United States Patent Application Publication	20250252749
Kind Code	A1
Publication Date	August 07, 2025
Inventor(s)	OAMI; Ryoma

INFORMATION PROCESSING APPARATUS, INFORMATION PROCESSING METHOD, AND INFORMATION PROCESSING PROGRAM

Abstract

To efficiency search for an object associated with a sensed event, an information processing apparatus includes a sensor that analyzes a captured video and senses whether a predetermined event has occurred, a determining unit that determines a type of an object to be used as query information based on a type of the event in response to sensing of the event occurrence, and a generator that detects the object of the determined type from the video and generates the query information based on the detected object.

Inventors: **OAMI; Ryoma (Tokyo, JP)**

Applicant: **NEC Corporation** (Tokyo, JP)

Family ID: **50387926**

Assignee: **NEC Corporation (Tokyo, JP)**

Appl. No.: **19/185428**

Filed: **April 22, 2025**

Foreign Application Priority Data

JP	2012-217592	Sep. 28, 2012

Related U.S. Application Data

parent US continuation 18229350 20230802 parent-grant-document US 12315260 child US 19185428
parent US continuation 18198022 20230516 parent-grant-document US 12183083 child US 18229350
parent US continuation 16285333 20190226 parent-grant-document US 11816897 child US

18198022
parent US continuation 14430416 20150323 parent-grant-document US 10248868 WO continuation PCT/JP13/74191 20130909 child US 16285333

Publication Classification

Int. Cl.: **G06V20/52** (20220101); **G06F16/732** (20190101); **G06F16/78** (20190101); **H04N5/91** (20060101); **H04N7/18** (20060101)

U.S. Cl.:

CPC **G06V20/52** (20220101); **G06F16/7335** (20190101); **G06F16/78** (20190101); **H04N5/91** (20130101); H04N7/18 (20130101)

Background/Summary

[0001] The present application is a Continuation application of Ser. No. 18/229,350 filed on Aug. 2, 2023, which is a Continuation application of Ser. No. 18/198,022 filed on May 16, 2023, which issued as U.S. Pat. No. 12,183,083, which is a Continuation application of Ser. No. 16/285,333 filed on Feb. 26, 2019, which issued as U.S. Pat. No. 11,816,897, which is a Continuation application of Ser. No. 14/430,416 filed on Mar. 23, 2015, which issued as U.S. Pat. No. 10,248,868, which is a National Stage Entry of PCT/JP2013/074191 filed on Sep. 9, 2013, which claims priority from Japanese Patent Application JP2012-217592 filed on Sep. 28, 2012, the contents of all of which are incorporated herein by reference, in their entirety.

TECHNICAL FIELD
[0002] The present invention relates to a technique of searching for an object in a video.
BACKGROUND ART
[0003] In the above-described technical field, patent literature 1 discloses a technique of causing a guard to designate an image and a person region including a person supposed to be a criminal, tracking the designated person in image data before and after the acquisition time of the designated image, and calculating the features of the person from the obtained tracking result.
CITATION LIST
Patent Literature
[0004] Patent literature 1: Japanese Patent Laid-Open No. 2009-027393
SUMMARY OF THE INVENTION
Technical Problem
[0005] In the technique described in the literature, however, only when the guard has designated an image and issued an instruction, the person search is performed in other images. Hence, the search speed and accuracy largely change depending on the ability of the guard. Especially, since the guard does not always take optimum action depending on the event to be detected, the search accuracy for a necessary object may lower.
[0006] The present invention enables to provide a technique of solving the above-described problems.
Solution to Problem
[0007] One aspect of the present invention provides an information processing apparatus comprising: [0008] a sensor that analyzes a captured video and senses whether a predetermined event has occurred; [0009] a determining unit that determines a type of an object to be used as query information based on a type of the event in response to sensing of the event occurrence; and

[0010] a generator that detects the object of the type determined by the determining unit from the video at a time other than the time of the event occurrence and generates the query information based on the object.
[0011] Another aspect of the present invention provides an information processing method comprising: [0012] analyzing a captured video and sensing whether a predetermined event has occurred; [0013] deciding a type of query information to be searched for in accordance with on a type of the event in response to sensing of the event occurrence; and [0014] generating the query information of the type determined in the deciding from the video at a time other than the time of the event occurrence.
[0015] Still other aspect of the present invention provides an information processing program for causing a computer to execute: [0016] analyzing a captured video and sensing whether a predetermined event has occurred; [0017] deciding a type of query information to be searched for in accordance with on a type of the event in response to sensing of the event occurrence; and [0018] generating the query information of the type determined in the deciding from the video at a time other than the time of the event occurrence.
Advantageous Effects of Invention
[0019] According to the present invention, it is possible to very efficiently search for an object associated with a sensed event.

Description

BRIEF DESCRIPTION OF THE DRAWINGS
[0020] FIG. **1** is a block diagram showing the arrangement of an information processing apparatus according to the first embodiment of the present invention;
[0021] FIG. **2** is a block diagram showing the arrangement of an information processing apparatus according to the second embodiment of the present invention;
[0022] FIG. **3** is a block diagram showing the arrangement of a search query generator according to the second embodiment of the present invention;
[0023] FIG. **4** is a flowchart showing the procedure of processing of the information processing apparatus according to the second embodiment of the present invention;
[0024] FIG. **5** is a block diagram showing the arrangement of a search query generator according to the third embodiment of the present invention;
[0025] FIG. **6** is a block diagram showing the arrangement of a search query generator according to the fourth embodiment of the present invention;
[0026] FIG. **7** is a view showing the arrangement of a table used by the search query generator according to the fourth embodiment of the present invention;
[0027] FIG. **8** is a view for explaining the operation of the search query generator according to the fourth embodiment of the present invention;
[0028] FIG. **9** is a view for explaining the operation of the search query generator according to the fourth embodiment of the present invention;
[0029] FIG. **10** is a view for explaining the operation of the search query generator according to the fourth embodiment of the present invention;
[0030] FIG. **11** is a view for explaining the operation of the search query generator according to the fourth embodiment of the present invention;
[0031] FIG. **12** is a view for explaining the operation of the search query generator according to the fourth embodiment of the present invention;
[0032] FIG. **13** is a block diagram showing the arrangement of a search query generator according to the fifth embodiment of the present invention;
[0033] FIG. **14** is a block diagram showing the arrangement of a search query generator according

