[DESCRIPTION]

[Title of Invention]

TRAINING APPARATUS, CAMERA ADJUSTMENT APPARATUS, TRAINING METHOD, CAMERA ADJUSTMENT METHOD, AND STORAGE MEDIUM

[Technical Field]

[0001]

The present disclosure generally relates to training apparatus, camera adjustment apparatus, training method, camera adjustment method, and storage medium.

[Background Art]

[0002]

A system that updates a configuration of a camera based on image data generated by the camera has been developed. PTL1 discloses a system that takes a video frame generated by a camera as input and update color-related camera parameters (e.g., contrast) of the camera.

[Citation List]

[Patent Literature]

[0003]

[PTL1] International Patent Publication No. WO2022/256246

[Summary of Invention]

[Technical Problem]

[0004]

The system of PTL1 is aimed at updating the color-related camera parameters. An objective of the present disclosure is to provide a novel technique of updating camera parameters of a camera based on image data generated by the camera.

[Solution to Problem]

[0005]

The present disclosure provides a training apparatus comprising at least one memory that is configured to store instructions and at least one processor.

The at least one processor that is configured to execute the instructions to: acquire image data generated by one or more cameras; acquire one or more camera parameters from an agent, which is configured to take first two-dimensional features data including two-dimensional features extracted from the image data to compute the one or more camera parameters based on the first two-dimensional features data, the one or more camera parameters determine an orientation, a focal length, or a position of the camera; generate second two-dimensional features data that includes two-dimensional features predicted to be extracted from image data captured by the

camera that is configured with the acquired one or more camera parameters; evaluate the second two-dimensional features data to compute an evaluation score; and update trainable parameters of the agent based on the evaluation score.

[0006]

The present disclosure provides a camera adjustment apparatus comprising at least one memory that is configured to store instructions and at least one processor.

The at least one processor that is configured to execute the instructions to: acquire one or more video frames generated by one or more cameras; extracts two-dimensional features from the one or more video frames; and compute one of more camera parameters of the camera by inputting the two-dimensional features into an agent and acquiring the one or more camera parameters output by the agent, the one or more camera parameters determine an orientation, a focal length, or a position of the camera.

The agent is configured to take two-dimensional features and has been trained to output one or more camera parameters in response to the two-dimensional features being input thereinto.

[0007]

The present disclosure provides a training method comprising: acquiring image data generated by one or more cameras; acquiring one or more camera parameters from an agent, which is configured to take first two-dimensional features data including two-dimensional features extracted from the image data to compute the one or more camera parameters based on the first two-dimensional features data, the one or more camera parameters determine an orientation, a focal length, or a position of the camera; generating second two-dimensional features data that includes two-dimensional features predicted to be extracted from image data captured by the camera that is configured with the acquired one or more camera parameters; evaluating the second two-dimensional features data to compute an evaluation score; and updating trainable parameters of the agent based on the evaluation score.

[0008]

The present disclosure provides a camera adjustment method comprising: acquiring one or more video frames generated by one or more cameras; extracting two-dimensional features from the one or more video frames; and computing one of more camera parameters of the camera by inputting the two-dimensional features into an agent and acquiring the one or more camera parameters output by the agent, the one or more camera parameters determine an orientation, a focal length, or a position of the camera.

The agent is configured to take two-dimensional features and has been trained to output one or more camera parameters in response to the two-dimensional features being input thereinto.

[0009]

The present disclosure provides a storage medium storing a program that causes a computer to execute: acquiring image data generated by one or more cameras; acquiring one or more camera parameters from an agent, which is configured to take first two-dimensional features data including two-dimensional features extracted from the image data to compute the one or more camera parameters based on the first two-dimensional features data, the one or more camera parameters determine an orientation, a focal length, or a position of the camera; generating second two-dimensional features data that includes two-dimensional features predicted to be extracted from image data captured by the camera that is configured with the acquired one or more camera parameters; evaluating the second two-dimensional features data to compute an evaluation score; and updating trainable parameters of the agent based on the evaluation score.

[0010]

The present disclosure provides a storage medium storing a program that causes a computer to execute: acquiring one or more video frames generated by one or more cameras; extracting two-dimensional features from the one or more video frames; and computing one of more camera parameters of the camera by inputting the two-dimensional features into an agent and acquiring the one or more camera parameters output by the agent, the one or more camera parameters determine an orientation, a focal length, or a position of the camera.

The agent is configured to take two-dimensional features and has been trained to output one or more camera parameters in response to the two-dimensional features being input thereinto.

[Advantageous Effects of Invention]

[0011]

According to the present disclosure, a novel technique of updating camera parameters of a camera based on image data generated by the camera is provided.

[Brief Description of Drawings]

[0012]

[Fig. 1]

Fig. 1 illustrates an overview of a camera adjustment apparatus.

[Fig. 2]

Fig. 2 is a block diagram illustrating an example of the functional configuration of the camera adjustment apparatus.

[Fig. 3]

Fig. 3 is a block diagram illustrating an example of the hardware configuration of a computer 1000 realizing the camera adjustment apparatus 2000 of the example embodiment.

[Fig. 4]

Fig. 4 is a flowchart illustrating an example flow of processes performed by the camera adjustment apparatus.

[Fig. 5]

Fig. 5 illustrates an overview of the training apparatus.

[Fig. 6]

Fig. 6 illustrates an example of the functional configuration of the training apparatus.

[Fig. 7]

Fig. 7 illustrates an example flow of processes that are performed by the training apparatus.

[Fig. 8]

Fig. 8 illustrates an example of 3D features extracted by the first generating unit.

[Fig. 9]

Fig. 9 illustrates a process of simulating an original situation in which the image data is generated by the camera.

[Fig. 10]

Fig. 10 illustrates a process of simulating a situation where the configuration of the camera is updated with the camera parameters.

[Fig. 11]

Fig. 11 illustrates an example way of evaluating the camera parameters.

[Fig. 12]

Fig. 12 illustrates an example way of evaluating the camera parameters.

[Description of Embodiments]

[0013]

Example embodiments according to the present disclosure will be described hereinafter with reference to the drawings. The same numeral signs are assigned to the same elements throughout the drawings, and redundant explanations are omitted as necessary. In addition, predetermined information (e.g., a predetermined value or a predetermined threshold) is stored in advance in a storage device to which a computer using that information has access unless otherwise described.

[0014]

<Overview>

Fig. 1 illustrates an overview of a camera adjustment apparatus 2000. It is noted that the overview illustrated by Fig. 1 shows an example of operations of the camera adjustment apparatus 2000 to make it easy to understand the camera adjustment apparatus 2000, and does not limit or narrow the scope of possible operations of the camera adjustment apparatus 2000.

[0015]

The camera adjustment apparatus 2000 is used to adjust a monitoring area of a camera 10 (i.e., an area that is captured by the camera 10). The camera 10 is configured to capture the monitoring area to generate the video frame 20. The video frame 20 includes one or more objects of interest, such as humans, animals, or vehicles.

[0016]

The camera 10 can be configured with one or more camera parameters that affect the monitoring area of the camera 10. The camera parameter may include a pan parameter, a tilt parameter, and a zoom parameter, and a position parameter. The zoom parameter may also be called focal length parameter.

[0017]

The pan parameter is used to specify a horizontal orientation of the camera 10. The tilt parameter is used to specify a vertical orientation of the camera 10. The zoom parameter is used to specify a focal length of the camera 10. Cameras whose pan parameter, tilt parameter, and zoom parameter can be specified may be called Pan-Tilt-Zoom (PTZ) cameras.

[0018]

The position parameter of the camera 10 is used to specify the position of the camera 10. When the position parameter of the camera 10 can be specified, the camera 10 is deployed in a movable manner. For example, the camera 10 is integrated into, mounted on, or attached to a movable object, such as a human, a vehicle, or a mobile robot. In another example, the camera 10 is mounted on a hanging rail on the ceiling.

