure will become apparent from the following description of exemplary embodiments with reference to the attached drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0011] The accompanying drawings, which are incorporated in and constitute a part of the specification, illustrate embodiments of the disclosure and, together with the description, serve to explain principles of the disclosure.

[0012] FIG. **1**A is a block diagram illustrating an image capturing apparatus according to a first embodiment.

[0013] FIG. **1**B is a block diagram illustrating an example of the functional configuration of a dictionary switching unit in the image capturing apparatus.

[0014] FIG. **2** is a schematic diagram illustrating an example of the configuration of a CNN in an object detection unit of the first embodiment.

[0015] FIG. **3** is a schematic diagram illustrating an example of part of the configuration of the CNN according to the first embodiment.

[0016] FIG. **4** is a schematic diagram illustrating an example of the configuration of a CNN in an object categorization unit of the first embodiment.

[0017] FIG. **5** is a flowchart illustrating ML dictionary switching processing according to the first embodiment.

[0018] FIG. **6** is a flowchart illustrating association information generation processing according to the first embodiment.

[0019] FIG. **7**A is a flowchart illustrating target image selection processing according to the first embodiment.

[0020] FIG. **7**B is a flowchart illustrating target image selection processing according to the first embodiment.

[0021] FIG. **8**A is a schematic diagram illustrating an example of the display of a dictionary characteristic expression according to the first embodiment.

[0022] FIG. **8**B is a schematic diagram illustrating an example of the display of a dictionary characteristic expression according to the first embodiment.

[0023] FIG. **9** is a flowchart illustrating target image selection processing according to a second embodiment.

[0024] FIG. **10** is a flowchart illustrating association information generation processing according to a third embodiment.

[0025] FIG. **11** is a diagram illustrating an example of an object categorization result according to the third embodiment.

[0026] FIG. **12** is a flowchart illustrating target image selection processing according to the third embodiment.

[0027] FIG. **13**A is a schematic diagram illustrating an example of the display of a dictionary characteristic expression according to the third embodiment.

[0028] FIG. **13**B is a schematic diagram illustrating an example of the display of a dictionary characteristic expression according to the third embodiment.

[0029] FIG. **14** is a block diagram illustrating an example of the configuration of an image processing system according to a fourth embodiment.

[0030] FIG. **15** is a block diagram illustrating an example of the configuration of a cloud system according to a fourth embodiment.

[0031] FIG. **16**A is a flowchart illustrating processing according to the fourth embodiment.

[0032] FIG. **16**B is a flowchart illustrating processing according to the fourth embodiment.

[0033] FIG. **16**C is a flowchart illustrating processing according to the fourth embodiment.

DESCRIPTION OF THE EMBODIMENTS

[0034] Hereinafter, embodiments will be described in detail with reference to the attached drawings. Note, the following embodiments are not intended to limit the scope of the claims. Multiple features are described in the embodiments, but it is not the case that all such features are required, and multiple such features may be combined as appropriate. Furthermore, in the attached drawings, the same reference numerals are given to the same or similar configurations, and redundant description thereof is omitted.

First Embodiment

(Configuration of Image Capturing Apparatus)

[0035] FIG. **1**A is a block diagram illustrating an example of the configuration of an image capturing apparatus **100** according to a first embodiment. The image capturing apparatus **100** shoots an object and records the data of moving images, still images, or the like to various types of media, such as tape, a solid-state memory, an optical disk, a magnetic disk, or the like. The image capturing apparatus **100** is, for example, a digital still camera, a video camera, or the like, but is not limited thereto. The various units in the image capturing apparatus **100** are connected over a bus **160**. Each unit is controlled by a central processing unit (CPU) **151**. The image capturing apparatus **100** incorporates an image processing apparatus constituted by an image processing unit **152**, an image compression/decompression unit **153**, an object detection unit **162**, an object categorization unit **163**, a dictionary switching unit **164**, and the like.

[0036] A lens unit **101** is configured including a fixed one-group lens **102**, a zoom lens **111**, an aperture stop **103**, a fixed three-group lens **121**, and a focus lens **131**. An aperture control unit **105** adjusts the diameter of an opening in the aperture stop **103**, to adjust an amount of light during shooting, by driving the aperture stop **103** via an aperture motor **104** (AM) in accordance with commands from the CPU **151**. The aperture control unit **105** controls the exposure using a luminance value of a specific object region. A zoom control unit **113** changes a focal length by driving the zoom lens **111** using a zoom motor **112** (ZM). A focus control unit **133** determines a driving amount at which to drive a focus motor **132** (FM) on the basis of a shift amount in a focus direction of the lens unit **101**. The focus control unit **133** controls a focus adjustment state by driving the focus lens **131** using the focus motor **132** (FM). AF control for a specific object region, for example, is implemented by the focus control unit **133** and the focus motor **132** controlling the movement of the focus lens **131**. The focus lens **131** is a lens for adjusting the focus. While the focus lens **131** is illustrated as a single lens in FIG. **1**A for simplicity, the focus lens **131** is normally constituted by a plurality of lenses.

[0037] An image sensor **141** is a photoelectric conversion element that photoelectrically converts an optical image of an object into an electrical signal. Light-receiving elements constituted by m pixels in the horizontal direction and n pixels in the vertical direction are arranged in the image sensor **141**. An object image formed on an image sensor **141** via the lens unit **101** is converted into an electrical signal by the image sensor **141**. The image formed on the image sensor **141** and photoelectrically converted is processed into an image signal (image data) by a captured image signal processing unit **142**, and is obtained as an image of an image capturing plane. The image data output from the captured image signal processing unit **142** is sent to an image capturing control unit **143**, and is temporarily held in a random access memory (RAM) **154**. The image data held in the RAM **154** is compressed by the image compression/decompression unit **153** and then recorded into a recording medium **157**. At the same time, the image data held in the RAM **154** is sent to the image processing unit **152**.

[0038] The image capturing control unit **143** receives, from the CPU **151**, an instruction for an accumulation time of the image sensor **141** and a setting value for gain used in output from the image sensor **141** to the captured image signal processing unit **142**, and controls the image sensor **141**. The CPU **151** sets the accumulation time and gain setting value on the basis of instructions from an operator input through an operation switch **156**, or on the basis of the magnitude of a pixel signal in the image data temporarily held in the RAM **154**.

[0039] The image processing unit **152** processes image signals, performs processing for enlarging or reducing the image data to an optimal size, calculates the similarity of image data, and the like. The image data that has been processed to the optimal size is sent to a display **150** and displayed as appropriate in order to display a preview image, a through-the-lens image, or the like. The object detection result from the object detection unit **162** can also be displayed in a superimposed manner in an image display on a display **150**. For example, the object detection result can be displayed in a rectangle or the like on the display **150**. Additionally, using the RAM **154** as a ring buffer makes it possible to buffer a plurality of items of image data captured within a predetermined period and detection results from the object detection unit **162** corresponding to each item of image data. Similarly, image data used for learning by the

object detection unit **162**, and object detection results corresponding to the image data, can be buffered. The image processing unit **152** also performs gamma correction, white balance processing, and the like based on the object region.

[0040] The operation switch **156** is an input interface including a touch panel, buttons, and the like. Various operations can be performed by using the operation switch **156** to select various function icons displayed on the display **150**. For example, the user can select images to be used for machine learning, perform operations necessary for machine learning (for example, designating a two-dimensional correct answer region corresponding to the image), and the like while viewing a shot image displayed in the display **150**. The user can also select an ML dictionary to be downloaded, instruct the download, and the like while viewing a GUI of the cloud system obtained through a communication unit **161** and displayed on the display **150**.

[0041] The recording medium **157** is a recording medium such as an SD card, and image data output from the captured image signal processing unit **142**, a plurality of ML dictionaries applicable in the object detection unit **162** and the object categorization unit **163**, and the like are recorded therein. Training images used in respective instances of machine learning, association information (described later), and the like are also recorded in the ML dictionaries applied in the object detection unit **162**. Icon images corresponding to each category and each class in the object categorization results are also recorded in the ML dictionary applied in the object categorization unit **163**. The communication unit **161** communicates ML dictionaries, information related to ML dictionaries such as training images, and the like by connecting to a cloud system or the like over Ethernet or wirelessly.

[0042] The object detection unit **162** determines a region in which an object, such as a horse, is present using an image signal, by applying an ML dictionary selected from a plurality of ML dictionaries recorded in the recording medium **157**. The object detection processing by the object detection unit **162** is implemented by feature extraction processing using a Deep Neural Network (DNN). The configuration of the object detection unit **162** will be described in detail later.

[0043] The object categorization unit **163** categorizes to which of predetermined classes an object belongs using an image signal, by applying an ML dictionary recorded in the recording medium **157**. The categorization processing by the object categorization unit **163** is implemented by feature extraction processing using a Deep Neural Network (DNN). In the object categorization unit **163**, appropriately switching the ML dictionary to be applied makes it possible to make various multi-class categories, such as a species category of the object, a posture category, a category for the presence or absence of ornaments, a hair color category, and the like. The configuration of the object categorization unit **163** will be described in detail later.