to the sixth embodiment of the present invention; and
[0034] FIG. **15** is a block diagram showing the arrangement of a search query generator according to the seventh embodiment of the present invention.
DESCRIPTION OF THE EMBODIMENTS
[0035] Preferred embodiments of the present invention will now be described in detail with reference to the drawings. It should be noted that the relative arrangement of the components, the numerical expressions and numerical values set forth in these embodiments do not limit the scope of the present invention unless it is specifically stated otherwise.
First Embodiment
[0036] An information processing apparatus **100** according to the first embodiment of the present invention will be described with reference to FIG. **1**. As shown in FIG. **1**, the information processing apparatus **100** includes a sensor **101**, a determining unit **102**, and a generator **103**. The sensor **101** analyzes a captured video and senses whether a predetermined event has occurred. In response to the event occurrence sensing by the sensor **101**, the determining unit **102** determines, based on the event type, the type of an object to be used as query information. The generator **103** detects an object of the determined type from the video and generates query information based on the object.
[0037] When query information is generated by the above-described arrangement, and a search using the query information is performed, it is possible to very efficiently search for an object associated with the sensed event.
Second Embodiment
[0038] An information processing apparatus according to the second embodiment of the present invention will be described next with reference to FIG. **2**. FIG. **2** is a block diagram for explaining the arrangement and use method of an information processing apparatus **200** according to this embodiment.
[0039] The information processing apparatus **200** automatically detects an event that occurs in an airport lounge by processing a video obtained by causing a surveillance camera **250** to capture people, benches, and the like in the airport lounge, and alerts a security guard or the like. The present invention is not limited to utilization in the airport lounge, as a matter of course, and is also usable in, for example, a casino, a shopping center, and the like.
[0040] The video captured by the surveillance camera **250** is stored in a video storage **210** and also sent to an event sensor **201**. Alternatively, the video temporarily stored in the video storage **210** may be sent to the event sensor **201** a little later. The event sensor **201** analyzes the acquired video and confirms, by referring to an event database **220**, whether an event as an alert target has not occurred. If an event as an alert target has occurred, alert information is generated and sent to an object type determining unit **202**. The thus generated alert information is output to a search query generator **204**. The alert information includes the time at which the alert is generated, the alert type, and information representing the position of the object as the alert target. The object position information can be expressed by coordinate values representing a position on the screen or coordinate values on a coordinate system projected onto a real space and calculated using camera parameters (for example, coordinate values on a coordinate system using a given point on the floor as a reference). Alternatively, the position information may be the position information of the circumscribed rectangle of the object or coordinate information representing a specific position such as the centroid of the object. In addition, various kinds of position information for specifying the position of the object are usable.
[0041] The alert targets include events that should attract attention of a guard, for example, appearance of a person included in a watch list (so-called black list), suspicious behavior, and an dangerous behavior. For example, an alert is generated upon detecting a person registered in the black list. The event sensor **201** can detect such a person by detecting a face from an input video, extracting features from the detected face, and collating the extracted features with those of the

faces of persons registered in the black list in the event database **220**.
[0042] The event sensor **201** also generates an alert upon sensing a specific behavior such as a fall or a crouch. More specifically, such a behavior can be detected by, for example, extracting a person from an input video, tracking the person through the frames, and sensing that the person's height has abruptly decreased, and the state has continued for a while. Alternatively, a discriminator may be constructed by learning person's states such as a crouch and a fall in advance. The parameters of the discriminator generated by the learning may be stored in the event database **220**, and the states may be sensed using the discriminator.
[0043] The event sensor **201** also generates an alert when a person has entered a restricted area. This can be sensed by, for example, marking the floor area of the restricted area in an image and determining whether the feet of a detected/tracked person have entered the region. The event database **220** may store in advance which area becomes restricted at which timing, and which camera captures an object in what kind of manner.
[0044] The event sensor **201** also generates an alert upon sensing an abandoned object or theft of an object left behind. An abandoned object can be sensed by, for example, sensing placement of an object using background subtraction or the like, and detecting that the state has continued for a predetermined time or more.
[0045] The types of objects and their changes possible in a video and how to generate an alert in response to them are recorded in the event database **220** in advance.
[0046] The event database **220** also stores a table representing the correspondence concerning what kind of object should be found to generate a query in each event.
[0047] The object type determining unit **202** determines an object type serving as a search query for each of various kinds of events sensed in the above-described manner. For example, in the event of appearance of a person included in the black list, the clothing of the person is used as the search query. Hence, a full-length image capable of specifying the clothing (including take-alongs and a cap) is determined as the target object.
[0048] On the other hand, in the event of a person who has fallen, a face and clothing capable of specifying the person who has fallen are used as the search queries. Hence, a face image and a full-length image capable of specifying them are determined as objects.
[0049] In the event of finding of a person who has entered a restricted area, a face and clothing capable of specifying the person are used as the search queries. Hence, a face image and a full-length image capable of specifying them are determined as objects.
[0050] In the event of finding of an abandoned object or theft of an object left behind, a face and clothing capable of specifying the person who has abandoned the object or the person who owns the object left behind and stolen are used as the search queries. Hence, a face image and a full-length image capable of specifying them are determined as objects.
[0051] The search query generator **204** extracts the features of an object and generates query features that are features to be used as queries at the time of search. When selecting a query feature from features extracted in advance, the query feature is selected and acquired from features extracted from the camera video **250** by a video analyzer **235** and stored in a feature database **230**. The generated query features are output to a searcher **205**. The searcher **205** executes a search using the query features for videos stored in the video storage **210** or other video storages **240**, and outputs a search result.
[0052] The searcher **205** searches the videos stored in the other video storages **240** for an object having features equal or similar to the query features. That is, the searcher **205** performs the search by collating the query features with the features of objects extracted from videos as search targets stored in the past and stored in the feature database **230**. When the similarity between the features is sufficiently high because, for example, the similarity exceeds a predetermined threshold (or the similarity is sufficiently low because, for example, the distance between the features is equal to or smaller than a predetermined threshold), the search result includes the ID of the camera, time, and

additional information altogether. Examples of the additional information are a thumbnail image, time information of the video necessary for a cue search, and a byte offset from the start of the file. A search result is generated by arranging the obtained results based on a predetermined criterion (for example, in descending order of similarity between features) and output.

