US 20180260628A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2018/0260628 A1**

NAMIKI et al. (43) **Pub. Date: Sep. 13, 2018**

(54) **APPARATUS AND METHOD FOR IMAGE PROCESSING TO CALCULATE LIKELIHOOD OF IMAGE OF TARGET OBJECT DETECTED FROM INPUT IMAGE**

(71) Applicant: **FANUC CORPORATION**, Yamanashi (JP)

(72) Inventors: **Yuuta NAMIKI**, Yamanashi (JP); **Fumikazu WARASHINA**, Yamanashi (JP)

(21) Appl. No.: **15/899,471**

(22) Filed: **Feb. 20, 2018**

(30) **Foreign Application Priority Data**

Mar. 13, 2017 (JP) 2017-047444

Publication Classification

(51) **Int. Cl.**

G06K 9/00	(2006.01)
G06N 3/08	(2006.01)
G06T 7/73	(2006.01)

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

CPC ***G06K 9/00664*** (2013.01); *B25J 9/1697* (2013.01); ***G06T 7/75*** (2017.01); ***G06N 3/088*** (2013.01)

(57) **ABSTRACT**

An image processing apparatus, which receives an input image and detects an image of a target object based on a detection algorithm, includes a machine learning device which performs learning by using a plurality of partial images cut out from at least one input image, based on a result of detection of the image of the target object, and calculates a likelihood of the image of the target object.

