

DESCRIPTION

Title of Invention:

CROWD DENSITY CALCULATION APPARATUS, CROWD DENSITY
CALCULATION METHOD, AND CROWD DENSITY CALCULATION PROGRAM

5 **Technical Field**

[0001] The present invention relates to a crowd density calculation apparatus, a crowd density calculation method, and a crowd density calculation program.

Background Art

[0002] There is a technique of estimating the number or density of persons from a
10 camera recorded video. The technique of estimating the number of persons from the camera recorded video includes a technique, such as a method of counting the number of persons based on person detection or a method of estimating the number of persons from foreground area.

According to the method based on person detection, the number of persons can
15 be estimated with high accuracy when the crowd density is low. However, according to this method, an amount of arithmetic processing increases as the number of persons increases. Further, according to this method, the crowd density raises as the number of persons increases. This results in decrease in estimation accuracy, being influenced by mutual occlusion, that is, hiding, among persons.

20 According to the method of estimating the number of persons from foreground area, the estimation accuracy is lower compared with the method based on person detection when the crowd density is low. However, according to this method, the amount of arithmetic processing does not change even when the crowd density is high.

Note that the technique of estimating the density of persons is equivalent to a
25 technique of estimating the number of persons for each arbitrary region in a video

frame.

[0003] Patent Literatures 1 and 2 disclose techniques of acquiring a recorded video of a crowd of persons, and thereby estimating the number of persons on a screen from the area of a person region that is the foreground extracted by background differencing.

5 Patent Literature 1 calculates a weighting value that quantitatively indicates a level of contribution of each pixel in an image to the number of persons. The weighting value is calculated from apparent volume of a target object in the image. This resolves a problem that the foreground area per unit number of persons appears differently depending on difference in depth, and allows estimation of the number of
10 persons even from an image with depth.

In addition, the Patent Literature 2 generates computer graphics (CG) models that are imitations of a crowd of persons, in a plurality of congestion levels, in advance, and derives a relation between the foreground area and the number of persons considering occlusion among persons in the crowd. Then, Patent Literature 2 allows
15 the estimation of the number of persons, the influence of the occlusion being controlled.

Citation List

Patent Literature

[0004] Patent Literature 1: JP 2009-294755 A

Patent Literature 2: JP 2005-025328 A

20 Summary of Invention

Technical Problem

[0005] Techniques disclosed in Patent Literatures 1 and 2 calculate the number of persons present in a video frame and the density of persons in an arbitrary region in the video frame. However, they do not estimate a position of a person in a real-world
25 physical space. This is because that Patent Literatures 1 and 2 relate a point in the

physical space to a point in the video frame, but not vice versa.

[0006] An object of the present invention is to calculate positions of persons in crowd in a real-world physical space from a video frame and output them as density distribution of a crowd of persons.

5 **Solution to Problem**

[0007] A crowd density calculation apparatus according to the present invention includes:

a video acquisition unit to acquire a video frame from a video stream including a picked-up video image of a person; and

10 an analysis unit to relate three-dimensional coordinates to the video frame, acquire as each three-dimensional region of a plurality of three-dimensional regions, a region representing each three-dimensional space of a plurality of three-dimensional spaces that are acquired based on the three-dimensional coordinates in the video frame, and based on number of persons present in each three-dimensional region of the
15 plurality of three-dimensional regions, calculate density distribution of persons in the video frame as crowd density distribution.

Advantageous Effects of Invention

[0008] In a crowd density calculation apparatus according to the present invention, an analysis unit relates three-dimensional coordinates to a video frame, and acquires as
20 each three-dimensional region of a plurality of three-dimensional regions, a region representing each three-dimensional space of a plurality of three-dimensional spaces that are acquired based on the three-dimensional coordinates in the video frame. The analysis unit also calculates, based on number of persons present in each three-dimensional region of the plurality of three-dimensional regions, crowd density
25 distribution in the video frame. Therefore, the crowd density calculation apparatus

according to the present invention allows a quantitative grasp of crowd density distribution in a real-world physical space from a video frame.

Brief Description of Drawings

[0009] Fig. 1 is a configuration diagram of a crowd density calculation apparatus

5 according to Embodiment 1;

Fig. 2 is a detailed configuration diagram of an analysis unit according to Embodiment 1;

Fig. 3 is a diagram explaining a definition of crowd density distribution;

Fig. 4 is a diagram illustrating an image of crowd density distribution in a case
10 where it is assumed that sizes of ΔX and ΔY are fixed;

Fig. 5 is a diagram illustrating an image of projecting a point in a real-world physical space on a video frame coordinate system;

Fig. 6 is a flowchart of a crowd density calculation process according to Embodiment 1;

15 Fig. 7 is a flowchart of an analysis process according to Embodiment 1;

Fig. 8 is a diagram illustrating an image of converting foreground area into number of persons for each three-dimensional region;

Fig. 9 is a diagram illustrating an image of outputting provisional density distribution from number of persons in each three-dimensional region;

20 Fig. 10 is a diagram illustrating an image of a correct foreground image;

Fig. 11 is a diagram illustrating an image of numerical congestion levels of the correct foreground image;

Fig. 12 is a diagram illustrating an image of a presence determination process according to Embodiment 1;

25 Fig. 13 is a configuration diagram of a crowd density calculation apparatus

according to a modification of Embodiment 1;

Fig. 14 is a detailed configuration diagram of an analysis unit according to Embodiment 2;

Fig. 15 is a flowchart of an analysis process according to Embodiment 2;

5 Fig. 16 is a diagram illustrating an image of a position correction process according to Embodiment 2; and

Fig. 17 is a flowchart of an analysis process according to Embodiment 3.

Description of Embodiments

[0010] Embodiments of the present invention will be explained with drawings below.

10 Note that a same or corresponding part in each of the drawings is denoted by a same reference sign. In descriptions of the embodiments, description on the same or corresponding part is omitted or simplified according to circumstances.

[0011] Embodiment 1.