[0053] The other video storages **240** store not only videos captured by the same camera as that for the video from which the query features are extracted but also videos captured by different cameras. The videos may include those up to the time immediately before the alert generation. In this case, object extraction and feature extraction are performed in real time for search target images, and the searcher **205** can immediately use the result.

[0054] This makes it possible to alert a guard **262** who is watching a monitor **261** in a monitor room **260** and also offer detailed information associated with the alert. When theft of an object left behind has occurred, the guard **262** is notified of the occurrence via a screen **270** of the display. It is possible to simultaneously offer the face images and full-length images of the owner and the perpetrator, their profiles if the profile information is available, and the current position of the perpetrator derived by causing the searcher **205** to search other videos, and the like. An image of the person in question may be displayed on the hand-held terminal of the guard and used for pursuit.

[0055] It is possible to present the guard **262** the past location of an object by searching for the same object in videos captured by other cameras up to the time immediately before the alert generation time. If the video can be reproduced from the search time by a cue search, the video scene that has undergone the search is reproduced by the cue search so that the guard **262** can confirm the state and behavior of the object.

[0056] Note that the information processing apparatus **200** shown in FIG. **2** includes a CPU (Central Processing Unit), a storage, a RAM (Random Access Memory), and a communication controller (none are shown) as hardware components. The storage stores program modules corresponding to the event sensor **201**, the object type determining unit **202**, the search query generator **204**, and the searcher **205**. The CPU executes the program modules to implement their functions. Data generated by a program module, for example, alert information, object type, a query feature, or the like is temporarily stored in the RAM and then used by another program module or stored in various kinds of databases.

[0057] FIG. **3** is a block diagram showing the internal arrangement of the search query generator **204** according to this embodiment. The search query generator **204** includes an object detector **341** and an object feature extractor **342**. The object detector **341** detects an object from alert information and a video and outputs object information to the object feature extractor **342**. The object feature extractor **342** extracts object features from the object information and the video and outputs them as query features.

[0058] The operation in FIG. **3** will be described next.

(When Performing Object Detection for First Time by Alert Information)

[0059] Alert information and a video are input to the object detector **341**. Based on the position information and time information of an object included in the alert information, the object detector **341** detects the object located at a corresponding position in a corresponding frame of the video and outputs an object region information. For example, when the object is a person, and the alert is a black list collation alert, the alert information includes the object information of a face. The object detector **341** extracts the region of the whole person from the video, and generates the information of the region of the whole person as the object region information, including the position information of the face. In, for example, a behavior alert, the region where a feature of the object should be extracted (for example, if the feature to be extracted is a clothing feature, the clothing region of the person) is obtained and generated as the object region information while increasing the accuracy of the position information of the object included in the alert information. If the position information of the object included in the alert information can directly be used as the

object region information, the processing may be skipped without performing any processing in the object detector **341**. The generated object region information is output to the object feature extractor **342** as object information.

[0060] The object feature extractor **342** extracts the features of the object from the video based on the object information. For example, when the object is a person, a face feature and a clothing feature are extracted. If the video is not suitable for face feature extraction because, for example, the face size is too small, or the person is not facing the front, only the clothing feature may be extracted. When extracting the clothing feature, the clothing may be divided into several parts, and a feature may be extracted in each part. For example, the features may be extracted by dividing the clothing into the clothing of the upper half of the body and that of the lower half of the body. Alternatively, the features may be extracted by separating belongings other than the clothing or take-alongs such as a cap and glasses. When the object is a car, license plate information and the features of the car body are extracted. If the license plate cannot be recognized because it is invisible or too small, only the features (color, shape, and model) of the car body may be extracted. The thus extracted features are output as query features.

(When Performing Object Detection/Feature Extraction in Advance and Selecting Features from Them)

[0061] A case in which the search query generator **204** extracts the features of an object after alert generation has been described above. If object feature extraction has already been performed in real time, the object features need not be extracted again. In this case, a feature to be used as a query feature is simply selected from the already extracted object features. More specifically, the object detector **341** limits video frames based on the time information designated by the alert information, selects an object having close object position information in the frames, and outputs its region information as object information. The object feature extractor **342** selects the features of the object designated by the object information from the already extracted object features, and outputs the selected features as query features.

[0062] The procedure of processing according to this embodiment will be described with reference to the flowchart shown in FIG. **4**. First, in step S**401**, the video storage **210** or the event sensor **201** acquires a video from the surveillance camera **250**.

[0063] In step S**403**, the event sensor **201** senses that a predetermined event (alert target) has occurred in the video.

[0064] In step S**405**, based on the sensed event, the object type determining unit **202** determines the type of an object to be further searched for. When the object type is determined, the process advances to step S**407**, and the search query generator **204** extracts the object of the type from videos before and after the event occurrence. In step S**409**, the features of the detected object are obtained, and search queries are determined based on the features. In step S**411**, the searcher **205** searches the stored videos for the search queries.

[0065] With the above-described arrangement and operation, according to this embodiment, it is possible to immediately execute a search in response to alert issuance and track the target person or object.

Third Embodiment

[0066] An information processing apparatus according to the third embodiment of the present invention will be described next with reference to FIG. **5**. FIG. **5** is a block diagram for explaining the arrangement and use method of a search query generator according to this embodiment. This embodiment is different from the above-described second embodiment in that a search query generator **501** includes an object detector/tracker **511** and an object feature extractor **512**. The rest of the arrangement and operation is the same as in the second embodiment. The same arrangement and operation are denoted by the same reference numerals, and a detailed description thereof will be omitted.

[0067] The object detector/tracker **511** extracts an object from alert information and a video, tracks

the object through frames, and outputs object information to the object feature extractor **512**. The object feature extractor **512** extracts object features from the object information and the video and outputs them as query features.

(When Performing Object Detection for First Time by Alert Information)

[0068] When the object detector/tracker **511** tracks an object through frames, the tracking continues from before alert generation. The alert information and the video are input to the object detector/tracker **511**. The object detector/tracker **511** performs object detection and detection processing in the input video, compares the result with object position information included in the alert information based on time information, and determines the target object. Next, the tracking result of the object is checked, and a result suitable for feature extraction is selected from the object detection results included in the tracking result. How large the object will be detected at each position of images can be calculated in advance using the calibration information of the camera (camera parameter information). Hence, a result whose detected size is close to the assumed value and which has a posture and state suitable for feature extraction is selected. In case of, for example, a person, a person who is standing erect in posture and whose foot or upper half of the body is not cut is selected. This can be determined based on, for example, information such as the aspect ratio of the circumscribed rectangular shape of the detected object or the position on the screen. Alternatively, when the shape information of the object can simultaneously be acquired (for example, when the silhouette can be obtained by background subtraction), the appropriateness of the posture and state may be determined from the shape information. Note that a plurality of object detection results (that is, detection results at a plurality of different times) may be selected. The time and object region information of the selected object are output to the object feature extractor **512** as object information.