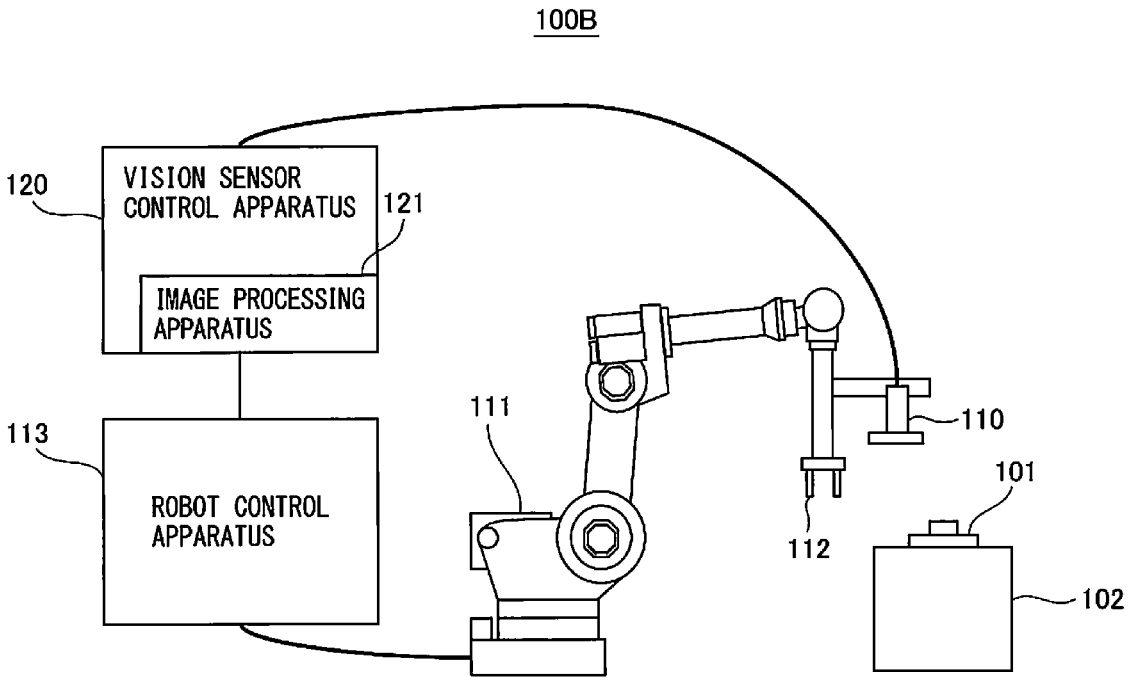


FIG. 1

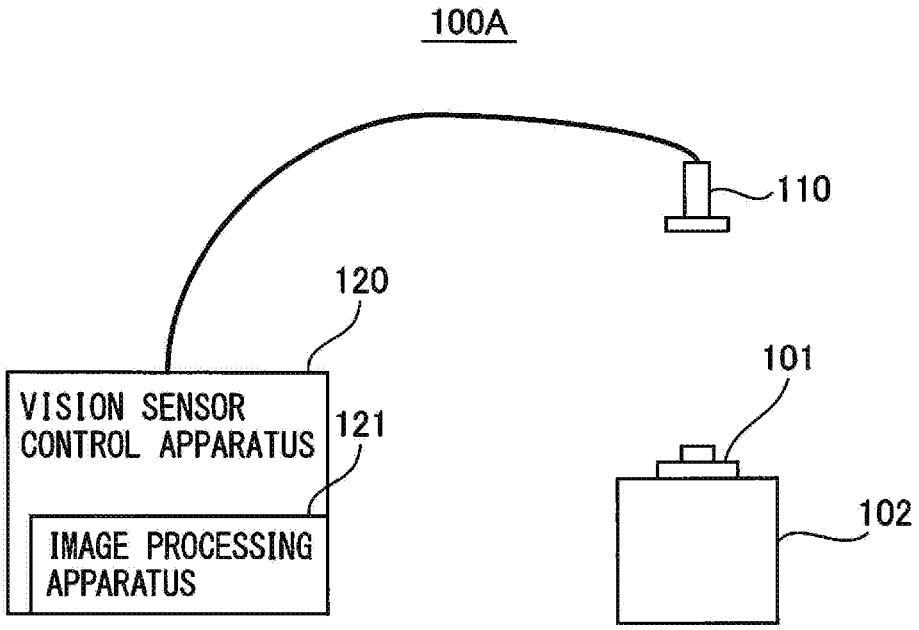

100A
110
120
VISION SENSOR CONTROL APPARATUS
121
IMAGE PROCESSING APPARATUS
101
102

FIG. 2

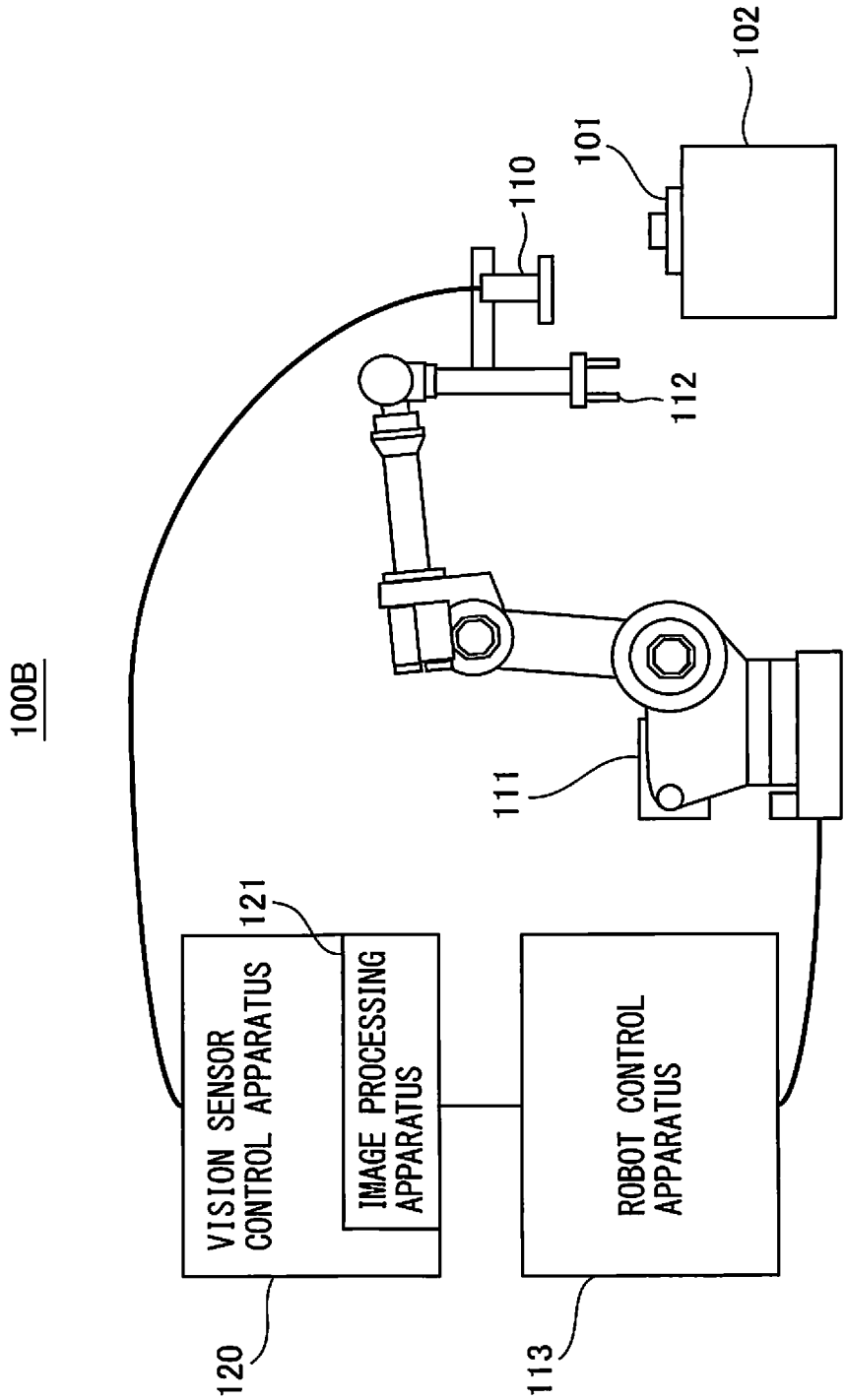

100B
102
101
110
112
111
121
VISION SENSOR CONTROL APPARATUS
IMAGE PROCESSING APPARATUS
ROBOT CONTROL APPARATUS
120
113

FIG. 3

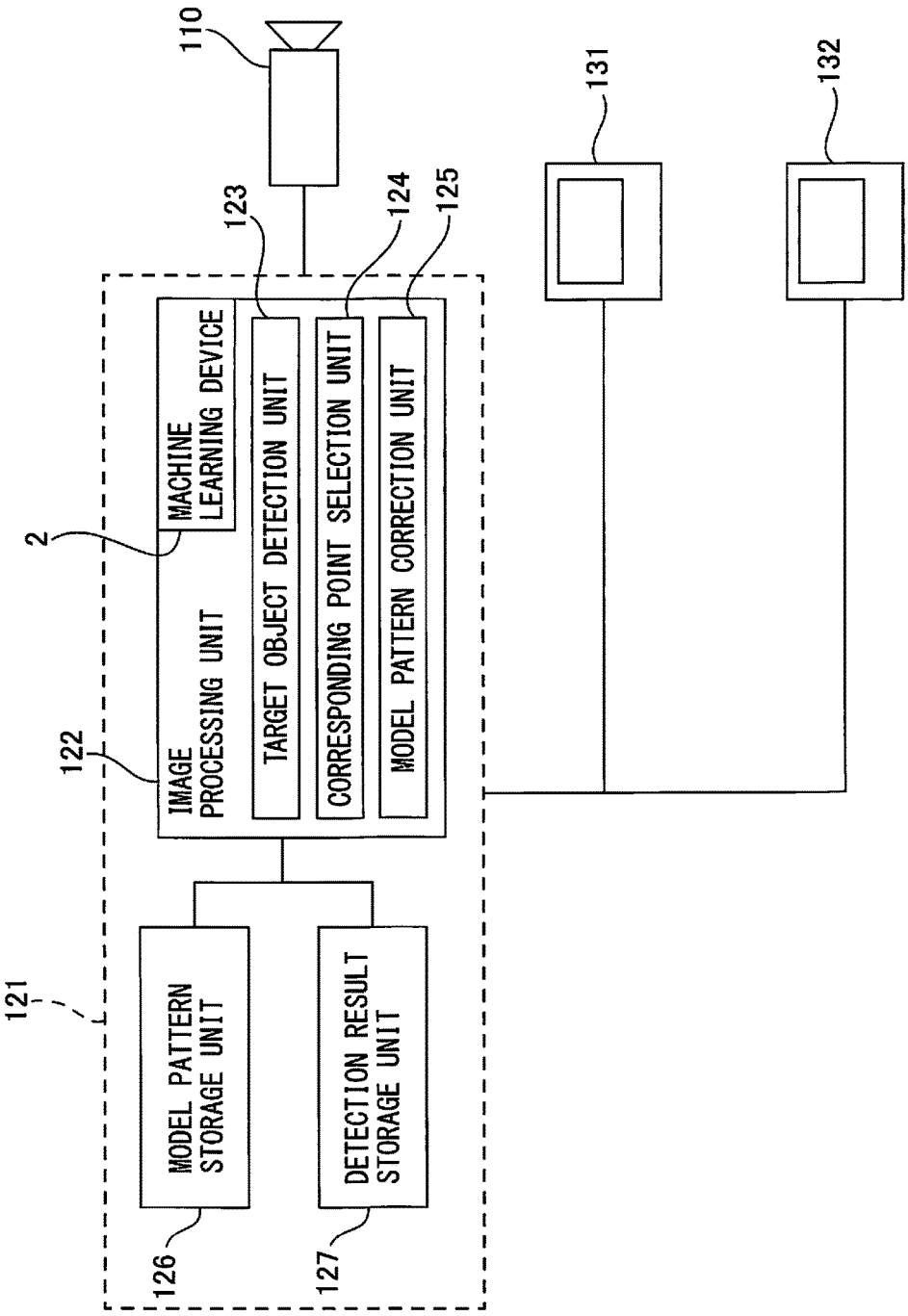

110
131
132
123
124
125
2
MACHINE LEARNING DEVICE
TARGET OBJECT DETECTION UNIT
CORRESPONDING POINT SELECTION UNIT
MODEL PATTERN CORRECTION UNIT
IMAGE PROCESSING UNIT
122
121
MODEL PATTERN STORAGE UNIT
DETECTION RESULT STORAGE UNIT
126
127

FIG. 4

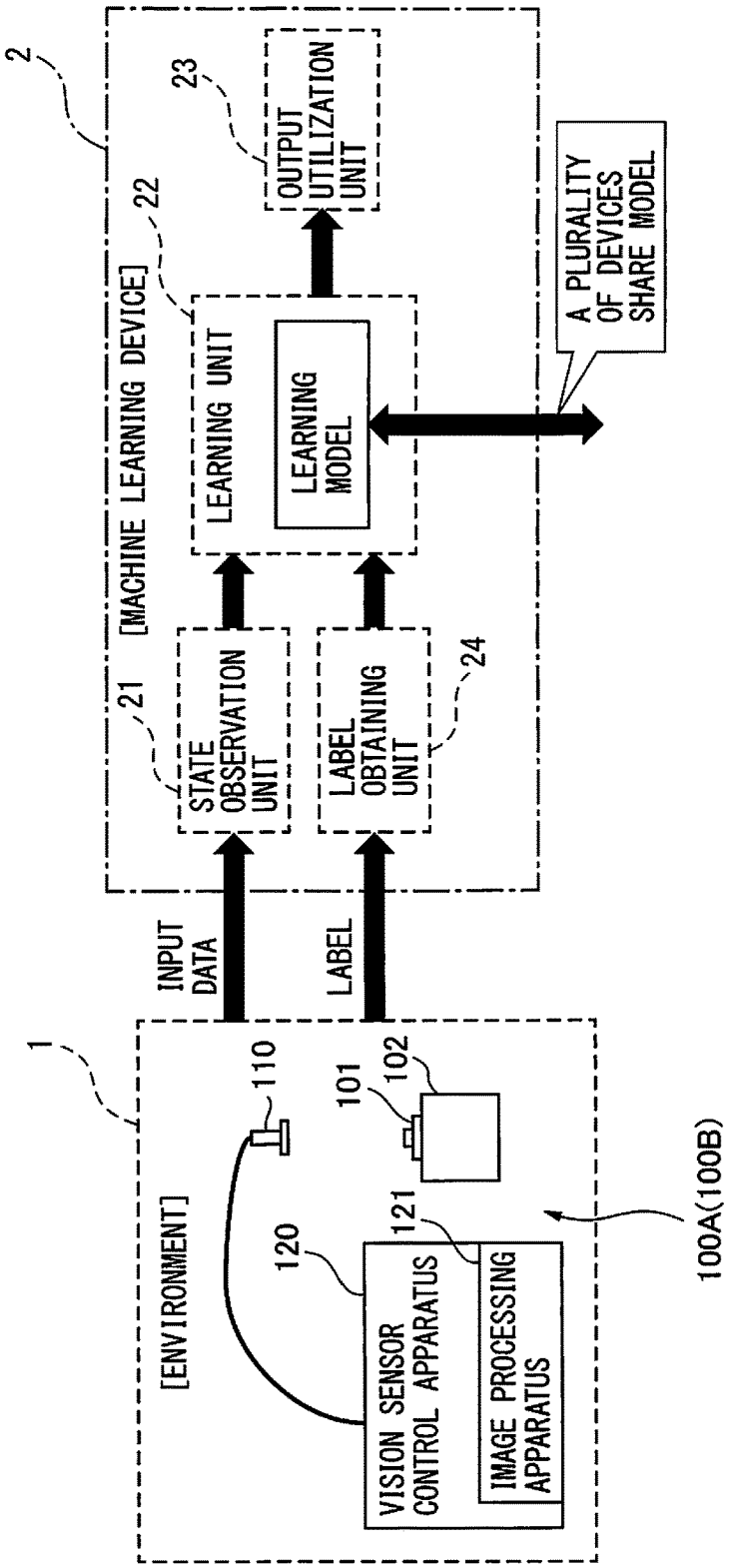

2
[MACHINE LEARNING DEVICE]
21
STATE OBSERVATION UNIT
22
LEARNING UNIT
LEARNING MODEL
23
OUTPUT UTILIZATION UNIT
24
LABEL OBTAINING UNIT
A PLURALITY OF DEVICES SHARE MODEL
INPUT DATA
LABEL
1
[ENVIRONMENT]
110
101
102
120
121
VISION SENSOR CONTROL APPARATUS
IMAGE PROCESSING APPARATUS
100A(100B)