[0019]

The video frame 20 is supposed to undergo an analytic task of interest: e.g., object detection or recognition; face detection or recognition; and pose detection or recognition. In some situation, the configuration of the camera 10 would not be suitable for capturing the objects of interest with respect to the analytic task of interest, resulting in lower accuracy of the analytic task of interest. For example, if the monitoring area of a camera includes only a small portion of the face of human, it is difficult to accurately perform a face recognition task on an image generated

by this camera. Thus, it is preferable to adjust the configuration of the camera 10 so that the camera 10 can capture the objects of interest in a manner suitable for the analytic task of interest.

[0020]

The camera adjustment apparatus 2000 is configured to compute one or more camera parameter 50 to be applied to the camera 10 based on the video frame 20. To do so, the camera adjustment apparatus 2000 acquires the video frame 20. Then, the camera adjustment apparatus 2000 extracts two-dimensional (2D) features of the objects of interest from the video frame 20 to generate a 2D features data 30, which includes the extracted 2D features.

[0021]

The 2D features of an object may include a 2D pose, a 2D mesh, and 2D landmarks of the object. The 2D pose of an object represents a pose of the object on a 2D plane with key-points of the object (e.g., joints of human) and connections between the key-points wherein a position of each key-point is represented by 2D coordinates. The 2D mesh of an object represents a shape of the object on a 2D plane with a collection of vertices and edges wherein a position of each vertex is represented by 2D coordinates. The 2D landmarks represents landmarks of the object on a 2D plane wherein a position of each landmark is represented by 2D coordinates.

[0022]

The 2D features of an object may be obtained for a whole of the object or for a part of the object that is related to the analytic task of interest. Suppose that the analytic task is a face detection task. In this case, objects are humans, and a part of human related to the analytic task of interest is face. Thus, the camera adjustment apparatus 2000 may extract 2D features of the face (e.g., face landmarks) for each human captured on the video frame 20.

[0023]

Suppose that a type of the object of interest is human, and a type of the 2D features to be included in the 2D features data 30 is 2D face landmark. In this case, the camera adjustment apparatus 2000 computes 2D face landmarks for each one of humans captured on the video frame 20, and generates the 2D features data 30 that include the computed 2D face landmarks.

[0024]

The camera adjustment apparatus 2000 includes an agent 40, and uses the agent 40 to compute the camera parameters 50. The agent 40 is an updatable (trainable, in other words) model that computes one or more camera parameters 50 that are suitable for the camera 10 based on the 2D features data 30. The agent 40 is configured to take the 2D features as input. The agent 40 has been trained to compute one or more camera parameters in response to the 2D features being input

thereinto and output those camera parameters. Example ways of training the agent 40 will be described later.

[0025]

The camera adjustment apparatus 2000 inputs the 2D features data 30 that is generated from the acquired video frame 20 into the agent 40. In response to the 2D features data 30 being input into the agent 40, the agent 40 computes and outputs the camera parameters. The camera adjustment apparatus 2000 acquires the camera parameters output by the agent 40, and handles those acquired camera parameters as the camera parameters 50.

[0026]

It is noted that, in some embodiments, there may be two or more cameras 10. In this case, the camera adjustment apparatus 2000 may acquire the video frames 20 generated by two or more cameras 10, and compute the camera parameters 50 of a particular camera 10 based on those video frames 20.

[0027]

<Example of Advantageous Effect>

The camera adjustment apparatus 2000 computes camera parameters 50 that determine an orientation, a focal length or a position of the camera 10 based on the video frame 20 generated by the camera 10. The orientation, the focal length, and the position of the camera 10 are camera parameters of types other than color-related parameters. Thus, according to the camera adjustment apparatus 2000, a novel technique of updating the camera parameters of the camera based on the image data generated by the camera is provided.

[0028]

In addition, according to the camera adjustment apparatus 2000, the camera parameters of the camera 10 are automatically updated so that the camera 10 becomes capable of capturing the objects of interest in a manner more suitable for the analytic task of interest. Thus, the camera adjustment apparatus 2000 enables a task executor of the analytic task of interest to more accurately perform the analytic task of interest on the video frames 20.

[0029]

Hereinafter, more detailed explanation of the camera adjustment apparatus 2000 will be described.

[0030]

<Example of Functional Configuration>

Fig. 2 is a block diagram illustrating an example of the functional configuration of the camera adjustment apparatus 2000. The camera adjustment apparatus 2000 includes a generating unit 2020 and a computing unit 2040. The generating unit 2020 acquires a video frame 20, and generates 2D features data 30 that includes the 2D features of the objects of interest captured on the video frame 20. The computing unit 2040 includes the agent 40. The computing unit 2040 computes the camera parameters 50 based on the 2D features data 30, by inputting the 2D features data 30 into the agent 40 and acquiring the camera parameters 50 output by the agent 40.

[0031]

<Example of Hardware Configuration>

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

[0032]

The camera adjustment apparatus 2000 may be realized by installing an application in the computer. The application is implemented with a program that causes the computer to function as the camera adjustment apparatus 2000. In other words, the program is an implementation of the functional units of the camera adjustment apparatus 2000.

[0033]

Fig. 3 is a block diagram illustrating an example of the hardware configuration of a computer 1000 realizing the camera adjustment apparatus 2000. In Fig. 3, the computer 1000 includes a bus 1020, a processor 1040, a memory 1060, a storage device 1080, an input/output (I/O) interface 1100, and a network interface 1120.

[0034]

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

computer 1000 and a network. The network may be a LAN (Local Area Network) or a WAN (Wide Area Network).

[0035]

The processor 1040 is configured to load instructions of the above-mentioned program from the storage device 1080 into the memory 1060 and executes those instructions so as to cause the computer 1000 to operate as the camera adjustment apparatus 2000.

[0036]

The hardware configuration of the computer 1000 is not restricted to that shown by Fig. 3. For example, as mentioned-above, the camera adjustment apparatus 2000 may be realized as a combination of multiple computers. In this case, those computers may be connected with each other through the network.

[0037]

<Flow of Process>

Fig. 4 is a flowchart illustrating an example flow of processes performed by the camera adjustment apparatus 2000. The generating unit 2020 acquires the video frame 20 (S102). The generating unit 2020 generates the 2D features data 30 based on the video frame 20 (S104). The computing unit 2040 computes the camera parameters 50 by inputting the 2D features data 30 into the agent 40 and acquiring the camera parameters 50 output by the agent 40 (S106).

[0038]

It is preferable that the configuration of the camera 10 is continuously updated so that the camera 10 can continuously capture objects of interest in a manner suitable for the analytic task of interest. To do so, the camera adjustment apparatus 2000 may repeatedly perform a series of processes shown by Fig. 4. For example, the camera adjustment apparatus 2000 performs the series of processes shown by Fig. 4 for each video frame generated by the camera 10. In another example, the camera adjustment apparatus 2000 performs the series of processes shown by Fig. 4 for every predefined number of the video frames (e.g., every 10 video frames) generated by the camera 10.

[0039]

In some embodiments, the camera adjustment apparatus 2000 stands by for signals relayed from auxiliary applications to perform the series of process. Such signals may take the form of a user input that requires the camera adjustment apparatus 2000 to conduct a self-examination on its current camera parameters. Alternatively, these signals can come from an active monitoring program that constantly evaluating the performance of the analytic task of

interest. In scenarios where this performance score falls below a predetermined threshold, the monitoring program sends a signal to the camera adjustment apparatus 2000 to perform the aforementioned series of process.

[0040]

<Acquisition of Video frame 20: S102>

The generating unit 2020 acquires the video frame 20 (S102). There are various ways to acquire the video frame 20. In some embodiments, the camera 10 is configured to, when it newly generates a video frame 20, send this newly generated video frame 20 to the camera adjustment apparatus 2000. In this case, the generating unit 2020 receives the video frame 20 sent by the camera 10 to acquire the video frame 20. In another example, the camera 10 is configured to put video frames 20 into a storage unit to which the camera adjustment apparatus 2000 has access. In this case, the camera adjustment apparatus 2000 acquires the video frame 20 from this storage unit 20.

