

TARGET TRACKING METHOD AND APPARATUS, ELECTRONIC DEVICE AND STORAGE MEDIUM

CROSS-REFERENCE TO RELATED APPLICATION

[0001] The present disclosure is based upon and claims priority to Chinese Patent Application No. 202011256801.6 filed on November 11, 2020, the disclosure of which is hereby incorporated by reference in its entirety.

TECHNICAL FIELD

[0002] The present disclosure relates to the field of computer vision technologies, and in particular to a target tracking method and apparatus, an electronic device and a storage medium.

BACKGROUND

[0003] Target tracking refers to tracking the same object (for example, a pedestrian, a vehicle, etc.) in consecutive video frames. With the development of computer vision technologies, target tracking algorithms are widely used in security and other fields, and are of great value in building intelligent life. However, in the case of dense target objects, the consistency of Identity document (ID) of the target objects is poor in a target tracking process.

SUMMARY

[0004] The present disclosure provides a technical solution of target tracking.

[0005] According to an aspect of the present disclosure, a target tracking method is provided, which includes the following operations.

- [0006] Multiple target objects conflicting with each other are determined according to distances between the target objects in a tracking area.
- [0007] In response to a situation that any target object in the multiple target objects does not conflict with other target objects in the multiple target objects, a current feature of the target object is extracted.
- [0008] A reference feature matching the current feature of the target object is determined based on multiple reference features corresponding to the multiple target objects, the multiple reference features being features extracted from the multiple target objects before the multiple target objects conflict with each other.
- [0009] Identification information corresponding to the reference feature matching the current feature of the target object is determined as identification information of the target object.
- [0010] In embodiments of the present disclosure, multiple target objects conflicting with each other are determined according to the distances between target objects in the tracking area; in response to the situation that any target object in the multiple target objects no longer conflicts with other target objects in the multiple target objects, the current feature of the target object is extracted; the reference feature matching the current feature of the target object is determined based on multiple reference features corresponding to the multiple target objects; and the identification information corresponding to the reference feature matching the current feature of the target object is determined as the identification information of the target object. In this way, the identification information of the target object can be kept consistent, that is, the identification information of the same target object can be kept consistent when the target object gets in and out of the conflict area, and thus the accuracy of multi-target tracking can be improved.
- [0011] In a possible implementation mode, the distance between any target object in the multiple target objects conflicting with each other and at least one of the other target objects in the multiple target objects is less than or equal to a distance threshold.
- [0012] The operation that in response to the situation that any target object in the

multiple target objects does not conflict with other target objects in the multiple target objects, the current feature of the target object is extracted may include the following operation.

[0013] For any target object in the multiple target objects, in response to the distances between the target object and other target objects in the multiple target objects being greater than the distance threshold, the current feature of the target object is extracted.

[0014] According to the implementation mode, a conflict relationship between the target objects is determined and target tracking is performed based on the relationship, which helps to improve the accuracy and efficiency of the target tracking.

[0015] In a possible implementation mode, the method may further include the following operations.

[0016] For any target object, first coordinates of the target object corresponding to multiple cameras are determined, the first coordinates of the target object corresponding to any camera being used for representing the first coordinates of the target object obtained according to an image acquired by the camera.

[0017] A sub-area where the target object is located in the tracking area is determined according to the first coordinates of the target object corresponding to the multiple cameras, the tracking area including multiple sub-areas.

[0018] According to the confidence of the multiple cameras for the sub-area where the target object is located, the first coordinates of the target object corresponding to the multiple cameras are fused to obtain fusion coordinates of the target object.

[0019] Before multiple target objects conflicting with each other are determined according to the distances between the target objects in the tracking area, the method may further include the following operation.

[0020] The distances between the target objects in the tracking area are determined according to the fusion coordinates of the target objects in the tracking area.

[0021] In the implementation mode, the fusion coordinates of the target object are

obtained in combination with the first coordinates of the target object corresponding to the multiple cameras, and target tracking is performed based on the fusion coordinates of the target object. In this way, target tracking can be performed based on more accurate coordinates, which helps to improve the accuracy of target tracking.

[0022] In a possible implementation mode, the operation that the current feature of the target object is extracted may include the following operations.

[0023] A sub-area where the target object is located in the tracking area is determined.

[0024] A camera extracting the current feature of the target object is determined according to the confidence of multiple cameras for the sub-area where the target object is located.

[0025] The current feature of the target object is extracted according to video frames acquired by the camera extracting the current feature of the target object.

[0026] In the implementation mode, the camera extracting the current feature of the target object is determined according to the confidence of multiple cameras for the sub-area where the target object is located, and the current feature of the target object is extracted according to the video frames acquired by the camera extracting the current feature of the target object. In this way, the extracted current feature of the target object can have richer visual information. Therefore, by performing target matching based on the current feature of the target object extracted in the implementation mode, the accuracy of target matching can be improved.

[0027] In a possible implementation mode, the operation that the camera extracting the current feature of the target object is determined according to the confidence of multiple cameras for the sub-area where the target object is located may include the following operation.

[0028] The camera extracting the current feature of the target object is determined according to the confidence of multiple cameras for the sub-area where the target object is located and overlapping information between a detection box of the target object and detection boxes of other target objects in the video frames acquired by the multiple

cameras.

[0029] In the implementation mode, the camera extracting the current feature of the target object is determined according to the confidence of multiple cameras for the sub-area where the target object is located and the overlapping information between the detection box of the target object and the detection boxes of other target objects in the video frames acquired by the multiple cameras, and the current feature of the target object is extracted according to the video frames acquired by the camera for extracting the current feature of the target object. In this way, the extracted current feature of the target object can have richer visual information. Therefore, by performing target matching based on the current feature of the target object extracted in the implementation mode, the accuracy of target matching can be further improved.