FIG. 5

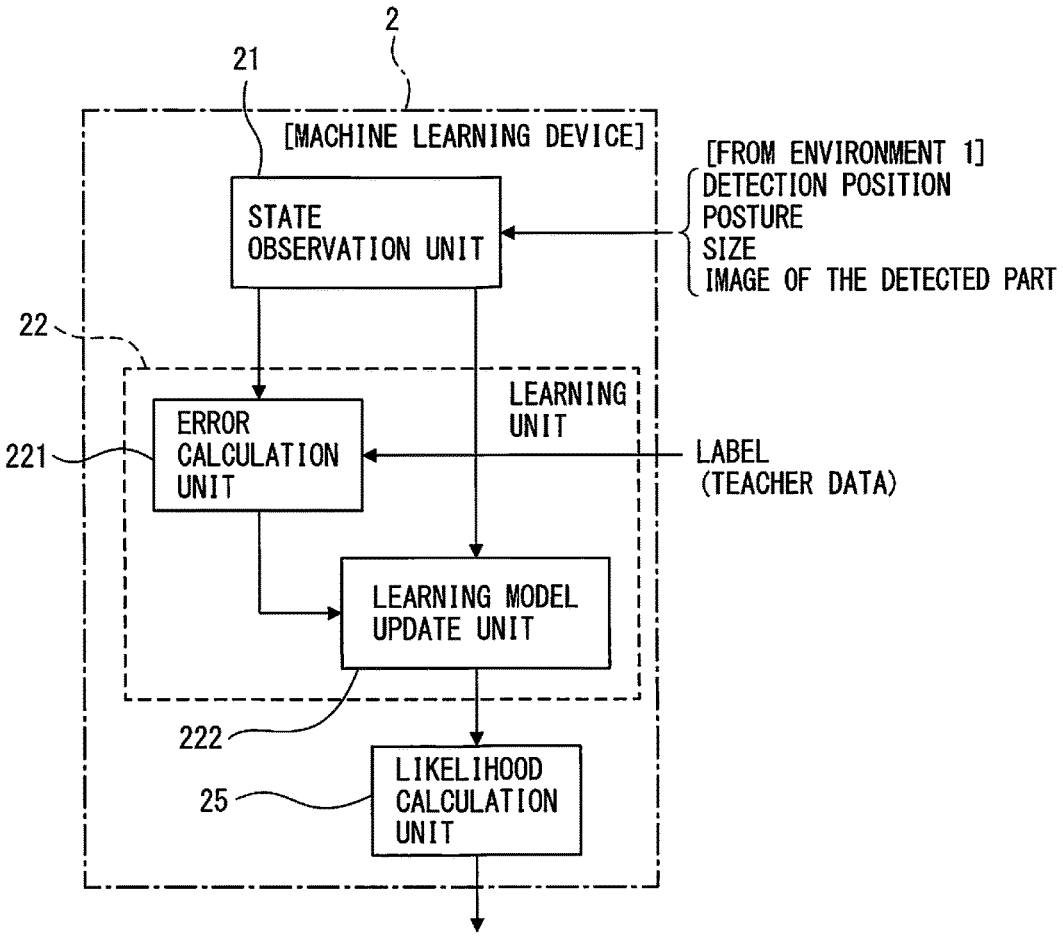

2
21
[MACHINE LEARNING DEVICE]
STATE OBSERVATION UNIT
[FROM ENVIRONMENT 1]
DETECTION POSITION
POSTURE
SIZE
IMAGE OF THE DETECTED PART
22
LEARNING UNIT
ERROR CALCULATION UNIT
221
LABEL (TEACHER DATA)
LEARNING MODEL UPDATE UNIT
222
LIKELIHOOD CALCULATION UNIT
25

FIG. 6

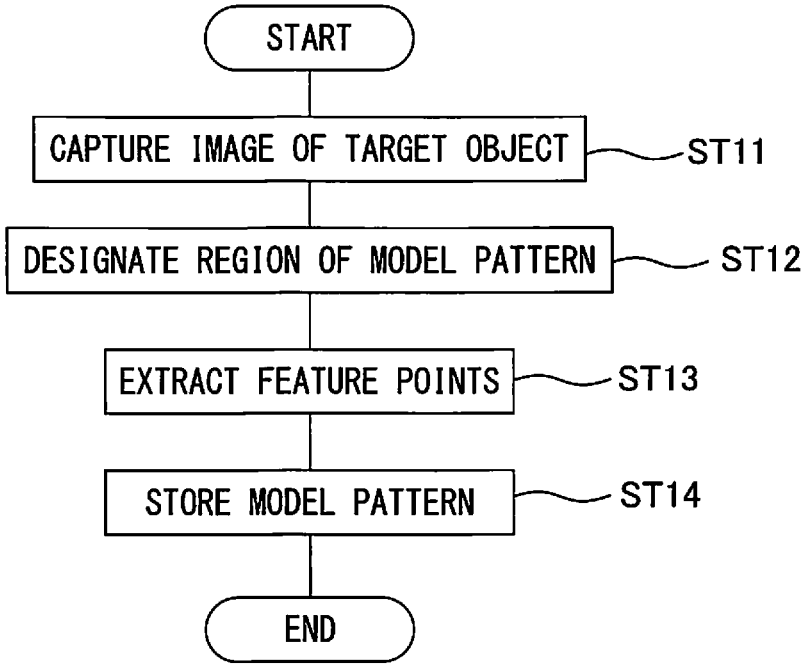

START
CAPTURE IMAGE OF TARGET OBJECT
ST11
DESIGNATE REGION OF MODEL PATTERN
ST12
EXTRACT FEATURE POINTS
ST13
STORE MODEL PATTERN
ST14
END

FIG. 7

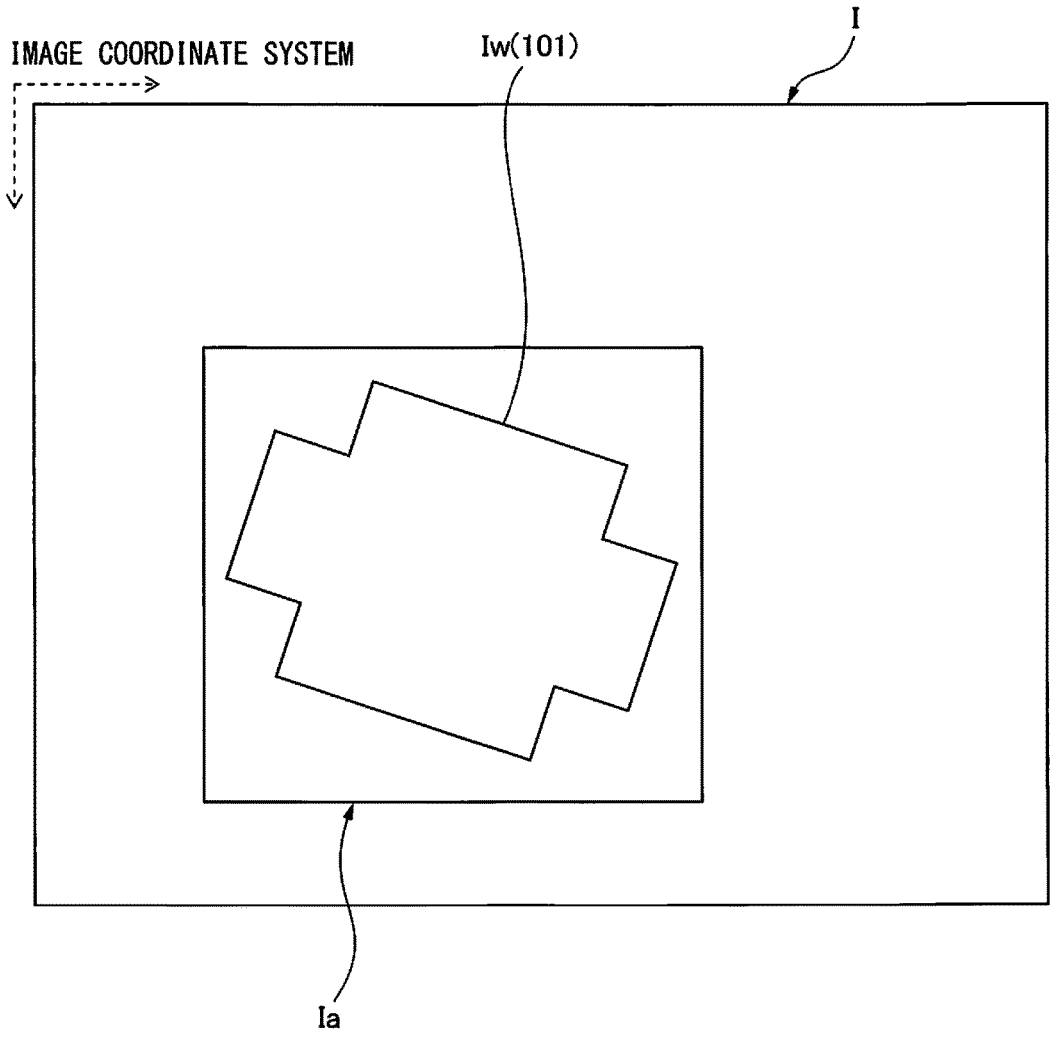


FIG. 8

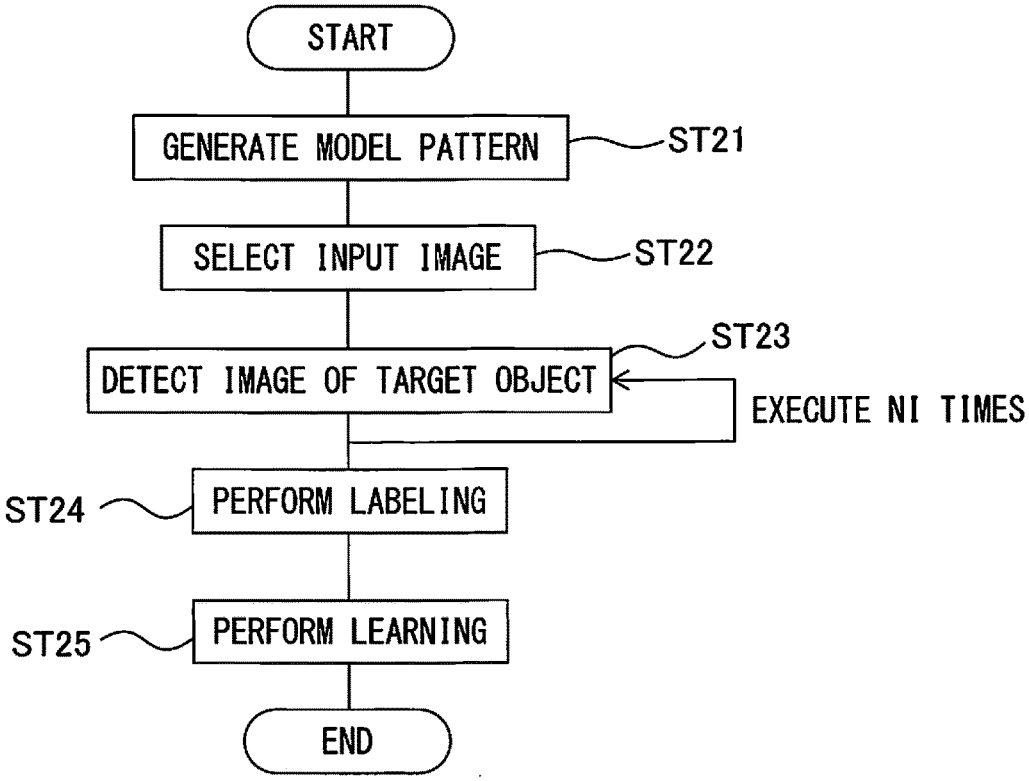

START
GENERATE MODEL PATTERN
ST21
SELECT INPUT IMAGE
ST22
ST23
DETECT IMAGE OF TARGET OBJECT
EXECUTE NI TIMES
ST24
PERFORM LABELING
ST25
PERFORM LEARNING
END