[0041]

The generating unit 2020 may acquire video frames 20 one by one or at once. In the latter case, for example, the generating unit 2020 periodically acquires the video frames 20 that are not acquired yet by the generating unit 2020.

[0042]

<Generation of 2D Feature Data 30: S104>

The generating unit 2020 generates the 2D features data 30 from the video frame 20 (S104). As mentioned above, there are various types of 2D features that are possible to be extracted from a 2D image: e.g., 2D pose, 2D mesh, 2D landmarks. In some embodiments, these 2D features can be extracted from a series of video frames: e.g., a sequence of 2D pose, 2D mesh, 2D landmarks. This implies an inherent temporal dimensionality to these 2D features, as they encapsulate data across a span of time.

[0043]

When a type of 2D features to be extracted from the video framed 20 is predefined, the generating unit 2020 can be configured to extract 2D features of the predefined type from the video frame 20. Then, the generating unit 2020 generates the 2D features data 30 that includes the extracted 2D features.

[0044]

In some embodiments, the generating unit 2020 dynamically determines a type of 2D features to be extracted from the video frame 20, and extracts the 2D features of the determined

type from the video frame 20. There are various possible ways to determine the type of 2D features to be extracted from the video frame 20. For example, a user of the camera adjustment apparatus 2000 is allowed to specify a type of 2D features to be extracted from the video frame 20. In this case, the generating unit 2020 acquires a user input that specifies the type of 2D features, and extracts 2D features of the specified type from the video frame 20.

[0045]

In another example, the user of the camera adjustment apparatus 2000 is allowed to specify the analytic task of interest. In this case, information (called "task information") that shows, for each one of two or more analytic tasks, an association between an analytic task and a type of 2D features to be extracted for that analytic task is prepared in advance. The generating unit 2020 acquires a user input that specifies the analytic task to be performed on the video frame 20, and determines the type of 2D features that is associated with the specified analytic task by the task information. Then, the generating unit 2020 extracts 2D features of the determined type from the video frame 20.

[0046]

In another example, information that specifies a type of 2D features to be extracted from video frame 20 is stored in advance in a storage unit to which the camera adjustment apparatus 2000 has access. In this case, the camera adjustment apparatus 2000 refers to that information to determine a type of 2D features to be extracted from the video frame 20.

[0047]

It is noted that there are various concrete ways (algorithms, in other words) to extract 2D features from image data, and any one of those ways can be applied to the generating unit 2020 to extract the 2D features from the video frame 20. For example, the generating unit 2020 includes a pre-trained machine learning-based model (hereinafter, 2D features extracting model) for each one of types of 2D features that are possible to be handed by the camera adjustment apparatus 2000. The 2D features extracting model corresponding to a particular type of 2D features is configured to take image data, extract that particular type of 2D features of the objects of interest from the image data, and output the extracted 2D features. The generating unit 2020 may inputs the video frame 20 to the 2D features extracting model corresponding to the type of 2D features to be fed into the analytic task of interest, thereby generating the 2D features data 30.

[0048]

<Computation of Camera Parameters 50: S106>

The computing unit 2040 computes the camera parameters 50 base on the 2D features data 30 (S106). Specifically, the computing unit 2040 inputs the 2D features data 30 that is generated by the generating unit 2020 into the agent 40. The agent 40 computes the camera parameters 50 in response to the 2D features data 30 being input into the agent 40, and outputs the camera parameters 50.

[0049]

The agent 40 is implemented by a trainable model: e.g., a machine learning-based model such as a neural network; or a mathematical function with trainable parameters. Example ways of training the agent 40 will be described later.

[0050]

<Usage of Camera Parameters 50>

The camera parameters 50 can be used to update the configuration of the camera 10. There are various ways to update the configuration of the camera 10. For example, the camera 10 is configured to receive an instruction (update instruction, hereinafter) of updating the configuration of the camera 10 and update the configuration of the camera 10 based on the update instruction. In this case, the camera adjustment apparatus 2000 may send an update instruction that indicates the camera parameters 50. The camera 10 receives this update instruction, and updates its own configuration based on the camera parameters 50 indicated by the update instruction.

[0051]

<Training of Agent 40 >

The agent 40 is required to have been trained in advance so that the agent 40 can provide the camera parameters 50 with which the camera 10 can capture the objects of interest on the video frame 20 in a manner suitable for the analytic task of interest. Hereinafter, example ways of training the agent 40 will be described.

[0052]

An apparatus that is configured to train the agent 40 is called “training apparatus”. The training apparatus may be implemented by the same computer as the computer that functions as the camera adjustment apparatus 2000, or may be implemented by a computer different from the computer that functions as the camera adjustment apparatus 2000. In the latter case, the training apparatus 3000 may be implemented with a computer that has a configuration similar to or same as the computer 1000 shown by Fig. 3.

[0053]

Fig. 5 illustrates an overview of the training apparatus 3000. It is noted that the overview illustrated by Fig. 5 shows an example of operations of the training apparatus 3000 to make it easy to understand the training apparatus 3000, and does not limit or narrow the scope of possible operations of the training apparatus 3000.

[0054]

The training apparatus 3000 acquires image data 110, causes the agent 40 to compute camera parameters 130 based on the acquired image data 110, and updates the agent 40 using the camera parameters 130. The image data 110 is a single image or a sequence of images that includes one or more objects of interest. The image data 110 may be generated by a camera 100, which is arbitrary camera. In some embodiments, the image data 110 includes two or more images generated by two or more cameras 100.

[0055]

The camera parameters 130 are camera parameters of the same types as the camera parameters 50. For example, when the camera parameters 50 include the pan parameter, the tilt parameter, and the zoom parameter, the camera parameters 130 also include the pan parameter, the tilt parameter, and the zoom parameter.

[0056]

The training apparatus 3000 may acquire the camera parameters 130 in the manner same as the manner in which the camera adjustment apparatus 2000 acquires the camera parameter 50. Specifically, the training apparatus 3000 computes 2D features of the objects of interest captured on the image data 110 to generate a first 2D features data 120, in the manner same as the manner in which the camera adjustment apparatus 2000 computes 2D features of the objects on interest captured on the video frame 20 to generate the 2D features data 30. The first 2D features data 120 is a data that represents 2D features of the objects of interest captured on the image data 110.

[0057]

Then, the training apparatus 3000 inputs the first 2D features data 120 into the agent 40 to acquire the camera parameters 130, in the manner same as the manner in which the camera adjustment apparatus 2000 inputs the 2D features data 30 into the agent 40 to acquire the camera parameter 50. It is noted that the training apparatus 3000 may acquire the camera parameters 130 by computing the camera parameters 130 by itself or by delegating the computation of the camera parameters 130 to the camera adjustment apparatus 2000 and acquiring the camera parameters 130 from the camera adjustment apparatus 2000.

[0058]

Conceptually, the training apparatus 3000 evaluates how effective the camera parameters 130 are to adjust the camera 100 to become more suitable for the analytic task of interest, and updates trainable parameters of the agent 40 based on the result of the evaluation. For this evaluation, the training apparatus 3000 simulates a camera (called virtual camera, hereinafter) that is virtually configured with the camera parameters 130, and evaluates a simulated result of capturing by this virtual camera.

[0059]

Specifically, the training apparatus 3000 evaluates the camera parameters 130 as follows. The training apparatus 3000 computes 3D features of the objects captured on the image data 110 to generate 3D features data 140. The 3D features of an object may include a 3D pose of the object, a 3D mesh of the object, and 3D landmarks of the object. In some embodiments, these 3D features can be extracted from a series of video frames: e.g., a sequence of 3D pose of the object, a sequence of 3D mesh of the object, and a sequence of 3D landmarks of the object. This implies an inherent temporal dimensionality to these 3D features, as they encapsulate data across a span of time.