[0044] The dictionary switching unit **164** switches the ML dictionary used by the object detection unit **162** to an ML dictionary selected from the plurality of ML dictionaries in response to a predetermined user operation made on the operation switch **156**. FIG. **1**B is a block diagram illustrating an example of the functional configuration of the dictionary switching unit **164**. A switching control unit **201** accepts user operations from the operation switch **156**, controls the various functional units of the dictionary switching unit **164**, and controls displays in the display **150** during the dictionary switching operation. A characteristic obtainment unit **202** provides a characteristic representation of the ML dictionary to the switching control unit **201** on the basis of the association information recorded in the recording medium **157**. An association information generation unit **203** selects a training image to be used in the characteristic representation from a plurality of training images used in the machine learning of the ML dictionary, generates association information associating the ML dictionary with the selected training image, and records the association information in a recording medium **157**. The dictionary switching unit **164** will be described in detail later.

[0045] Returning to FIG. **1**A, a battery **159** is managed by a power source managing unit **158**, and provides a stable supply of power to the image capturing apparatus **100** as a whole. Control programs necessary for the image capturing apparatus **100** to operate, parameters used in the operations of the various units, and the like are recorded in a flash memory **155**. When the image capturing apparatus **100** is started up in response to a user operation (i.e., when the apparatus transitions from a power off state to a power on state), the control programs, parameters, and the like stored in the flash memory **155** are partially loaded into the RAM **154**. The CPU **151** controls the operations of the image capturing apparatus **100** in accordance with control programs, parameters, and the like loaded into the RAM **154**.

(Configuration of Object Detection Unit **162**)

[0046] In the present embodiment, the object detection unit **162** is constituted by a Convolutional Neural Network (CNN), but is not limited thereto, and any DNN using machine learning technology serves as an embodiment of the present disclosure. The basic configuration of a CNN will be described with reference to FIGS. **2** and **3**. FIG. **2** illustrates the basic configuration of a CNN that detects an object from two-dimensional image data that has been input. In FIG. **2**, the left end corresponds to the input, and the processing advances with progress to the right. The CNN takes two layers, called a feature detection layer (an S layer) and a feature integration layer (a C layer) as a single set, and those sets are configured hierarchically.

[0047] With the CNN, first, in the S layer, the next feature is detected on the basis of a feature detected in the previous hierarchy level. The features detected in the S layer are integrated in the C layer, and the detection result in that hierarchy level is then sent to the next hierarchy level. The S layer is constituted by feature detection cell planes, and detects different features in each feature detection cell plane. The C layer is constituted by feature integration cell planes, and pools the detection results from the feature detection cell planes in the previous stage. Unless specified otherwise, the feature detection cell planes and the feature integration cell planes will be collectively referred to as "feature planes" in the following. In the present embodiment, an output layer, which is the final hierarchy level, is constituted only by S layers, and does not use C layers.

[0048] Feature detection processing in the feature detection cell plane and feature integration processing in the feature integration cell plane will be described in detail with reference to FIG. **3**. The feature detection cell plane is constituted by a plurality of feature detection neurons, and the feature detection neurons are integrated in the C layers of the previous hierarchy level according to a predetermined structure. Meanwhile, the feature integration cell plane is

constituted by a plurality of feature integration neurons, and the feature integration neurons are integrated in the S layers of the same hierarchy level according to a predetermined structure. In FIG. **3**, in an M-th cell plane of the S layer in an L-th hierarchy level, the output value of the feature detection neuron at a position (ξ, ζ) is denoted as $y^{LS}{}_{M}(\xi, \zeta)$, and in an M-th cell plane of the C layer in an L-th hierarchy level, the output value of the feature integration neuron at position (ξ, ζ) is denoted as $y^{LC}{}_{M}(\xi, \zeta)$. Then, assuming that the coupling coefficients of the respective neurons are represented by $w^{LS}{}_{M}(n, u, v)$ and $w^{LC}{}_{M}(u, v)$, each output value can be expressed as the following [Math. 1] and [Math. 2].

$$y_M^{LS}(\xi, \zeta) \equiv f\left(u_M^{LS}(\xi, \zeta)\right) \equiv f\left\{\sum_{n,u,v} w_M^{LS}(n, u, v) \cdot y_n^{L-1C}(\xi + u, \zeta + v)\right\} \quad \text{[Math. 1]}$$

$$y_M^{LC}(\xi, \zeta) \equiv u_M^{LC}(\xi, \zeta) \equiv \sum_{u,v} w_M^{LC}(u, v) \cdot y_M^{LS}(\xi + u, \zeta + v) \quad \text{[Math. 2]}$$

[0049] f in [Math. 1] is an activation function, such as any sigmoid function such as a logistic function or a hyperbolic tangent function, and may be realized by a tan h function, for example. $u^{LS}{}_{M}(\xi, \zeta)$ represents the internal state of the feature detection neuron at position (ξ, ζ) in the M-th cell plane of the S layer in the L-th hierarchy level. [Math. 2] finds a simple linear sum without using an activation function. When an activation function is not used, as in [Math. 2], the internal state $u^{LC}{}_{M}(\xi, \zeta)$ and the output value $y^{LC}{}_{M}(\xi, \zeta)$ of the neuron are equal. $y^{L-1C}{}_{n}(\xi+u, \zeta+v)$ in [Math. 1] and $y^{LS}{}_{M}(\xi+u, \zeta+v)$ in [Math. 2] are called "integration destination output values" of the feature detection neuron and the feature integration neuron, respectively.

[0050] ξ, ζ, u, v, and n in [Math. 1] and [Math. 2] will be described next. The position (ξ, ζ) corresponds to positional coordinates in the input image, and for example, when $y^{LS}{}_{M}(\xi, \zeta)$ is a high output value, it is highly likely that a feature to be detected in the M-th cell plane of the S layer in the L-th hierarchy level is present at a pixel position (ξ, ζ) in the input image. n in [Math. 2] indicates an n-th cell plane in the C layer of an L−1-th hierarchy level, and is called an "integration destination feature number". Basically, a product-sum operation is performed for all cell planes present in the C layer of the L−1-th hierarchy level. (u,v) represents relative positional coordinates of the coupling coefficient, and a product-sum operation is performed in a limited range (u, v) in accordance with the size of the feature to be detected. This limited range of (u, v) is called a "receptive field". The size of the receptive field will be called the "receptive field size" hereinafter, and is expressed as a number of horizontal pixels×a number of vertical pixels in an integrated range.

[0051] In [Math. 1], in the S layer where L=1, i.e., the first S layer, $y^{L-1C}{}_{n}(\xi+u\ \zeta+v)$ is an input image $y^{in\text{-}image}(\xi+u, \zeta+v)$ or an input position map $y^{in\text{-}posi\text{-}map}(\xi+u, \zeta+v)$. Incidentally, neurons, pixels, and the like are distributed discretely, and the integration destination feature number is also discrete. ξ, ζ, u, v, and n are thus discrete values rather than continuous variables. Here, ξ and ζ are non-negative integers, n is a natural number, and u and v are integers, and all have limited ranges.

[0052] $w^{LS}{}_{M}(n, u, v)$ in [Math. 1] represents a coupling coefficient distribution for detecting a predetermined feature, and the predetermined feature can be detected by adjusting that coupling coefficient distribution to an appropriate value. The adjustment of the coupling coefficient distribution is training, and in the CNN architecture, the coupling coefficient is adjusted by presenting a variety of test patterns to repeatedly and gradually correct the coupling coefficient such that $y^{LS}{}_{M}(\xi, \zeta)$ becomes an appropriate output value.

[0053] Next, $w^{LC}{}_{M}(u, v)$ in [Math. 2] uses a two-dimensional Gaussian function, and can be expressed as indicated by the following [Math. 3].

$$w_M^{LC}(u, v) = \frac{1}{2\pi\sigma_{L,M}^2} \cdot \exp\left(-\frac{u^2 + v^2}{2\sigma_{L,M}^2}\right) \quad \text{[Math. 3]}$$

[0054] Here, (u, v) is a limited range, and as with the feature detection neuron, the limited range will be called a "receptive field" and the size of the range will be called a "receptive field size". The receptive field size may be set to any suitable value in accordance with the size of the M-th feature in the S layer of the L-th hierarchy level. In [Math. 3], σ represents a feature size factor, and may be set to any suitable constant in accordance with the receptive field size. Specifically, σ may be set such that values furthest on the outside of the receptive field are values that can substantially be treated as zero.

[0055] The object detection unit **162** performing object detection in the S layer of the final hierarchy level by performing computations such as those described above in each hierarchy level is the configuration of the object detection unit **162** according to the present embodiment.