FIG. 9

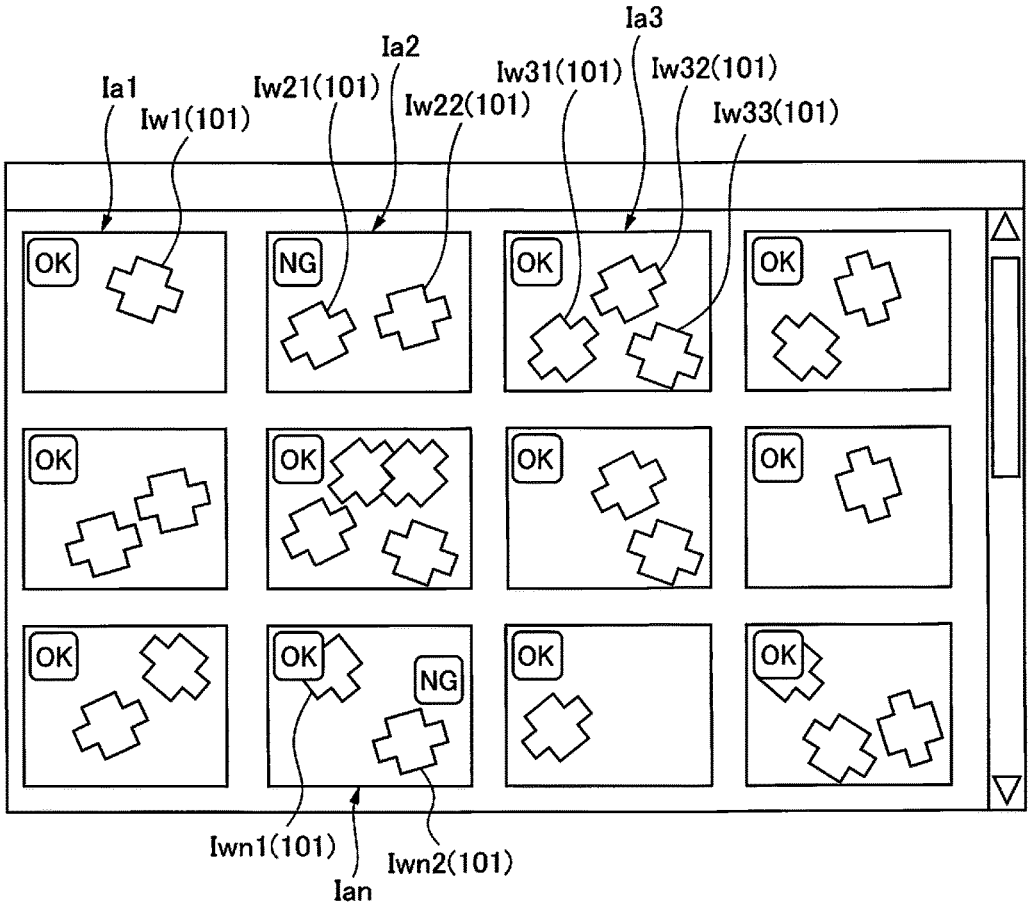

Ia1
Iw1(101)
Ia2
Iw21(101)
Iw22(101)
Ia3
Iw31(101)
Iw32(101)
Iw33(101)
OK
NG
Iwn1(101)
Ian
Iwn2(101)

FIG. 10

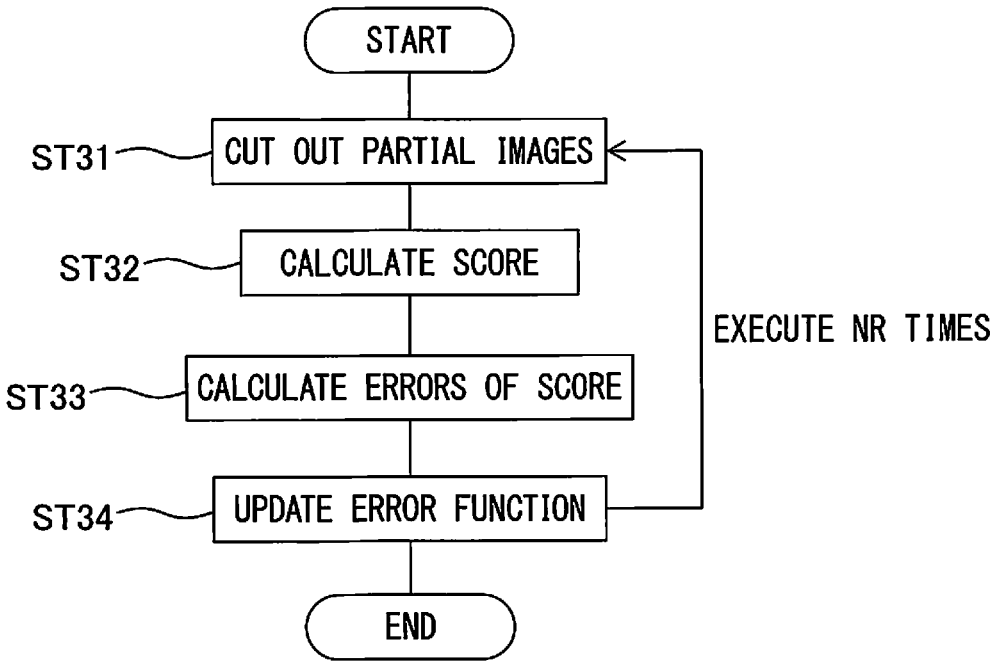

START
ST31
CUT OUT PARTIAL IMAGES
ST32
CALCULATE SCORE
ST33
CALCULATE ERRORS OF SCORE
ST34
UPDATE ERROR FUNCTION
END
EXECUTE NR TIMES

FIG. 11

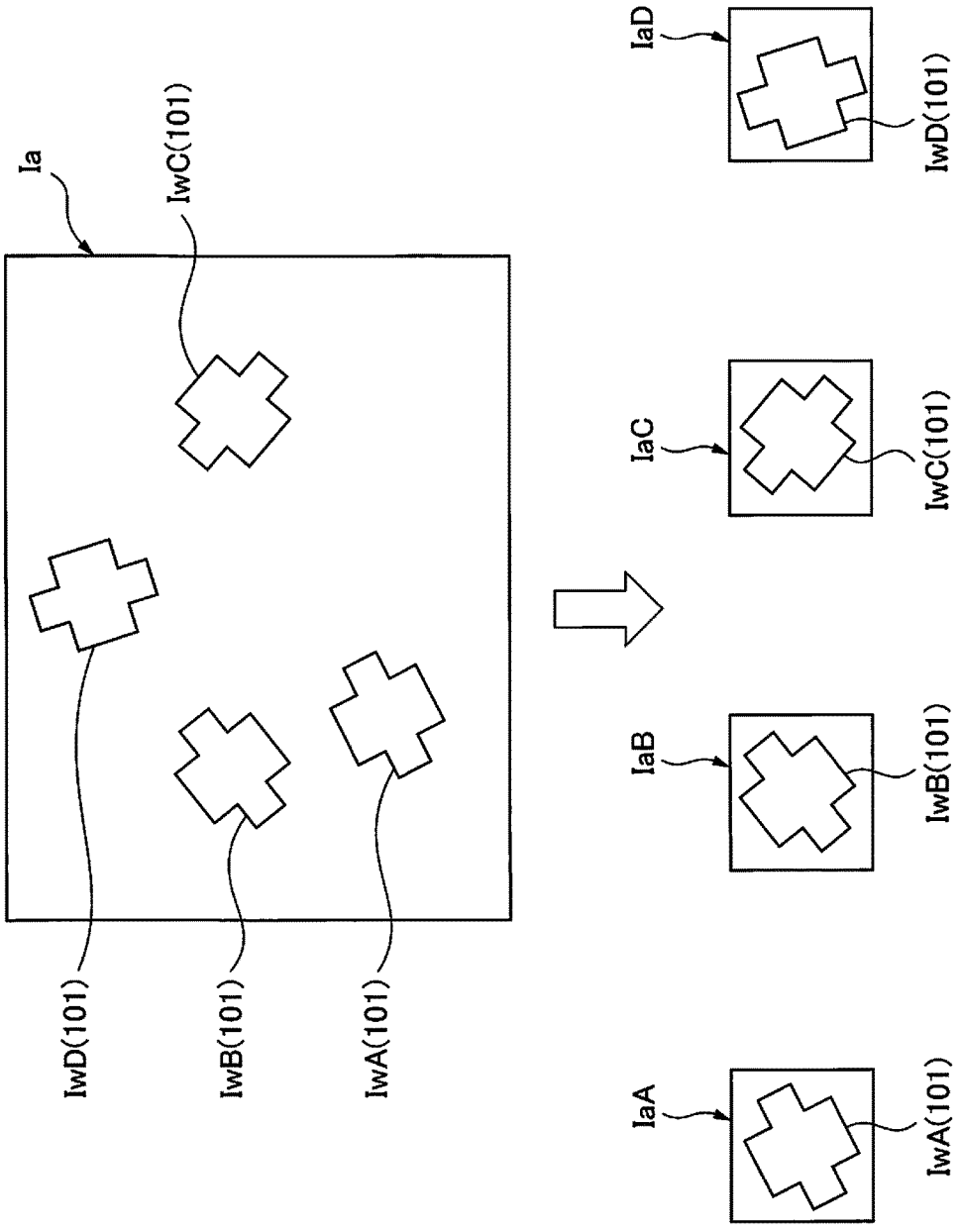

Ia
IwC(101)
IwD(101)
IwB(101)
IwA(101)
IaD
IwD(101)
IaC
IwC(101)
IaB
IwB(101)
IaA
IwA(101)