[0069] The object feature extractor **512** extracts object features from the video based on the object information, like the object feature extractor **342** shown in FIG. **3**. In this case, however, the object information may include a previous time as the time of feature extraction. In this case, the object features are extracted from the video frame of the corresponding time. In addition, there may exist a plurality of feature extraction times included in the object information. In this case, object features are extracted from the respective frames of the designated times, and the set of the extracted features is output as a query feature.

[0070] Note that if object detection/tracking processing is performed at the same time as the alert generation, the object detector/tracker **511** may use object detection/tracking information obtained there. When the features have already been extracted at the time of object detection/tracking, the object feature extractor **512** may select the feature at the designated time from the already extracted features and generate the query feature.

[0071] As described above, according to this embodiment, an object is detected and tracked. This makes it possible to more properly generate a search query and improve the search accuracy.

Fourth Embodiment

[0072] An information processing apparatus according to the fourth embodiment of the present invention will be described with reference to FIG. **6**. FIG. **6** is a block diagram for explaining the arrangement and use method of a search query generator according to this embodiment. This embodiment is different from the above-described third embodiment in that a search query generator **501** includes a time difference determining unit **613** and an object detector/tracker **611**. The rest of the arrangement and operation is the same as in the second embodiment. The same arrangement and operation are denoted by the same reference numerals, and a detailed description thereof will be omitted.

[0073] The time difference determining unit **613** determines feature extraction time information serving as the reference of query feature extraction from alert information and outputs it to the object detector/tracker **611**. The object detector/tracker **611** obtains object information from the alert information, a video, and the feature extraction time information and outputs it to an object

feature extractor **512**. The object feature extractor **512** extracts object features from the object information and the video and outputs them as query features.

[0074] The alert information is input to the time difference determining unit **613**. Based on event type information included in the alert information, the time difference determining unit **613** determines time difference information that designates how long the time goes back from the alert generation time to use video data as a query. This time difference changes depending on the event type and is therefore determined based on an event table **700** set as shown in FIG. **7**. The event table **700** holds time difference information for each event type, and the time difference information is selected in accordance with input alert information.

[0075] The thus determined time difference information is output to the object detector/tracker **611**. In addition to the time difference information, the alert information and the video are also input to the object detector/tracker **611**. The object detector/tracker **611** determines the target object, like the object detector/tracker **511** shown in FIG. **5**. The time (query generation reference time) that goes back from the alert generation time by the time designated by the time difference information is obtained. An object detection result at that time is obtained from the tracking result of the object. The time and the object detection result are generated as object information and output. Alternatively, an object detection result suitable for feature extraction may be selected from the tracking results before and after the query generation reference time, generated as object information, and output, like the object detector/tracker **511** shown in FIG. **5**. The generated object information is output to the object feature extractor **512**.

[0076] The object feature extractor **512** generates and outputs object features, as described with reference to FIG. **5**.

[0077] Note that if object detection/tracking processing is performed at the same time as the alert generation, the object detector/tracker **611** may use object detection/tracking information obtained there. When the features have already been extracted at the time of object detection/tracking, the object feature extractor **512** may select the feature at the designated time from the already extracted features and generate the query feature.

[0078] In this embodiment, when an alert is generated, an image suitable for a search is selected based on tracking information at that time and alert type information, and the search is automatically executed. It is therefore possible to implement a quick search in real time when an alert is generated. As a possible use, for example, when a suspicious person alert is generated, the video of a camera located near the camera that has issued the alert is searched to track the person.

[0079] The search is performed using the features obtained in the above-described way as the query features. If the similarity between the features extracted within the tracked range is low, the search may be performed after presenting the features to the guard and causing him/her to determine whether they indicate the same person. When a person is tracked before and after the video designated by the guard, and the search is performed by designating features supposed to be most reliable in the video, the search accuracy is expected to be higher than when performing the search using only images designated by the guard.

[0080] FIG. **7** is a view showing the event table **700** used to set an object type, a predetermined time difference, a predetermined time width, and the number of object search images in accordance with an event type. This will be explained below in detail for each event type.

(1) Black List Collation

Objects to be Searched for and Queries to be Generated

[0081] A face and clothing are searched for as objects. A clothing feature is extracted in addition to a face feature. For this reason, the features are extracted at a person position suitable for clothing feature extraction (including the upper and lower halves of the body).

Time Difference

[0082] The clothing features cannot sufficiently be extracted from an image including a face in a large size. It is therefore necessary to extract the features from images different from the image at

the time of alert. The person position suitable for extraction is determined by the camera arrangement (viewing angle). Basically, the image preferably includes the person from head to foot in a size as large as possible. In a camera arrangement for capturing a person approaching from far away, as shown in FIG. **8**, the moving time elapsed from an image **801** including a foot reaching the lower edge of the screen to an image **802** in which face collation is actually possible is designated as the time difference for query generation. That is, the average moving time (or a representative value such as the median or mode of the moving time) from the image **801** that includes the person in full length and facilitates extraction of clothing features to the image **802** in which the person is located at a position where the face is captured so large that the black list collation is possible is obtained in advance. A value almost equal to or larger than the moving time is determined as the time difference information. As for the clothing pattern features, more detailed information can be obtained by locating the person as close as possible. For this reason, in an image **803** including the person at a shorter distance, the features of the upper portion of the clothing may be extracted in more detail to create the search queries.

Number of Images to be Used for Query Generation (Number of Time Differences)

[0083] Basically, one image in which the person is located in full length before alert generation suffices, as described above. However, a plurality of images may be acquired near the position, and a plurality of reliable features among them may be used. The reliability of clothing feature extraction can be determined based on the aspect ratio of the circumscribed rectangle of the person or the degree of person likelihood of the person silhouette shape. The degree of person likelihood can be calculated using a discriminator that has learned person silhouettes. As for a color feature, the reliability can be evaluated by analyzing the magnitude of saturation and that of brightness of a color (if the saturation or brightness is low, the influence of shadow or the like is supposedly present) and the presence/absence of a portion where the pixel values are saturated.