[0060]

The 3D pose of an object represents a pose of the object in a 3D space with key-points of the object (e.g., joints of human) and connections between the key-points wherein a position of each key-point is represented by 3D coordinates. The 3D mesh of an object represents a shape of the object in a 3D space with a collection of vertices and edges wherein a position of each vertex is represented by 3D coordinates. The 3D landmarks represent landmarks of the object in a 3D space wherein a position of each landmark is represented by 3D coordinates. The 3D features of an object may be obtained for a whole of the object or for a part of the object that is related to the analytic task of interest.

[0061]

The training apparatus 3000 generates second 2D features data 150 using the 3D features data 140 and the camera parameters 130. The second 2D features data 150 represents 2D features of the objects of interest that are expected to be extracted from image data if the above-mentioned virtual camera generates this image data by capturing the scene same as the scene captured on the image data 110. This means that the training apparatus 3000 simulates a camera configured with the camera parameters 130 capturing the scene as the scene captured on the image data 110.

[0062]

The training apparatus 3000 evaluates the second 2D features data 150 to compute an

evaluation score 160. The evaluation score 160 represents how suitable the second 2D features data 150 is for the analytic task of interest. The more suitable the second 2D features data 150 is for the analytic task, the more effective the camera parameters 130 are to adjust the camera.

[0063]

Based on the evaluation score 160, the training apparatus 3000 computes new values of updatable parameters of the agent 40. Then, the training apparatus 3000 updates the agent 40 by applying those new values to the updatable parameters of the agent 40.

[0064]

It can be seen that the agent 40 is trained in a reinforcement learning-based fashion. In a reinforcement learning, an agent determines an action to take based on the current state of an environment. The state of the environment is updated as a result of the action taken by the agent. The agent determines what action to take based on the current state of the environment and a rule (e.g., a policy or a value function) that takes into consideration cumulative rewards that are predicted using the current state.

[0065]

In a training phase of the agent, the agent is caused to take an action, a new state of the environment is determined based on the taken action, and the cumulative rewards are computed based on the new state. The rule is updated based on the computed cumulative rewards.

[0066]

From the viewpoint of reinforcement learning, the agent 40 can be seen as an agent that handles the 2D features extracted from the video frame 20 as data representing the current state of the environment, and that determines the camera parameters 50 as an action to take. The updatable parameters of the agent 40 can be seen as data that defines the rule to determine what action to take.

[0067]

In the training phase of the agent 40, the training apparatus 3000 causes the agent 40 to compute the camera parameters 130. This process can be seen as the process of causing the agent to take an action. Then, the training apparatus 3000 computes the second 2D features data 150 by computing the 3D features data 140 and simulating a situation where the objects of interested are captured by the camera 100 that is configured with the camera parameters 130. This process can be seen as the process of determining a new state of the environment based on the taken action. Next, the training apparatus 3000 evaluates the second 2D features data 150 to compute the evaluation score 160. This process can be seen as the process of computing the cumulative

rewards based on the new state. Lastly, the training apparatus 3000 update the updatable parameters of the agent 40 based on the evaluation score 160. This process can be seen as the process of updating the rule based on the cumulative rewards.

[0068]

<<Example of Advantageous Effect>>

According to the training apparatus 3000, the agent 40 is trained to compute the camera parameters 50 that determine an orientation, a focal length or a position of the camera 10 based on the video frame 20 generated by the camera 10. The orientation, the focal length, and the position of the camera 10 are camera parameters of types other than color-related parameters. Thus, according to the training apparatus 3000, a novel technique of updating the camera parameters of the camera based on the image data generated by the camera is provided.

[0069]

In addition, according to the training apparatus 3000, the agent 40 is trained to become capable of automatically updating the camera parameters 50 so that the camera 10 becomes capable of capturing the objects of interest in a manner more suitable for the analytic task of interest. Thus, the training apparatus 3000 helps a task executor of the analytic task of interest to more accurately perform the analytic task of interest on the video frames 20.

[0070]

<<Example of Functional Configuration of Training Apparatus>>

Fig. 6 illustrates an example of the functional configuration of the training apparatus 3000. The training apparatus 3000 includes an acquiring unit 3020, a first generating unit 3040, a second generating unit 3060, an evaluating unit 3080, and an updating unit 3100. The acquiring unit 3020 acquires image data 110, and camera parameters 130 based on the image data 110. Specifically, a first 2D features data 120 is generated by extracting 2D features of the objects of interest from the image data 110. Then, the first 2D features data 120 is input into the agent 40, thereby the camera parameters 130 are output by the agent 40.

[0071]

The first generating unit 3040 generates a 3D features data 140 based on the image data 110. The second generating unit 3060 generates a second 2D features data 150 based on the 3D features data 140 and the camera parameters 130. The evaluating unit 3080 evaluates the second 2D features data 150 to compute an evaluation score 160. The updating unit 3100 updates updatable parameters of the agent 40 based on the evaluation score 160.

[0072]

<<Flow of Process>>

Fig. 7 illustrates an example flow of processes that are performed by the training apparatus 3000. The acquiring unit 3020 acquires the image data 110 (S202). The acquiring unit 3020 acquires the camera parameters 130 based on the image data 110 (S204). The first generating unit 3040 generates the 3D features data based on the image data 110 (S206). The second generating unit 3060 generates the second 2D features data 150 based on the 3D features data 140 and the camera parameters 130 (S208). The evaluating unit 3080 evaluates the second 2D features data 150 to compute the evaluation score 160 (S210). The updating unit 3100 updates the agent 40 based on the evaluation score 160 (S212).

[0073]

To make the agent 40 capable of accurately computing the camera parameters 130 that are suitable for the analytic task of interest, it is preferable to update the agent 40 multiple times. Thus, the training apparatus 3000 may repeatedly perform a series of the processes shown by Fig. 7 for each one of multiple image data 110 different from each other.

[0074]

<<Acquisition of Image Data 110: S202>>

The acquiring unit 3020 acquires the image data 110 (S202). The acquiring unit 3020 may acquire the image data 110 in the way similar to the way in which the generating unit 2020 acquires the video frame 20.

[0075]

It is noted that when the training apparatus 3000 acquires two or more image data 110, those image data 110 are not necessarily generated by the same camera as each other. For example, two or more cameras 100 are deployed, and two or more image data 110 may be acquired from those cameras 100. It is also noted that the image data 110 is not necessarily a video frame, and may be generated by a still camera.

[0076]

<<Computation of Camera Parameters 130: S204>>

The generating unit 2020 acquires the camera parameters 130 (S204). In some embodiments, the training apparatus 3000 is configured to compute the camera parameters 130 by itself. In this case, the acquiring unit 3020 generates the first 2D features data 120 based on the image data 110. Then, the acquiring unit 3020 inputs the first 2D features data 120 into the agent 40, thereby acquiring the camera parameters 130 from the agent 40.

[0077]

In some embodiments, the training apparatus 3000 is configured to delegate the computation of the camera parameters 130 to the camera adjustment apparatus 2000. In this case, the training apparatus 3000 inputs (e.g., sends) the image data 110 to the camera adjustment apparatus 2000, thereby causing the camera adjustment apparatus 2000 to compute the camera parameters 130. Then, the training apparatus 3000 acquires the camera parameters 130 from the camera adjustment apparatus 2000.

[0078]

<<Generation of 3D Features Data 140: S206>

The first generating unit 3040 generates the 3D features data 140 based on the image data 110 (S206). There are various techniques to compute the 3D features of the objects using image data on which those objects are captured, and any one or more of those techniques can be applied to the first generating unit 3040.

[0079]

Fig. 8 illustrates an example of 3D features extracted by the first generating unit 3040. In this example, it is assumed that the analytic task of interest is a face recognition task, and features to be extracted from the image data 110 are face landmarks of human. In this case, the first generating unit 3040 extracts a set 190 of 3D face landmarks for each human captured on the image data 110.