FIG. 12

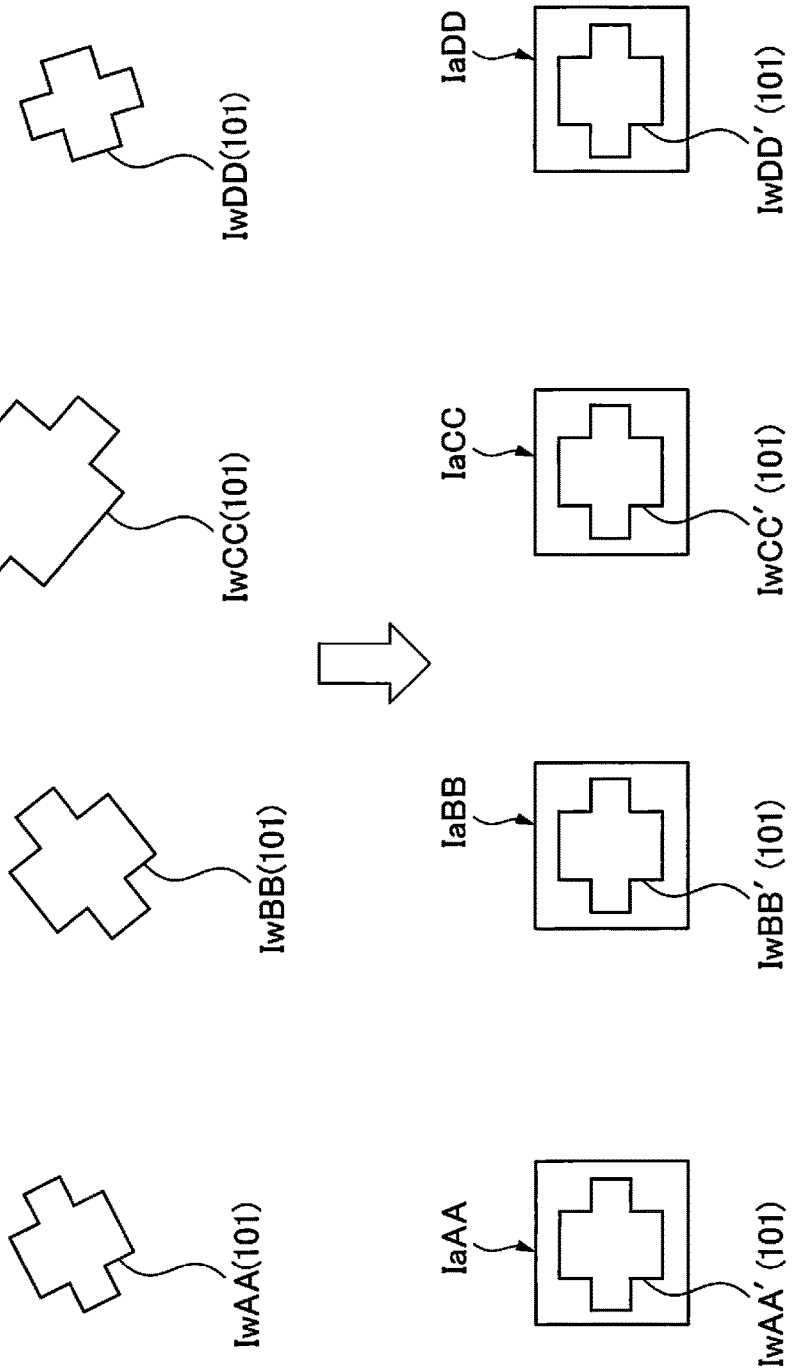

IwAA(101)
IwBB(101)
IwCC(101)
IwDD(101)
IaAA
IaBB
IaCC
IaDD
IwAA' (101)
IwBB' (101)
IwCC' (101)
IwDD' (101)

FIG. 13A

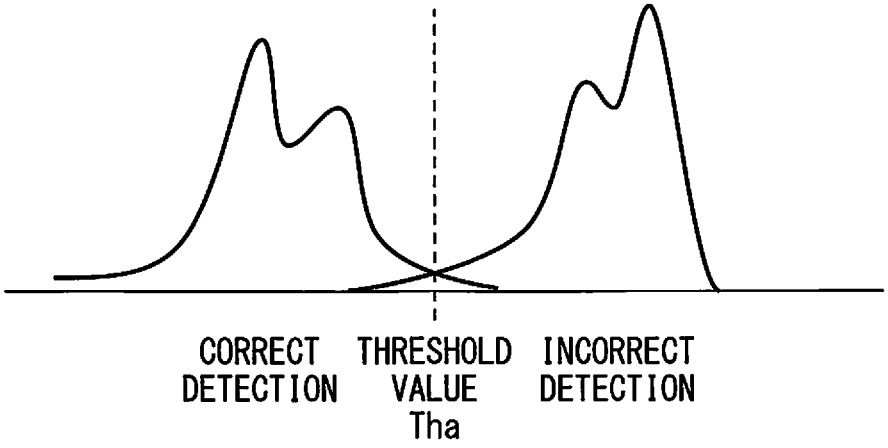
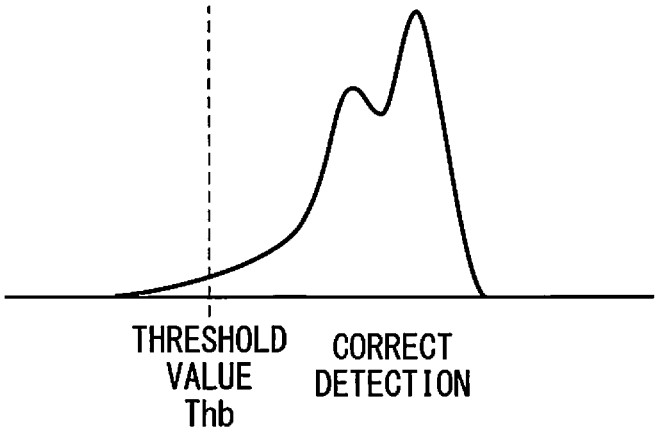


FIG. 13B



APPARATUS AND METHOD FOR IMAGE PROCESSING TO CALCULATE LIKELIHOOD OF IMAGE OF TARGET OBJECT DETECTED FROM INPUT IMAGE

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0001] The present invention relates to an apparatus and a method for image processing to calculate a likelihood of an image of a target object detected from an input image.

2. Description of the Related Art

[0002] Conventionally, for example, when an image processing apparatus is used to detect an image of a certain target object from the images in the field of vision of an image capture apparatus, feature quantity matching is performed between a piece of reference information representing the target object (generally called a model pattern, a template, or the like) and the input images obtained by the image capture apparatus. It is common to judge that an image of the target object has been successfully detected when, for example, the degree of correspondence between the model pattern and the image of the target object exceeds a designated level (threshold value).

[0003] When the threshold value of the degree of correspondence is too low, images not representing the target object are detected (false positive detection) and when the threshold value of the degree of correspondence is too high, images ought to be detected as the target object are not detected (false negative detection). To deal with this problem, adjustments of detection parameters including the threshold value of the degree of correspondence are conducted through trial and error, for example, by an operator repeating detections many times.

[0004] Conventionally, for example, Japanese Laid-open Patent Publication No. 2016-057918 discloses an image processing apparatus to learn segmentation from a plurality of images for learning and region teacher data to perform segmentation based on the result of the learning, and then to generate new images for learning by combining the properly segmented regions with the regions not properly segmented as a result of the segmentation, to perform renewed learning of segmentation by using the new images.

[0005] Conventionally, for example, Japanese Laid-open Patent Publication No. 2010-191772 discloses a learning apparatus to use positive images with the target object appearing therein and negative images without the target object appearing therein as generative images, from which to extract feature points; then to apply filtering to the generative images using a plurality of filters to obtain a filtered images; and to calculate a statistic of pixel values for each of the sub-regions around the feature points of the filtered images, to learn to correctly recognize the object to be recognized using the calculated statistics as feature quantities of the feature points.

[0006] Conventionally, for example, Japanese Laid-open Patent Publication No. 2016-062524 discloses a data processing system to segment an input image to regions and put a teacher label to each segmented region to perform a learning, based on images cut out from the regions and the teacher labels, to detect regions containing an image of the target object from the image, based on the model obtained by the learning.

[0007] Conventionally, for example, Japanese Examined Patent Publication No. 5234833 discloses a facial expression recognition apparatus to extract, from the image data, the regions containing an image of human face with an expression that matches a specified keyword to perform learning by inputting the feature quantities obtained from the regions into a learning device, thereby generating a discriminator to discriminate human facial expressions.

[0008] Conventionally, for example, International Publication Pamphlet WO2014/084218 discloses a target object detection apparatus to segment an input image into regions and apply a mask to the segmented sub-regions that contain no image of the target object, to perform learning by extracting features from the parts other than the masked regions and inputting the features to a learning device, and to determine whether or not specified regions in the image contain an image of the target object using a discriminator obtained by the learning.

[0009] As described above, for example, adjustments of detection parameters including the threshold value of the degree of correspondence between the model pattern and the target object are conducted through trial and error, for example, by an operator repeating detections many times. The definition of the degree of correspondence is dependent upon, for example, the detection algorithm. In other words, the degree of correspondence is calculated using only the image features used in a particular detection algorithm. For example, when the algorithm used is one for detecting a target object by using features of edge points in the image, such as the Generalized Hough Transform, the degree of correspondence is calculated based on the degree of correspondence of the edge points and so images having corresponding edges may be detected even when it is obvious to human eye that they are wrong ones.

[0010] In other words, it takes trouble to adjust the detection parameters so that only the objects intended by the operator may be detected. In addition, it may not be possible to achieve the detection of only the objects intended by the operator by using the degree of correspondence calculated according to the detection algorithm. An option in addressing such a case is to calculate a likelihood (probability of a rightly detected image of the target object or plausibility of an image of the target object, likelihood of successful detection) based on the result of detection according to the detection algorithm and determine whether or not the object is detected, based on the likelihood. For example, it may be preferable to calculate a likelihood by giving greater weights to some parts and smaller weights to other parts. Further, there may be a case in which it is preferable to calculate (compute) a likelihood by using features other than those used in the detection algorithm.