***** Description of Configuration *****

15 With Fig. 1, a configuration of a crowd density calculation apparatus 100 according to the present embodiment will be explained.

The crowd density calculation apparatus 100 is a computer. The crowd density calculation apparatus 100 includes a processor 910, and also includes other hardware, such as a memory 921, an auxiliary storage device 922, an input interface 930,
20 an output interface 940, and a communication device 950. The processor 910 is connected with these other hardware via a signal line, and controls these other hardware.

[0012] The crowd density calculation apparatus 100 includes, as functional components, a video acquisition unit 110, an analysis unit 120, a result output unit 130,
25 and a storage unit 140. The storage unit 140 stores an analysis parameter 141 that is

used in an analysis process by the analysis unit 120.

[0013] Functions of the video acquisition unit 110, the analysis unit 120, and the result output unit 130 are realized by software. The storage unit 140 is included in a memory 921. The storage unit 140 may be included in the auxiliary storage device
5 922.

[0014] The processor 910 is a device that executes a crowd density calculation program. The crowd density calculation program is a program that realizes the functions of the video acquisition unit 110, the analysis unit 120, and the result output unit 130.

10 The processor 910 is an integrated circuit (IC) that performs an arithmetic processing. A specific example of the processor 910 is a CPU, a digital signal processor (DSP), or a graphics processing unit (GPU).

[0015] The memory 921 is a storage device that temporarily stores data. A specific example of the memory 921 is a static random access memory (SRAM), or a dynamic
15 random access memory (DRAM).

[0016] The auxiliary storage device 922 is a storage device that saves data. A specific example of the auxiliary storage device 922 is an HDD. Alternatively, the auxiliary storage device 922 may be a portable storage medium, such as an SD (registered trademark) memory card, CF, NAND flash, a flexible disk, an optical disc, a
20 compact disc, a Blue-ray (registered trademark) disc, or a DVD. Note that HDD is an abbreviation of hard disk drive. SD (registered trademark) is an abbreviation of Secure Digital. CF is an abbreviation of CompactFlash. DVD is an abbreviation of digital versatile disk.

[0017] The input interface 930 is a port that is connected with an input device, such as
25 a mouse, keyboard, or touch panel. Alternatively, the input interface 930 may be a

port that is connected with a camera 200. The input interface 930 is, in specific, a Universal Serial Bus (USB) terminal. Note that the input interface 930 may be a port that is connected with a local area network (LAN). Via the input interface 930, the crowd density calculation apparatus 100 may acquire a video stream 21 from the camera
 5 200.

[0018] The output interface 940 is a port with which a cable of an output device, such as a display, is connected. The output interface 940 is, in specific, a USB terminal or a High Definition Multimedia Interface (HDMI, registered trademark) terminal. The display is, in specific, a liquid crystal display (LCD). Via the output interface 940, the
 10 crowd density calculation apparatus 100 presents on the display, an analysis result outputted by the result output unit 130.

[0019] The communication device 950 communicates with other device via a network. The communication device 950 has a receiver and a transmitter. The communication device 950 is connected with a communication network, such as a LAN, Internet, or a
 15 telephone line, through wire or by radio. The communication device 950 is, in specific, a communication chip or a network interface card (NIC). Via the communication device 950, the crowd density calculation apparatus 100 may receive the video stream 21 from the camera 200. Via the communication device 950, the crowd density calculation apparatus 100 may also transmit to an external device, the analysis result
 20 outputted by the result output unit 130.

[0020] A crowd density calculation program is loaded into the processor 910, and executed by the processor 910. The memory 921 stores not only the crowd density calculation program, but also an operating system (OS). The processor 910 executes the crowd density calculation program while executing the OS. The crowd density
 25 calculation program and the OS may be stored in the auxiliary storage device 922. The

crowd density calculation program and the OS stored in the auxiliary storage device 922 are loaded into the memory 921, and executed by the processor 910. Note that the crowd density calculation program may be installed partly or entirely in the OS.

[0021] The crowd density calculation apparatus 100 may include a plurality of
 5 processors that substitute for the processor 910. These plurality of processors share a task to execute the crowd density calculation program among the plurality of processors. Each of the processors is a device that executes the crowd density calculation program as the processor 910 is.

[0022] Data, information, a signal value, and a variable value that are used, processed,
 10 or outputted by the crowd density calculation program are stored in the memory 921, the auxiliary storage device 922, or a register or a cache memory in the processor 910.

[0023] Replacing “unit” of each of the video acquisition unit 110, the analysis unit 120, and the result output unit 130 by “process”, “procedure”, or “step”, the crowd density calculation program causes a computer to execute the each process, procedure,
 15 or step. And, a crowd density calculation method is a method that is carried out by the crowd density calculation apparatus 100 executing the crowd density calculation program.

The crowd density calculation program may be provided as being stored in a computer-readable storage medium. The crowd density calculation program may be
 20 provided also as a program product.

[0024] With Fig. 2, a detailed configuration of the analysis unit 120 according to the present embodiment will be explained.

The analysis unit 120 includes a foreground extraction unit 121, a provisional density calculation unit 122, a presence determination unit 123, a standardization unit
 25 124, and a distribution output unit 125. In other words, each unit of the video

acquisition unit 110, the analysis unit 120, and the result output unit 130 is each unit of the video acquisition unit 110, the foreground extraction unit 121, the provisional density calculation unit 122, the presence determination unit 123, the standardization unit 124, the distribution output unit 125, and the result output unit 130.

5 [0025] With Figs. 1 and 2, a summary of each functional component of the crowd density calculation apparatus 100 will be explained.

The crowd density calculation apparatus 100 is connected with the camera 200 that picks up a video image of an object and distributes it as a video stream 21. The object is, in specific, a person. In other words, the video stream 21 is a video of a crowd of persons.

The video acquisition unit 110 acquires, via the input interface 930, the video stream 21 that is distributed from the camera 200. The video acquisition unit 110 acquires the video frame 22 from the video stream 21. In specific, the video acquisition unit 110 decodes the video stream 21, and transforms it into the video frame 22.