[0080]

<<Generation of Second 2D Features Data 150: S208>>

The second generating unit 3060 generates the second 2D features data 150 based on the 3D features data 140 and the camera parameters 130 (S208). Conceptually, the second generating unit 3060 simulates a situation where a camera to which the camera parameters 130 are applied photographs the objects of interest whose 3D features are represented by the 3D features data 140. By doing so, it can obtain 2D features of the objects of interest that are expected to be captured by a real camera to which the camera parameters 130 are applied.

[0081]

To do so, the second generating unit 3060 may operate as follows. The second generating unit 3060 acquires a configuration (hereinafter, current configuration) of the camera 100 at the time when the camera 100 generates the image data 110. The second generating unit 3060 may determine a configuration of the virtual camera, which is a simulated version of the camera 100 to which the camera parameters 130 are applied.

[0082]

The second generating unit 3060 determines an orientation of the virtual camera, an angle of view of the virtual camera, and a position of the virtual camera based on the camera parameters 130 and the current configuration of the camera 100. By doing so, the second generating unit 3060 determines a monitoring area of the virtual camera in a virtual 3D space in which the 3D features indicated by the 3D features data 140 can be located.

[0083]

The second generating unit 3060 determines the orientation of the virtual camera based on a default orientation of the camera 100, the pan parameter included in the camera parameters 130, and the tilt parameter included in the camera parameters 130. When the camera parameters 130 includes neither the pan parameter nor the tile parameter, the orientation of the virtual camera is set to be the same as the current orientation of the camera 100.

[0084]

When the camera parameters 130 include the zoom parameter, the second generating unit 3060 determines the angle of view of the virtual camera based on a default angle of view of the camera 100 and the zoom parameter included in the camera parameters 130. When the camera parameters 130 do not include the zoom parameter, the angle of view of the virtual camera is set to be the same as the current angle of view of the camera 100.

[0085]

When the camera parameters 130 include the position parameter, the second generating unit 3060 determines the position of the virtual camera based on the position parameter included in the camera parameters 130. When the camera parameters 130 do not include the position parameter, the position of the virtual camera is set to be the same as the current position of the camera 100.

[0086]

After determining the configuration of the virtual camera, the second generating unit 3060 generates the second 2D features data 150 based on the 3D features data 140 and the configuration of the virtual camera. For example, the second generating unit 3060 determines a 2D image plane in the virtual 3D space based on the configuration of the virtual camera, and projects the 3D features of the objects of interest indicated by the 3D features data 140 onto the 2D image plane. By doing so, the second generating unit 3060 computes the 2D features of the objects of interest seen from the virtual camera. Then, the second generating unit 3060 generates the second 2D features data 150 that includes the 2D features of the objects of interest computed by the projection mentioned above.

[0087]

Hereinafter, an example process of generating the second 2D features data 150 is explained with referring to Fig. 9 and Fig. 10. In this example, it is assumed that the analytic task of interest is a face recognition task, and features to be extracted from the image data 110 are face landmarks of human.

[0088]

Fig. 9 illustrates a process of simulating an original situation in which the image data 110 is generated by the camera 100. In Fig. 9, the 3D face landmarks are generated from the image data 110. The configuration of the virtual camera 170 is the same as the configuration of the camera 100 at the time of generating the image data 110.

[0089]

Fig. 10 illustrates a process of simulating a situation where the configuration of the camera 100 is updated with the camera parameters 130. In Fig. 10, the monitoring area of the virtual camera 170 is different from that in Fig. 9 since the configuration of the virtual camera 170 is set based on the camera parameters 130. Then, the second 2D features data 150 is generated by projecting the 3D face landmarks onto a 2D image plane 180, which is determined based on the configuration of the virtual camera 170.

[0090]

As described above, the second generating unit 3060 predicts the 2D features of objects of interest through the simulation. Thus, the action taken by the agent 40 (i.e., the camera parameters 130) can be evaluated without actually configuring the camera 100 with the camera parameters 130 to capture the objects of interest. This fact facilitates the training of the agent 40 and reduces time required for the training of the agent 40.

[0091]

<<Evaluation of Second 2D Features Data 150: S210>>

The evaluating unit 3080 evaluates the second 2D features data 150 to compute the evaluation score 160 (S212). A way of evaluating the 2D features of the objects of interest depends on the analytic task of interest. For example, when the 2D features of the objects are supposed to undergo a face recognition task, the 2D features of the objects should include detailed features of each person's face whereas the 2D features of the objects need not to include of features of body, arms, or legs of each person. On the other hand, when the 2D features of the objects are supposed to undergo a pose estimation task, features of body, arms, and legs of each person should also be included in the 2D features of the objects.

[0092]

Thus, the evaluating unit 3080 evaluates the second 2D features data 150 with respect to the analytic task of interest. In some embodiments, the evaluating unit 3080 includes an evaluator for each one of two or more types of analytic task. The evaluator corresponding to a particular type of analytic task may be configured to take a 2D features data and compute the evaluation score 160 that represents how suitable the 2D features data are for that particular type of analytic task. In this case, the evaluating unit 3080 choses the evaluator corresponding to the type of the analytic task of interest, and computes the evaluation score 160 of the second 2D features data 150 using the chosen evaluator.

[0093]

There are various ways to determine the type of the analytic task of interest. For example, the evaluating unit 3080 acquires an identifier of the type of analytic task of interest that is specified by the user of the training apparatus 3000. Then, the evaluating unit 3080 chooses the evaluator corresponding to the type of the specified identifier. In another example, information that indicates a type of analytic task of interest is stored in a storage unit in advance. In this case, the evaluating unit 3080 acquires this information from this storage unit, and chooses the evaluator corresponding to the type indicated by this information.

[0094]

In some embodiments, the type of the analytic task of interest is predefined. In this case, the evaluating unit 3080 may be configured to include only the evaluator corresponding to the predefined type of analytic task.

[0095]

In some embodiments, the evaluator corresponding to a particular type of analytic task can be implemented with one of conventional algorithms to perform that particular type of analytic task. For example, a human detector may be configured to take 2D features of objects that are extracted from image data, and detect one or more bounding boxes with a confidence score that is a likelihood of the corresponding bounding box representing a human.

[0096]

Thus, for example, when the analytic task of interest is human detection, the evaluating unit 3080 may feed the second 2D features data 150 to the human detector to obtain the confidence score for each human represented by the second 2D features data 150, and computes the evaluation score based on the confidence scores. When the second 2D features data 150 includes the 2D features of a single person, the evaluating unit 3080 obtains a single confidence

score from the human detector. In this case, the evaluating unit 3080 may use this single value as the evaluation score. On the other hand, when the second 2D features data 150 includes the 2D features of two or more persons, the evaluating unit 3080 obtains two or more confidence scores from the human detector. In this case, the evaluating unit 3080 may aggregate those confidence scores (e.g., take an average of those confidence scores) into a single value to use this single value as the evaluation score 160.

[0097]

The above-mentioned way of computing the evaluation score 160 can be applied not only to the case where the analytic task of interest is human detection but also various cases where the analytic task of interest computes a score representing its confidence (or likelihood) of the result of the computation: e.g., detection or recognition of human, detection or recognition of face, or detection of recognition of pose.

[0098]

There are various other ways to compute the evaluation score 160. Hereinafter, examples of other ways to compute the evaluation score 160 will be explained.

[0099]

<<<Example 1>>>

Fig. 11 illustrates an example way of evaluating the camera parameters 130. In this example the evaluating unit 3080 includes a task executor 200 and an evaluating model 220.

[0100]

The task executor 200 performs an analytic task of interest on image data 110. The task executor 200 is configured to take image data, perform the analytic task of interest on the image data, and output a confidence score 210 that represents how confident the task executor 200 is of the result of the analytic task of interest. For example, when the task executor is a face detector, the confidence score may represent a likelihood of a detected face really representing a face.

