DESCRIPTION

Title of Invention:

IMAGE PROCESSING DEVICE, IMAGE PROCESSING METHOD, AND IMAGE PROCESSING PROGRAM

Technical Field

[0001] The present invention relates to an image processing device, an image processing method, and an image processing program.

Background Art

[0002] Conventionally, there has been known a technique of estimating the number, density, flow rate, or the like of persons from camera images. As a technique of estimating the number of persons from camera images, there is a method of counting the number of persons based on person detection, a method of estimating the number of persons from a foreground area, or the like. The former has an advantage in the aspect of precision when the density of crowd is low, but when the density of crowd becomes high, there is a problem that the precision of detection decreases due to the influence of occlusion (covering) of persons from each other. The latter is inferior to the former in the precision of analysis when the density is low, but is capable of processing with a low arithmetic operation amount even when the density is high.

[0003] For example, in Patent Literature 1, the latter technique is disclosed. Specifically, in Patent Literature 1, a foreground extracted by a background subtraction from an image in which a crowd is captured is specified as a person region. In Patent Literature 1, the number of persons in the image is then estimated based on the area of the person region. In Patent Literature 1, a CG (Computer Graphics) model simulating a crowd in advance is generated with a plurality of congestion degrees. Subsequently, a relational equation of a foreground area and the number of persons having occlusion

of a crowd with each other taken into consideration is derived, thereby making it possible to estimate the number of persons while suppressing the influence of occlusion.

Citation List

Patent Literature

[0004] Patent Literature 1: JP 2005-25328 A

Summary of Invention

Technical Problem

[0005] The foreground area per person obtained by a background difference varies depending on actually measured dimensions of persons, the precision of foreground extraction due to illumination conditions, and the like. Accordingly, the technique of Patent Literature 1 has a problem that a CG model generated in advance and a foreground extracted from actual camera images do not match, and thus errors in person-number estimation result occur.

[0006] It is a main object of the present invention to solve the problems described above. Specifically, it is a main object of the present invention to improve the precision in estimating the number of subjects shown in a captured image.

Solution to Problem

[0007] An image processing device according to the present invention includes:

a reference-image generating unit to extract, from a first captured image that is a captured image in which an image-capturing space where a subject exists is captured and the subject is shown, a subject image that is an image of the subject, and to generate a plurality of reference images in which a plurality of subjects exist in the image-capturing space based on the extracted subject image while varying number of subjects existing in the image-capturing space for each reference image; and

a subject-number estimating unit to compare a second captured image that is a

captured image of the image-capturing space at an image-capturing time different from an image-capturing time of the first captured image with the plurality of reference images, and to estimate number of subjects shown in the second captured image.

Advantageous Effects of Invention

[0008] According to the present invention, it is possible to improve the precision in estimating the number of subjects shown in a captured image.

Brief Description of Drawings

[0009] Fig. 1 is a diagram illustrating a functional configuration example of a crowd monitoring device according to a first embodiment.

Fig. 2 is a diagram illustrating an internal configuration example of a parameter obtaining unit according to the first embodiment.

Fig. 3 is a diagram illustrating an internal configuration example of a person-number analyzing unit according to the first embodiment.

Fig. 4 is a flowchart for illustrating an operation example of the crowd monitoring device according to the first embodiment.

Fig. 5 is a flowchart for illustrating an operation example of the parameter obtaining unit according to the first embodiment.

Fig. 6 is a flowchart for illustrating an operation example of the person-number analyzing unit according to the first embodiment.

Fig. 7 is a diagram illustrating an example of a person detection result according to the first embodiment.

Fig. 8 is a diagram illustrating an example of a foreground extraction result according to the first embodiment.

Fig. 9 is a diagram illustrating an example of information stored for each detected person according to the first embodiment.

Fig. 10 is an explanatory diagram of a depression-angle estimating method of a camera according to the first embodiment.

Fig. 11 is an explanatory diagram of a deriving method of a horizontal line height according to the first embodiment.

Fig. 12 is an explanatory diagram of a deriving method of vertical point coordinates according to the first embodiment.

Fig. 13 is a diagram illustrating an example of a crowd foreground image of a congestion level 1 according to the first embodiment.

Fig. 14 is a diagram illustrating an example of a crowd foreground image of a congestion level 2 according to the first embodiment.

Fig. 15 is a diagram illustrating an example of a crowd foreground image of a congestion level 3 according to the first embodiment.

Fig. 16 is a diagram illustrating an example of a foreground area amount of the crowd foreground image of the congestion level 1 according to the first embodiment.

Fig. 17 is a diagram illustrating an example of a foreground area amount of the crowd foreground image of the congestion level 2 according to the first embodiment.

Fig. 18 is a diagram illustrating an example of a foreground area amount of the crowd foreground image of the congestion level 3 according to the first embodiment.

Fig. 19 is an explanatory diagram of a person-number estimating process according to the first embodiment.

Fig. 20 is an explanatory diagram of the person-number estimating process according to the first embodiment.