(2) Fall Sensing

Objects to be Searched for and Queries to be Generated

[0084] A face and clothing are searched for as objects. It is difficult to accurately separate and extract the information of clothing of the upper and lower halves of the body in a fall state. For this reason, the features are extracted from an image of a person standing erect. To specify a person or his/her history from the clothing information of the person who has fallen down, an image of a normal erect posture is necessary for extracting appropriate features. In particular, to do a search with a distinction between the clothing of the upper half of the body and that of the lower half, the features need to be extracted while separating the upper half and lower half of the body. Hence, an image of an erect posture is important in this case as well.

[0085] To more properly obtain features suitable for a search using a tracking result, the tracking time needs to be prolonged so that a video with the person before the fall is included. In this case, since tracking processing is time-consuming, no quick search can be performed. Hence, the efficiency can be raised by searching for not the tracking result but an image within a designated range.

Time Difference

[0086] Fall sensing is basically set to issue an alert after the elapse of a predetermined time from a fall. For this reason, a time slightly longer than a set determination time is set as the time difference (for example, if the alert is set to be issued when a person has fallen down and remained immovable for 30 sec, the time difference is set to, for example, 35 sec).

Number of Images to be Used for Query Generation (Number of Time Differences)

[0087] As shown in FIG. **9**, when a fall is sensed from an image **903**, basically, using one full-length image **901** going back sufficiently long suffices to extract an object. However, a long time may elapse before the fall in a state in which the person crouches down a little while facing downward, like an image **902**. In this situation, the clothing features may badly be extracted. Hence, to select features a little before, a plurality of images at longer time differences may be used

(for example, in the above-described example, two images 45 sec and 55 sec before are used together). Queries may be generated using all of the plurality of selected images. Alternatively, the reliabilities may be calculated, and only images of high reliabilities (for example, reliabilities higher than a predetermined threshold) may be used, as in black list collation.

(3) Abandonment/Carry Away

Objects to be Searched for and Queries to be Generated

[0088] Although an alert itself is issued upon sensing abandoned luggage (or a region that has newly appeared after luggage has been carried away), the features of the person who has abandoned/carried away the object are extracted. In this case as well, clothing features are extracted because face features are often difficult to extract.

Time Difference

[0089] Abandonment is basically sensed when the rest time of an object has exceeded a predetermined threshold, as in images **1002** and **1003** shown in FIG. **10**. For this reason, a time slightly longer than the determination time is set as the time difference. That is, to issue an alert upon sensing abandonment, a value equal to or more than the time until an abandoned object is determined is determined as time difference information, and a person near the object as in an image **1001** is searched for. Note that in this case, not the features of the object but those of the person who has abandoned the object are extracted.

[0090] Carry away is sensed upon finding a person who has carried away an object that has remained immovable for a predetermined time, as in images **1102** and **1103** shown in FIG. **11**. For the person who has carried away the object, object extraction is performed without a time difference, and the features of the face and clothing of the person are generated. On the other hand, for the person from whom the object has been carried away, a value equal to or more than the carry away determination time is determined as time difference information, and a person who existed first near the object as in an image **1101** is searched for. The features of the face and clothing of the person are extracted. The reliability of carry away determination may be raised by collating the person in the image **1101** with the person in the image **1103**.

Number of Images to be Used for Query Generation (Number of Time Differences)

[0091] An action to place an object on the floor is expected to take a predetermined time. During the placing action, the person may be in a posture unsuitable for clothing feature extraction because he/she crouches or bends. The person may abandon an object by, for example, dropping it without largely changing the posture. Hence, a plurality of times slightly longer than the set determination time are set as the time differences (for example, if the alert is set to be issued when an object has remained immovable for 60 sec, three time differences are set to, for example, 65 sec, 75 sec, and 85 sec).

(4) Intrusion Sensing

Objects to be Searched for and Queries to be Generated

[0092] Basically, in a situation where intrusion sensing is performed, the face is not always visible and is unsuitable for feature extraction. Hence, clothing features are extracted.

Time Difference

[0093] As shown in FIG. **12**, the clothing features in an image **1202** at the time of alert generation may be obtained. However, the person's position suitable for the clothing features is determined by the viewing angle of the camera, as in black list collation. Assume that the restricted area is set on the far side of the camera, as shown in FIG. **12**. In this case, before entering the restricted area, the person is located close to the camera, and more detailed clothing features can be extracted. Hence, a timing like an image **1201** including the person in full length as large as possible is determined, and the time until movement to the restricted area is set as the time difference. Reversely, assume that the restricted area is located on the near side of the camera, and intrusion is sensed when a person has moved closer from far away. In this case, the time difference is set such that the clothing features are extracted when the person has moved a little longer from the timing immediately after

entering the restricted area.
Number of Images to be Used for Query Generation (Number of Time Differences)
[0094] Basically, one image suffices, as in black list collation.
(5) Hangout/Detour Behavior Sensing
Objects to Be Searched for and Queries to Be Generated
[0095] Although an alert itself is generated upon sensing occurrence of hangout or occurrence of a detour behavior, the extraction targets are the features of the person who is the cause of the hangout or detour behavior. In this case as well, clothing features are extracted because face features are often difficult to extract.
[0096] “Hangout” indicates a long-time stop of a plurality of persons. “Detour behavior” indicates people's behavior of avoiding a spilled liquid, suspicious substance, soil, or the like on the floor, or a suspicious (dangerous) person.
[0097] The search target is, for example, a person (for example, street performer) who has caused the hangout or a suspicious (dangerous) person who has caused the detour behavior.
Time Difference
[0098] Basically, a time slightly longer than the determination time necessary to sense the hangout or detour behavior is set as the time difference (for example, if the alert is set to be issued when a plurality of persons have stopped for 30, 60, or 120 sec etc., the time difference is set to, for example, 35, 65, or 125 sec).
Number of Images to Be Used for Query Generation (Number of Time Differences)
[0099] Basically, one image before alert generation suffices, as described above. However, a plurality of images may be acquired in the vicinity, and a plurality of reliable features in the images may be used. As described above, the reliability of clothing feature extraction can be determined based on the aspect ratio of the circumscribed rectangle of a person or the degree of person likelihood of a person silhouette shape.
[0100] As described above, a search can be performed very accurately because object detection and search query generation are performed at a time difference corresponding to the event that has occurred.
Fifth Embodiment
[0101] An information processing apparatus according to the fifth embodiment of the present invention will be described with reference to FIG. **13**. FIG. **13** is a block diagram for explaining the arrangement and use method of a search query generator **1301** according to this embodiment. This embodiment is different from the above-described fourth embodiment in that the search query generator **1301** includes an object identity determiner **1312**, and includes an object detector **1311** in place of the object detector/tracker **611**. The rest of the arrangement and operation is the same as in the second embodiment. The same arrangement and operation are denoted by the same reference numerals, and a detailed description thereof will be omitted.
[0102] When the object detector **1311** does not perform tracking, the object identity determiner **1312** determines, after object feature generation, whether an object is identical to that at the time of alert generation. In this case, the object detector **1311** outputs two object detection results: the object detection result at the time of alert generation and that at the time given the time difference given by a time difference determining unit **613**. An object feature extractor **512** extracts object features for each result and outputs them to the object identity determiner **1312**. The object identity determiner **1312** determines the similarity between the features. Upon determining that the objects are identical, the features given the time difference are output as query features. On the other hand, upon determining that the objects are not identical, the features at the time of alert generation are output as query features.
[0103] Note that alert information may also be input to the object identity determiner **1312**, and the identity determination criterion may be changed in accordance with the alert type. For example, in fall sensing, since the similarity between the features lowers depending on the difference in the