The analysis unit 120 relates three-dimensional coordinates to the video frame 22. The analysis unit 120 acquires as each three-dimensional region of a plurality of three-dimensional regions, a region representing each three-dimensional space of a plurality of three-dimensional spaces that are acquired based on the three-dimensional coordinates on the video frame 22. Then, the analysis unit 120 calculates, based on number of persons present in each three-dimensional region of the plurality of three-dimensional regions, crowd density distribution in the video frame 22. In other words, the analysis unit 120 calculates, using the video frame 22, positions of persons in crowd in three-dimensional coordinates in a physical space as crowd density distribution 225.

The result output unit 130 outputs to an output device, such as a display, via the output interface 940, the crowd density distribution 225 outputted from the analysis unit 120.

[0026] Next, a schema of each functional component that the analysis unit 120 has will be explained.

The foreground extraction unit 121 extracts from the video frame 22, a portion that has a characteristic of foreground as a foreground image 221. Using the foreground image 221 and the analysis parameter 141 stored in the storage unit 140, the provisional density calculation unit 122 calculates provisional density distribution 222, which is apparent crowd density distribution corresponding to each position in a physical space. By determining a position where there is no person present in the physical space, the presence determination unit 123 corrects the provisional density distribution 222. The presence determination unit 123 outputs the corrected provisional density distribution 222 as corrected density distribution 223. Using a total number of persons present in an image represented by the video frame 22, the standardization unit 124 standardizes the corrected density distribution 223, and then outputs fixed density distribution 224. The distribution output unit 125 transforms the fixed density distribution 224 into an output manner, and finally outputs it as the crowd density distribution 225.

[0027] With Fig. 3, a definition of the crowd density distribution will be explained. The crowd density distribution is distribution of number of persons in a real-world physical space.

As illustrated in Fig. 3, it is assumed that, on a physical space coordinate system X_r - Y_r - Z_r , a surrounded region with a width of ΔX and a depth of ΔY at one position (X_i, Y_j) on a plane X_r - Y_r is ΔS_{ij} , the physical space coordinate system X_r - Y_r - Z_r

being real-world three-dimensional coordinates, the plane X_r - Y_r being corresponding to real-world ground. And, it is assumed that a region of a surrounded three-dimensional space of a prism with base ΔS_{ij} and a height of H is three-dimensional space V_{ij} . It is assumed that number of persons present in the three-dimensional space V_{ij} is h_{ij} . The number of persons h_{ij} being related to ΔS_{ij} and such h_{ij} for all analyzed regions being laid out makes the crowd density distribution. It is assumed that the height of H is an approximate height of a human.

[0028] Fig. 4 is a diagram illustrating an image of crowd density distribution in a case where it is assumed that sizes of ΔX and ΔY are fixed. The crowd density distribution in a case where there is one person each in region ΔS_{02} and at an intermediate point between region ΔS_{10} and region ΔS_{20} is as described next. In this case, the crowd density distribution shows distribution of one person in a position of ΔS_{02} , 0.5 person in a position of ΔS_{10} , 0.5 person in a position of ΔS_{20} , and zero in remaining positions of ΔS_{ij} .

Note that the sizes of ΔX and ΔY are not prescribed. The sizes of ΔX and ΔY may be variable.

[0029] Fig. 5 is a diagram illustrating an image of projecting a point in a real-world physical space on a video frame coordinate system. In other words, Fig. 5 is an image of relating three-dimensional coordinates to a video frame. It is assumed that a point on ground in a real-world physical space is $P_{gij} = (X_{ij}, Y_{ij}, 0)$. It is assumed that a point on a plane at a height of H in the real-world physical space is $P_{hij} = (X_{ij}, Y_{ij}, H)$. It is assumed that P_{gij} , the point on the ground, and P_{hij} , the point on the plane at the height of H , being projected on a video frame coordinate system x_{img} - y_{img} are points p_{gij} and p_{hij} . Information that relates P_{gij} to p_{gij} and P_{hij} to p_{hij} is stored in the storage unit 140 as the analysis parameter 141. It is also assumed that a projected region of three-dimensional

space V_{ij} in the physical space on the video frame coordinate system is
 three-dimensional region v_{ij} . The three-dimensional region v_{ij} is a two-dimensional
 region illustrated with oblique lines in Fig. 5. In other words, the three-dimensional
 region v_{ij} is a two-dimensional region in the video frame drawn by a perimeter of the
 5 three-dimensional space V_{ij} when the three-dimensional space V_{ij} on the
 three-dimensional coordinates is drawn in the video frame. The three-dimensional
 region v_{ij} is called a prismatic region or a cuboid region.

[0030] When a person is standing in each three-dimensional space V_{ij} of a plurality of
 three-dimensional spaces, each three-dimensional region v_{ij} of a plurality of
 10 three-dimensional regions has a head region to which a head of the person corresponds
 and a ground region corresponding to ground on which the person is standing. The
 head region is a region surrounded by p_{hij} , p_{hi+1j} , p_{hij+1} , and $p_{hi+1j+1}$ at a height
 equivalent to a position of the head in the three-dimensional region v_{ij} . And, the
 ground region is a region surrounded by p_{gij} , p_{gi+1j} , p_{gij+1} , and $p_{gi+1j+1}$ equivalent to a
 15 position of the ground in the three-dimensional region v_{ij} .

[0031] Information that relates coordinates in a physical space and coordinates on a
 video frame to each other may be a coordinate transformation formula or a pair of the
 coordinates in the physical space and the coordinates on the video frame corresponding
 to each other.

20 Also, it is not necessary that all P_{gij} or all P_{hij} are on a same plane. As long as
 region V_{ij} in a three-dimensional space is definable, a face represented by each P_{gij} or
 each P_{hij} may be curved or stepped.

[0032] *** Description of Operation ***

With Fig. 6, a crowd density calculation process S100 by the crowd density
 25 calculation apparatus 100 according to the present embodiment will be explained.