Fig. 21 is an explanatory diagram of the person-number estimating process according to the first embodiment.

Fig. 22 is a diagram illustrating a hardware configuration example of the crowd

monitoring device according to the first embodiment.

Fig. 23 is a diagram illustrating a relation between a functional configuration and a hardware configuration of the crowd monitoring device according to the first embodiment.

Description of Embodiments

[0010] Embodiments of the present invention will be explained below with reference to the drawings. In the following descriptions of the embodiments and the drawings, elements denoted by the same reference signs indicate the same or corresponding parts.

[0011] First embodiment.

Descriptions of configurations

Fig. 1 illustrates a functional configuration example of a crowd monitoring device 20 according to a first embodiment.

Further, Fig. 22 illustrates a hardware configuration example of the crowd monitoring device 20 according to the first embodiment.

The crowd monitoring device 20 corresponds to an image processing device. Operations performed by the crowd monitoring device 20 correspond to an image processing method and an image processing program.

As illustrated in Fig. 1, the crowd monitoring device 20 is connected to a camera 10.

The camera 10 is installed at a position looking down an image-capturing space where a subject exists. The image-capturing space is a space of a monitoring target of the camera 10. In the present embodiment, the subject is a person. Further, in the present embodiment, it is assumed that a plurality of persons exist in the image-capturing space. In the following descriptions, a plurality of persons existing in the image-capturing space are also referred to as crowd.

[0012] Prior to the descriptions of the functional configuration example of the crowd monitoring device 20 of Fig. 1, the hardware configuration example of the crowd monitoring device 20 is described with reference to Fig. 22.

[0013] The crowd monitoring device 20 is a computer.

The crowd monitoring device 20 includes, as hardware, a processor 1101, a memory 1102, a network interface 1103, and a storage device 1104.

The storage device 1104 stores therein a program for implementing functions of an image receiving and decoding unit 201, a changeover switch 202, a mode-changeover control unit 203, a parameter obtaining unit 204, a person-number analyzing unit 206, and an analysis-result output unit 207 illustrated in Fig. 1. The program is loaded from the storage device 1104 to the memory 1102. The processor 1101 reads the program from the memory 1102. The processor 1101 then executes the program to perform operations of the image receiving and decoding unit 201, the changeover switch 202, the mode-changeover control unit 203, the parameter obtaining unit 204, the person-number analyzing unit 206, and the analysis-result output unit 207, which will be described later.

A parameter storage 205 illustrated in Fig. 1 is implemented by the storage device 1104.

Fig. 23 illustrates a relation between the functional configuration illustrated in Fig. 1 and the hardware configuration illustrated in Fig. 22. Fig. 23 illustrates that the image receiving and decoding unit 201, the changeover switch 202, the mode-changeover control unit 203, the parameter obtaining unit 204, the person-number analyzing unit 206, and the analysis-result output unit 207 are implemented by the processor 1101. Further, Fig. 23 illustrates that the parameter storage 205 is implemented by the storage device 1104.

The network interface 1103 receives a compressed image stream from the camera 10.

[0014] Next, the functional configuration example of the crowd monitoring device 20 illustrated in Fig. 1 is described.

[0015] The image receiving and decoding unit 201 decodes a compressed image stream delivered from the camera 10 to convert the compressed image stream into an image frame. The image frame is a captured image of an image-capturing space.

[0016] The mode-changeover control unit 203 controls the operation mode of the crowd monitoring device 20. As the operation mode of the crowd monitoring device 20, there are a parameter obtaining mode and a person-number counting mode. The mode-changeover control unit 203 outputs a mode control signal to the changeover switch 202.

[0017] The changeover switch 202 switches an output destination of the image frame according to the mode control signal from the mode-changeover control unit 203. More specifically, when the operation mode of the crowd monitoring device 20 is the parameter obtaining mode, the changeover switch 202 outputs the image frame to the parameter obtaining unit 204. On the other hand, when the operation mode of the crowd monitoring device 20 is the person-number counting mode, the changeover switch 202 outputs the image frame to the person-number analyzing unit 206. The image frame output from the changeover switch 202 to the parameter obtaining unit 204 corresponds to a first captured image. The image frame output from the changeover switch 202 to the person-number analyzing unit 206 corresponds to a second captured image. The image frame output from the changeover switch 202 to the person-number analyzing unit 206 is a captured image of the image-capturing space at an image-capturing time different from an image-capturing time of the image frame output

from the changeover switch 202 to the parameter obtaining unit 204.

[0018] When the operation mode of the crowd monitoring device 20 is the parameter obtaining mode, the parameter obtaining unit 204 obtains an analysis parameter for person number analysis by using the image frame.