[0101]

The evaluating model 220 may be a machine learning-based model (e.g., neural network). The evaluating model 220 is configured to take 2D features of an object, compute a suitability score 230 that represents how suitable the 2D features of the object are for the analytic task of interest, and output the suitability score 230.

[0102]

The evaluating model 220 is trained in advance with a plurality of training data that associates an input data and a ground-truth data. The input data represents 2D features of an object.

The ground-truth data represents a score that should be output by the evaluating model in response to the input data associated with that ground-truth data being input into the evaluating model 220. The evaluating model 220 can be trained using one of various loss functions, such as mean square error or mean absolute error.

[0103]

In some embodiments, the training data may be generated using the task executor 200. Specifically, the training apparatus 3000 acquires image data, extracts the 2D features from the image data, and generates the input data that represents the extracted 2D features. In addition, the training apparatus 3000 inputs the image data into the task executor 200, thereby obtaining the confidence score 210 corresponding to that image data. Then, the training apparatus 3000 associates the input data generated from the image data with the confidence score 210 output by the task executor 200 to generate the training data.

[0104]

Suppose that the analytic task of interest is face detection. In this case, the first generating unit 3040 generates the 3D features data 140 that includes 3D face features (e.g., 3D face landmarks or 3D head pose) for each human captured on the image data 110. Then, the second generating unit 3060 generates the second 2D features data 150 that includes the 2D face features (e.g., 2D face landmarks or 2D head pose where depth information is dropped) for each human.

[0105]

In this case, the evaluating model 220 may be pre-trained to output continuous score between 0 and 1, indicating how well 2D face features are suitable for face detection. The score 1 means that the 2D face features are 100% suitable for face detection.

[0106]

The evaluating unit 3080 uses the task executor 200 and the evaluating model 220 to compute the evaluation score 160. Specifically, the evaluating unit 3080 feeds the image data 110 into the task executor 200, thereby obtaining the confidence score 210. When the image data 110 includes two or more objects of interest, the evaluating unit 3080 aggregates the confidence scores of those objects: e.g., taking an average of the confidence scores of those objects.

[0107]

In addition, the evaluating unit 3080 feeds the 2D features of each object of interest included in the second 2D features data 150 into the evaluating model 220, thereby obtaining the suitability score 230 for each object. When the second 2D features data 150 includes 2D features

for two or more objects of interest, the evaluating unit 3080 aggregates the suitability scores of those objects: e.g., taking an average of the suitability scores of those objects.

[0108]

Then, the evaluating unit 3080 computes the evaluation score 160 that represents how much the suitability score 230 is greater than the confidence score 210. For example, the evaluation score 160 is defined as the suitability score 230 subtracted by the confidence score 210, or defined as the suitability score 230 divided by the confidence score 210.

[0109]

<<<Example 2>>>

Fig. 12 illustrates an example way of evaluating the camera parameters 130. In this example, it is assumed that the analytic task of interest is a pose-related task, such as a pose detection task or a pose estimation task. Thus, the 3D features data 140 includes a 3D pose for each object of interest captured on the image data 110 while the 2D features data 150 includes a 2D pose for each object of interest.

[0110]

The evaluating unit 3080 includes a pose lifter 240 that converts a 2D pose into a 3D pose. The pose lifter 240 may be a machine learning-based model, such as a neural network. The pose lifter 240 is configured to take a 2D pose of an object as input, converts the 2D pose of the object into a 3D pose of the object, and output this 3D pose of the object. The pose lifter 240 has been trained in advance.

[0111]

For each 2D pose included in the second 2D features data 150, the evaluating unit 3080 feeds the 2D pose into the pose lifter 240, thereby obtaining a 3D pose for each 2D pose. Then, the evaluating unit 3080 computes a degree of similarity between the 3D poses obtained from the 2D features data 150 using the pose lifter 240 and the 3D poses included in the 3D features data 140, and uses this degree of similarity as the evaluation score 160. This means that the more similar the 3D poses obtained from the second 2D features data 150 are to the 3D poses obtained from the image data 110, the larger the evaluation score 160 is.

[0112]

When the second 2D features data 150 includes two or more 2D poses of the object, the evaluating unit 3080 computes a degree of similarity between the 3D pose of the object obtained from the second 2D features data 150 and the 3D pose of the object included in the 3D features data 140 for each object, and aggregates the computed degrees of similarity into a single value

(e.g., takes average of the computed degrees of similarity) to use this single value as the evaluation score 160.

[0113]

There are various ways to compute a degree of similarity between 3D poses, and one of those ways can be applied to the evaluating unit 3080. For example, the evaluating unit 3080 computes mean per joint position error between the 3D pose included in the 3D features data 140 and the 3D pose obtained from the second 2D features data 150 as the degree of similarity between those 3D poses.

[0114]

<<Update of Agent 40: S212>>

The updating unit 3100 updates the updatable parameters of the agent 40 based on the evaluation score 160 (S212). Conceptually, the evaluation score 160 can be handled as cumulative rewards from the viewpoint of reinforcement learning. Thus, the updating unit 3100 updates the updatable parameters of the agent 40 in a reinforcement learning-based fashion using the evaluation score 160 as the cumulative rewards.

[0115]

There are many algorithms to update updatable parameters of an agent based on cumulative rewards in reinforcement learning, and one of those algorithms can be applied to the updating unit 3100 to update the updatable parameters of the agent 40. Examples of those algorithms are Gradient Descent, Adam Optimizer, RMSProp, Q-learning, Actor-Critic, Proximal Policy Optimization (PPO) and Trust Region Policy Optimization (TRPO).

[0116]

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

provide the program to a computer via a wired communication line (e.g., electric wires, and optical fibers) or a wireless communication line.

[0117]

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

[0118]

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

<Supplementary notes>

(Supplementary Note 1)

A training apparatus comprising:

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

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

acquire image data generated by one or more camera;

acquire one or more camera parameters from an agent, which is configured to take first two-dimensional features data including two-dimensional features extracted from the image data to compute the one or more camera parameters based on the first two-dimensional features data, the one or more camera parameters determine an orientation, a focal length, or a position of the camera;

generate second two-dimensional features data that includes two-dimensional features predicted to be extracted from image data captured by the camera that is configured with the acquired one or more camera parameters;

evaluate the second two-dimensional features data to compute an evaluation score; and

update trainable parameters of the agent based on the evaluation score.

(Supplementary Note 2)

The training apparatus according Claim 1,

wherein the generation of the second two-dimensional features data includes:

extracting three-dimensional features from the acquired image data; and

projecting the three-dimensional features onto a two-dimensional image plane based on the acquired one or more camera parameters to compute the two-dimensional features to be included in the second two-dimensional features data.

(Supplementary Note 3)

The training apparatus according Claim 2,

wherein the two-dimensional features represent one or more two-dimensional poses of objects,

wherein the three-dimensional features represent one or more three-dimensional poses of objects, and

wherein the evaluation of the second two-dimensional features data includes:

converting the one or more two-dimensional poses of objects included in the second two-dimensional features data into one or more three-dimensional poses of objects; and

compute degree of similarity between the one or more three-dimensional poses of objects included in the three-dimensional features data and the one or more three-dimensional poses of objects that are acquired from the second two-dimensional features data to compute the evaluation score.

(Supplementary Note 4)

The training apparatus according Claim 1 or 2,

wherein the evaluation of the second two-dimensional features data includes:

inputting the image data into a task executor that is configured to execute a task of interest on the image data, thereby acquiring a first score that represents how confident the task executor is of a result of the execution of the task of interest;

inputting the second two-dimensional features data into an evaluating model that is configured to compute a second score that represents how suitable the two-dimensional features included in the second two-dimensional features data are for the task of interest, thereby acquiring the second score; and

comparing the first score and the second score to compute the evaluation score.

(Supplementary Note 5)

The training apparatus according any one of Claims 1 to 3,

wherein the update of the agent includes taking the evaluation score as cumulative rewards to compute new values of the trainable parameters of the agent that define a policy or value function with which the one or more camera parameters are determined.