[0033] <Analysis Parameter Load Process>

In step ST01, the crowd density calculation apparatus 100 loads an analysis parameter 141 into the storage unit 140. The analysis parameter 141 may be stored in the auxiliary storage device 922, or may be inputted from outside via the input interface
 5 930 or the communication device 950. The loaded analysis parameter 141 is used by the analysis unit 120.

[0034] <Video Acquisition Process>

In step ST02, the video acquisition unit 110 stands ready to receive a video stream 21 from the camera 200. Upon receiving the video stream 21 from the camera
 10 200, the video acquisition unit 110 decodes at least one frame of the received video stream 21. The video stream to be received here is, for example, video coded data, which is compressed by a video compression coding format, being IP-delivered by a video delivery protocol. A specific example of the video compression coding format is H.262/MPEG-2 video, H.264/AVC, H.265/HEVC, or JPEG. A specific example of
 15 the video delivery protocol is MPEG-2 TS, RTP/RTSP, MMT, or DASH. MPEG-2 TS is an abbreviation of Moving Picture Experts Group 2 Transport Stream. RTP/RTSP is an abbreviation of Real-time Transport Protocol/Real Time Streaming Protocol. MMT is an abbreviation of MPEG Media Transport. DASH is an abbreviation of Dynamic Adaptive Streaming over HTTP. The video stream to be
 20 received may be in a coding or delivery format other than the above, or may be of an uncompressed transmission standard, such as SDI or HD-SDI. SDI is an abbreviation of serial digital interface. HD-SDI is an abbreviation of high-definition serial digital interface.

[0035] <Analysis Process>

25 In step ST03, the analysis unit 120 acquires a video frame 22 from the video

acquisition unit 110. The analysis unit 120 analyses the video frame 22, using the analysis parameter 141. By analyzing the video frame 22, the analysis unit 120 calculates crowd density distribution of a crowd of persons in the video frame 22. The analysis unit 120 also transforms the calculated crowd density distribution into an output manner, and outputs it as crowd density distribution 225.

[0036] <Result Output Process>

In step ST04, the result output unit 130 outputs the crowd density distribution 225 outputted from the analysis unit 120 outside the crowd density calculation apparatus 100 via the output interface 940. An output manner can be, for example, displaying on a monitor, outputting to a log file, outputting to an externally connected device, or transmitting to a network. The result output unit 130 may output the crowd density distribution 225 in a manner other than the above. The result output unit 130 may also output the crowd density distribution 225 outside every time the crowd density distribution 225 is outputted from the analysis unit 120. Alternatively, the result output unit 130 may perform intermittent output by outputting the crowd density distribution 225 after performing aggregation or statistical processing thereof for a certain period or a certain number. After step ST04, the crowd density calculation apparatus 100 returns to step ST02, and carries out processes for a next video frame 22.

[0037] <<Detail of Analysis Process>>

With Fig. 7, a detail of the analysis process according to the present embodiment will be explained.

In step ST11, the foreground extraction unit 121 extracts an image of a person, that is, foreground, in the video frame 22 as a foreground image 221. The foreground extraction unit 121 outputs the foreground image 221 to the provisional density calculation unit 122.

[0038] Fig. 8 is a diagram illustrating an image of converting foreground area into number of persons for each three-dimensional region. Fig. 9 is a diagram of outputting provisional density distribution from the number of persons in each three-dimensional region. In Fig. 8, an image of two persons is extracted as the foreground image 221.

5 As to a technique for a foreground extraction process, there is a background differencing method that pre-registers a background image, and calculates difference between the pre-registered background image and an inputted image. There is also an adaptive background differencing method that automatically updates a background image based on image frames that are continuously inputted, using a model called a
10 mixture of Gaussian distribution (MOG). There is also a dense optical flow deviation algorithm that acquires motion information in an image in pixels.

[0039] In step ST12, the provisional density calculation unit 122 calculates as provisional density distribution 222, number of persons apparently present in each three-dimensional region of a plurality of three-dimensional regions based on the
15 foreground image 221. In specific, the provisional density calculation unit 122 projects a point in a physical space on a video frame coordinate system, and after that, using the foreground image 221 and a relation 142, calculates the number of persons present in each three-dimensional region v_{ij} . The provisional density calculation unit 122 uses provisional crowd density estimation as a method to calculate the number of
20 persons present in each three-dimensional region v_{ij} .

[0040] As illustrated in Fig. 8, the provisional density calculation unit 122 aggregates the foreground area for each three-dimensional region v_{ij} in the video frame 22. The provisional density calculation unit 122 calculates the number of persons in each three-dimensional region v_{ij} based on the foreground area in each three-dimensional
25 region v_{ij} . In doing so, the provisional density calculation unit 122, using a relation

142 of the foreground area and the number of persons that is found in advance,
calculates the number of persons in each three-dimensional region v_{ij} . The provisional
density calculation unit 122 regards the number of persons in each three-dimensional
region v_{ij} as number of persons h_{ij} in region ΔS_{ij} corresponding to each

5 three-dimensional region v_{ij} . The provisional density calculation unit 122 outputs
crowd density distribution representing number of persons h_{ij} calculated for all regions
 ΔS_{ij} as the provisional density distribution 222, as illustrated in Fig. 9.

[0041] <<How to Find Relation of Foreground Area and Number of Persons >>

Here, it is assumed that the relation 142 of the foreground area and the number
10 of persons considers occlusion among persons in crowd. It is assumed that the relation
142 of the foreground area and the number of persons is stored in the storage unit 140.
How to derive the relation 142 of the foreground area and the number of persons will be
explained below.

Fig. 10 is a diagram illustrating an image of a correct foreground image.

15 Fig. 11 is a diagram illustrating an image of numerical congestion levels of the
correct foreground image.

A correct image representing number of persons in a video frame that is known
and a grounding point of a person in a physical coordinate system that is also known is
prepared. The correct image is processed by foreground extraction to generate a
20 correct foreground image as illustrated in Fig. 10.