More specifically, the parameter obtaining unit 204 estimates a depression angle of the camera 10 as an external parameter of the camera 10. Further, the parameter obtaining unit 204 extracts a subject image that is an image of a subject from the image frame as the first captured image, and generates a plurality of foreground maps based on the extracted subject image and the estimated depression angle of the camera 10. The parameter obtaining unit 204 generates the plurality of foreground maps while changing the number of subjects existing in the image-capturing space in each foreground map. The foreground map is sectioned into a plurality of partial regions. The foreground map indicates the number of subjects (a regional subject number) shown in a partial region and a foreground area amount in the partial region for each partial region. The foreground map is an image to be compared with the image frame at the time of person number analysis by the person-number analyzing unit 206. The foreground map corresponds to a reference image.

The parameter obtaining unit 204 stores the generated plurality of foreground maps as analysis parameters in the parameter storage 205.

The parameter obtaining unit 204 corresponds to a reference-image generating unit. Further, a process performed by the parameter obtaining unit 204 correspond to a reference-image generating process.

[0019] The parameter storage 205 stores therein analysis parameters generated by the parameter obtaining unit 204.

[0020] The person-number analyzing unit 206 performs person number analysis when

the operation mode of the crowd monitoring device 20 is the person-number counting mode.

More specifically, the person-number analyzing unit 206 compares an image frame as a second captured image with the plurality of foreground maps generated by the parameter obtaining unit 204, and estimates the number of subjects shown in the image frame.

The person-number analyzing unit 206 corresponds to a subject-number estimating unit. A process performed by the person-number analyzing unit 206 correspond to a subject-number estimating process.

[0021] The analysis-result output unit 207 outputs a result of the person number analysis by the person-number analyzing unit 206 to outside.

[0022] Fig. 2 illustrates an internal configuration example of the parameter obtaining unit 204.

The parameter obtaining unit 204 is constituted by a person detecting unit 2041, a foreground extracting unit 2042, a depression-angle estimating unit 2043, a foreground-map generating unit 2044, and a person information storage 2045.

[0023] The person detecting unit 2041 detects an image of a person as a subject from the image frame as the first captured image.

[0024] The foreground extracting unit 2042 extracts a foreground image of a person identified by the person detecting unit 2041.

[0025] The depression-angle estimating unit 2043 estimates a depression angle of the camera 10 based on person-detection result information of a large number of persons and the foreground image.

[0026] The foreground-map generating unit 2044 generates a plurality of foreground maps based on the depression angle of the camera 10, the person-detection result

information of a large number of persons, and the foreground image.

The foreground-map generating unit 2044 then stores the foreground map as an analysis parameter in the parameter storage 205.

[0027] The person information storage 2045 stores therein the person-detection result information and the foreground image.

[0028] Fig. 3 illustrates an internal configuration example of the person-number analyzing unit 206.

The person-number analyzing unit 206 is constituted by a foreground extracting unit 2061 and a person-number estimating unit 2062.

[0029] The foreground extracting unit 2061 extracts a foreground image from the image frame as the second captured image.

[0030] The person-number estimating unit 2062 performs person number estimation based on the foreground image by using the analysis parameter stored in the parameter storage 205.

[0031] ***Descriptions of operations***

Next, an operation example of the crowd monitoring device 20 according to the first embodiment is described.

Fig. 4 is a flowchart for illustrating an operation example of the crowd monitoring device 20.

[0032] First, immediately after the crowd monitoring device 20 is activated, the mode-changeover control unit 203 sets the operation mode as the parameter obtaining mode (Step ST01). That is, the mode-changeover control unit 203 outputs a mode control signal to notify the changeover switch 202 of the parameter obtaining mode.

[0033] Next, the mode-changeover control unit 203 refers to the parameter storage 205 and confirms whether an analysis parameter is already obtained (Step ST02). That is,

the mode-changeover control unit 203 confirms whether the analysis parameter is stored in the parameter storage 205.

[0034] When the analysis parameter is not obtained yet (NO in ST02), the process proceeds through Step ST04 and Steps ST05, ST06, and the analysis parameter is stored in the parameter storage 205.

In Step ST02, which is after the analysis parameter is stored in the parameter storage 205, it is determined that the analysis parameter is already obtained, and the process proceeds to Step ST03.

[0035] Meanwhile, when the analysis parameter is already obtained (YES in Step ST02), the mode-changeover control unit 203 changes the operation mode to the person-number counting mode (Step ST03).

When the analysis parameter is already stored in the parameter storage 205 before the crowd monitoring device 20 is activated, Step ST02 is determined as YES. Step ST02, which is after the process proceeds through Step ST04 and Steps ST05, ST06 and the analysis parameter is stored in the parameter storage 205 as described above, is also determined as YES.

[0036] Next, the image receiving and decoding unit 201 receives a compressed image stream from the camera 10, and decodes an image frame of at least one frame of the received compressed image stream (Step ST04). For example, the image receiving and decoding unit 201 receives, as the compressed image stream, image-encoded data compressed by an image compression encoding method such as H.262/MPEG-2 video, H.264/AVC, H.265/HEVC, or JPEG. Further, as the compressed image stream, for example, the image receiving and decoding unit 201 may receive image data that is IP-delivered by an image delivery protocol such as MPEG-2 TS (Moving Picture Experts Group 2 Transport Stream), RTP/RTSP (Real-time Transport Protocol/Real

Time Streaming Protocol), MMT (MPEG Media Transport), or DASH (Dynamic Adaptive Streaming over HTTP).