(Supplementary Note 6)

The training apparatus according any one of Claims 1 to 3,

wherein the two-dimensional features include a two-dimensional pose of object, a two-dimensional mesh of object, or two-dimensional landmarks of object.

(Supplementary Note 7)

A camera adjustment apparatus comprising:

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

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

acquire one or more video frames generated by one or more cameras;

extracts two-dimensional features from the one or more video frames; and

compute one of more camera parameters of the camera by inputting the two-dimensional features into an agent and acquiring the one or more camera parameters output by the agent, the one or more camera parameters determine an orientation, a focal length, or a position of the camera,

wherein the agent is configured to take two-dimensional features and has been trained to output one or more camera parameters in response to the two-dimensional features being input thereinto.

(Supplementary Note 8)

A training method comprising:

acquiring image data generated by one or more cameras;

acquiring one or more camera parameters from an agent, which is configured to take first two-dimensional features data including two-dimensional features extracted from the image data to compute the one or more camera parameters based on the first two-dimensional features data, the one or more camera parameters determine an orientation, a focal length, or a position of the camera;

generating second two-dimensional features data that includes two-dimensional features predicted to be extracted from image data captured by the camera that is configured with the acquired one or more camera parameters;

evaluating the second two-dimensional features data to compute an evaluation score; and

updating trainable parameters of the agent based on the evaluation score.

(Supplementary Note 9)

The training method according Claim 8,

wherein the generation of the second two-dimensional features data includes:

extracting three-dimensional features from the acquired image data; and

projecting the three-dimensional features onto a two-dimensional image plane based on the acquired one or more camera parameters to compute the two-dimensional features to be included in the second two-dimensional features data.

(Supplementary Note 10)

The training method according Claim 9,

wherein the two-dimensional features represent one or more two-dimensional poses of objects,

wherein the three-dimensional features represent one or more three-dimensional poses of objects, and

wherein the evaluation of the second two-dimensional features data includes:

converting the one or more two-dimensional poses of objects included in the second two-dimensional features data into one or more three-dimensional poses of objects; and

compute degree of similarity between the one or more three-dimensional poses of objects included in the three-dimensional features data and the one or more three-dimensional poses of objects that are acquired from the second two-dimensional features data to compute the evaluation score.

(Supplementary Note 11)

The training method according Claim 8 or 9,

wherein the evaluation of the second two-dimensional features data includes:

inputting the image data into a task executor that is configured to execute a task of interest on the image data, thereby acquiring a first score that represents how confident the task executor is of a result of the execution of the task of interest;

inputting the second two-dimensional features data into an evaluating model that is configured to compute a second score that represents how suitable the two-dimensional features included in the second two-dimensional features data are for the task of interest, thereby acquiring the second score; and

comparing the first score and the second score to compute the evaluation score.

(Supplementary Note 12)

The training method according any one of Claims 8 to 10,

wherein the update of the agent includes taking the evaluation score as cumulative rewards to compute new values of the trainable parameters of the agent that define a policy or value function with which the one or more camera parameters are determined.

(Supplementary Note 13)

The training method according any one of Claims 8 to 10,

wherein the two-dimensional features include a two-dimensional pose of object, a two-dimensional mesh of object, or two-dimensional landmarks of object.

(Supplementary Note 14)

A camera adjustment method comprising:

acquiring one or more video frames generated by one or more cameras;

extracting two-dimensional features from the one or more video frames; and

computing one of more camera parameters of the camera by inputting the two-dimensional features into an agent and acquiring the one or more camera parameters output by the agent, the one or more camera parameters determine an orientation, a focal length, or a position of the camera,

wherein the agent is configured to take two-dimensional features and has been trained to output one or more camera parameters in response to the two-dimensional features being input thereinto.

(Supplementary Note 15)

A storage medium storing a program that causes a computer to execute:

acquiring image data generated by one or more cameras;

acquiring one or more camera parameters from an agent, which is configured to take first two-dimensional features data including two-dimensional features extracted from the image data to compute the one or more camera parameters based on the first two-dimensional features data, the one or more camera parameters determine an orientation, a focal length, or a position of the camera;

generating second two-dimensional features data that includes two-dimensional features predicted to be extracted from image data captured by the camera that is configured with the acquired one or more camera parameters;

evaluating the second two-dimensional features data to compute an evaluation score; and

updating trainable parameters of the agent based on the evaluation score.

(Supplementary Note 16)

The storage medium according Claim 15,

wherein the generation of the second two-dimensional features data includes:

extracting three-dimensional features from the acquired image data; and

projecting the three-dimensional features onto a two-dimensional image plane based on the acquired one or more camera parameters to compute the two-dimensional features to be included in the second two-dimensional features data.

(Supplementary Note 17)

The storage medium according Claim 16,

wherein the two-dimensional features represent one or more two-dimensional poses of objects,

wherein the three-dimensional features represent one or more three-dimensional poses of objects, and

wherein the evaluation of the second two-dimensional features data includes:

converting the one or more two-dimensional poses of objects included in the second two-dimensional features data into one or more three-dimensional poses of objects; and

compute degree of similarity between the one or more three-dimensional poses of objects included in the three-dimensional features data and the one or more three-dimensional poses of objects that are acquired from the second two-dimensional features data to compute the evaluation score.

(Supplementary Note 18)

The storage medium according Claim 15 or 16,

wherein the evaluation of the second two-dimensional features data includes:

inputting the image data into a task executor that is configured to execute a task of interest on the image data, thereby acquiring a first score that represents how confident the task executor is of a result of the execution of the task of interest;

inputting the second two-dimensional features data into an evaluating model that is configured to compute a second score that represents how suitable the two-dimensional features included in the second two-dimensional features data are for the task of interest, thereby acquiring the second score; and

comparing the first score and the second score to compute the evaluation score.

(Supplementary Note 19)

The storage medium according any one of Claims 15 to 17,

wherein the update of the agent includes taking the evaluation score as cumulative rewards to compute new values of the trainable parameters of the agent that define a policy or value function with which the one or more camera parameters are determined.

(Supplementary Note 20)

The training method according any one of Claims 15 to 17,

wherein the two-dimensional features include a two-dimensional pose of object, a two-dimensional mesh of object, or two-dimensional landmarks of object.

(Supplementary Note 21)

A storage medium storing a program that causes a computer to execute:

acquiring one or more video frames generated by one or more cameras;

extracting two-dimensional features from the one or more video frames; and

computing one of more camera parameters of the camera by inputting the two-dimensional features into an agent and acquiring the one or more camera parameters output by the agent, the one or more camera parameters determine an orientation, a focal length, or a position of the camera,

wherein the agent is configured to take two-dimensional features and has been trained to output one or more camera parameters in response to the two-dimensional features being input thereinto.

[Reference Signs List]

[0119]

10 camera

20 video frame

30 two-dimensional features data

40 agent

50 camera parameter

100 camera

110 image data

120 first two-dimensional features data

130 camera parameter

140 three-dimensional features data

150 second two-dimensional features data

160 evaluation score

170 virtual camera

180 two-dimensional image plane

190 set

200 task executor

210 confidence score

220 evaluating model

230 suitability score

240 pose lifter

1000 computer

1020 bus

1040 processor

1060 memory

1080 storage device

1100 input/output interface

1120 network interface

2000 camera adjustment apparatus

2020 generating unit

2040 computing unit

3000 training apparatus

3020 acquiring unit

3040 first generating unit

3060 second generating unit

3080 evaluating unit

3100 updating unit

[CLAIMS]

[Claim 1]

A training apparatus comprising:

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

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

acquire image data generated by one or more cameras;

acquire one or more camera parameters from an agent, which is configured to take first two-dimensional features data including two-dimensional features extracted from the image data to compute the one or more camera parameters based on the first two-dimensional features data, the one or more camera parameters determine an orientation, a focal length, or a position of the camera;

generate second two-dimensional features data that includes two-dimensional features predicted to be extracted from image data captured by the camera that is configured with the acquired one or more camera parameters;

evaluate the second two-dimensional features data to compute an evaluation score; and

update trainable parameters of the agent based on the evaluation score.