[0011] An object of the present invention is to provide an apparatus and a method for image processing to properly calculate a likelihood of the image of the target object to correctly detect the target object.

SUMMARY OF INVENTION

[0012] According to a first aspect of the present invention, there is provided an image processing apparatus which receives an input image and detects an image of a target object based on a detection algorithm, including a machine

learning device which performs learning by using a plurality of partial images cut out from at least one input image, based on a result of detection of the image of the target object, and calculates a likelihood of the image of the target object.

[0013] According to a second aspect of the present invention, there is provided an image processing method for receiving an input image and detecting an image of a target object based on a detection algorithm, including performing machine learning by using a plurality of partial images cut out from at least one input image based on a detection result of the image of the target object to calculate a likelihood of the image of the target object.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] The present invention will be understood more clearly by referring to the following accompanying drawings.

[0015] FIG. **1** is a schematic diagram illustrating an example of a machine tool system including a vision sensor control apparatus to which an image processing apparatus of the present invention is applied;

[0016] FIG. **2** is a schematic diagram illustrating an example of a robot system including a vision sensor control apparatus to which the image processing apparatus of the present invention is applied;

[0017] FIG. **3** is a block diagram illustrating the present embodiment of the image processing apparatus;

[0018] FIG. **4** is a schematic diagram illustrating an example of a machine learning device provided for the image processing apparatus of the present embodiment;

[0019] FIG. **5** is a block diagram illustrating an example of the machine learning device illustrated in FIG. **4**;

[0020] FIG. **6** is a flowchart illustrating an example of processing of model pattern generation by the image processing apparatus according to the present embodiment;

[0021] FIG. **7** is a diagram illustrating an example of a model pattern designation region in a captured image.

[0022] FIG. **8** is a flowchart illustrating an example of learning processing by the image processing apparatus according to the present embodiment;

[0023] FIG. **9** is a diagram for a description of the labeling processing in the flowchart illustrated in FIG. **8**;

[0024] FIG. **10** is a flowchart for a description of an example of the learning processing in the flowchart illustrated in FIG. **8**;

[0025] FIG. **11** is a diagram for a description of an example of the processing of cutting out partial images in the flowchart illustrated in FIG. **10**;

[0026] FIG. **12** is a diagram for a description of a modified example of the processing in the image processing apparatus of the present embodiment; and

[0027] FIGS. **13**A and **13**B are graphs for a description of another modified example of the processing in the image processing apparatus of the present embodiment.

DETAILED DESCRIPTION

[0028] An embodiment of an apparatus and a method for image processing according to the present invention will be described below in detail with reference to the attached drawings. FIG. **1** is a schematic diagram illustrating an example of a machine tool system **100**A including a vision sensor control apparatus to which an image processing apparatus of the present invention is applied and FIG. **2** is a schematic diagram illustrating an example of a robot system **100**B including a vision sensor control apparatus to which the image processing apparatus of the present invention is applied.

[0029] FIG. **1** illustrates a target object **101**, a bench **102**, a vision sensor **110**, a vision sensor control apparatus **120**, and an image processing apparatus **121**. The target object (workpiece) **101** may be made of various materials such as metal, wood, stone, resin, or the like and may be in various forms. The machine tool (the machine tool itself is not illustrated) is a machine to apply processing on the target object **101**, such as cutting, drilling, grinding, polishing, rolling, forging, or bending and may be, for example, a lathe, a driller, a boring machine, a miller, a gear cutting machine, a grinder, or the like and, needless to say, may be a CNC (computer numerical control) machine tool.

[0030] As illustrated in FIG. **1**, the vision sensor control apparatus **120** includes the image processing apparatus **121** and is configured to receive image data (input images) from the vision sensor **110** provided in a fixed position from which it is possible to capture the image of the target object **101** placed on the bench **102**. The machine tool performs certain processing on the target object **101** in accordance with the output of the vision sensor control apparatus **120** and a certain inputted program.

[0031] The vision sensor **110** may be an electronic camera having an image pickup device to capture greyscale images or color images, such as a CCD (charge coupled device) and may be, for example, a stereo camera or 3D sensor that can obtain range images or 3D point groups. The 3D point groups may be on the outline of the target object **101** or may be on the surface of the target object **101**.

[0032] FIG. **2** illustrates a robot **111**, a hand **112**, and a robot control apparatus **113**. As illustrated in FIG. **2** the vision sensor control apparatus **120** includes the image processing apparatus **121** and is configured to receive image data (input images) of the image of the target object **101** placed on the bench **102**, the image captured, for example, by the vision sensor **110** attached to a certain position such as an end of the arm of the robot **111**. The robot control apparatus **113** controls the robot **111**, the hand **112**, and the like in accordance with the output of the vision sensor control apparatus **120** and a certain inputted program to perform a certain operation to the target object **101**.

[0033] FIG. **3** is a block diagram illustrating the present embodiment of the image processing apparatus. As illustrated in FIG. **3**, the image processing apparatus **121** includes an image processing unit **122** including a machine learning device **2**, a model pattern storage unit **126**, and a detection result storage unit **127**. The image processing apparatus **121** is connected to the vision sensor **110**, a control panel **131**, a display device **132**, and the like. The image processing unit **122** includes a target object detection unit **123**, a corresponding point selection unit **124**, and a model pattern correction unit **125** in addition to the machine learning device **2**. The target object detection unit **123** detects an image of the target object from input data (input images) using a model pattern taught thereto and the corresponding point selection unit **124** selects corresponding points by evaluating the appropriateness of extracted feature points as the corresponding points to the feature points in the model pattern. The model pattern correction unit **125** performs correction processing of the model pattern.

[0034] The model pattern storage unit **126** stores, for example, a taught model pattern. The detection result storage unit **127** stores the result of detection of images of the target object from the input data (input images) by using the taught model pattern. The units included in the image processing apparatus **121** are implemented by software in a computer including an MPU (micro-processing unit), a ROM (read only memory), a RAM (random access memory), and the like. The machine learning device **2** will be described in detail later with reference to the drawings.

[0035] The vision sensor **110** is connected with the image processing apparatus **121** by a communication cable and outputs, for example, captured image data to the image processing apparatus **121**. The control panel **131** is connected with the image processing apparatus **121** by a communication cable and used, for example, for making a setting or the like for the image processing apparatus **121** to detect images of the target object **101**. The display device **132** is connected with the image processing apparatus **121** by a communication cable and displays, for example, images captured by the vision sensor **110** and the content of the setting made by using the control panel **131**.

[0036] The vision sensor **110**, the control panel **131**, and the display device **132** may be integrated with the image processing apparatus **121** and, needless to say, various modifications and changes can be made. The machine learning device **2** (image processing apparatus **121**) may be, for example, installed in the robot control apparatus **113** of the robot system **100**B in FIG. **2** or may be an independent apparatus separate from the vision sensor control apparatus **120** and the robot control apparatus **113**.

[0037] FIG. **4** is a schematic diagram illustrating an example of a machine learning device provided for the image processing apparatus of the present embodiment. As illustrated in FIG. **4**, the machine learning device **2** includes a state observation unit **21**, a learning unit **22**, an output utilization unit **23**, and a label obtaining unit **24**. The state observation unit **21** receives input data from an environment **1** such as, for example, detection position, posture, size, and images of detected parts and the label obtaining unit **24** receives inputs indicating, for example, a success or failure in detection. The output utilization unit **23** calculates (computes), for example, a likelihood (probability of a rightly detected image of the target object or plausibility of an image of the target object, likelihood of successful detection). The output utilization unit **23** may have various functions to utilize the output of the machine learning unit **22** in addition to the calculation of a likelihood.

[0038] In other words, the label obtaining unit **24** obtains labels (teacher data) and, for example, receives at least one partial image cut out from an input image (image data), with teacher labels (OK/NG, Correct Detection/Incorrect Detection, integer, real number) attached to each partial image. The region from which to cut out a partial image may be, for example, a region surrounding the model pattern or may be a region predefined relative to the model pattern. The learning unit **22** receives the partial images cut out from the region and the teacher labels, performs learning (machine learning, in particular, supervised learning) to generate a learning model and, based on the generated learning model, calculates from the partial images a likelihood of an image of the target object **101**.

[0039] The learning unit **22** performs "supervised learning" and generates a learning model based on the state variable (detection position, posture, size, and a partial image of the detected part) observed by the state observation unit **21** and the labels (success or failure in detection: OK/NG) obtained by the label obtaining unit **24**. Although the environment **1** has been described in FIG. **4** using as an example the image processing apparatus **121** (vision sensor control apparatus **120**) provided in the machine tool system **100**A, needless to say, the environment **1** is not limited thereto.

[0040] FIG. **5** is a block diagram illustrating an example of the machine learning device illustrated in FIG. **4**, in which the learning unit **22** is configured with an error calculation unit **221** and a learning model update unit **222**. As illustrated in FIG. **5**, the learning unit **22** includes an error calculation unit **221** to receive the output (state variable) of the state observation unit **21** and the labels (teacher data) to calculate errors and a learning model update unit **222** to receive the output of the state observation unit and the output of the error calculation unit **221** to update the learning model. The machine learning device **2** (learning unit **22**) is not limited to those that perform supervised learning and, for example, may perform learning by using images labeled as Correct Detection (OK) and output as likelihood the degree of being different from the partial images used for the learning or may regard all detected images as having been correctly detected and perform learning (unsupervised learning) based on these. In other words, the machine learning device **2** can perform unsupervised learning by, for example, allowing narrower (stricter) ranges for the parameters to be used in the detection algorithm to obtain only the image data almost certain to turn out to be correct.