As illustrated in Fig. 11, the correct foreground image is divided into a
plurality of small regions to calculate an amount of the foreground area per person in
each of the small regions for each congestion level. In addition, a same process is
applied to a large number of correct foreground images with various congestion levels
25 and arrangement patterns to aggregate the foreground area per person in each of the

small regions. This allows derivation of relationship between the number of persons and the foreground in each of the small regions for each congestion level as the relation 142 of the foreground area and the number of persons. The relation 142 of the foreground area and the number of persons is saved in the storage unit 140. An
 5 occupancy ratio of foreground area in a small region at each congestion level is saved in the storage unit 140, as a level threshold 143 for determination of a congestion level.

When using the relation 142 of the foreground area and the number of persons, the provisional density calculation unit 122 determines a congestion level for each small region, and uses the relation 142 corresponding to the congestion level, calculates the
 10 number of persons from the foreground area. The provisional density calculation unit 122 determines a congestion level by comparing an occupancy ratio of foreground in a small region with a level threshold 143.

[0042] <<Presence Determination Process>>

In step ST13, the presence determination unit 123 determines whether there is
 15 a person present in each three-dimensional region v_{ij} of the plurality of three-dimensional regions. In other words, the presence determination unit 123 determines whether there is a person present in each three-dimensional space V_{ij} of a plurality of three-dimensional spaces, and thereby determines that there is no person present regarding three-dimensional region v_{ij} corresponding to a three-dimensional
 20 space for which it has been determined that there is no person present. The presence determination unit 123 makes determination on presence of a person for each three-dimensional region v_{ij} , and changes to zero, the number of persons in the three-dimensional region v_{ij} for which it has been determined that there is no person present. In other words, the presence determination unit 123 changes to zero, the
 25 number of persons in region ΔS_{ij} corresponding to the three-dimensional region v_{ij} for

which the presence determination unit 123 has determined that there is no person present. In other words, the presence determination unit 123 outputs the provisional density distribution 222 as corrected density distribution 223 after the number of persons in the three-dimensional region v_{ij} being corrected to zero, the

5 three-dimensional region v_{ij} being corresponding to the three-dimensional space for which it has been determined that there is no person present.

[0043] Fig. 12 is a diagram of an image of the presence determination process according to the present embodiment.

As described above, when a person is standing in the three-dimensional space

10 V_{ij} , each three-dimensional region v_{ij} of the plurality of three-dimensional regions has a head region to which a head of the person corresponds. And, each three-dimensional region v_{ij} of the plurality of three-dimensional regions has a ground region corresponding to ground on which the person is standing. When there is a person present in both regions of the head region and the ground region in the

15 three-dimensional region v_{ij} , the presence determination unit 123 determines that there is a person present in the three-dimensional space V_{ij} corresponding to the three-dimensional region v_{ij} . In specific, as illustrated in Fig. 12, when foreground area in each of the head region and the ground region of each three-dimensional region v_{ij} is equal to or less than a prescribed amount, the presence determination unit 123

20 determines that there is no person present, and therefore that $h_{ij} = 0$. Also, only for three-dimensional region v_{ij} with foreground area in both regions of the head region and the ground region that is equal to or more than the prescribed amount, the presence determination unit 123 uses a value of the provisional density distribution 222 without change.

25 The presence determination unit 123 outputs the provisional density

distribution 222 after implementing the presence determination process, as the corrected density distribution 223.

[0044] <<Standardization Unit>>

In step ST14, the standardization unit 124 acquires a total number of persons in the video frame 22 based on the foreground image 221. The standardization unit 124 standardizes number of persons in each three-dimensional region v_{ij} of a plurality of three-dimensional regions in the corrected density distribution 223. In specific, the standardization unit 124 standardizes crowd density by a total number of persons h_{total} present in a video frame, using Formula (1) and Formula (2) below. The standardization unit 124 calculates the total number of persons h_{total} in the video frame 22 by applying the relation 142 of the foreground area and the number of persons to the foreground image 221 overall. Rows in Formula (2) is a total number of i . And, Cols in Formula (2) is a total number of j .

[Expression 1]

$$h_{tij} = \frac{h_{total}}{h'_{total}} h_{ij} \quad (\text{Formula 1})$$

15

[Expression 2]

$$h'_{total} = \sum_{i=1}^{Rows} \sum_{j=1}^{Cols} h_{ij} \quad (\text{Formula 2})$$

[0045] The standardization unit 124 implements a standardization process for all three-dimensional regions v_{ij} in corrected density distribution, and then outputs it as fixed density distribution 224.

[0046] <<Distribution Output Process>>

5 In step ST15, the distribution output unit 125 acquires the fixed density distribution 224 from the standardization unit 124. The distribution output unit 125 transforms the fixed density distribution 224 into an output manner, and outputs it as the crowd density distribution 225 to the result output unit 130.

[0047] *** Other Configuration ***

10 In the present embodiment, functions of the video acquisition unit 110, the analysis unit 120, and the result output unit 130 are realized by software. Hereinafter, the functions of the video acquisition unit 110, the analysis unit 120, and the result output unit 130 may be realized by hardware, as a modification.

[0048] Fig. 13 is a diagram illustrating a configuration of a crowd density calculation apparatus 100 according to a modification of the present embodiment.

The crowd density calculation apparatus 100 includes electronic circuit 909, a memory 921, an auxiliary storage device 922, an input interface 930, an output interface 940, and a communication device 950.

[0049] The electronic circuit 909 is dedicated electronic circuit to realize the functions of the video acquisition unit 110, the analysis unit 120, and the result output unit 130.

20 The electronic circuit 909 is, in specific, a single circuit, a composite circuit, a programmed processor, a parallel-programmed processor, a logic IC, a GA, an ASIC, or an FPGA. GA is an abbreviation of gate array. ASIC is an abbreviation of application-specific integrated circuit. FPGA is an abbreviation of field-programmable gate array.