(Learning Method of Object Detection Unit **162**)

[0056] A specific learning method for the object detection unit **162** will be described next. In the present embodiment, the coupling coefficient is adjusted through supervised learning. In supervised learning, an actual neuron output value is found by supplying a test pattern, and based on a relationship between the output value and a target signal (a desired output value to be output by that neuron), the coupling coefficient $w^{LS}{}_{M}(n, u, v)$ may be adjusted. In the learning according to the present embodiment, the coupling coefficient is corrected by using the least-squares method in the final feature detection layer and using an error back propagation method in intermediate feature detection layers. Details of the method for correcting the coupling coefficient, such as the least-squares method and the error back propagation method, can be found in NPL 1.

[0057] In the object detection unit **162**, many specific patterns to be detected and patterns not to be detected are prepared as test patterns for the learning. Each test pattern takes an image and a target signal as a single set. If the tan h function is used for the activation function, when the specific pattern to be detected is presented, the target signal is supplied so that an output of 1 is obtained for the neurons, in the feature detection cell plane in the final layer, located in the regions where the specific pattern is present. Conversely, when a pattern not to be detected is presented, the target signal is supplied so that an output of −1 is obtained for the neurons located in the regions of that pattern. In actual detection, computations are performed using the coupling coefficient $w^{LS}{}_{M}(n, u, v)$ constructed through learning, and if the neuron output in the feature detection cell

plane of the final layer is at least a predetermined value, the object is determined to be present. In this manner, the object detection unit **162** is constructed such that an object can be detected from a two-dimensional image.

(Configuration of Object Categorization Unit **163**)

[0058] In the present embodiment, the object categorization unit **163** is constituted by a CNN, but is not limited thereto, and any DNN using machine learning technology serves as an embodiment of the present disclosure. The basic configuration of the CNN will be described with reference to FIG. **4**. The CNN of the object categorization unit **163** is configured such that the number of outputs k of the n layers corresponds to the number of classes to be categorized. A softmax function is used for the activation function. Other parts are the same as in the configuration of the object detection unit **162** and will therefore not be described. Performing computations such as those described above in each hierarchy level, and performing object categorization in the S layer of the final hierarchy level of the object categorization unit **163**, is the CNN configuration of the object categorization unit **163** according to the present embodiment.

(Learning Method of Object Categorization Unit **163**)

[0059] A specific learning method for the object categorization unit **163** will be described next. In the object categorization unit **163**, many specific patterns to be categorized are prepared for each of many classes. Each test pattern takes an image and a target signal as a single set. When a softmax function is used for the activation function, the target signal is supplied such that the output of a correct answer class is 1, and outputs other than the correct answer class is 0. In actual categorization, computations are performed using the coupling coefficient $w^{LS}{}_{M}(n,u,v)$ constructed through learning, and the object is determined to belong to a class for which the neuron output in the feature detection cell plane of the final layer is at least a predetermined value.

[0060] A plurality of ML dictionaries, for which learning such as that described above has been performed for each of the plurality of multi-class categories, are prepared in advance and recorded in the recording medium **157**. Other parts are the same as in the method for the object detection unit **162** and will therefore not be described. In this manner, the object categorization unit **163** is constructed such that an object can be categorized into multiple classes from a two-dimensional image.

(Flow of Overall Processing by Dictionary Switching Unit **164**)

[0061] FIG. **5** is a flowchart illustrating the overall processing performed by the dictionary switching unit **164** according to the present embodiment. Note that each functional unit of the dictionary switching unit **164** illustrated in FIG. **1**B may be realized by the CPU **151** executing predetermined software, by dedicated hardware, or by software and hardware operating cooperatively.

[0062] In S**500**, the switching control unit **201** displays an ML dictionary switching screen on the display **150** in response to a predetermined user operation made using the operation switch **156** (an ML dictionary switching operation). A specific example of the switching screen will be described later with reference to FIGS. **8**A and **8**B. In S**501**, the switching control unit **201** determines whether an operation for switching the ML dictionary displayed on the switching screen has been made by the user through the operation switch **156**. If the switching control unit **201** determines that an operation for switching the ML dictionary has been made (YES in S**501**), the sequence moves to S**502**, whereas if the switching control unit **201** determines that such an operation has not been made (NO in S**501**), the sequence moves to S**505**.

[0063] In S**502**, the characteristic obtainment unit **202** determines whether association information for the new ML dictionary to be displayed, which was switched to in S**501**, is recorded in the recording medium **157**. If the characteristic obtainment unit **202** determines that association information for the ML dictionary to be displayed is not recorded in the recording medium **157** (NO in S**502**), the sequence moves to S**503**. However, if the characteristic obtainment unit **202** determines that association information for the ML dictionary to be displayed is recorded in the recording medium **157** (YES in S**502**), the sequence moves to S**504**.

[0064] In S**503**, the association information generation unit ("generation unit" hereinafter) **203** reads out the ML dictionary to be displayed after the display switching operation in S**501**, and a plurality of training images to be used to train the ML dictionary to be displayed, from the recording medium **157**, and generates the association information for the ML dictionary. The generation unit **203** records the obtained association information into the recording medium **157**. Here, the "association information" is information expressing a pair including an ML dictionary and a target image selected to express the characteristics of the ML dictionary, and includes the target image itself. The processing for generating the association information performed in S**503** will be described in detail later with reference to FIGS. **6**, **7**A, and **7**B.

[0065] Next, in S**504**, the characteristic obtainment unit **202** reads out the association information pertaining to the ML dictionary to be displayed after the switching operation has been performed in S**501** from the recording medium **157**, and obtains a training image associated by the read-out association information. The switching control unit **201** expresses the characteristics of the ML dictionary to be displayed after the switching by displaying the ML dictionary to be displayed, and a plurality of training images obtained by the characteristic obtainment unit **202**, on the display **150**. The display on the display **150** and the method for expressing the characteristics of the ML dictionary will be described in detail later with reference to FIGS. **8**A and **8**B.

[0066] In S**505**, the switching control unit **201** determines whether an operation for switching the ML dictionary to be applied in the object detection unit **162** has been made by the user through the operation switch **156** for the ML dictionary currently displayed on the switching screen. If the switching control unit **201** determines that an operation for switching the ML dictionary has been made (YES in S**505**), the sequence moves to S**506**. However, if the switching control unit **201** determines that an operation for switching the ML dictionary has not been made (NO in S**505**), the sequence moves to S**507**. In S**506**, the switching control unit **201** switches the ML dictionary applied in the object detection unit **162** to the ML dictionary to be displayed on the current switching screen in accordance with the switching operation

made by the user. Thereafter, object detection processing using the ML dictionary switched to in S**506** can be performed on through-the-lens images, captured images, and the like in the image capturing apparatus **100**. In S**507**, the switching control unit **201** determines whether an end operation has been made by the user through the operation switch **156**. If the switching control unit **201** determines that an end operation has been made (YES in S**507**), the sequence ends. However, if the switching control unit **201** determines that an end operation has not been made (NO in S**507**), the sequence returns to S**501**, and the foregoing processing is repeated.

(Flow of Association Information Generation Processing)

[0067] FIG. **6** is a flowchart illustrating processing for generating the association information, executed in S**503** of the first embodiment.

[0068] In S**600**, the generation unit **203** calculates a feature vector for each of the plurality of training images ("corresponding image group" hereinafter) used in the machine learning for the ML dictionary to be displayed after the switching in S**500**. These corresponding image groups are recorded in the recording medium **157** in correspondence with the ML dictionary. Although it is desirable that the corresponding image group include all the training images used in the machine learning for the ML dictionary, the corresponding image group may be constituted by training images randomly extracted from all the training images used in the machine learning, to the extent that the effects of the present disclosure are not affected. Note that a publicly-known method for calculating a feature vector from an image can be used to calculate the feature vector in S**600**. Alternatively, each training image may be input to the object detection unit **162**, and the output of a feature integration layer n-**1**, which is intermediate data obtained by the object detection unit **162**, may be used as the feature vector. Alternatively, each training image may be input to the object categorization unit **163** in which any one of the ML dictionaries recorded in the recording medium **157** is applied, and the output of the feature integration layer n-**1**, which is intermediate data obtained by the object categorization unit **163**, may be used as the feature vector.

[0069] In S**601**, the generation unit **203** calculates an average vector using the feature vectors calculated in S**600**. In S**602**, the generation unit **203** determines whether a first target image has been selected. If the generation unit **203** determines that the first target image has not been selected (NO in S**602**), the sequence moves to S**603**, whereas if the generation unit **203** determines that the first target image has been selected (YES in S**602**), the sequence moves to S**604**.