[0041] The labels inputted to the error calculation unit **221** correspond to the output of the label obtaining unit **24** in FIG. **4**. The likelihood calculation unit **25** corresponds, for example, to the above-described output utilization unit **23** in FIG. **4** and calculates (computes) and outputs a likelihood (probability of a rightly detected image of the target object or plausibility of an image of the target object, likelihood of successful detection) based on the output of the learning unit **22** (learning model update unit **222**). As described above, the image processing apparatus of the present embodiment allows appropriate calculation of a likelihood of an image of the target object and allows correct detection of the target object.

[0042] For the machine learning device **2**, general-purpose computers or processors may be used and GPGPU (general-purpose computing on graphics processing units) or large-scale PC clustering, for example, may be applied for processing at a higher speed. The machine learning device **2** is communicable with at least one other machine learning device and can mutually exchange or share the learning model generated by the learning unit **22** of the machine learning device **2** with the at least one other machine learning device. It is needless to say that the machine learning device **2** (learning unit **22**) includes a neural network, for example, constituted by the afore-mentioned GPGPU or the like.

[0043] To prevent, for example, a false positive detection of a background image as an image (Iw) of the target object **101**, parts randomly obtained from the background image may be added to the teacher data as indicating unwanted detection. Further, the learning calculation and the calculation of a likelihood after the learning may be performed using the images converted by scaling, shear transformation,

or the like based on the size in the detection result, as the images in the predefined regions with respect to the detection position and posture. More specifically, for example, when the target object is detected twice as large as the model pattern used for teaching, learning can be performed faster by magnifying the images of certain regions twofold.

[0044] In addition, the learning calculation and the calculation of the likelihood after the learning may be performed using features extracted by the same feature extraction method as used in the detection algorithm (for example, edge) or features extracted by a feature extraction method different from the methods used in the detection algorithm (for example, brightness or luminance gradient) from the images contained in predefined regions with respect to the detection position and posture, to perform learning faster than by using the images themselves.

[0045] Further, the learning may be performed step by step by generating teacher data allowing relatively large ranges for detection parameters (with large margins of error) at the initial stage and, as the learning progresses, gradually allowing smaller ranges for the detection parameters (with smaller margins of error). After the first stage, a likelihood is calculated by the calculation method newly learned and the results are used for subsequent stages. This allows the learning to be continued without too many instances of false detection.

[0046] By using the likelihood calculated by the method learned, a threshold value against which to determine whether or not an image of the target object has been detected can be established automatically. A likelihood may be a scalar quantity and may be a vector quantity. Alternatively, the machine learning device **2** does not output a likelihood but, for example, determines a success or failure in detection (OK/NG) and outputs labels directly. A processing by the image processing apparatus according to the present embodiment will be described in detail below.

[0047] FIG. **6** is a flowchart illustrating an example of processing of model pattern generation by the image processing apparatus according to the present embodiment and FIG. **7** is a diagram illustrating an example of a model pattern designation region in a captured image. As illustrated in FIG. **6**, when the processing of model pattern generation starts, for example, the vision sensor (camera) **110** captures an image of the target object **101** on the bench **102** in Step ST**11**, from which to generate a model pattern to be used for the teaching, and the processing proceeds to Step ST**12**. The vision sensor **110** is preferably disposed in the same position relative to the target object **101** as when the target object **101** is detected during the actual work and the like.

[0048] In Step ST**12**, as illustrated in FIG. **7**, the region Ia containing the image Iw of the target object **101** in the captured image is defined by a rectangle and designated as a model pattern designation region. The processing of designating the model pattern region Ia in Step ST**12** is not limited to a processing using a rectangular region as illustrated in FIG. **7** but, needless to say, it can be done by using a circle, an ellipse, or the like. In other words, in FIG. **7**, an image coordinate system is defined for the captured image and a rectangular model pattern designation region Ia is designated in such a way as to include the image Iw of the target object **101** in the image coordinate system. The processing of designating the model pattern designation region Ia can be done by, for example, an operator (user) looking at the image on the display device **132** and inputting instructions to the control panel **131** but, for example, the image processing unit **122** may identify the parts with a high luminance gradient as forming the outline of the image Iw of the target object **101** and automatically designate a model pattern designation region Ia in such a way as to contain the image Iw within.

[0049] Proceeding to Step ST**13**, for example, edge points in the model pattern designation region Ia are extracted as feature points, and physical quantities such as the positions of the edge points, posture (direction of the luminance gradient), and the steepness of luminance gradient are calculated. Further, for example, a model pattern coordinate system may be defined in the designated region to convert the positions of the edge points and the posture expressed in values according to the image coordinate system to those expressed in values according to the model pattern coordinate system.

[0050] Proceeding to Step ST**14**, the physical quantities of the extracted edge points are stored in the model pattern storage unit **126** as feature points constituting the model pattern. Although edge points are used as feature points in the description above, needless to say, feature points are not limited to edge points. In addition, instead of extracting edge points, SIFT (scale-invariant feature transform) feature points, or the like from the image Iw of the target object **101** to obtain feature points for constituting a model pattern, geometric figures such as a line segment, a rectangle, and a circle may be disposed along the outline of the target object **101** appearing in the image Iw to generate a model pattern. In such a case, for example, feature points are established at appropriate intervals on the geometric figures forming the outline.

[0051] Further, a model pattern can be generated based on CAD data or the like. When 2D CAD data is used, a model pattern is generated in the same way as when geometric figures are used and, when 3D CAD is used, feature points are obtained by projecting the shape of the target object expressed by the CAD data onto an image and extracting feature points from the projected image.

[0052] FIG. **8** is a flowchart illustrating an example of learning processing by the image processing apparatus according to the present embodiment. As illustrated in FIG. **8**, when the learning processing starts, first, a model pattern is generated (taught) in Step ST**21**, and then in Step ST**22**, input images are selected. In Step ST**21**, the model pattern is taught by using an image captured by the camera (vision sensor) **110** and, in Step ST**21**, a plurality of images containing an image Iw of the target object **101** are prepared as input images.

[0053] In Step ST**23**, detection is performed for the target object (**101**) for each of the input images and the processing then proceeds to Step ST**24**. It is desirable to obtain both correct instances and false instances of detection and so detection is performed allowing relatively large ranges for the detection parameters (with large margins of error). In Step ST**23**, the detection processing of the target object is performed, for example, NI times, NI being the number of the images. The detection parameters may include, for example, various elements such as range of sizes relative to the model pattern, range of shear transformation, range of detection positions, range of angles, ratio of edge points matching the edge points of the model pattern, threshold value of distance for assuming edge points of an image to be

matching the edge points of the model pattern, and threshold value of edge points contrast.

[0054] In Step ST**24**, for example, the detection results are displayed on the display device **132** and the operator conducts visual checks and places labels of Correct Detection (OK) or Incorrect Detection (NG). FIG. **9** is a diagram for a description of the labeling processing in the flowchart illustrated in FIG. **8** and, as illustrated in FIG. **9**, in the labeling processing in Step ST**24**, a label of either OK or NG is attached to the images Iw**1**, Iw**21**, Iw**22**, Iw**31** to Iw**33**, . . . of the target object **101** contained in a plurality of model pattern designation regions Ia**1**, Ia**2**, Ia**3**, ... The learning may be performed using only the correct detection results by performing detection with relatively narrow ranges for the detection parameters. In such a case, visual checks (attaching labels of OK or NG) by the operator, for example, is not necessary. It is needless to say that the images Iw**1**, Iw**21**, Iw**22**, Iw**31** to Iw**33** . . . of the target object **101** may be taken out separately to attach a label of either OK or NG to each of the Iw**1**, Iw**21**, Iw**22**, Iw**31** to Iw**33**, . . . that are so taken out.

[0055] In FIG. **9**, the labels of OK, NG, and OK are respectively attached to the image Iw**1**, the images Iw**21** and Iw**22**, and the images Iw**31** to Iw**33** of the target object **101** contained in the model pattern designation regions Ia**1**, Ia**2**, Ia**3**, the labels being attached to the image or the images as a whole in each model pattern designation region, but it is needless to say that, for example, as in the model pattern designation region Ian, the image Iwn**1** of the target object **101** may be labeled as OK while Iwn**2** may be labeled as NG. Labeling as OK or NG may be automatically performed, for example, based on a certain threshold value: a detection result indicating a value equal to or greater than the threshold value is labeled OK (Correct Detection) while a detection result indicating a value smaller than the threshold value is labeled NG (Incorrect Detection); however, for example, after labeling having been performed automatically, the operator may correct the labels.

[0056] Further, in Step ST**25**, learning (machine learning, for example, supervised learning) is performed and the processing ends. The learning in Step ST**25** is performed by, for example, the machine learning device **2** described above with reference to FIG. **4** and FIG. **5**. The learning processing in Step ST**25** may be performed, for example, NR times, NR being the number of the model patterns, or a predefined number of times. For example, when detection is performed after the learning, a likelihood is calculated with the learned learning model to determine whether the detection is performed correctly. A likelihood may be calculated before the learning, for example, to calculate the errors to be used for backpropagation during the learning. A likelihood may be calculated after the learning, for example, to determine whether or not the images of the target object actually detected are the correct ones. For calculating the likelihood of an images of the target object before or after performing the learning, the images obtained by processing and converting the images contained in predefined regions with respect to the detected position and posture of the detected target object may be used.

[0057] FIG. **10** is a flowchart for a description of an example of the learning processing in the flowchart illustrated in FIG. **8** and for a description of an example of the learning (supervised learning) processing in Step ST**25**. FIG. **11** is a diagram for a description of an example of the processing of cutting out partial images in the flowchart illustrated in FIG. **10**.

[0058] As illustrated in FIG. **10**, when the learning processing starts, partial images in the predefined regions at the detected position and posture are cut out from the input image in Step ST**31**, and the processing proceeds to Step ST**32**. In other words, in Step ST**31**, when, for example, the input image (the image in the region Ia) contains four images IwA to IwD of the target object **101**, four regions IaA, IaB, IaC, IaD, respectively containing one partial image IwA, IwB, IwC, IwD, are cut out as illustrated in FIG. **11**.