Further, the image receiving and decoding unit 201 may receive, as the compressed image stream, image data encoded by an encoding method other the above ones or image data delivered in a delivery format other than the above ones. The image receiving and decoding unit 201 may receive, as the compressed image stream, image data delivered with non-compression transmission standards such as SDI (Serial Digital Interface), HD (High-Definition)-SDI, or the like.

[0037] Next, the changeover switch 202 outputs the image frame output from the image receiving and decoding unit 201 to the parameter obtaining unit 204 or the person-number analyzing unit 206 according to the operation mode set by the mode-changeover control unit 203 (Step ST05).

That is, if a mode control signal to notify of the parameter obtaining mode is output from the mode-changeover control unit 203, the changeover switch 202 outputs the image frame to the parameter obtaining unit 204. On the other hand, if a mode control signal to notify of the person-number counting mode is output from the mode-changeover control unit 203, the changeover switch 202 outputs the image frame to the person-number analyzing unit 206.

In the following descriptions, a configuration of exclusively operating the parameter obtaining unit 204 and the person-number analyzing unit 206 is assumed. However, it is also possible to employ a configuration of operating the parameter obtaining unit 204 while simultaneously operating the person-number analyzing unit 206, thereby updating the analysis parameter as needed.

[0038] When the operation mode is the parameter obtaining mode, the parameter obtaining unit 204 performs a parameter obtaining process (Step ST06). That is, the

parameter obtaining unit 204 sequentially processes the image frame from the changeover switch 202, generates an analysis parameter, and stores the generated analysis parameter in the parameter storage 205. The parameter obtaining unit 204 generates the analysis parameter by processing the image frame of a certain period. Upon completion of the generation of the analysis parameter, the parameter obtaining unit 204 stores the analysis parameter in the parameter storage 205. When the analysis parameter is stored in the parameter storage 205, the process returns to Step ST02. Details of operations of the parameter obtaining unit 204 are described later.

[0039] When the operation mode is the person-number counting mode, the person-number analyzing unit 206 performs a person-number counting process (Step ST07). That is, the person-number analyzing unit 206 analyzes the image frame from the changeover switch 202 by using the analysis parameter stored in the parameter storage 205, thereby analyzing the number of persons shown in this mage frame. Details of operations of the person-number analyzing unit 206 are described later.

[0040] After the number of persons is analyzed by the person-number analyzing unit 206, the analysis-result output unit 207 outputs a person-number analysis result indicating an analysis result of the person-number analyzing unit 206 to outside (Step ST08).

The analysis-result output unit 207 outputs the person-number analysis result by displaying it on a monitor, outputting it to a log file, outputting it to an externally connected device, or transmitting it to a network, for example. Further, the analysis-result output unit 207 may output the person-number analysis result by any other format. Furthermore, the person-number analysis result may be output to outside each time the person-number analysis result is output from the person-number analyzing unit 206, or may be output intermittently such as outputting a person-number analysis

result of a specific period or a specific number after being totalized or statistically processed. After Step ST08, the process returns to Step ST04, and processing of the next image frame is performed.

[0041] In the following descriptions, detailed operations of the parameter obtaining unit 204 are explained.

Fig. 5 is a flowchart for illustrating operations of the parameter obtaining unit 204.

[0042] First, the person detecting unit 2041 performs a person detection process on an image frame input, and stores person-detection result information in the person information storage 2045 (Step ST11).

The person detecting unit 2041 detects a person satisfying that the person is standing, that a contact position between toes and the ground and a head top part are visible, and that no occlusion with another person is occurring or the ratio of occlusion with another person is small.

[0043] The person detecting unit 2041 outputs person-detection result information to the person information storage 2045 for each detected person. For example, the person-detection result information includes information indicating a frame having a rectangular, elliptic, or another shape that sufficiently surrounds a detected person, coordinate information indicating a grounding position between the ground and the toes of a detected person, coordinate information indicating the head top part of a person, and a moving speed of a grounding point.

Fig. 7 illustrates an image of a person detection result by the person detecting unit 2041.

As a method for person detection, a method by using a feature amount of still image basis, such as HOG (Histogram of Oriented Gradients), ICF (Integral Channel

Features), and ACF (Aggregate Channel Features) may be used. Further, as a method for person detection, a method by using a feature amount of video basis by using a plurality of image frames that are temporally close may also be used.

Further, in the present embodiment, the person detecting unit 2041 uses a person detection method on an assumption that the size of a person shown in the image frame is unknown. Instead of this, the person detecting unit 2041 may perform a plurality of times of person detection, and when a tendency toward the scale of a person of each image area is obtained, the method may be switched to the person detection method assuming this scale, so as to further accelerate the process.