[0070] In S**603**, the generation unit **203** selects a first target image from the corresponding image group. The processing for selecting the first target image will be described in detail later with reference to the flowchart in FIG. **7**A. In S**604**, the generation unit **203** selects a second or subsequent target image from the corresponding image group. The processing for selecting the second target image will be described in detail later with reference to the flowchart in FIG. **7**B. Next, in S**605**, the generation unit **203** determines whether an ending condition is met. If the generation unit **203** determines that the ending condition is met (YES in S**605**), the sequence moves to S**606**. However, if the generation unit **203** determines that the ending condition is not met (NO in S**605**), the sequence returns to S**602**, and the foregoing processing is repeated. Here, in the present embodiment, a target image selection number is used as the ending condition in S**605**. The target image selection number is set in consideration of a size that can be viewed by the user when the training images are displayed on the display **150** in the display expressing the characteristics of the ML dictionary (S**504**). In the present embodiment, this number is set to four images (the first target image, and three images among the second and subsequent target images), for example. In S**606**, the generation unit **203** records information specifying the training images selected in S**603** and S**604** into the recording medium **157** as the association information pertaining to the ML dictionary to be displayed.

(Target Image Selection Processing)

[0071] FIG. **7**A is a flowchart illustrating processing for selecting the first target image (S**603**). The generation unit **203** selects, as the first target image, the training image having the feature vector which, of the plurality of feature vectors obtained from the corresponding image group (the plurality of training images), has the smallest distance from the average vector of the plurality of feature vectors.

[0072] In S**700**, the generation unit **203** selects a training image of interest from the corresponding image group. In S**701**, the generation unit **203** determines whether the training image of interest selected in S**700** has been selected as the target image. If the generation unit **203** determines that the training image of interest has been selected as the target image (YES in S**701**), the sequence returns to S**700**, and the next training image of interest is selected. However, if the generation unit **203** determines that the training image of interest has not been selected as the target image (NO in S**701**), the sequence moves to S**702**. Because the selection of the target image is performed repeatedly in the flowchart illustrated in FIG. **6**, the determination in this step is made with the intention of ensuring the target image is not selected twice. Note that S**701** can be omitted when the target image selected using the processing of S**603** is a single image as indicated in FIG. **6**. However, S**701** is required when the processing for selecting the plurality of target images is performed in order from the shortest distance with respect to the average vector (e.g., when a plurality of target images are selected through the processing of S**603**).

[0073] In S**702**, the generation unit **203** determines whether the training image of interest selected in S**700** meets a condition for being subject to association processing. If the generation unit **203** determines that the training image of interest meets the condition for being subject to the association processing (YES in S**702**), the sequence moves to S**703**. However, if the generation unit **203** determines that the training image of interest does not meet the condition for being subject to the association processing (NO in S**702**), the sequence returns to S**700**. In machine learning, even if an image is included in the training images, the image may not have a sufficient impact on the performance of the ML dictionary. As such, as one of the conditions for being subject to the association processing (conditions for being selected as a target image), a condition is set to ensure that training images which have little impact on the characteristics of the ML dictionary are not selected as target images. More specifically, the condition for being subject to the association processing is that an object can be detected when the training image of interest is input to the object detection unit **162**. Note that at this time, the ML dictionary to be

displayed is temporarily set in the object detection unit **162**. Alternatively, the condition may be that of the feature vectors calculated in S**600**, the percentage of the feature vectors for which the distance from the feature vector calculated from the training image of interest is sufficiently small (feature vectors for which the distance is no greater than a predetermined threshold) is at least a predetermined percentage. Furthermore, here, the "distance" may be any numerical value representing the similarity of feature vectors, e.g., a Euclidean distance or a cosine similarity between the vectors.

[0074] In S**703**, the generation unit **203** calculates a distance between the feature vector, among the feature vectors calculated in S**600**, that has been calculated from the training image of interest, and the average vector calculated in S**601**. The distance between the vectors is the same as the distance described with reference to S**702**. In S**704**, the generation unit **203** determines whether a candidate image has been selected in S**705**, which will be described later. If the generation unit **203** determines that a candidate image has not been selected (NO in S**704**), the sequence moves to S**705**, whereas if the generation unit **203** determines that a candidate image has been selected (YES in S**704**), the sequence moves to S**706**.

[0075] In S**705**, the generation unit **203** selects the training image of interest as a candidate image. Meanwhile, in S**706**, the generation unit **203** determines whether the distance between the feature vector of the training image of interest and the average vector is smaller than the distance between the feature vector of the candidate image selected in S**705** or S**707** and the average vector. If the generation unit **203** determines that the distance between the feature vector of the training image of interest and the average vector is smaller than the distance between the feature vector of the candidate image and the average vector (YES in S**706**), the sequence moves to S**707**. However, if the generation unit **203** determines that the distance between the feature vector of the training image of interest and the average vector is not smaller than the distance between the feature vector of the candidate image and the average vector (NO in S**706**), the sequence moves to S**708**. The determination in S**706** is made with the intention of selecting, as the candidate image, a training image from which a feature vector having a smaller distance from the average vector can be obtained. In S**707**, the generation unit **203** changes the candidate image to a current training image of interest.

[0076] In S**708**, the generation unit **203** determines whether a training image that has not been selected as the training image of interest is present in the corresponding image group. If the generation unit **203** determines that a training image that has not been selected as the training image of interest is present (YES in S**708**), the sequence returns to S**700**, and the foregoing processing is repeated. However, if the generation unit **203** determines that a training image that has not been selected as the training image of interest is not present (NO in S**708**), the sequence moves to S**709**. In S**709**, the generation unit **203** selects a training image of interest that is a candidate image as the target image.

[0077] According to the processing of FIG. **7**A as described above, a training image from which the feature vector having the smallest distance from the average vector is obtained is selected as the target image from the corresponding image group. In other words, a representative training image representing the performance of the ML dictionary instructed to be switched to in S**501** is selected.

[0078] FIG. **7**B is a flowchart illustrating processing for selecting the second and subsequent target images, executed in S**604** of the first embodiment. In this processing, the generation unit **203** selects a predetermined number of feature vectors from the plurality of feature vectors obtained from the corresponding image group, in order from the feature vector having the greatest distance from the average vector of the plurality of feature vectors. Therefore, by repeating S**604** a predetermined number of times, a predetermined number of training images are selected in order from the image for which the distance between the feature vector and the average vector is greatest. In FIG. **7**B, the processes of S**710** to S**719**, excluding S**716**, are the same as the processes of S**700** to S**709**, excluding S**706**, in FIG. **7**A. In S**716**, the generation unit **203** determines whether the distance between the feature vector of the training image of interest and the average vector is greater than the distance between the feature vector of the candidate image selected in S**715** or S**717** and the average vector. If the generation unit **203** determines that the distance between the feature vector of the training image of interest and the average vector is greater than the distance between the feature vector of the candidate image and the average vector (YES in S**716**), the sequence moves to S**717**. However, if the generation unit **203** determines that the distance between the feature vector of the training image of interest and the average vector is not greater than the distance between the feature vector of the candidate image and the average vector (NO in S**716**), the sequence moves to S**718**. The determination in S**716** is made with the intention of selecting, as the candidate image, a training image from which a feature vector having a greater distance from the average vector can be calculated.

[0079] According to the processing of FIG. **7**B as described above, a training image from which the feature vector having the greatest distance from the average vector is obtained is selected as the target image from the corresponding image group. In other words, a special training image representing the performance of the ML dictionary instructed to be switched to in S**501** is selected.

(Example of Display of Dictionary Characteristics)

[0080] FIGS. **8**A and **8**B are diagrams illustrating examples of the switching screen displayed on the display **150** by the switching control unit **201** through the processes of S**500** and S**504**. When a dictionary switching mode is designated, the switching control unit **201** displays a switching screen **8***a* illustrated in FIG. **8**A, for example. Items **800** and **801** are GUIs for switching the ML dictionary to be displayed by operating the operation switch **156**. When the item **800** or **801** is manipulated, in S**501**, the switching control unit **201** determines that an operation for switching the ML dictionary to be displayed has been made. An item **802***a* is the ML dictionary to be displayed, and is indicated by "001", which is the filename of that ML dictionary recorded in the recording medium **157**.

[0081] An area **803***a* is an area for expressing the characteristics of the ML dictionary, and is displayed by the processing of S**504**. In the example in FIG. **8**A, training images **804***a* to **807***a*, which are the target images selected to express the characteristics of the ML dictionary "001", are displayed. The training image **804***a* is the training image

selected as the first target image in S**603**. In S**603**, the representative training image is easily selected, and the training image **804***a* is an example of a training image in which one horse is standing on four legs. The training images **805***a*, **806***a*, and **807***a* are training images selected as the second and subsequent target images in S**604**. In S**604**, a special training image is easily selected, and the training images **805***a* to **807***a* are examples of training images in which a rider, ornaments, a plurality of objects, and the like are present. A switching button **808** is a GUI that accepts user operations for switching the ML dictionary applied in the object detection unit **162**. When the switching button **808** is operated, switching the ML dictionary applied in S**505** is determined to have been instructed.

[0082] As described above, the switching control unit **201** generates display information for displaying a dictionary and at least one training image in association with each other on the basis of the association information, and functions as a display information generation unit that displays those items on the display **150**. The dictionary characteristics of each dictionary are then presented to the user according to this display information. From the display illustrated in FIG. **8**A, for example, the user can understand in advance that the ML dictionary "001" is an ML dictionary suitable for shooting images at a racecourse or the like, for example.