posture, the similarity determination criterion may be lenient. Alternatively, by attaching importance to the object position information, if the object positions are almost the same, the features may be output, regarding the objects as identical even if the similarity is low. In abandonment sensing as well, the determination may be performed in consideration of the position information.

Sixth Embodiment

[0104] An information processing apparatus according to the sixth embodiment of the present invention will be described with reference to FIG. **14**. FIG. **14** is a block diagram for explaining the arrangement and use method of a search query generator **1401** according to this embodiment. This embodiment is different from the above-described fourth embodiment in that the search query generator **1401** includes a time difference learner **1414**. The rest of the arrangement and operation is the same as in the fourth embodiment. The same arrangement and operation are denoted by the same reference numerals, and a detailed description thereof will be omitted.

[0105] When the time difference is determined depending on the camera arrangement, the search query generator **1401** shown in FIG. **14** learns the time difference from an object detection/tracking result. Based on event type information included in alert information, a time difference determining unit **1411** determines time difference information that designates how long the time goes back from the alert generation time to use video data as a query, like the time difference determining unit **613**. This time difference is determined based on an event table **700** set as shown in FIG. **7**. The time difference information in the event table **700** here can be updated in accordance with an instruction from the time difference learner **1414**. The time difference learner **1414** measures, from the object detection/tracking result, a time needed for a person to move from the position at the time of alert generation to the position where clothing features are easily extracted, and calculates the representative value of the time difference by statistically processing the measurement result. The time difference learner **1414** updates the time difference stored in the time difference determining unit **1411** by the representative value. As the method of calculating the representative value of the time difference, not only the average but also the median or mode, or an average calculated while excluding, as outliers, portions of a predetermined percentage or more on both sides of the distribution is usable.

[0106] The time difference may be calculated by inputting, to the time difference learner **1414**, not only the simple object size but also information representing whether the illumination condition or background is suitable for feature extraction. For example, if there is a place that is slightly small in size but is capable of stably extracting features under the illumination, the time of movement from the position to the position at the time of alert generation may be calculated. If a plurality of such places exist, the time difference may be obtained for each place, and the time difference representative value may be calculated for each place. Alternatively, if there is a place where the background has a color different from the color of clothing, and the features can easily be extracted (for example, a place with a dark background color is preferable for a person dressed in white), that place may be used. However, the easily extractable background color changes depending on the color of clothing. Hence, a plurality of positions may be obtained, and the representative value of the time difference may be obtained for each position. In this case, a suitable time difference is selected in accordance with the color actually extracted by an object feature extractor **512** at the time of query feature generation.

[0107] As described above, according to this embodiment, since object detection is performed while learning the time differences, the object extraction image selection accuracy is improved along with the use. As a result, the accuracy of query features is improved.

Seventh Embodiment

[0108] An information processing apparatus according to the seventh embodiment of the present invention will be described with reference to FIG. **15**. FIG. **15** is a block diagram for explaining the arrangement and use method of a search query generator **1501** according to this embodiment. This

embodiment is different from the above-described sixth embodiment in that a time difference learner **1514** learns the time difference in consideration of an object feature extraction result as well. The rest of the arrangement and operation is the same as in the sixth embodiment. The same arrangement and operation are denoted by the same reference numerals, and a detailed description thereof will be omitted.

[0109] The time difference learner **1514** determines the time difference in consideration of the reliability of an extracted object feature as well. The reliability calculation method is the same as described above. The reliability of clothing feature extraction can be determined based on the aspect ratio of the circumscribed rectangle of a person or the degree of person likelihood of a person silhouette shape. As for a color feature, the reliability can be evaluated by analyzing the magnitude of saturation and that of brightness of a color (if the saturation or brightness is low, the influence of shadow or the like is supposedly present) and the presence/absence of a saturated portion. Learning is performed by selecting a time difference of a reliable feature.

[0110] As described above, according to this embodiment, since the time difference is learned in consideration of the reliability of a feature, the object extraction image selection accuracy is improved along with the use. As a result, the accuracy of query features is improved.

Other Embodiments

[0111] While the present invention has been described with reference to exemplary embodiments, it is to be understood that the invention is not limited to the disclosed exemplary embodiments. The scope of the following claims is to be accorded the broadest interpretation so as to encompass all such modifications and equivalent structures and functions.

[0112] The present invention is applicable to a system including a plurality of devices or a single apparatus. The present invention is also applicable even when a program for implementing the functions of the embodiments is supplied to the system or apparatus directly or from a remote site. Hence, the present invention also incorporates the program installed in a computer to implement the functions of the present invention on the computer, a storage medium storing the program, and a WWW (World Wide Web) server that causes a user to download the program.

Other Expressions of Embodiments

[0113] Some or all of the above-described embodiments can also be described as in the following supplementary notes but are not limited to the followings.

(Supplementary Note 1)

[0114] There is provided an information processing apparatus comprising: [0115] a sensor that analyzes a captured video and senses whether a predetermined event has occurred; [0116] a determining unit that determines a type of an object to be used as query information based on a type of the event in response to sensing of the event occurrence; and [0117] a generator that detects the object of the type determined by the determining unit from the video and generates the query information based on the object.