[Claim 2]

The training apparatus according Claim 1,

wherein the generation of the second two-dimensional features data includes:

extracting three-dimensional features from the acquired image data; and

projecting the three-dimensional features onto a two-dimensional image plane based on the acquired one or more camera parameters to compute the two-dimensional features to be included in the second two-dimensional features data.

[Claim 3]

The training apparatus according Claim 2,

wherein the two-dimensional features represent one or more two-dimensional poses of objects,

wherein the three-dimensional features represent one or more three-dimensional poses of objects, and

wherein the evaluation of the second two-dimensional features data includes:

converting the one or more two-dimensional poses of objects included in the second two-dimensional features data into one or more three-dimensional poses of objects; and

compute degree of similarity between the one or more three-dimensional poses of objects included in the three-dimensional features data and the one or more three-dimensional poses of objects that are acquired from the second two-dimensional features data to compute the evaluation score.

[Claim 4]

The training apparatus according Claim 1 or 2,

wherein the evaluation of the second two-dimensional features data includes:

inputting the image data into a task executor that is configured to execute a task of interest on the image data, thereby acquiring a first score that represents how confident the task executor is of a result of the execution of the task of interest;

inputting the second two-dimensional features data into an evaluating model that is configured to compute a second score that represents how suitable the two-dimensional features included in the second two-dimensional features data are for the task of interest, thereby acquiring the second score; and

comparing the first score and the second score to compute the evaluation score.

[Claim 5]

The training apparatus according any one of Claims 1 to 3,

wherein the update of the agent includes taking the evaluation score as cumulative rewards to compute new values of the trainable parameters of the agent that define a policy or value function with which the one or more camera parameters are determined.

[Claim 6]

The training apparatus according any one of Claims 1 to 3,

wherein the two-dimensional features include a two-dimensional pose of object, a two-dimensional mesh of object, or two-dimensional landmarks of object.

[Claim 7]

A camera adjustment apparatus comprising:

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

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

acquire one or more video frames generated by one or more cameras;

extracts two-dimensional features from the one or more video frames; and

compute one of more camera parameters of the camera by inputting the two-dimensional features into an agent and acquiring the one or more camera parameters output by the agent, the one or more camera parameters determine an orientation, a focal length, or a position of the camera,

wherein the agent is configured to take two-dimensional features and has been trained to output one or more camera parameters in response to the two-dimensional features being input thereinto.

[Claim 8]

A training method comprising:

acquiring image data generated by one or more cameras;

acquiring one or more camera parameters from an agent, which is configured to take first two-dimensional features data including two-dimensional features extracted from the image data to compute the one or more camera parameters based on the first two-dimensional features data, the one or more camera parameters determine an orientation, a focal length, or a position of the camera;

generating second two-dimensional features data that includes two-dimensional features predicted to be extracted from image data captured by the camera that is configured with the acquired one or more camera parameters;

evaluating the second two-dimensional features data to compute an evaluation score; and

updating trainable parameters of the agent based on the evaluation score.

[Claim 9]

The training method according Claim 8,

wherein the generation of the second two-dimensional features data includes:

extracting three-dimensional features from the acquired image data; and

projecting the three-dimensional features onto a two-dimensional image plane based on the acquired one or more camera parameters to compute the two-dimensional features to be included in the second two-dimensional features data.

[Claim 10]

The training method according Claim 9,

wherein the two-dimensional features represent one or more two-dimensional poses of objects,

wherein the three-dimensional features represent one or more three-dimensional poses of objects, and

wherein the evaluation of the second two-dimensional features data includes:

converting the one or more two-dimensional poses of objects included in the second two-dimensional features data into one or more three-dimensional poses of objects; and

compute degree of similarity between the one or more three-dimensional poses of objects included in the three-dimensional features data and the one or more three-dimensional poses of objects that are acquired from the second two-dimensional features data to compute the evaluation score.

[Claim 11]

The training method according Claim 8 or 9,

wherein the evaluation of the second two-dimensional features data includes:

inputting the image data into a task executor that is configured to execute a task of interest on the image data, thereby acquiring a first score that represents how confident the task executor is of a result of the execution of the task of interest;

inputting the second two-dimensional features data into an evaluating model that is configured to compute a second score that represents how suitable the two-dimensional features included in the second two-dimensional features data are for the task of interest, thereby acquiring the second score; and

comparing the first score and the second score to compute the evaluation score.

[Claim 12]

The training method according any one of Claims 8 to 10,

wherein the update of the agent includes taking the evaluation score as cumulative rewards to compute new values of the trainable parameters of the agent that define a policy or value function with which the one or more camera parameters are determined.

[Claim 13]

The training method according any one of Claims 8 to 10,

wherein the two-dimensional features include a two-dimensional pose of object, a two-dimensional mesh of object, or two-dimensional landmarks of object.

[Claim 14]

A camera adjustment method comprising:

acquiring one or more video frames generated by one or more cameras;

extracting two-dimensional features from the one or more video frames; and

computing one of more camera parameters of the camera by inputting the two-dimensional features into an agent and acquiring the one or more camera parameters output by the agent, the one or more camera parameters determine an orientation, a focal length, or a position of the camera,

wherein the agent is configured to take two-dimensional features and has been trained to output one or more camera parameters in response to the two-dimensional features being input thereinto.

[Claim 15]

A storage medium storing a program that causes a computer to execute:

acquiring image data generated by one or more cameras;

acquiring one or more camera parameters from an agent, which is configured to take first two-dimensional features data including two-dimensional features extracted from the image data to compute the one or more camera parameters based on the first two-dimensional features data, the one or more camera parameters determine an orientation, a focal length, or a position of the camera;

generating second two-dimensional features data that includes two-dimensional features predicted to be extracted from image data captured by the camera that is configured with the acquired one or more camera parameters;

evaluating the second two-dimensional features data to compute an evaluation score; and

updating trainable parameters of the agent based on the evaluation score.

[Claim 16]

The storage medium according Claim 15,

wherein the generation of the second two-dimensional features data includes:

extracting three-dimensional features from the acquired image data; and

projecting the three-dimensional features onto a two-dimensional image plane based on the acquired one or more camera parameters to compute the two-dimensional features to be included in the second two-dimensional features data.

[Claim 17]

The storage medium according Claim 16,

wherein the two-dimensional features represent one or more two-dimensional poses of objects,

wherein the three-dimensional features represent one or more three-dimensional poses of objects, and

wherein the evaluation of the second two-dimensional features data includes:

converting the one or more two-dimensional poses of objects included in the second two-dimensional features data into one or more three-dimensional poses of objects; and

compute degree of similarity between the one or more three-dimensional poses of objects included in the three-dimensional features data and the one or more three-dimensional poses of objects that are acquired from the second two-dimensional features data to compute the evaluation score.

[Claim 18]

The storage medium according Claim 15 or 16,

wherein the evaluation of the second two-dimensional features data includes:

inputting the image data into a task executor that is configured to execute a task of interest on the image data, thereby acquiring a first score that represents how confident the task executor is of a result of the execution of the task of interest;

inputting the second two-dimensional features data into an evaluating model that is configured to compute a second score that represents how suitable the two-dimensional features included in the second two-dimensional features data are for the task of interest, thereby acquiring the second score; and

comparing the first score and the second score to compute the evaluation score.

[Claim 19]

A storage medium storing a program that causes a computer to execute:

acquiring one or more video frames generated by one or more cameras;

extracting two-dimensional features from the one or more video frames; and

computing one of more camera parameters of the camera by inputting the two-dimensional features into an agent and acquiring the one or more camera parameters output by the agent, the one or more camera parameters determine an orientation, a focal length, or a position of the camera,

wherein the agent is configured to take two-dimensional features and has been trained to output one or more camera parameters in response to the two-dimensional features being input thereinto.