[0083] FIG. **8**B will be described next. The items **800** and **801** and the switching button **808** are the same as in FIG. **8**A. An item **802***b* indicates the ML dictionary to be displayed in response to the item **800** or **801** being operated, and is indicated by "002", which is the filename of that ML dictionary recorded in the recording medium **157**. Training images **804***b* to **807***b*, which are the target images selected to express the characteristics of the ML dictionary "002", are displayed in an area **803***b*. The training image **804***b* is an example of the training image selected as the first target image in S**603**. The training images **805***b*, **806***b*, and **807***b* are examples of training images selected as the second and subsequent target images in S**604**. In S**604**, a special training image is easily selected, and the training images **805***b* to **807***b* are examples of characteristic training images such as a horse with its head down, in a posture such as prone, a similar species such as a zebra, and the like.

[0084] By displaying dictionary characteristics as described with reference to FIG. **8**B above, the user can understand in advance that the dictionary is an ML dictionary suitable for shooting images at a zoo or the like, for example.

[0085] As described above, according to the first embodiment, training images having representative features and special features with respect to an ML dictionary are displayed. Accordingly, by confirming those images, the user can select the ML dictionary to be applied having understood, in advance, the overall characteristics of the dictionary. Note that in the foregoing embodiment, the total number of selected target images is four, the number of target images selected through the processing of S**603** (FIG. **7**A) (representative training images) is one, and the number of target images selected through the processing of S**604** (FIG. **7**B) (special training images) is three. However, the present disclosure is not limited thereto, and the total number of target images, the number of representative training images, and the number of special training images may be set as desired. However, because using special training images makes it possible to express a broader range of the characteristics of the ML dictionary, it is desirable that the number of special training images be higher than the number of representative training images, from the standpoint of understanding the comprehensive characteristics.

Second Embodiment

[0086] In the first embodiment, the distance between the feature vector of the training image of interest and the average vector was used to determine whether to select the training image of interest as the second or subsequent target image. In a second embodiment, the distance from the feature vector of the first target image (a representative training image) is used to determine whether to select the training image of interest as the second or subsequent target image. The configuration, functions, and processing of the image capturing apparatus **100** of the second embodiment are the same as in the first embodiment, with the exception of the processing for selecting the second and subsequent target images. The processing for selecting the second and subsequent target images according to the second embodiment (the processing in S**604** of FIG. **6**) will be described hereinafter with reference to FIG. **9**.

(Processing for Selecting Association Target Image)

[0087] FIG. **9** is a flowchart illustrating processing for selecting the second and subsequent target images, executed in S**604** of the second embodiment. The processing of S**900** to S**909**, excluding S**903** and S**906**, is the same as the processing of S**710** to S**719**, excluding S**713** and S**716**, in the first embodiment (FIG. **7**B).

[0088] In S**903**, the generation unit **203** calculates a distance between the feature vector calculated from the training image of interest and the feature vector calculated from the first target image selected in S**603**. Note that these feature vectors are calculated in S**600**. In S**906**, the generation unit **203** determines whether the distance between the feature vector of the training image of interest and the feature vector of the first target image is greater than the distance between the feature vector of the candidate image selected in S**905** or S**907** and the feature vector of the first target image. If the generation unit **203** determines that the distance between the feature vector of the training image of interest and the feature vector of the first target image is greater than the distance between the feature vector of the selected candidate image and the feature vector of the first target image (YES in S**906**), the sequence moves to S**907**. If the generation unit **203** determines that the distance between the feature vector of the training image of interest and the feature vector of the first target image is not greater than the distance between the feature vector of the selected candidate image and the feature vector of the first target image (NO in S**906**), the sequence moves to S**908**. The determination in S**906** is made with the intention of selecting, as the candidate image, a training image from which a feature vector having a greater distance from the feature vector of the first selected image can be calculated.

[0089] As described above, according to the second embodiment, training images having representative features and special features are displayed, and by confirming those images, the user can select an ML dictionary to be applied having understood the overall dictionary characteristics in advance.

Third Embodiment

[0090] In the first and second embodiments, the training image used to express the characteristics of the ML dictionary applied in the object detection unit **162** (i.e., the training image associated with the ML dictionary) was selected on the basis of the feature vector of the training image. In a third embodiment, the training image used to express the characteristics of the ML dictionary applied in the object detection unit **162** is selected on the basis of a result of the object categorization unit **163** categorizing the plurality of training images used in the machine learning for the ML dictionary into a plurality of classes. The overall processing by the dictionary switching unit **164** according to the third embodiment is the same as in the first embodiment and the second embodiment (FIG. **5**), with the exception of the association information generation processing in S**503** and the method for displaying the dictionary characteristics in S**504**.

(Association Information Generation Processing (S**503**))

[0091] FIG. **10** is a flowchart illustrating association information generation processing according to the third embodiment. First, in S**1000**, the generation unit **203** selects an ML dictionary of interest from a plurality of ML dictionaries applied in the object categorization unit **163** and recorded in the recording medium **157** ("corresponding ML dictionary group" hereinafter), and applies the selected ML dictionary in the object categorization unit **163**. Note that the corresponding ML dictionary group may be selected in accordance with the object to which the ML dictionary applied in the object detection unit **162** (the ML dictionary instructed to be switched to in S**501**) corresponds. For example, if the ML dictionary applied in the object detection unit **162** is a dictionary for detecting "horses", the corresponding ML dictionary group may be constituted by ML dictionaries pertaining to the category of "horses." Next, in S**1001**, the generation unit **203** selects a training image of interest from a plurality of training images recorded in the recording medium **157** ("corresponding image group" hereinafter), in correspondence with the ML dictionary instructed to be switched to in S**501** (the ML dictionary applied in the object detection unit **162**). Although it is desirable that the corresponding image group include all the training images used in the machine learning for the ML dictionary here, training images extracted randomly may be used, to the extent that the effects of the present disclosure are not affected. In S**1002**, the generation unit **203** causes the object categorization unit **163** to process the training image of interest and perform object categorization in which the ML dictionary of interest is applied. In S**1003**, the generation unit **203** determines whether a training image that has not been selected as the training image of interest in S**1001** is present in the corresponding image group. If the generation unit **203** determines that a training image that has not been selected as the training image of interest is present in the corresponding image group (YES in S**1003**), the sequence returns to S**1001**, and the foregoing processing is repeated. However, if the generation unit **203** determines that a training image that has not been selected as the training image of interest is not present in the corresponding image group (NO in S**1003**), the sequence moves to S**1004**.

[0092] In S**1004**, the generation unit **203** determines whether an ML dictionary that has not been selected as the ML dictionary of interest in S**1000** is present in the corresponding ML dictionary group. If the generation unit **203** determines that an ML dictionary that has not been selected as the ML dictionary of interest is present in the corresponding ML dictionary group (YES in S**1004**), the sequence returns to S**1000**, and the foregoing processing is repeated. However, if the generation unit **203** determines that an ML dictionary that has not been selected as the ML dictionary of interest is not present in the corresponding ML dictionary group (NO in S**1004**), the sequence moves to S**1005**. In S**1005**, the generation unit **203** selects a target image for expressing the characteristics of the ML dictionary from the training images and the icon images recorded in the recording medium **157**, on the basis of the object categorization results obtained in S**1000** to S**1004**. The selection of the target image will be described later with reference to the flowchart in FIG. **12**. In S**1006**, the generation unit **203** records information specifying the target image selected in S**1005** into the recording medium **157** as the association information of the ML dictionary to be displayed. Here, the "association information" is information on a combination of the target image selected in S**1005** in correspondence with the ML dictionary to be displayed for which the switching operation was performed in S**501**, and the target image itself.

(Example of Object Categorization Result)

[0093] FIG. **11** illustrates an example of an object categorization result obtained through the processing of S**1000** to S**1004** of the present embodiment. In the example in FIG. **11**, five object categories, pertaining to "species", "hair color", "posture", "mask", and "saddle", are made, and the meaning and number of training images for each class into which the training image group has been categorized are described. Classes of "species" are, for example, "horse", "zebra", "donkey", "other", and the like. Classes of "hair color" are "bay hair", "black hair", "chestnut hair", "other", and the like. Classes of "posture" are "standing on four legs", "standing on two legs", "prone", "sleeping", and the like. Classes of "mask" and "saddle" are "yes" and "no".

(Target Image Selection Processing)

[0094] FIG. **12** is a flowchart illustrating processing for selecting the target image performed in the third embodiment.

[0095] In S**1200**, the generation unit **203** selects a category of interest from the object categorization results obtained in S**1000** to S**1004**. Here, "category" is a category obtained by the ML dictionary of interest selected in S**1000**, and in the example in FIG. **11**, is either "species", "hair color", "posture", "mask", or "saddle". In S**1201**, the generation unit **203** selects a class of interest from the classes included in the category of interest. For example, in the example in FIG. **11**, if the category of interest is "species", the class of interest is determined from "horse", "zebra", "donkey", or "other".