The person detecting unit 2041 obtains information of the moving speed of a grounding point by tracking a characteristic point across a plurality of image frames, tracking a person, and the like. Further, when a motion vector for compression coding is included in the compressed image stream input from the camera 10, the person detecting unit 2041 may use motion vector information thereof as it is, so as to obtain the information of the moving speed of the grounding point.

[0044] The parameter obtaining unit 204 may be configured to store all detection results as person-detection result information in the person information storage 2045, or may be configured to suppress the amount of data stored in the person information storage 2045 to be a certain amount or less, in order to save the capacity of the person information storage 2045. For example, when the person detecting unit 2041 detects another person again in an image area in which a person is already detected, the person detecting unit 2041 discards any of the detection results to suppress the amount of data accumulated in the person information storage 2045 to be a certain amount or less. The person detecting unit 2041 may take an average value of detection results to integrate the detection results, thereby suppressing the amount of data accumulated in

the person information storage 2045 to be a certain amount or less.

[0045] Next, the foreground extracting unit 2042 performs, based on the person-detection result information, foreground extraction within a region of a frame surrounding the detected person, and stores an image of the foreground extraction result thereof in the person information storage 2045 by associating the image with the person-detection result information (Steps ST12, ST13).

The foreground image extracted by the foreground extracting unit 2042 corresponds to a subject image.

[0046] Fig. 8 illustrates an image of a foreground extraction result.

Further, Fig. 9 illustrates an image of a pair of person-detection result information and a foreground image of a person corresponding to each piece of the person-detection result information, which is stored in the person information storage 2045. Note that the person-detection result information and the foreground image are referred to as person information in combination.

[0047] For example, the foreground extracting unit 2042 extracts a foreground image by using a background subtraction method, an adaptive-type background subtraction method, and a dense optical flow derivation algorithm.

The background subtraction method is a method of registering a background image in advance and calculating a difference between an input image and the background image. The adaptive-type background subtraction method is a method of automatically updating a background image by using a model such as MOG (Mixture of Gaussian) Distribution based on sequentially input image frames. The dense optical flow derivation algorithm is a method of obtaining motion information in an image in pixel units.

[0048] Next, the foreground extracting unit 2042 determines whether sufficient person

information is obtained by the processes of Steps ST11 to ST13 (Step ST14).

When sufficient person information is obtained, the process proceeds to Step ST15. On the other hand, when sufficient person information is not obtained, the process is ended.

For example, determination as to whether sufficient person information is obtained is made on measures such as the number of pieces of person information that have been obtained hitherto and an elapsed time since the parameter obtaining process is started. It is also possible that a parameter estimating process of Step ST15 and thereafter is performed, and the determination as to whether sufficient person information is obtained may be performed by the reliability of a parameter estimated by the parameter estimating process.

[0049] When sufficient person information is obtained, the depression-angle estimating unit 2043 performs estimation of a depression angle of the camera 10 by using the person information stored in the person information storage 2045 (Step ST15).

Figs. 10, 11, and 12 illustrate an outline of a depression-angle estimating process.

As illustrated in Fig. 10, in the present embodiment, it is assumed that a camera is installed downward from horizontal. It is also assumed that the lens distortion of the camera is corrected, and the camera can be considered as a pinhole camera. It is also assumed that internal parameters of the camera are known, and each angle formed between a direction of each coordinate and an optical axis in an image is known.

In this case, if an image coordinate corresponding to a horizontal line height is known in the image, an angle α[rad] formed between an optical axis direction and a horizontal direction is uniquely obtained. The angle formed between the optical axis direction and the ground, that is, a depression angle θ[rad] can be obtained by $\theta=\alpha$.

Alternatively, if image coordinates corresponding to a vertical point in the image are known, an angle β[rad] formed between the optical axis direction and the vertical direction is uniquely obtained. The depression angle θ[rad] is obtained by $\theta=\pi/2-\beta$.

[0050] Fig. 11 illustrates an image of deriving a horizontal line height.

The depression-angle estimating unit 2043 extends a direction of movement of a grounding point of a moving person based on person information, and assumes the point where the most extended lines of moving directions of a plurality of persons are crossing as a point on the horizontal line. This is based on an assumption that a plurality of pedestrians are walking in parallel and straight. Due to the influence such as precision errors of moving direction detection and a person not walking in parallel, respective lines do not always cross at one point. Therefore, the depression-angle estimating unit 2043 uses many pieces of person information to estimate a more likely result as the horizontal height.

[0051] Fig. 12 illustrates an image of deriving a vertical point coordinate.

The depression-angle estimating unit 2043 extends a line connecting coordinates of a head top part and a grounding point, that is, a line in a normal direction based on person information, and assumes the point where the most lines in the normal direction of a plurality of persons are crossing as a vertical point. This is based on an assumption that pedestrians are standing upright on the ground. Due to the influence such as detection errors of coordinates of the head top part and the grounding point and a pedestrian not standing upright, respective lines do not always cross at one point. Therefore, the depression-angle estimating unit 2043 uses many pieces of person information to estimate a more likely result as vertical point coordinates.

