IMAGE PROCESSING APPARATUS, IMAGE PROCESSING METHOD, AND STORAGE MEDIUM

[Technical Field]

[0001]

The present invention relates to an image processing apparatus, an image processing method, and a program.

[Background Art]

[0002]

In recent years, in a surveillance system and the like, a technique for detecting and searching for a state such as a pose and behavior of a person from an image of a surveillance camera is used. For example, Patent Documents 1 and 2 have been known as related techniques. Patent Document 1 discloses a technique for searching for a similar pose of a person, based on a key joint of a head, a hand, a foot, and the like of the person included in a depth video. Patent Document 2 discloses a technique for searching for a similar image by using pose information such as a tilt provided to an image, which is not related to a pose of a person. Patent Document 3 discloses a technique for receiving an input that specifies a person desired to be searched on an image, and searching for an image including the specified person. Patent Document 4 discloses a technique for acquiring, from an image, pose information about a search target being formed of a plurality of feature points, and searching for an image including a pose similar to a pose determined in the pose information. Note that, in addition, Non-Patent Document 1 has been known as a technique related to a skeleton estimation of a person.

[Related Document]

[Patent Document]

[0003]

[Patent Document 1] Japanese Patent Application Publication (Translation of PCT Application) No. 2014-522035

[Patent Document 2] Japanese Patent Application Publication No. 2006-260405

[Patent Document 3] Japanese Patent Application Publication No. 2020-13290

standing pose. The pose region WA2 is, for example, a display region of a sitting pose, and displays a skeleton structure and a person that are classified into the cluster C2 and are similar to the sitting pose. The pose region WA3 is, for example, a display region of a sleeping pose, and displays a skeleton structure and a person that are classified into the cluster C3 and are similar to the sleeping pose.

[0086]

Subsequently, as illustrated in Fig. 3, the image processing apparatus 100 performs the search processing (S105). In the search processing, as illustrated in Fig. 5, the image processing apparatus 100 receives an input of a search condition (S121).

[0087]

One example of the processing in S121 will be described by using Fig. 49. The image acquisition unit 101 acquires a query image, based on any of the acquisition examples 1 to 3 described above (S301). Next, the selection unit 109 selects at least one person included in the query image, based on the processing example 1 or 2 of selecting at least one person included in the query image described above (S302). Next, the feature value computation unit 103 computes a feature value of a two-dimensional skeleton structure of the person selected in S302 (S303). The feature value of the two-dimensional skeleton structure is a search query.

[0088]

Another example of the processing in S121 will be described by using Fig. 50. The image acquisition unit 101 acquires a query image, based on any of the acquisition examples 1 to 3 described above (S311). Next, the image processing apparatus 100 performs processing of detecting a person on the query image (S312). For example, the selection unit 109 may perform processing of detecting a person on a whole query image (whole region in an image), based on processing of detecting a known person. In addition, the skeleton structure detection unit 102 may perform processing of detecting a two-dimensional skeleton structure of a person on a whole query image (whole region in an image).

[0089]

Then, when the number of persons included in the image determined

based on a result of the detection processing in S312 is plural (Yes in S313), the selection unit 109 selects at least one person from the plurality of persons included in the image (S314). Subsequently, the feature value computation unit 103 computes a feature value of a two-dimensional skeleton structure of the at least one person selected in S314 (S316). The feature value of the two-dimensional skeleton structure is a search query. The processing in S314 and S316 is similar to the processing in S302 and S303 in Fig. 49.

[0090]

On the other hand, when the number of persons included in the image determined based on a result of the detection processing in S312 is one (No in S313), the selection unit 109 selects the detected one person (S315). Then, the feature value computation unit 103 computes a feature value of a two-dimensional skeleton structure of the one person selected in S315 (S316). The feature value of the two-dimensional skeleton structure is a search query.

[0091]

Note that, the "number of persons included in the image determined based on a result of the detection processing in S312" may be the number of persons detected in S312, or may be the number of "selectable persons" described above. A concept of the "selectable person" is as described in "Processing Examples 1 and 2 of Detecting Person and Deciding Person to be Displayed as Selectable Person".

[0092]

Returning to Fig. 5, subsequently, the search unit 105 searches for a skeleton structure, based on the search condition (S122). With the skeleton structure specified by the user as the search query, the search unit 105 searches for a skeleton structure having a high degree of similarity of a feature value from among the skeleton structures that are search targets and are stored in the database 110. The search unit 105 computes a degree of similarity between a feature value of the skeleton structure being the search query and a feature value (feature value of the skeleton structure detected from the analysis target image) of the skeleton structure being the search target, and extracts a skeleton structure having the computed degree of similarity higher than a predetermined threshold value. The feature value of the skeleton structure being the search query may use a feature value being computed in advance, or may use a feature

CLAIMS

1. An image processing apparatus comprising:

image acquisition means for acquiring an image;

selection means for selecting at least one person included in the image, based on a user input;

skeleton structure detection means for detecting a two-dimensional skeleton structure of a person included in the image;

feature value computation means for computing a feature value of the detected two-dimensional skeleton structure; and

search means for searching, from analysis target images, for the analysis target image including a person in a state similar to a state of the selected person, based on a degree of similarity to a feature value of the two-dimensional skeleton structure.

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

the selection means

receives a user input that specifies at least a partial region in the image, and

selects a person detected in the specified partial region.

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

the selection means

displays, on the image, a processed image in which a person detected in the image is displayed as a selectable person in an identifiable manner, and

receives a user input that selects at least one person from the selectable person.

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

the selection means displays the processed image in which a person whose detection result of the two-dimensional skeleton structure satisfies a predetermined condition is displayed as the selectable person in an identifiable manner.

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

the skeleton structure detection means extracts a plurality of keypoints of a body in processing of detecting the two-dimensional skeleton structure, and

a person who satisfies the predetermined condition is

a person having a number of the extracted keypoints equal to or more than a reference value, or

a person in which an evaluation value computed based on at least one of a number of the extracted keypoints and reliability of each of the extracted keypoints is equal to or more than a reference value.

6. The image processing apparatus according to any one of claims 1 to 5, wherein

the selection means selects a part of a body of the selected person, based on a user input that specifies at least a partial region in the image, and

the search means searches for the analysis target image including a person in a state similar to a state of the selected person by setting a weight of the selected part of the body to be greater than that of another part.

7. The image processing apparatus according to any one of claims 1 to 6, wherein

the selection means

determines a number of persons included in the image being acquired by the image acquisition means,

selects, when a determined number of persons is plural, at least one person from a plurality of persons included in the image, based on a user input, and

selects, when a determined number of persons is one, the one person.

8. An image processing method comprising,

by a computer:

acquiring an image;

selecting at least one person included in the image, based on a user

input;

detecting a two-dimensional skeleton structure of a person included in the image;

computing a feature value of the detected two-dimensional skeleton structure; and

searching, from analysis target images, for the analysis target image including a person in a state similar to a state of the selected person, based on a degree of similarity to a feature value of the two-dimensional skeleton structure.

9. A storage medium storing a program causing a computer to function as:

image acquisition means for acquiring an image;

selection means for selecting at least one person included in the image, based on a user input;

skeleton structure detection means for detecting a two-dimensional skeleton structure of a person included in the image;

feature value computation means for computing a feature value of the detected two-dimensional skeleton structure; and

search means for searching, from analysis target images, for the analysis target image including a person in a state similar to a state of the selected person, based on a degree of similarity to a feature value of the two-dimensional skeleton structure.