25

The functions of the video acquisition unit 110, the analysis unit 120, and the result output unit 130 may be realized by one electronic circuit, or by being divided into and realized by a plurality of electronic circuitries.

As another modification, a part of the functions of the video acquisition unit
 5 110, the analysis unit 120, and the result output unit 130 may be realized by electronic circuit, and remaining of the functions may be realized by software.

[0050] Each of the processor and the electronic circuit is called also processing
 circuitry. In other words, in the crowd density calculation apparatus 100, the functions
 of the video acquisition unit 110, the analysis unit 120, and the result output unit 130
 10 are realized by the processing circuitry.

[0051] In the crowd density calculation apparatus 100, “units” of the video acquisition
 unit 110, the foreground extraction unit 121, the provisional density calculation unit 122,
 the presence determination unit 123, the standardization unit 124, the distribution output
 unit 125, and the result output unit 130 may be replaced by “steps” or “processes”.
 15 Also, “processes” of the video acquisition process, the foreground extraction process,
 the provisional density calculation process, the presence determination process, the
 standardization process, the distribution output process, and the result output process
 may be replaced by “programs”, “program products”, or “computer-readable storage
 media which stores a program”.

20 [0052] *** Description of Effect of Present Embodiment ***

In the crowd density calculation apparatus 100 according to the present
 embodiment, a foreground extraction unit extracts a foreground image from a video
 frame of a video recorded by a camera. A provisional density calculation unit
 calculates number of persons present in a region that is a projected region of a region in
 25 a physical space on the video frame, that is, apparent number of persons, and generates

provisional density distribution. The presence determination unit determines a region in the physical space where there is no person present, and based on such determination result, corrects the provisional density distribution. The result output unit outputs to an output device, such as a display, the corrected and standardized provisional density distribution as crowd density distribution.

Thus, the crowd density calculation apparatus 100 according to the present embodiment allows calculation of a position where a person is present in a real-world physical space from a video frame inputted from a camera and output thereof as crowd density distribution.

[0053] Embodiment 2.

In the present embodiment, differences from Embodiment 1 will be explained. Note that a component that is same as that of Embodiment 1 is denoted by a same reference sign, and description thereon may be omitted.

[0054] In Embodiment 1, each three-dimensional region v_{ij} of a plurality of

three-dimensional regions is overlapped by one another. Therefore, in calculating provisional density distribution, there is a case where foreground appears even in three-dimensional region v_{ij} other than three-dimensional region v_{ij} in which a person is actually present. This may result in inaccuracy in the provisional density distribution to be calculated. In the present embodiment, an embodiment to improve high-accuracy of crowd density distribution by eliminating influence of duplication of three-dimensional regions will be explained.

According to the present embodiment, the standardization unit 124 according to Embodiment 1 is excluded, and a position correction unit 126 is newly added between the provisional density calculation unit 122 and the presence determination unit 123.

[0055] With Fig. 14, a detailed configuration of the analysis unit 120a according to the

present embodiment will be explained.

The analysis unit 120a includes the foreground extraction unit 121, the provisional density calculation unit 122, the position correction unit 126, the presence determination unit 123, and the distribution output unit 125.

5 [0056] With Fig. 15, a detail of an analysis process by the analysis unit 120a according to the present embodiment will be explained.

In Fig. 15, steps ST11 and ST12 are same as those of Embodiment 1.

In step ST16, the position correction unit 126 acquires provisional density distribution 222 from the provisional density calculation unit 122. The position
10 correction unit 126 corrects the provisional density distribution 222 based on number of persons in a duplicate region representing a duplicate part of neighboring three-dimensional spaces of a plurality of three-dimensional spaces, and outputs the corrected provisional density distribution 222. In other words, the position correction unit 126 corrects, using the provisional density distribution 222, a position of a person,
15 considering the influence of the duplication of three-dimensional regions v_{ij} .

[0057] Fig. 16 is a diagram illustrating an image of a position correction process by the position correction unit 126 according to the present embodiment.

A duplicate region A_{duplm} is, as illustrated in Fig. 16, a region where three-dimensional region v_l and three-dimensional region v_m are overlapped by each
20 other.

The position correction process by the position correction unit 126 works, as illustrated in Fig. 16, as a certain kind of filter that makes values dispersed around sharp.

Subscripts l and m here are replacements of a two-variable subscript ij into a
25 one-variable subscript for convenience. Hereinafter in the description of the present

embodiment, the subscripts l and m are used. Relationship between the subscripts l and ij falls under $l = i \times \text{Cols} + j$. The same applies to relationship between the subscripts m and ij .

[0058] Here, it is assumed that vector representation of number of persons in each region ΔS_l to be calculated is h_l , a coefficient matrix indicating a duplication relationship between cells is A , and vector representation of output acquired by the provisional density distribution 222 is h . The number of persons in each region ΔS_l is calculable by Formula (3) and Formula (4).

Using Formula (3) and Formula (4), the position correction unit 126 is able to output the high accurate crowd density distribution 225 after the influence of the duplicate region being eliminated.

[0059] A detail of derivation of Formula (3) will be explained below. It is assumed that number of persons present in three-dimensional space V_l is h_{tl} and number of persons appearing in the three-dimensional region v_l is h_l . Number of persons appearing in the three-dimensional region v_m out of the number of persons h_{tl} present in the three-dimensional space V_l is $h_{com_l \rightarrow m}$. Accordingly, apparent number of persons h_l appearing in the three-dimensional region v_l can be expressed by Formula (5). The $h_{com_l \rightarrow m}$ can be expressed as product of number of persons h_{tl} present in cell l times coefficient α_{lm} . Therefore, a coefficient matrix representing a duplicate relationship between cells is Formula (4). If it is assumed that $\alpha_{lm} = 1$ when $l = m$, the apparent number of persons can be expressed by Formula (7), which is multiplication of the number of persons present by the coefficient, according to Formula (5) and Formula (6).