[0052] The method of using a horizontal line height is suitable for a case where the depression angle of the camera is shallow, that is, the depression angle of the camera is

close to a horizontal direction. The method of using a vertical point is suitable for a case where the depression angle of the camera is deep, that is, the depression angle of the camera is closer to a vertical direction. The depression-angle estimating unit 2043 may select a depression angle obtained by a more suitable method out of depression angles obtained by using the both methods, or may select an average value of the both angles. Alternatively, the depression-angle estimating unit 2043 may use other depression-angle estimating methods.

[0053] Next, the foreground-map generating unit 2044 generates a foreground map used by the person-number analyzing unit 206, by using depression angle information obtained by the depression-angle estimating unit 2043 and person information stored in the person information storage 2045 (Step ST16).

[0054] More specifically, the foreground-map generating unit 2044 first calculates a two-dimensional coordinate system (hereinafter, referred to as road surface coordinate system) on the road surface from depression angle information and movement information of a person grounding point. For example, the foreground-map generating unit 2044 defines a main moving direction of the person as an X direction, and defines a direction orthogonal to the X direction as a Y direction.

[0055] Next, the foreground-map generating unit 2044 obtains an absolute scale of road surface coordinates as coordinates on the road surface coordinate system. For example, the foreground-map generating unit 2044 assumes a moving width in a unit time on road surface coordinates as a moving width at an average moving speed of a pedestrian given in advance, and obtains the absolute scale of the road surface coordinates.

[0056] Next, after the road surface coordinate system is obtained, the foreground-map generating unit 2044 combines foreground images of each person obtained in Step ST13

to generate a foreground image of a crowd expected to be observed with a specific congestion degree (hereinafter, referred to as crowd foreground image).

Figs. 13, 14, and 15 illustrate examples of a crowd foreground image. Fig. 13 illustrates an example of a crowd foreground image of a congestion level 1. Fig. 14 illustrates an example of a crowd foreground image of a congestion level 2. Fig. 15 illustrates an example of a crowd foreground image of a congestion level 3.

As illustrated in Figs. 13 to 15, the foreground-map generating unit 2044 generates a plurality of crowd foreground images in which a plurality of persons exist in an image-capturing space while varying the number of persons existing in the image-capturing space for each crowd foreground image.

Further, as illustrated in Figs. 13 to 15, in the crowd foreground images, occlusion occurs between persons depending on the density and positions of the persons.

The foreground-map generating unit 2044 converts image coordinates of a grounding point of person information obtained hitherto into road surface coordinates, and then attaches the foreground image at a plurality of positions where the density of persons becomes a predetermined value on the road surface coordinate system, thereby generating a crowd foreground image. For example, it is assumed that the grid of the road surface coordinates illustrated in Fig. 13 is at an interval of 50 cm on the road surface coordinate system. That is, the entire grid is 4 [m^2]. When the crowd density of the congestion level 1 is 1 [person/m^2], the foreground-map generating unit 2044 randomly locates four persons in a grid. When the crowd density of the congestion level 2 is 2 [person/m^2], the foreground-map generating unit 2044 randomly locates eight persons in the grid. When the crowd density of the congestion level 3 is 4 [person/m^2], the foreground-map generating unit 2044 randomly locates 16 persons in

the grid. In any case, for example, the foreground-map generating unit 2044 randomly locates persons so as to keep a shoulder width average value of persons and a distance equal to or more than a personal-space radius average value per density, so that the persons are not too close to each other.

[0057] Next, after generating a crowd foreground image, as illustrated in Figs. 16, 17, and 18, the foreground-map generating unit 2044 sections the crowd foreground image into a plurality of partial regions, and calculates a foreground area amount per person (example: 150 pixels/person) for each congestion level. In Figs. 17 and 18, for the drawing purposes, the foreground area amount is not displayed on some of the persons.

The foreground-map generating unit 2044 calculates a foreground area amount included in a partial region for each partial region. For example, in a partial region denoted by reference sign 161 in Fig. 16, for example, 90 pixels/partial region is obtained. Since the crowd foreground image is generated by combining foreground images of a plurality of persons, the foreground-map generating unit 2044 can obtain a true value of the number of persons shown in each partial region. The foreground-map generating unit 2044 counts the number of persons shown in a partial region for each partial region. When the whole foreground image of one person is included in one partial region, the number of person shown in this partial region is one (1 person/partial region). When only a foreground image of a part of a person is included in one partial region, the foreground-map generating unit 2044 divides the total area of the foreground image of the person by the area of the part of the person included in the partial region to obtain the number of persons included in this partial region in units after a decimal point. For example, the foreground area amount of the person included in the partial region of reference sign 161 is 150 pixels. When it is assumed that 90 pixels are included in the partial region of reference sign 161, the foreground-map generating unit 2044 defines

the number of persons included in the partial region of reference sign161 as 0.6 person (90 ÷ 150 = 0.6) (0.6 person/partial region).

Subsequently, the foreground-map generating unit 2044 adds, to each partial region, the foreground area amount (example: 90 pixels/partial region) of a person included in a partial region and the number of persons (example: 1 person/partial region, 0.6 person/partial region) included in the partial region.