(Supplementary Note 2)

[0118] There is provided the information processing apparatus according to supplementary note 1, wherein the generator generates a feature of the object of the type determined by the determining unit as the query information.

(Supplementary Note 3)

[0119] There is provided the information processing apparatus according to supplementary note 1 or 2, wherein the generator generates the query information of the type determined by the determining unit using the video at a timing different by a predetermined time difference from a time of the event occurrence, and the determining unit further determines the predetermined time difference in accordance with the type of the event.

(Supplementary Note 4)

[0120] There is provided the information processing apparatus according to supplementary note 1, 2, or 3, wherein the generator generates the query information using the video in a predetermined

time width at the timing different from the time of the event occurrence, and [0121] the determining unit further determines the predetermined time width in accordance with the type of the event.

(Supplementary Note 5)

[0122] There is provided the information processing apparatus according to any one of supplementary notes 1 to 4, wherein the generator generates the query information using a predetermined region in the video, and [0123] the determining unit further determines the predetermined region in accordance with the type of the event.

(Supplementary Note 6)

[0124] There is provided the information processing apparatus according to any one of supplementary notes 1 to 5, wherein the generator generates the query information using a predetermined number of images including at least one image in the video, and [0125] the determining unit further determines the predetermined number in accordance with the type of the event.

(Supplementary Note 7)

[0126] There is provided the information processing apparatus according to any one of supplementary notes 1 to 6, further comprising: [0127] a tracker that tracks the object associated with the event; and [0128] an updater that updates the time difference determined by the determining unit by calculating, from a result of the tracking of the object by the tracker, the time difference from the time of the event occurrence to the timing of the image to generate the query information.

(Supplementary Note 8)

[0129] There is provided the information processing apparatus according to any one of supplementary notes 1 to 7, wherein the event is finding of a person included in a predetermined list, the sensor determines, based on a face image of the person, whether the person is a person included in the list, and the determining unit determines clothing of the person included in the list as the query information.

(Supplementary Note 9)

[0130] There is provided the information processing apparatus according to any one of supplementary notes 1 to 8, wherein the event is finding of a fall of a person in the video, and [0131] the determining unit determines clothing of the person who has fallen before the fall as the query information.

(Supplementary Note 10)

[0132] There is provided the information processing apparatus according to any one of supplementary notes 1 to 9, wherein the event is finding of one of abandonment and carry away of luggage, and [0133] the determining unit determines clothing of a person who has abandoned or carried away the luggage as the query information.

(Supplementary Note 11)

[0134] There is provided the information processing apparatus according to any one of supplementary notes 1 to 10, wherein the event is finding of a person who has entered a restricted area in the video, and [0135] the determining unit determines clothing of the person who has entered the restricted area as the query information.

(Supplementary Note 12)

[0136] There is provided an information processing method comprising: analyzing a captured video and sensing whether a predetermined event has occurred; [0137] deciding a type of query information to be searched for in accordance with on a type of the event in response to sensing of the event occurrence; and [0138] generating the query information of the type determined in the deciding from the video.

(Supplementary Note 13)

[0139] There is provided the information processing method according to supplementary note 12,

wherein in the generating, a feature of the object of the type determined in the deciding is generated as the query information.

(Supplementary Note 14)

[0140] There is provided the information processing method according to supplementary note 12 or 13, wherein in the generating, the query information of the type determined in the deciding is generated using the video at a timing different by a predetermined time difference from a time of the event occurrence, and [0141] in the deciding, the predetermined time difference is further determined in accordance with the type of the event.

(Supplementary Note 15)

[0142] There is provided the information processing method according to supplementary note 12, 13, or 14, wherein in the generating, the query information is generated using the video in a predetermined time width at the timing different from the time of the event occurrence, and [0143] in the deciding, the predetermined time width is further determined in accordance with the type of the event.

(Supplementary Note 16)

[0144] There is provided the information processing method according to any one of supplementary notes 12 to 15, wherein in the generating, the query information is generated using a predetermined region in the video, and [0145] in the deciding, the predetermined region is further determined in accordance with the type of the event.

(Supplementary Note 17)

[0146] There is provided the information processing method according to any one of supplementary notes 12 to 16, wherein in the generating, the query information is generated using a predetermined number of images including at least one image in the video, and [0147] in the deciding, the predetermined number is further determined in accordance with the type of the event.

(Supplementary Note 18)

[0148] There is provided the information processing method according to any one of supplementary notes 12 to 17, further comprising: [0149] tracking the object associated with the event; and [0150] updating the time difference determined in the deciding by calculating, from a result of the tracking of the object in the tracking, the time difference from the time of the event occurrence to the timing of the image to generate the query information.

(Supplementary Note 19)

[0151] There is provided the information processing method according to any one of supplementary notes 12 to 18, wherein the event is finding of a person included in a predetermined list, [0152] in the analyzing the captured video and sensing whether the predetermined event has occurred, it is determine, based on a face image of the person, whether the person is a person included in the list, and [0153] in the deciding, clothing of the person included in the list is determined as the query information.

(Supplementary Note 20)

[0154] There is provided the information processing method according to any one of supplementary notes 12 to 19, wherein the event is finding of a fall of a person in the video, and [0155] in the deciding, clothing of the person who has fallen before the fall is determined as the query information.

(Supplementary Note 21)

[0156] There is provided the information processing method according to any one of supplementary notes 12 to 20, wherein the event is finding of one of abandonment and carry away of luggage, and [0157] in the deciding, clothing of a person who has abandoned or carried away the luggage is determined as the query information.

(Supplementary Note 22)

[0158] There is provided the information processing method according to any one of supplementary notes 12 to 21, wherein the event is finding of a person who has entered a restricted

area in the video, and [0159] in the deciding, clothing of the person who has entered the restricted area is determined as the query information.

(Supplementary Note 23)

[0160] There is provided an information processing program for causing a computer to execute:

[0161] analyzing a captured video and sensing whether a predetermined event has occurred; [0162] deciding a type of query information to be searched for in accordance with on a type of the event in response to sensing of the event occurrence; and [0163] generating the query information of the type determined in the deciding from the video.

(Supplementary Note 24)

[0164] There is provided the information processing program according to supplementary note 23, wherein in the generating, a feature of the object of the type determined in the deciding is generated as the query information.