[0059] In Step ST**32**, pixel values of the partial images are inputted to the learning model to calculate a likelihood. Scores may be set on a scale of, for example, from 0 to 1. In Step ST**33**, errors for the calculated likelihood are calculated by, for example, giving 1.0 for a Correct Detection (OK) label attached to a detection result and 0.0 for an Incorrect Detection (NG) label, and the processing proceeds to Step ST**34**. In Step ST**34**, the learning model (parameters of the learning model) is updated. In other words, in Step ST**34**, the errors are used for backpropagation in the learning model to update the parameters of the learning model. The processing in Steps ST**31** to ST**34** is performed NR times, NR being the number of results used for the learning.

[0060] For example, when the learning is completed, the machine learning device **2** that has completed learning can be used to calculate a likelihood. In other words, the detection of the target object from a newly obtained input image is performed using an existing detection algorithm, and the partial images containing an image of the target object are cut out from the input image. The partial images are then inputted to the machine learning device **2** to calculate a likelihood. Based on the calculated likelihood, it is possible to determine whether or not the images of the target object are the results of correct detection. This allows for avoiding to use incorrect detection results and also for determining whether a detected target object is a good or defective product.

[0061] FIG. **12** is a diagram for a description of a modified example of the processing in the image processing apparatus of the present embodiment. After cutting out from the input image the partial images of the predefined regions with respect to the detected position and posture, the partial images may be transformed, for example, to adjust their positions, postures, sizes, distortions to the model pattern. In other words, as illustrated in FIG. **12**, four partial images IwAA, IwBB, IwCC, IwDD of the target object **101**, which are obtained by cutting out from the input image (Ia) the partial images of predefined region with respect to the detected position and posture, are transformed (standardized) to adjust their positions, postures, sizes, and distortions to the model pattern, which results in four regions IaAA, IaBB, IaCC, IaDD respectively containing standardized partial images IwAA', IwBB', IwCC', IwDD'.

[0062] This ensures that the pixel values inputted to the learning model for the same input category are the values of the same parts (corresponding parts). When, for example, the detection position can be obtained in a unit smaller than one pixel (picture element), the partial images may be shifted in a unit smaller than one pixel.

[0063] Further, after cutting out the partial images, feature quantities are extracted from the partial images and inputted to the learning model. The feature quantities may be more

than one kind and may be, for example, luminance gradient, direction of luminance gradient, edge points, and the like. In addition, the feature quantities extracted from the partial images and the partial images may be inputted to the learning model at the same time.

[0064] Further, in performing learning, exclusion regions may be designated, which are not to be used for the learning. Designating exclusion regions on the model pattern as described above enables designation regions to be shifted according to the detection results. This allows exclusion regions to be designated for the partial images and, for example, when the partial images have undergone rotation and position adjustment, exclusion regions can be used as they are.

[0065] Further, in another modified example, the learning processing may be performed step by step. In other words, teacher data are generated allowing relatively large ranges for the detection parameters (with large margins of error) initially (at the initial stage) and, as the learning progresses (proceeding to subsequent stages), allowing smaller ranges for the detection parameters (with smaller margins of error) to perform the learning further.

[0066] More specifically, when teacher data are generated at first, for example, the size is set at 0.9 to 1.1 and the ratio of edge correspondence at not less than 60%. When the teacher data for the next round of learning are generated, for example, the size is set at from 0.95 to 1.05 and the ratio of edge correspondence is set at 80%. It is needless to say that these values are presented merely as an example.

[0067] FIGS. **13**A and **13**B are graphs for a description of another modified example of the processing in the image processing apparatus of the present embodiment and FIG. **13**A illustrates a case in which a classification is made between instances of correct detection (OK) and those of incorrect detection (NG) by a threshold value Tha while FIG. **13**B illustrates a case in which classification of instances of correct detection is made by a threshold value Thb. In still another example of processing in the image processing apparatus of the present embodiment, for example, the threshold value may be automatically established for determining whether or not an image of the target object has been detected according to the likelihood outputted by the machine learning device **2**. This can be done, for example, by the following procedure.

[0068] First, the partial images in the learning data are inputted one by one to the machine learning device **2** to obtain degrees of correspondence. The obtained degrees of correspondence are stored after being classified based on whether the partial images are labeled Correct Detection or Incorrect Detection. Further, the probability distribution is calculated for a set consisting of the degrees of correspondence with Correct Detection labels and for a set consisting of the degrees of correspondence with Incorrect Detection labels. For calculating the probability distribution, for example, normal mixed distribution may be used. A threshold value for classifying instances of correct detection and incorrect detection may then be calculated from the calculated probability distribution.

[0069] As illustrated in in FIG. **13**A, for example, the threshold value may be set at a value Tha, at which two probability distribution curves for correct detection and incorrect detection intersect, or may be set at a value Thb, over which the probability of an instance of correct detection exceeds a predefined value in the probability distribution.

[0070] Further, regions that are not the detection target in the input image may be added to the teacher data as examples of incorrect detection. This can be done, for example, by the following procedure. First, generate detection results randomly. Next, confirm that the detection results are away from the correct results with a certain range or more. For example, when the detection results are given in position, angle, and scale, confirm that they are away at a certain distance or more in the position, angle, and scale spaces. Then store the detection results with an Incorrect Detection label attached.

[0071] This will, for example, make it difficult to detect regions that are not the image of the target object in the input image. In addition, by performing detection with relatively narrow ranges for the detection parameters and treating the detection result as being correct and by performing learning based on these correct detection results together with the incorrect detection results automatically added, learning (supervised learning) can be performed, for example, without an operator attaching labels.

[0072] The apparatus and a method for image processing according to the present invention has an advantageous effects of properly calculating a likelihood of an image of the target object to correctly detect the target object.

[0073] All examples and conditional language provided herein are intended for the pedagogical purposes of aiding the reader in understanding the invention and the concepts contributed by the inventor to further the art, and are not to be construed as limitations to such specifically recited examples and conditions, nor does the organization of such examples in the specification relate to a showing of the superiority and inferiority of the invention. Although one or more embodiments of the present invention have been described in detail, it should be understood that various changes, substitutions, and alterations could be made hereto without departing from the spirit and scope of the invention.

What is claimed is:

1. An image processing apparatus which receives an input image and detects an image of a target object based on a detection algorithm, comprising:

a machine learning device which performs learning by using a plurality of partial images cut out from at least one input image, based on a result of detection of the image of the target object, and calculates a likelihood of the image of the target object.

2. The image processing apparatus according to claim **1**, wherein

the machine learning device comprises:

a state observation unit which observes, as a state variable, at least one of detection position, posture, size, and an image of a detected part related to the target object; and

a learning unit which generates a learning model for calculating the likelihood of the image of the target object, based on the state variable observed by the state observation unit, and

the machine learning device performs unsupervised learning.

3. The image processing apparatus according to claim 1, wherein

the machine learning device comprises:

a state observation unit which observes, as a state variable, at least one of detection position, posture, size, and an image of a detected part related to the target object;

a label obtaining unit which obtains labels indicating correct detection or incorrect detection to be attached to each of the plurality of partial images; and

a learning unit which generates a learning model for calculating the likelihood of the image of the target object, based on the state variable observed by the state observation unit and the labels obtained by the label obtaining unit, and

the machine learning device performs supervised learning.

4. The image processing apparatus according to claim 3, wherein

the learning unit comprises:

an error calculation unit which calculates errors based on the state variable observed by the state observation unit and the labels obtained by the label obtaining unit; and

a learning model update unit which updates the learning model based on outputs of the state observation unit and the error calculation unit.

5. The image processing apparatus according to claim 3, wherein the learning unit attaches a label indicating incorrect detection to a partial image cut out from a region in the input image which region contains no image of the target object.

6. The image processing apparatus according to claim 1, wherein the machine learning device uses an image obtained by converting an image contained in a predefined region with respect to a position and posture of the detected target object to calculate the likelihood of the image of the target object before or after performing the machine learning.

7. The image processing apparatus according to claim 1, wherein the machine learning device receives features extracted by a same feature extraction method as used in the detection algorithm from an image contained in a predefined region with respect to a position and posture of the detected target object to calculate the likelihood of the image of the target object before or after performing the machine learning, and calculates the likelihood of the image of the target object.

8. The image processing apparatus according to claim 1, wherein the machine learning device receives features extracted by a feature extraction method different from methods used in the detection algorithm from an image contained in a predefined region with respect to a position and posture of the detected target object to calculate the likelihood of the image of the target object before or after performing the machine learning.

9. The image processing apparatus according to claim 1, wherein the machine learning device performs the machine learning step by step, allowing relatively large ranges for detection parameters at an initial stage and gradually allowing smaller ranges for the detection parameters as the learning progresses to subsequent stages.

10. The image processing apparatus according to claim 1, wherein a threshold value against which to determine whether or not the image has been detected is established automatically based on the likelihood outputted by the machine learning device.

11. The image processing apparatus according to claim 1, wherein the machine learning device is communicable with at least one other machine learning device and mutually exchange or share a learning model generated by the machine learning device with the at least one other machine learning device.

12. An image processing method for receiving an input image and detecting an image of a target object based on a detection algorithm, comprising:

performing machine learning by using a plurality of partial images cut out from at least one input image based on a detection result of the image of the target object to calculate a likelihood of the image of the target object.

13. The image processing method according to claim 12,

wherein the performing of machine learning to calculate the likelihood of the image of the target object comprises:

observing, as a state variable, at least one of detection position, posture, size, and an image of a detected part related to the target object; and

generating a learning model with which to calculate the likelihood of the image of the target object, based on the state variable, to perform unsupervised learning.

14. The image processing method according to claim 12,

wherein the performing of machine learning to calculate the likelihood of the image of the target object comprises:

observing, as a state variable, at least one of detection position, posture, size, and an image of a detected part related to the target object;

obtaining labels indicating correct detection or incorrect detection to be attached to each of the plurality of partial images; and

generating a learning model with which to calculate the likelihood of the image of the target object, based on the state variable and the labels, to perform supervised learning.

15. The image processing method according to claim 14, wherein

the generating of the learning model comprises:

calculating errors based on the state variable and the labels; and

updating the learning model based on the state variable and the calculated errors.

* * * * *