In the following descriptions, the foreground area amount (example: 90 pixels/partial region) of a person included in a partial region is referred to as regional person area amount. The number of persons (example: 1 person/partial region, 0.6 person/partial region) included in a partial region is referred to as regional person number.

A crowd foreground image in which the regional person area amount and the regional person number are added to each partial region is referred to as foreground map. As described above, the foreground map corresponds to a reference image.

[0058] The foreground-map generating unit 2044 may generate a plurality of crowd foreground images for each congestion level. That is, the foreground-map generating unit 2044 may generate a plurality of crowd foreground images while randomly varying the arrangement of persons with respect to one congestion level (for example, the congestion level 3). In this manner, when a plurality of crowd foreground images are generated for one congestion level, the foreground-map generating unit 2044 may store all of a plurality of foreground maps obtained from the plurality of crowd foreground images in the parameter storage 205. Further, the foreground-map generating unit 2044 may obtain an average value of the plurality of foreground maps obtained from the plurality of crowd foreground images to integrate the plurality of foreground maps into one foreground map, thereby storing only the one foreground map in the parameter

storage 205. Due to this process, an effect of being capable of handling various arrangement patterns of persons is expected.

[0059] Finally, the person information storage 2045 stores the generated foreground map in the parameter storage 205 (Step ST17). By storing the foreground map, the parameter obtaining process is completed.

[0060] In the following descriptions, detailed operations of the person-number analyzing unit 206 are described. Fig. 6 is a flowchart for illustrating operations of the person-number analyzing unit 206.

[0061] First, the foreground extracting unit 2061 performs foreground extraction with respect to a whole image frame input or a region of interest (Step ST21). This foreground extraction process is the same process as the foreground extraction process of the foreground extracting unit 2042 of the parameter obtaining unit 204.

[0062] Next, the person-number estimating unit 2062 performs person number estimation by using foreground maps for each congestion level stored in the parameter storage 205 on a foreground image extracted by the foreground extracting unit 2061 (Step ST22).

Figs. 19, 20, and 21 explain a person-number estimating process.

As illustrated in Fig. 19, the person-number estimating unit 2062 extracts a foreground image from an image frame. The person-number estimating unit 2062 then sections the foreground image into partial regions same as those of the foreground maps. In the following descriptions, the partial regions in the foreground image illustrated in Fig. 19 are referred to as estimation-target partial regions.

The person-number estimating unit 2062 determines a foreground area amount for each estimation-target partial region as illustrated in Fig. 20.

Next, for each estimation-target partial region, the person-number estimating

unit 2062 extracts partial regions at the same positions from a plurality of foreground maps for each congestion level, and compares the foreground area amount of each of the extracted partial regions with a foreground area amount in the estimation-target partial region. The person-number estimating unit 2062 then selects a partial region of a foreground map having a foreground area amount that is most similar to the foreground area amount in the estimation-target partial region.

For example, the person-number estimating unit 2062 extracts a partial region of the first row and the first column from each of foreground maps of the congestion levels 1 to 3 with respect to the estimation-target partial region of the first row and the first column (the left end of the top row) of the foreground image of Fig. 20. Next, the person-number estimating unit 2062 compares the foreground area amount in each of three partial regions extracted from the foreground maps of the congestion levels 1 to 3 with a foreground area amount in the estimation-target partial region of Fig. 20. The person-number estimating unit 2062 then selects, from the three extracted partial regions, a partial region to which the foreground area amount that is most similar to the foreground area amount in the estimation-target partial region of Fig. 20 is added. The person-number estimating unit 2062 performs the above process for all the estimation-target partial regions of Fig. 20. In the following descriptions, the partial region selected for the estimation-target partial region is referred to as selected partial region.

In this manner, the person-number estimating unit 2062 selects a selected partial region from the foreground maps of the congestion levels 1 to 3 for each estimation-target partial region. Therefore, the congestion level of the selected partial region may be different depending on the estimation-target partial region. For example, there may be a case where a partial region of the foreground map of the

congestion level 3 is selected for the estimation-target partial region of nth row and mth column of the foreground image of Fig. 20, a partial region of the foreground map of the congestion level 2 is selected for the estimation-target partial region of nth row and m+1th column, and a partial region of the foreground map of the congestion level 1 is selected for the estimation-target partial region of nth row and m+2th column.

Next, the person-number estimating unit 2062 obtains the number of persons included in an estimation-target partial region for each estimation-target partial region. Specifically, the person-number estimating unit 2062 divides the foreground area amount of the estimation-target partial region by the regional person area amount of the selection partial region to multiply the divided value by the regional person number of the selected partial region. That is, the person-number estimating unit 2062 performs a calculation of (foreground area amount of estimation-target partial region) ÷ (regional person area amount of selected partial region) × (regional person number of selected partial region), thereby obtaining the number of persons included in the estimation-target partial region.