[0096] In S**1202**, the generation unit **203** determines whether a class of interest meets a condition for being subject to association processing. The "condition" here is that the percentage of training images belonging to the class of interest of all the training images is at least a predetermined value. For example, if the predetermined value is 5%, in the example in FIG. **11**, the classes of "zebra", "donkey", and "other" in the category of "species" are determined not to satisfy the condition. In CNN-based machine learning, even if an image is included in the training images, the

image may not have a sufficient impact on the performance of the CNN. Accordingly, the determination in S**1202** is made with the intention of ensuring that training images in a low-impact class are not selected as the target image. If the generation unit **203** determines that the class of interest meets the condition for being subject to the association processing (YES in S**1202**), the sequence moves to S**1203**. However, if the generation unit **203** determines that the class of interest does not meet the condition for being subject to the association processing (NO in S**1202**), the sequence moves to S**1205**.

[0097] In S**1203**, the generation unit **203** determines a training image, in the corresponding image group, that has been categorized into a class of interest as the target image. The target image may be any training image that has been categorized into a class of interest. For example, in FIG. **11**, if the category of interest is "species" and the class of interest is "horse", any of the 49,000 training images categorized into that class may be used as the target image. In S**1204**, the generation unit **203** additionally selects an icon image, from the icon images recorded in the recording medium **157**, that corresponds to the class of interest selected in S**1201** as a further target image. Additionally, if an icon image corresponding to a plurality of classes in the category of interest is selected as the target image, an icon image that covers that plurality of classes may be selected. For example, in the example in FIG. **11**, if the category of interest is "hair color", all the classes of "bay hair", "black hair", "chestnut hair", and "other color hair" meet the conditions for being subject to association processing. Accordingly, an icon image that represents a class of all the colors of hair may be selected instead of icon images corresponding to the individual classes.

[0098] In S**1205**, the generation unit **203** determines whether a class that has not been selected as a class of interest in the category determined as the category of interest in S**1200** is present. If the generation unit **203** determines that a class that has not been selected as a class of interest is present (YES in S**1205**), the sequence returns to S**1201**, and the foregoing processing is repeated. However, if the generation unit **203** determines that a class that has not been selected as the class of interest is not present (NO in S**1205**), the sequence moves to S**1206**. In S**1206**, the generation unit **203** determines whether a category that has not been selected as the category of interest is present in the object categorization results obtained in S**1000** to S**1004**. If the generation unit **203** determines that a category that has not been selected as the category of interest is present (YES in S**1206**), the sequence returns to S**1200**, and the foregoing processing is repeated. However, if the generation unit **203** determines that a category that has not been selected as the category of interest is not present (NO in S**1206**), the sequence ends.

[0099] As described above, according to the third embodiment, the training images and icon images corresponding to all categories and all classes that meet the condition for being subject to the association processing are selected as target images. Because icon images can express many dictionary characteristics within the narrow display range of the display **150**, it is desirable that all icon images that meet the conditions for the association processing be selected as target images. However, when the display range is limited, the determination in S**1206** may be changed to a method that determines whether an ending condition is met, such as the determination made in S**605** in the first and second embodiments. In this case, it is desirable to prioritize categories and classes that meet predetermined criteria in the selection of the target image. For example, in the processing of selecting the first target image, it is desirable to prioritize a category having a large number of classes that meet the condition in S**1202**, and furthermore, to prioritize a class, among the classes belonging to that category, having a large number of categorized training images. In the processing of selecting the second and subsequent target images, it is desirable to prioritize a category having a small number of classes that meet the condition for processing in S**1202**, and furthermore, to prioritize a class, among the classes belonging to that category, having a small number of categorized training images. According to such criteria, representative images (the training image and the icon image) are easily selected in the selection of the first target image, and special images (the training image and the icon image) are easily selected in the selection of the second or subsequent target image.

(Example of Display of Dictionary Characteristics)

[0100] The dictionary characteristics can be displayed using the training image determined as the target image in S**1203** described above. An example of the display in this case is the same as in the first embodiment (FIGS. **8**A and **8**B). On the other hand, in the third embodiment, the dictionary characteristics can be displayed using the icon image determined as the target image in S**1204**. FIGS. **13**A and **13**B are diagrams illustrating an example of the display of the switching screen displayed on the display **150** by the switching control unit **201** in S**504** according to the third embodiment. FIGS. **13**A and **13**B illustrate an example of the display expressing the dictionary characteristics using icon images, according to the third embodiment. It should be noted that the display expressing the dictionary characteristics using training images, illustrated in FIGS. **8**A and **8**B, and the display expressing the dictionary characteristics using icon images, illustrated in FIGS. **13**A and **13**B, may be switched and displayed as desired by the user.

[0101] FIG. **13**A will be described first. In a switching screen **13***a*, the items **800** and **801** and the switching button **808** are the same as those in the first embodiment (FIGS. **8**A and **8**B). An item **1002***a* is the ML dictionary to be displayed, and is indicated by "001", which is the filename of that ML dictionary recorded in the recording medium **157**. An area **1003***a* is an area expressing the dictionary characteristics, and a plurality of icon images expressing the dictionary characteristics, represented by an icon image **1004***a*, are displayed therein. The icon image **1004***a* is an example of the icon image selected as the target image in S**1204**. In the target image selection processing according to the third embodiment, the icon images corresponding to all categories and all classes are determined, without any limitations on the number of images, and thus the characteristics of the dictionary can therefore be expressed more comprehensively. FIG. **13**B is an example of a switching screen **13***b* that displays dictionary characteristics for an ML dictionary that is different from the switching screen **13***a*, according to the third embodiment. An icon image **1004***b* indicating "all postures" in the area **1003***a* is an icon image that collectively represents all the classes of "standing on four legs", "standing on two legs", "prone", and "sleeping" in FIG. **11**. Note that the icon image may be an image that enables the user to understand the class, such as a character string indicating the

class, a graphic or a photograph indicating the class, or the like, as illustrated in FIGS. **13**A and **13**B.

[0102] As described above, according to the third embodiment, a plurality of icon images are displayed, and by confirming those images, the user can select the ML dictionary to be applied having understood, in advance, the overall dictionary characteristics.

Fourth Embodiment

(Overall Configuration)

[0103] The first embodiment to the third embodiment described a configuration in which the characteristic representation of the ML dictionary is obtained by an image processing apparatus (the object detection unit **162**, the object categorization unit **163**, the dictionary switching unit **164**, and the like) within the image capturing apparatus **100**. The fourth embodiment will describe a configuration in which the characteristic representation of the ML dictionary is obtained by an external image processing apparatus (e.g., a cloud system) and provided to the image capturing apparatus. FIG. **14** is a diagram illustrating an example of the overall configuration of an image processing system **1400** according to the fourth embodiment. The image processing system **1400** includes an image capturing apparatus **1401**, an image capturing apparatus **1402**, a cloud system **1403**, and a network **1404**. The image capturing apparatus **1401**, the image capturing apparatus **1402**, and the cloud system **1403** are connected to each other in a communication-enabling manner over the network **1404**. The specific configurations of the image capturing apparatus **1401** and the image capturing apparatus **1402** are the same as that of the image capturing apparatus **100** according to the first embodiment. However, the dictionary switching unit **164** need not have a function for obtaining the target image for expressing the characteristics of the ML dictionary. The cloud system **1403** is an example of an image processing apparatus capable of communicating with the image capturing apparatus **1401** and the image capturing apparatus **1402**. Furthermore, the image processing apparatus serving as the apparatus external to the image capturing apparatus is not limited to the cloud system **1403**, and may be implemented by a server apparatus on a LAN, for example.

[0104] The image capturing apparatus **1401** is configured in the same manner as the image capturing apparatus **100**, is used by a user to shoot training images, and performs machine learning for ML dictionaries applicable in the object detection unit **162**. The ML dictionary obtained through machine learning and the plurality of training images used in the machine learning ("corresponding image group" hereinafter) are uploaded to the cloud system **1403** over the network **1404**. Although it is desirable that the corresponding image group include all the training images used in the machine learning, training images extracted randomly may be used, to the extent that the effects achieved by the embodiment of the present disclosure are not affected. The image capturing apparatus **1402** has the same configuration as the image capturing apparatus **100**, and can utilize an environment provided by the cloud system **1403**. For example, the image capturing apparatus **1402** downloads the ML dictionary from the cloud system **1403**, applies the dictionary to the object detection unit **162**, and uses the object detection function for the captured image. The cloud system **1403** records a plurality of ML dictionaries, and a plurality of training images for each ML dictionary, uploaded by the user. The cloud system **1403** also provides an environment that expresses the performance of the ML dictionary, and in which the user can download a plurality of ML dictionaries. For the sake of convenience, FIG. **14** illustrates one each of the image capturing apparatus **1401** which uploads and the image capturing apparatus **1402** which downloads, but the configuration is not limited thereto. For example, a plurality of ML dictionaries from a plurality of image capturing apparatuses may be uploaded to the cloud system **1403**.