If it is assumed that a total number of l is N , a formula is N linear equations in N unknowns, according to Formula (7), and thus, a replacement thereof in a matrix representation is Formula (8). Formula (3) can be derived by multiplying an inverse

matrix of coefficient matrix A obtained from both sides of Formula (8).

[Expression 3]

$$\mathbf{h}_t = \mathbf{A}^{-1} \mathbf{h} \quad (\text{Formula 3})$$

5 [Expression 4]

$$\begin{aligned} \mathbf{h}_t &= \begin{pmatrix} h_{t1} \\ h_{t2} \\ \vdots \\ h_{tN} \end{pmatrix} \\ \mathbf{h} &= \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{pmatrix} \\ \mathbf{A} &= \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{N1} & \alpha_{N2} & \dots & \alpha_{NN} \end{pmatrix} \end{aligned} \quad (\text{Formula 4})$$

[Expression 5]

$$h_l = h_{com_1 \rightarrow l} + h_{com_2 \rightarrow l} + h_{com_3 \rightarrow l} + \dots + h_{com_{l-1} \rightarrow l} + h_{tl} \\ + h_{com_{l+1} \rightarrow l} + \dots + h_{com_N \rightarrow l} \quad (\text{Formula 5})$$

[Expression 6]

$$h_{com_i \rightarrow j} = \alpha_{ij} h_{ti} \quad (\text{Formula 6})$$

[Expression 7]

$$h_l = \sum_{k=1}^N \alpha_{lk} h_{tk} \quad (\text{Formula 7})$$

5

[Expression 8]

$$\begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_N \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{N1} & \alpha_{N2} & \dots & \alpha_{NN} \end{pmatrix} \begin{pmatrix} h_{t1} \\ h_{t2} \\ \vdots \\ h_{tN} \end{pmatrix} \\ \Leftrightarrow \mathbf{h} = \mathbf{A} \mathbf{h}_t \quad (\text{Formula 8})$$

[0060] Note that how to calculate the coefficient α_{lm} is not limited. For example, the

coefficient α_{lm} may be calculated as Formula (9), using area A_l of the three-dimensional region v_l and the duplicate region A_{duplm} of the three-dimensional region v_l and the three-dimensional region v_m , illustrated in Fig. 16.

5 [Expression 9]

$$\alpha_{lm} = \frac{A_{duplm}}{A_l} \quad (\text{Formula 9})$$

[0061] Note that it is assumed that calculation of the coefficient matrix A using Formula (9) is carried out only once in an analysis parameter load process in step ST01.

Steps ST13 and ST15 hereafter are same as those of Embodiment 1.

10 [0062] As described above, a crowd density calculation apparatus with the analysis unit 120a according to the present embodiment allows calculation of a crowd density distribution with higher accuracy by eliminating influence of a duplicate region.

[0063] Embodiment 3.

In the present embodiment, differences from Embodiment 2 will be explained.

15 Note that a component that is same as those of Embodiments 1 and 2 is denoted by a same reference sign, and description thereon may be omitted.

[0064] In Embodiment 2, a coefficient matrix A indicating duplicate relationship calculated using Formula (9) is based on an assumption that foreground appears evenly in three-dimensional region v_{ij} . Actually, the coefficient matrix A indicating the

20 duplicate relationship is influenced by a position where a person is actually present in

three-dimensional space V_{ij} . Therefore, there is a chance that calculated crowd density distribution includes an error depending on the position where the person is actually present in the three-dimensional space V_{ij} . The present embodiment optimizes the coefficient matrix A by a numerical calculation. In doing so, a total number of persons
 5 present on a screen is used as a total number of persons h_{total} of the crowd density distribution. The total number of persons h_{total} on the screen is calculated by applying a relation of foreground area and number of persons to a foreground image overall.
 [0065] With Fig. 17, a detail of an analysis process according to the present embodiment will be explained.

10 In Fig. 17, steps ST11, ST12, and ST16 are same as those of Embodiment 1.
 [0066] In step ST17, the position correction unit 126 optimizes the coefficient matrix A by recalculation. The position correction unit 126 also repeats correction of provisional density distribution until an error in a total number of persons in a video frame falls equal to or below a threshold.

15 [0067] Formula (2) and Formula (7) are used to calculate h'_{total} .

An evaluation function regarding the correction of the provisional density distribution is decided by Formula (10). The position correction unit 126 repeats calculation using a method of steepest descent until error E that is calculated by Formula (10) falls equal to or below the threshold.

20 [Expression 10]

$$E = (h_{total} - h'_{total})^2 \quad (\text{Formula 10})$$

[0068] Note that a method of optimization by the position correction unit 126 is not

limited to the method of steepest descent.

[0069] As described above, crowd density calculation apparatus according to the present embodiment updates the coefficient matrix A for every frame, using Formula (10). Therefore, the crowd density calculation apparatus according to the present
5 embodiment allows calculation of the crowd density distribution with higher accuracy compared with Embodiment 2.

[0070] In Embodiments 1 to 3 above, each unit of the crowd density calculation apparatus is explained as an independent functional block. However, a configuration of the crowd density calculation apparatus does not need to be a configuration such as
10 those of the embodiments described above. The functional block of the crowd density calculation apparatus can be any configuration as long as it allows realization of functions explained in the embodiments described above. In addition, the crowd density calculation apparatus does not need to be one apparatus, but can be a system composed of a plurality of apparatuses.

15 In addition, a plurality of parts of Embodiments 1 to 3 may be implemented in combination. Alternatively, one part of these embodiments may be implemented. These embodiments may be implemented otherwise entirely or partly, or in any combination.

In other words, for Embodiments 1 to 3, it is possible to freely combine each
20 embodiment, or arbitrarily modify or omit a component of or in each embodiment.

[0071] Essentially, the embodiments described above are preferable exemplifications, and are not intended to limit a scope, a scope of an applicable object, and a scope of usage of the present invention. It is possible to make various changes in the
embodiments described above, as necessary. The crowd density calculation apparatus
25 according to the embodiments described above is applicable to a crowd density

estimation apparatus and a crowd density estimation system to estimate crowd density.