Subsequently, the person-number estimating unit 2062 totalizes the number of persons obtained for each estimation-target partial region to obtain the number of persons included in the whole image frame or a region of interest.

[0063] ***Descriptions of effects of embodiment***

As described above, in the present embodiment, the parameter obtaining unit 204 obtains a regional person area amount and a regional person number in consideration of occlusion by using an actually measured foreground image. The person-number analyzing unit 206 then applies the regional person area amount and the regional person number to the foreground image of a person-number estimation target to perform person number estimation. Therefore, highly precise person number

estimation is possible without obtaining average dimensions of persons in advance.

Further, in the present embodiment, since the parameter obtaining unit 204 estimates the depression angle of a camera that is an external parameter of the camera, it is not necessary to measure the external parameter of the camera in advance. In the technique disclosed in Patent Literature 1, it is necessary to measure and obtain camera external parameters in advance for generating a CG model. For example, the camera external parameters are parameters such as a depression angle of a camera and a distance from each of scattered persons to the camera. In this manner, in the technique disclosed in Patent Literature 1, since it is necessary to measure and obtain the camera external parameters in advance, there is a problem of large cost at the time of installing the camera. In the present embodiment, as described above, there is an effect where it is not necessary to measure and obtain the external parameters of a camera in advance.

[0064] In the above descriptions, while explanations have been made as a person is exemplified as the subject, the subject is not limited to a person. For example, the subject may be a living matter such as a wild animal and an insect, or may be a mobile object other than a person, such as a vehicle.

[0065] Further, in the invention of the present application, within the scope of the invention, arbitrary combinations, modifications, and omissions of the constituent elements or procedures described in the embodiment can be made.

[0066] ***Descriptions of hardware configuration***

Finally, supplementary descriptions of a hardware configuration of the crowd monitoring device 20 are provided.

A processor 901 illustrated in Fig. 22 is an IC (Integrated Circuit) that performs processing.

The processor 1101 is a CPU (Central Processing Unit), a DSP (Digital Signal

Processor), or the like.

The memory 1102 illustrated in Fig. 22 is a RAM (Random Access Memory).

The storage device 1104 illustrated in Fig.22 is a ROM (Read Only Memory), a flash memory, an HDD (Hard Disk Drive), or the like.

The network interface 1103 illustrated in Fig. 22 includes a receiver that receives data and a transmitter that transmits data.

The network interface 1103 is a communication chip or a NIC (Network Interface Card), for example.

[0067] An OS (Operating System) is also stored in the storage device 1104.

At least a part of the OS is executed by the processor 1101.

The processor 1101 executes programs for realizing functions of the image receiving and decoding unit 201, the changeover switch 202, the mode-changeover control unit 203, the parameter obtaining unit 204, the person-number analyzing unit 206, and the analysis-result output unit 207 while executing at least a part of the OS.

The processor 1101 executes the OS, and task management, memory management, file management, communication control, and the like are performed.

Further, at least any one of pieces of information, data, signal values, and variable values indicating the results of processing performed by the image receiving and decoding unit 201, the changeover switch 202, the mode-changeover control unit 203, the parameter obtaining unit 204, the person-number analyzing unit 206, and the analysis-result output unit 207 is stored at least in any one of the memory 1102, the storage device 1104, a register and a cache memory in the processor 1101.

Further, the programs for realizing the functions of the image receiving and decoding unit 201, the changeover switch 202, the mode-changeover control unit 203, the parameter obtaining unit 204, the person-number analyzing unit 206, and the

analysis-result output unit 207 can be stored in portable memory media such as a magnetic disk, a flexible disk, an optical disk, a compact disk, a Blue-ray (registered trademark) disk, and a DVD.

[0068] "Unit" of each of the image receiving and decoding unit 201, the changeover switch 202, the mode-changeover control unit 203, the parameter obtaining unit 204, the person-number analyzing unit 206, and the analysis-result output unit 207 can be replaced with "circuit", "step", "procedure", or "process".

The crowd monitoring device 20 can be realized by an electronic circuit such as a logic IC (Integrated Circuit), a GA (Gate Array), an ASIC (Application Specific Integrated Circuit), and an FPGA (Field-Programmable Gate Array).

In this case, each of the image receiving and decoding unit 201, the changeover switch 202, the mode-changeover control unit 203, the parameter obtaining unit 204, the person-number analyzing unit 206, and the analysis-result output unit 207 is realized as a part of an electronic circuit.

The processor and the electronic circuit described above are also collectively referred to as processing circuitry.

Reference Signs List

[0069] 10: camera; 20: crowd monitoring device; 201: image receiving and decoding unit; 202: changeover switch; 203: mode-changeover control unit; 204: parameter obtaining unit; 205: parameter storage; 206: person-number analyzing unit; 207: analysis-result output unit; 1101: processor; 1102: memory; 1103: network interface; 1104: storage device; 2041: person detecting unit; 2042: foreground extracting unit; 2043: depression-angle estimating unit; 2044: foreground-map generating unit; 2045: person information storage; 2061: foreground extracting unit; 2062: person-number estimating unit.