(Supplementary Note 25)

[0165] There is provided the information processing program according to supplementary note 23 or 24, wherein in the generating, the query information of the type determined in the deciding is generated using the video at a timing different by a predetermined time difference from a time of the event occurrence, and [0166] in the deciding, the predetermined time difference is further determined in accordance with the type of the event.

(Supplementary Note 26)

[0167] There is provided the information processing program according to supplementary note 23, 24, or 25, wherein in the generating, the query information is generated using the video in a predetermined time width at the timing different from the time of the event occurrence, and in the deciding, the predetermined time width is further determined in accordance with the type of the event.

(Supplementary Note 27)

[0168] There is provided the information processing program according to any one of supplementary notes 23 to 26, wherein in the generating, the query information is generated using a predetermined region in the video, and [0169] in the deciding, the predetermined region is further determined in accordance with the type of the event.

(Supplementary Note 28)

[0170] There is provided the information processing program according to any one of supplementary notes 23 to 27, wherein in the generating, the query information is generated using a predetermined number of images including at least one image in the video, and in the deciding, the predetermined number is further determined in accordance with the type of the event.

(Supplementary Note 29)

[0171] There is provided the information processing program according to any one of supplementary notes 23 to 28, further comprising: [0172] tracking the object associated with the event; and [0173] updating the time difference determined in the deciding by calculating, from a result of the tracking of the object in the tracking, the time difference from the time of the event occurrence to the timing of the image to generate the query information.

(Supplementary Note 30)

[0174] There is provided the information processing program according to any one of supplementary notes 23 to 29, wherein the event is finding of a person included in a predetermined list, [0175] in the analyzing the captured video and sensing whether the predetermined event has occurred, it is determine, based on a face image of the person, whether the person is a person included in the list, and [0176] in the deciding, clothing of the person included in the list is determined as the query information.

(Supplementary Note 31)

[0177] There is provided the information processing program according to any one of supplementary notes 23 to 30, wherein the event is finding of a fall of a person in the video, and

[0178] in the deciding, clothing of the person who has fallen before the fall is determined as the query information.
(Supplementary Note 32)
[0179] There is provided the information processing program according to any one of supplementary notes 23 to 31, wherein the event is finding of one of abandonment and carry away of luggage, and [0180] in the deciding, clothing of a person who has abandoned or carried away the luggage is determined as the query information.
(Supplementary Note 33)
[0181] There is provided the information processing program according to any one of supplementary notes 23 to 32, wherein the event is finding of a person who has entered a restricted area in the video, and [0182] in the deciding, clothing of the person who has entered the restricted area is determined as the query information.
[0183] This application claims the benefit of Japanese Patent Application No. 2012-217592 filed on Sep. 28, 2012, which is hereby incorporated by reference herein in its entirety.

Claims

1. An information processing apparatus comprising: a memory configured to store program instructions; one or more processors configured to read the program instructions to: analyze videos which are captured by a plurality of cameras to detect at least one type of event among predetermined plural types of events, determine a type of a target object to be detected; detect the target object from a second part of video which is different from a first part of video in which the at least one type of event is detected and provide an operator with the second part of video.
2. The information processing apparatus according to claim 1, wherein the memory configured to store a table representing a correspondence between the type of event and the type of the target object to be found out.
3. The information processing apparatus according to claim 2, wherein the table represents a correspondence between the type of event and time difference between the first and second part of video, and wherein the one or more processors are configured to read the program instructions to determine the second part of video according to the time difference.
4. The information processing apparatus according to claim 3, wherein the processors configured to read the program instructions to update the time difference included in the table when the target object is detected from a third part of video.
5. The information processing apparatus according to claim 2, wherein the table represents a correspondence between the type of event and two or more time differences between the first part of video and the second parts of video, and wherein the one or more processors are configured to read the program instructions to determine the second parts of video according to the time difference.
6. The information processing apparatus according to claim 2, wherein the table represents a correspondence between the type of event and time width of the second part of video.
7. The information processing apparatus according to claim 2, wherein the table represents a correspondence between the type of event and an area in the second part of video, and wherein the one or more processors are configured to read the program instructions to detect the target object in the area of the second part of video.
8. An information processing method comprising: analyzing videos which are captured by a plurality of cameras to detect at least one type of event among predetermined plural types of events, determining a type of a target object to be detected; and detecting the target object from a second part of video which is different from a first part of video in which the at least one type of event is detected and provide an operator with the second part of video.
9. The information processing method according to claim 8, wherein the type of a target object is

determined with reference to a table storing a correspondence between the type of event and the type of the target object to be detected.

10. The information processing method according to claim 9, wherein the table represents a correspondence between the type of event and time difference between the first and second part of video, and wherein the second part of video is determined according to the time difference.

11. The information processing method according to claim 9, further comprising updating the time difference when the target object is detected from a third part of video.

12. The information processing method according to claim 9, wherein the table represents a correspondence between the type of event and two or more time differences between the first part of video and the second parts of video, and wherein the second parts of video are determined according to the time difference.

13. The information processing method according to claim 9, wherein the table represents a correspondence between the type of event and time width of the second part of video.

14. The information processing method according to claim 9, wherein the table represents a correspondence between the type of event and an area in the second part of video, and the method comprises detecting the target object in the area of the second part of video.

15. A non-transitory computer readable medium storing a program causing a computer to execute a process comprising: analyzing videos which are captured by a plurality of cameras to detect at least one type of event among predetermined plural types of events, determining a type of a target object to be detected; and detecting the target object from a second part of video which is different from a first part of video in which the at least one type of event is detected and provide an operator with the second part of video.

16. The non-transitory computer readable medium according to claim 15, wherein the type of a target object is determined with reference to a table storing a correspondence between the type of event and the type of the target object to be detected.

17. The non-transitory computer readable medium according to claim 16, wherein the table represents a correspondence between the type of event and time difference between the first and second part of video, and wherein the second part of video is determined according to the time difference.

18. The non-transitory computer readable medium according to claim 16, further comprising updating the time difference when the target object is detected from a third part of video.

19. The non-transitory computer readable medium according to claim 16, wherein the table represents a correspondence between the type of event and two or more time differences between the first part of video and the second parts of video, and wherein the second parts of video are determined according to the time difference.

20. The non-transitory computer readable medium according to claim 16, wherein the table represents a correspondence between the type of event and time width of the second part of video.