(Configuration of Cloud System)

[0105] FIG. **15** is a block diagram illustrating an example of the configuration of the cloud system **1403** according to the fourth embodiment. The various functional units in the cloud system **1403** are connected to each other over a bus **1500**. A control unit **1501** controls the various functional units. A recording unit **1502** is a high-capacity recording medium such as an HDD, and the plurality of ML dictionaries uploaded from the image capturing apparatus **1401** and the plurality of ML dictionaries applicable in an object categorization unit **1506** are recorded therein. Training images used in respective instances of machine learning, association information, and the like are also recorded in the plurality of ML dictionaries uploaded from the image capturing apparatus **1401**. Icon images corresponding to each category and each class in the object categorization results are also recorded in the ML dictionary applied in the object categorization unit **1506**. A communication unit **1503** communicates ML dictionaries, information related to ML dictionaries such as training images, and the like by connecting to the image capturing apparatus **1401** and the image capturing apparatus **1402** over Ethernet or wirelessly.

[0106] A display image generation unit **1504** generates display images, and provides a GUI used for uploading and downloading ML dictionaries to the users of the image capturing apparatus **1401** and the image capturing apparatus **1401** via the communication unit **1503**. An object detection unit **1505** determines a region in which an object, such as a horse, is present in an image, by applying an ML dictionary recorded in the recording unit **1502**. The object detection unit **1505** is constituted by the same CNN as the object detection unit **162** in the image capturing apparatus **1401** and the image capturing apparatus **1402**. The object categorization unit **1506** categorizes to which of predetermined classes an object in an image belongs, by applying an ML dictionary recorded in the recording unit **1502**. The object categorization unit **1506** is constituted by the same CNN as the object categorization unit **163** in the image capturing apparatus **1401** and the image capturing apparatus **1402**.

(Flow of Overall Processing)

[0107] FIGS. **16**A to **16**C are flowcharts illustrating overall processing performed by the cloud system **1403** according to the fourth embodiment. Processing performed by the image capturing apparatus **1401**, the cloud system **1403**, and the image capturing apparatus **1402** will be described with reference to FIGS. **16**A, **16**B, and **16**C, respectively.

[0108] FIG. **16**A is a flowchart illustrating the overall processing performed by the image capturing apparatus **1401** according to the fourth embodiment. In S**1600**, the switching control unit **201** determines whether an operation

for uploading an ML dictionary has been made by the user through the operation switch **156**. If the switching control unit **201** determines that an upload operation has been made (YES in S**1600**), the sequence moves to S**1601**, whereas if the switching control unit **201** determines that an upload operation has not been made (NO in S**1600**), the sequence moves to S**1603**.

[0109] In S**1601**, the switching control unit **201** makes an upload request to the cloud system **1403**, which is an external apparatus, through the communication unit **161**. Next, in S**1602**, the switching control unit **201** sends the ML dictionary recorded in the recording medium **157** and the plurality of training images used in the training of the ML dictionary to the cloud system **1403** through the communication unit **161**. In S**1603**, the switching control unit **201** determines whether an ending instruction has been made by the user through the operation switch **156**. If the switching control unit **201** determines that no ending instruction has been made (NO in S**1603**), the sequence returns to S**1600**, and the foregoing processing is repeated. However, if the switching control unit **201** determines that an ending instruction has been made (YES in S**1603**), the sequence ends.

[0110] FIG. **16**B is a flowchart illustrating overall processing performed by the cloud system **1403** according to the fourth embodiment. First, in S**1610**, the control unit **1501** determines whether an ML dictionary upload request has been made from an external apparatus (the image capturing apparatus **1401**, in this example) through the communication unit **1503**. If the control unit **1501** determines that an upload request has been made (YES in S**1610**), the sequence moves to S**1611**. However, if the control unit **1501** determines that no upload request has been made (NO in S**1610**), the sequence moves to S**1613**.

[0111] In S**1611**, the control unit **1501** receives, through the communication unit **1503**, the ML dictionary and the plurality of training images used to train the ML dictionary from the image capturing apparatus **1401** that made the upload request, and records those items in the recording unit **1502**. In S**1612**, the control unit **1501** reads out the ML dictionary and the plurality of training images used to train the ML dictionary, recorded in the recording unit **1502** in S**1611**, and generates the association information. The details of the processing for generating the association information in S**1612** are similar to those in the first embodiment or the second embodiment (S**503**), with the exception that the processing is executed by the corresponding units in the cloud system **1403**.

[0112] In S**1613**, the control unit **1501** determines whether a request to display the ML dictionary characteristics has been made from an external apparatus (the image capturing apparatus **1402**, in this example) through the communication unit **1503**. If the control unit **1501** determines that the display request has been made (YES in S**1613**), the sequence moves to S**1614**, whereas if the control unit **1501** determines that no display request has been made (NO in S**1613**), the sequence returns to S**1610**. In S**1614**, the display image generation unit **1504** generates a characteristic representation image expressing the characteristics of the ML dictionary, on the basis of the association information generated in S**1612**, the ML dictionary, and the plurality of training images corresponding to the ML dictionary. The display image generation unit **1504** then sends the generated characteristic representation image to the image capturing apparatus **1402** from which the display request was made, through the communication unit **1503**. Note that the details of the processing performed in S**1614** are similar to those in the first embodiment or the second embodiment (S**504**), with the exception that the processing is executed by the corresponding units in the cloud system **1403**.

[0113] In S**1615**, the control unit **1501** determines whether a request to download the ML dictionary has been received from an external apparatus (the image capturing apparatus **1402**, in this example) through the communication unit **161**. If the control unit **1501** determines that a download request has been received (YES in S**1615**), the sequence moves to S**1616**, whereas if the control unit **1501** determines that a download request has not been received (NO in S**1615**), the sequence returns to S**1610**. In S**1616**, the control unit **1501** sends the ML dictionary, among the plurality of ML dictionaries recorded in the recording unit **1502**, the ML dictionary expressing the dictionary characteristics in S**1614**, and the association information generated in S**1612** for that ML dictionary, to the image capturing apparatus **1402** that made the download request, through the communication unit **1503**.

[0114] FIG. **16**C is a flowchart illustrating the overall processing performed by the image capturing apparatus **1402** according to the fourth embodiment. In S**1620**, whether an operation for displaying the dictionary characteristics has been made by the user through the operation switch **156** is determined. If the switching control unit **201** determines that the display operation has been made (YES in S**1620**), the sequence moves to S**1621**, whereas if the switching control unit **201** determines that no display operation has been made (NO in S**1620**), the sequence moves to S**1623**.

[0115] In S**1621**, the switching control unit **201** sends an ML dictionary characteristic display request to the cloud system **1403**, which is an external apparatus, through the communication unit **161**. In S**1622**, the switching control unit **201** receives a characteristic representation image of the ML dictionary from the cloud system **1403** via the communication unit **161**, and displays the characteristic representation image on the display **150**. The characteristic representation image is the same as the example of the display of the switching screen illustrated in the first embodiment (FIGS. **8**A and **8**B), for example.

[0116] In S**1623**, the switching control unit **201** determines whether a download operation for the ML dictionary for which the characteristics have been expressed in the characteristic representation image has been made by the user through the operation switch **156**. If the switching control unit **201** determines that a download operation has been made (YES in S**1623**), the sequence moves to S**1624**, whereas if the switching control unit **201** determines that no download operation has been made (NO in S**1623**), the sequence moves to S**1626**. In S**1624**, the switching control unit **201** sends an ML dictionary download request to the cloud system **1403**, which is an external apparatus, through the communication unit **161**. In S**1625**, through the communication unit **161**, the switching control unit **201** receives the ML dictionary and the association information (e.g., the target image) generated for the ML dictionary from the cloud system **1403**, and records the association information (e.g., the target image) in the recording medium **157**.

[0117] In S**1626**, the switching control unit **201** determines whether an operation for switching the ML dictionary, recorded in the recording medium **157** in S**1625**, to be

applied in the object detection unit **162** has been made by the user through the operation switch **156**. If the switching control unit **201** determines that the switching operation has been made (YES in S**1626**), the sequence moves to S**1627**, whereas if the switching control unit **201** determines that the switching operation has not been made (NO in S**1626**), the sequence moves to S**1628**. In S**1627**, the switching control unit **201** switches the ML dictionary applied in the object detection unit **162** on the basis of the switching operation made by the user. Thereafter, object detection processing using the ML dictionary switched to in S**1627** can be performed on through-the-lens images, captured images, and the like in the image capturing apparatus **1402**. In S**1628**, the switching control unit **201** determines whether an ending instruction has been made by the user through the operation switch **156**. If the switching control unit **201** determines that an ending instruction has been made (YES in S**1628**), the sequence ends. However, if the switching control unit **201** determines that no ending instruction has been made (NO in S**1628**), the sequence returns to S**1620**, and the foregoing processing is repeated.