Reference Signs List

[0072] 21: video stream; 22: video frame; 100: crowd density calculation apparatus;
 110: video acquisition unit; 120 and 120a: analysis unit; 121: foreground extraction
 5 unit; 122: provisional density calculation unit; 123: presence determination unit; 124:
 standardization unit; 125: distribution output unit; 126: position correction unit; 130:
 result output unit; 140: storage unit; 141: analysis parameter; 142: relation; 143: level
 threshold; 200: camera; 221: foreground image; 222: provisional density distribution;
 223: corrected density distribution; 224: fixed density distribution; 225: crowd density
 10 distribution; 909: electronic circuit; 910: processor; 921: memory; 922: auxiliary
 storage device; 930: input interface; 940: output interface; 950: communication device;
 S100: crowd density calculation process

CLAIMS

[Claim 1] A crowd density calculation apparatus comprising:

a video acquisition unit to acquire a video frame from a video stream including a picked-up video image of a person; and

5 an analysis unit to relate three-dimensional coordinates to the video frame, acquire as each three-dimensional region of a plurality of three-dimensional regions, a region representing each three-dimensional space of a plurality of three-dimensional spaces that are acquired based on the three-dimensional coordinates on the video frame, and based on number of persons present in each three-dimensional region of the

10 plurality of three-dimensional regions, calculate density distribution of persons in the video frame as crowd density distribution,

wherein the analysis unit includes:

a foreground extraction unit to extract an image of a person in the video frame as a foreground image;

15 a provisional density calculation unit to calculate as provisional density distribution, number of persons apparently present in each three-dimensional region of the plurality of three-dimensional regions based on the foreground image; and

a presence determination unit to determine whether there is a person present in each three-dimensional space of the plurality of three-dimensional spaces and output the

20 provisional density distribution as corrected density distribution after number of persons in a three-dimensional region being corrected to zero, the three-dimensional region being corresponding to a three-dimensional space for which it has been determined that there is no person present.

[Claim 2] The crowd density calculation apparatus according to claim 1,

25 wherein each three-dimensional region of the plurality of three-dimensional

regions comprises, when the person is standing in each three-dimensional space of the plurality of three-dimensional spaces, a head region to which a head of the person corresponds and a ground region corresponding to ground on which the person is standing, and

5 wherein the presence determination unit determines, when there is a person present in both regions of the head region and the ground region in a three-dimensional region, that there is a person in a three-dimensional space corresponding to the three-dimensional region.

[Claim 3] The crowd density calculation apparatus according to claim 1 or 2,

10 wherein the analysis unit comprises:

 a standardization unit to acquire a total number of persons in the video frame based on the foreground image, and based on the total number of persons, standardize number of persons in each three-dimensional region of the plurality of three-dimensional regions in the corrected density distribution; and

15 a distribution output unit to acquire from the standardization unit, the standardized corrected density distribution as fixed density distribution, and transform the fixed density distribution into an output manner.

[Claim 4] The crowd density calculation apparatus according to any one claim of claims 1 to 3,

20 wherein the analysis unit comprises a position correction unit to correct the provisional density distribution based on number of persons in a duplicate region representing a duplicate part of neighboring three-dimensional regions of the plurality of three-dimensional regions, and output the corrected provisional density distribution.

[Claim 5] The crowd density calculation apparatus according to claim 4,

25 wherein the position correction unit repeats the correction of the provisional

density distribution until an error in the total number of persons in the video frame falls equal to or below a threshold.

[Claim 6] A crowd density calculation method comprising:

by a video acquisition unit, acquiring a video frame from a video stream

5 including a picked-up video image of a person;

by an analysis unit, relating three-dimensional coordinates to the video frame,

acquiring as each three-dimensional region of a plurality of three-dimensional regions, a region representing each three-dimensional space of a plurality of three-dimensional spaces that are acquired based on the three-dimensional coordinates on the video frame,

10 and based on number of persons present in each three-dimensional region of the plurality of three-dimensional regions, calculating crowd density distribution that is density distribution of persons in the video frame;

by a foreground extraction unit of the analysis unit, extracting an image of a person in the video frame as a foreground image;

15 by a provisional density calculation unit of the analysis unit, calculating as provisional density distribution, number of persons apparently present in each three-dimensional region of the plurality of three-dimensional regions based on the foreground image; and

by a presence determination unit of the analysis unit, determining whether
20 there is a person present in each three-dimensional space of the plurality of three-dimensional spaces and outputting the provisional density distribution as corrected density distribution after number of persons in a three-dimensional region being corrected to zero, the three-dimensional region being corresponding to a three-dimensional space for which it has been determined that there is no person
25 present.

[Claim 7] A crowd density calculation program that causes a computer to execute:

 a video acquisition process of acquiring a video frame from a video stream

5 including a picked-up video image of a person; and

 an analysis process of relating three-dimensional coordinates to the video frame, acquiring as each three-dimensional region of a plurality of three-dimensional regions, a region representing each three-dimensional space of a plurality of three-dimensional spaces that are acquired based on the three-dimensional coordinates
10 on the video frame, and based on number of persons present in each three-dimensional region of the plurality of three-dimensional regions, calculating crowd density distribution that is density distribution of persons in the video frame,

 wherein the analysis process includes:

 a foreground extraction process of extracting an image of a person in the video
15 frame as a foreground image;

 a provisional density calculation process of calculating as provisional density distribution, number of persons apparently present in each three-dimensional region of the plurality of three-dimensional regions based on the foreground image; and

 a presence determination process of determining whether there is a person
20 present in each three-dimensional space of the plurality of three-dimensional spaces and outputting the provisional density distribution as corrected density distribution after number of persons in a three-dimensional region being corrected to zero, the three-dimensional region being corresponding to a three-dimensional space for which it has been determined that there is no person present.