[0118] As described above, according to the fourth embodiment, an image capturing apparatus can obtain, from a cloud system, training images having representative features and special features as information expressing the characteristics of an ML dictionary to be displayed, and can display the training images. Accordingly, by confirming those images, the user of the image capturing apparatus can download and apply the ML dictionary having understood, in advance, the overall dictionary characteristics, without placing a load on the image capturing apparatus.

[0119] Although the configuration according to the first embodiment or the second embodiment is employed as the configuration for the association information generation processing in S**1612** and the obtainment of the characteristic representation image in S**1614** (S**1622**), the configuration is not limited thereto. The configuration of the third embodiment may be employed as the configuration for the association information generation processing in S**1612** and the obtainment of the characteristic representation image in S**1614** (S**1622**). In this case, the processing for generating the association information in S**1612** is similar to that of the third embodiment (FIGS. **10** and **12**), with the exception that the processing is executed by the corresponding units in the cloud system **1403**. The characteristic representation image of the ML dictionary generated by the cloud system **1403** in S**1614**, received through the communication unit **161** of the image capturing apparatus **1402** in S**1622**, and displayed is the same as the example described with reference to FIGS. **13**A and **13**B in the third embodiment.

[0120] As described above, according to the fourth embodiment, icon images are displayed by the cloud system, and by confirming those images, the user can download and apply the ML dictionary having understood, in advance, the overall dictionary characteristics, without placing a load on the image capturing apparatus.

[0121] According to one aspect of the present disclosure, information enabling a user to understand the characteristics of a dictionary obtained through machine learning can be generated.

OTHER EMBODIMENTS

[0122] Embodiment(s) of the present disclosure can also be realized by a computer of a system or apparatus that reads out and executes computer executable instructions (e.g., one or more programs) recorded on a storage medium (which may also be referred to more fully as a 'non-transitory computer-readable storage medium') to perform the functions of one or more of the above-described embodiment(s) and/or that includes one or more circuits (e.g., application specific integrated circuit (ASIC)) for performing the functions of one or more of the above-described embodiment(s), and by a method performed by the computer of the system or apparatus by, for example, reading out and executing the computer executable instructions from the storage medium to perform the functions of one or more of the above-described embodiment(s) and/or controlling the one or more circuits to perform the functions of one or more of the above-described embodiment(s). The computer may comprise one or more processors (e.g., central processing unit (CPU), micro processing unit (MPU)) and may include a network of separate computers or separate processors to read out and execute the computer executable instructions. The computer executable instructions may be provided to the computer, for example, from a network or the storage medium. The storage medium may include, for example, one or more of a hard disk, a random-access memory (RAM), a read only memory (ROM), a storage of distributed computing systems, an optical disk (such as a compact disc (CD), digital versatile disc (DVD), or Blu-ray Disc (BD)™), a flash memory device, a memory card, and the like.

[0123] While the present disclosure has been described with reference to exemplary embodiments, it is to be understood that the present disclosure is not limited to the disclosed exemplary embodiments. The scope of the following claims is to be accorded the broadest interpretation so as to encompass all such modifications and equivalent structures and functions.

1. An image processing apparatus comprising:

an obtaining unit configured to obtain a dictionary obtained through machine learning and a plurality of training images used in the machine learning of the dictionary;

a selecting unit configured to select from the plurality of training images, at least one training image to be used as information expressing a characteristic of the dictionary; and

a generating unit configured to generate association information that associates the at least one training image selected by the selecting unit with the dictionary.

2. The image processing apparatus according to claim **1**, wherein the selecting unit selects the at least one training image on the basis of a plurality of feature vectors obtained from the plurality of training images.

3. The image processing apparatus according to claim **2**, further comprising:

a detecting unit configured to detect an object from an image using one of a plurality of dictionaries obtained through machine learning,

wherein as the plurality of feature vectors, the selecting unit obtains intermediate data obtained from the detecting unit as a result of inputting the plurality of training images into the detecting unit.

4. The image processing apparatus according to claim **2**, wherein the selecting unit selects a training image having a feature vector, among the plurality of feature vectors, having a smallest distance from an average vector that is an average of the plurality of feature vectors.

5. The image processing apparatus according to claim 2, wherein the selecting unit selects a predetermined number of feature vectors from the plurality of feature vectors in order from a feature vector having a greatest distance from an average vector that is an average of the plurality of feature vectors, and selects a predetermined number of training images corresponding to the predetermined number of feature vectors.

6. The image processing apparatus according to claim 4, wherein the selecting unit selects a predetermined number of feature vectors from the plurality of feature vectors in order from a feature vector having a greatest difference from the feature vector having the smallest distance from the average vector, and selects a predetermined number of training images corresponding to the predetermined number of feature vectors.

7. The image processing apparatus according to claim 3, wherein the selecting unit selects the at least one training image from training images satisfying a condition that the detecting unit is capable of detecting an object from the training image using the dictionary.

8. The image processing apparatus according to claim 2, wherein the selecting unit selects the at least one training image from training images satisfying a condition that other training images having a feature vector for which a distance from an obtained feature vector is smaller than a predetermined value occupy at least a predetermined percentage of the plurality of training images.

9. The image processing apparatus according to claim 1, further comprising:

a display information generating unit configured to generate display information for displaying the dictionary and the at least one training image in association with each other on the basis of the association information.

10. The image processing apparatus according to claim 1, further comprising:

a categorizing unit configured to categorize the plurality of training images into a plurality of classes,

wherein the selecting unit selects the training image on the basis of a categorizing result obtained as a result of inputting the plurality of training images into the categorizing unit.

11. The image processing apparatus according to claim 10, wherein the selecting unit selects a training image belonging to a class, among the plurality of classes, into which a greatest number of training images have been categorized in the categorizing result.

12. The image processing apparatus according to claim 10, wherein the selecting unit selects a predetermined number of training images from a predetermined number of classes, among the plurality of classes, selected in order from a class having a smallest number of training images categorized in the categorizing result.

13. The image processing apparatus according to claim 10, wherein the selecting unit selects the at least one training image from training images belonging to a class in which a ratio of a total number of categorized training images to a total number of the plurality of training images is at least a predetermined value.

14. The image processing apparatus according to claim 10, wherein:

the selecting unit further selects at least one icon image, from a plurality of icon images corresponding to the plurality of classes, that corresponds to a class to which each of the at least one training images belongs, and

the generating unit includes information associating the dictionary with the at least one icon image in the association information.

15. The image processing apparatus according to claim 14, further comprising:

a display information generating unit configured to generate display information for displaying the dictionary and the at least one training image or the at least one icon image in association with each other on the basis of the association information.

16. An image processing apparatus comprising:

an obtaining unit configured to obtain a dictionary obtained through machine learning and a plurality of training images used in the machine learning of the dictionary;

a categorizing unit configured to categorize the plurality of training images into a plurality of classes;

a selecting unit configured to select at least one icon image, from a plurality of icon images corresponding to the plurality of classes, to be used as information expressing a characteristic of the dictionary, on the basis of a result of the categorizing unit categorizing the plurality of training images; and

a generating unit configured to generate association information that associates the at least one icon image selected by the selecting unit with the dictionary.

17. The image processing apparatus according to claim 1, further comprising:

a receiving unit configured to receive the dictionary and the plurality of training images from an external apparatus; and

a sending unit configured to send the dictionary and the association information to the external apparatus.

18. An image capturing apparatus comprising:

an image capturing unit;

the image processing apparatus according to claim 1;

a switching unit configured to switch, in response to a user operation, a dictionary to be applied in detecting unit that detects an object from an image captured by the image capturing unit; and

a display unit configured to display a characteristic of the dictionary switched to by the switching unit on the basis of the association information.

19. An image capturing apparatus comprising:

an image capturing unit;

a communicating unit configured to communicate with the image processing apparatus according to claim 17;

a switching unit configured to switch, in response to a user operation, a dictionary to be applied in detecting unit that detects an object from an image captured by the image capturing unit; and

a display unit configured to display a characteristic of the dictionary switched to by the switching unit on the basis of the association information, the association information being received from the image processing apparatus by the communicating unit.

20. A control method of an image processing apparatus, the control method comprising:

obtaining a dictionary obtained through machine learning and a plurality of training images used in the machine learning of the dictionary;

selecting, from the plurality of training images, at least one training image to be used as information expressing a characteristic of the dictionary; and

generating association information that associates the selected training image(s) with the dictionary.

21. A control method of an image processing apparatus, the control method comprising:

obtaining a dictionary obtained through machine learning and a plurality of training images used in the machine learning of the dictionary;

categorizing the plurality of training images into a plurality of classes;

selecting at least one icon image, from a plurality of icon images corresponding to the plurality of classes, to be used as information expressing a characteristic of the dictionary, on the basis of a result of the plurality of training images being categorized in the categorizing step; and

generating association information that associates the at least one icon image selected in the selecting step with the dictionary.

22. A non-transitory computer-readable storage medium which stores a program for causing a computer to execute the method according to claim **20**.

* * * * *