[0030] In a possible implementation mode, the overlapping information includes Intersection-over-Union (IoU) between the detection box of the target object and the detection boxes of the other target objects.

[0031] The operation that the camera extracting the current feature of the target object is determined according to the confidence of multiple cameras for the sub-area where the target object is located and the overlapping information between the detection box of the target object and the detection boxes of other target objects in the video frames acquired by the multiple cameras may include the following operation.

[0032] From cameras that meet an overlapping condition, a camera with the highest confidence for the sub-area where the target object is located is determined as the camera extracting the current feature of the target object. The overlapping condition indicates that in the video frames acquired currently, the IoU between the detection box of the target object and the detection boxes of the other target objects is less than a predetermined threshold.

[0033] In the implementation mode, the camera with the highest confidence for the sub-area where the target object is located is determined from the cameras whose currently acquired video frames meet the overlapping condition in the multiple cameras, as the camera extracting the current feature of the target object, and the current feature of

the target object is extracted according to the video frames acquired by the camera extracting the current feature of the target object. In this way, the extracted current feature of the target object can have richer visual information. Therefore, by performing target matching based on the current feature of the target object extracted in the implementation mode, the accuracy of target matching can be further improved.

[0034] In a possible implementation mode, the method may further include the following operation.

[0035] For any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas is determined according to the distances between the camera and the multiple sub-areas of the tracking area.

[0036] In the implementation mode, for any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas is determined according to the distances between the camera and the multiple sub-areas of the tracking area, and coordination fusion is performed and/or the camera for extracting the current feature of the target object is determined based on the confidence determined, which helps to improve the accuracy of target tracking.

[0037] In a possible implementation mode, after the confidence of the multiple cameras for the multiple sub-areas is determined, the method may further include the following operation.

[0038] For any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas is adjusted according to the average distance between the target objects in the video frames acquired by the camera.

[0039] In the implementation mode, for any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas is dynamically adjusted according to the average distance between the target objects in the video frames acquired by the camera, and coordination fusion is performed and/or the camera for extracting the current feature of the target object is determined based on the confidence determined, which helps to improve the accuracy of target tracking.

- [0040] In a possible implementation mode, the operation that the reference feature matching the current feature of the target object is determined based on multiple reference features corresponding to the multiple target objects may include the following operations.
- [0041] The reference feature that has the maximum similarity to the current feature of the target object is determined from the multiple reference features corresponding to the multiple target objects.
- [0042] In response to the similarity between the reference feature with the maximum similarity and the current feature of the target object being greater than or equal to a similarity threshold, the reference feature with the maximum similarity is determined as the reference feature matching the current feature of the target object.
- [0043] In the implementation mode, the reference feature that has the maximum similarity to the current feature of the target object is determined from multiple reference features corresponding to the multiple target objects, and in response to the similarity between the reference feature with the maximum similarity and the current feature of the target object being greater than or equal to the similarity threshold, the reference feature with the maximum similarity is determined as the reference feature matching the current feature of the target object. In this way, when the target object no longer conflicts with other target objects in multiple target objects, first the reference feature matching the current feature of the target object is searched based on the multiple reference features corresponding to the multiple target objects, which helps to improve the speed and accuracy of searching the reference feature matching the current feature of the target object.
- [0044] In a possible implementation mode, after the reference feature that has the maximum similarity to the current feature of the target object is determined, the method may further include the following operation.
- [0045] In response to the similarity between the reference feature with the maximum similarity and the current feature of the target object being less than the similarity threshold, the reference feature matching the current feature of the target object is

determined based on the reference feature corresponding to a target object that does not conflict with the target object and conflicts with any of the other target objects.

[0046] In the implementation mode, in response to the situation that the similarity between the reference feature with the maximum similarity and the current feature of the target object is less than the similarity threshold, the reference feature matching the current feature of the target object is determined based on the reference feature corresponding to a target object that does not conflict with the target object and conflicts with any of the other target objects. In this way, the accuracy of the determined reference feature matching the current feature of the target object can be further improved on the premise of improving the speed of searching the reference feature matching the current feature of the target object.

[0047] In a possible implementation mode, the method may further include the following operations.

[0048] For any target object, in response to a situation that the target object does not conflict with other target objects at either the current moment or at the previous moment of the current moment, target objects that do not conflict with other target objects at the previous moment and do not match other target objects at the current moment is determined as the remaining target objects at the previous moment; or, in response to a situation that the target object conflicts with the same other target objects at the current moment and the previous moment, target objects that do not match other target objects of the current moment in a conflict area to which the target object belongs at the previous moment is determined as the remaining target objects at the previous moment.

[0049] A target object closest to the target object at the current moment is determined from the remaining target objects of the previous moment as a target object matching the target object at the current moment, and the identification information of the matched target object is taken as the identification information of the target object.

[0050] According to the implementation mode, computation amount can be reduced, and time cost can be reduced to meet a real-time requirement of target tracking.

[0051] According to an aspect of the present disclosure, a target tracking apparatus is

provided, which includes: a first determining portion, an extracting portion, a second determining portion, and a third determining portion.

[0052] The first determining portion is configured to determine multiple target objects conflicting with each other according to distances between target objects in a tracking area.

[0053] The extracting portion is configured to extract, in response to the situation that any target object in the multiple target objects does not conflict with other target objects in the multiple target objects, a current feature of the target object.

[0054] The second determining portion is configured to determine a reference feature matching the current feature of the target object based on multiple reference features corresponding to the multiple target objects, the multiple reference features being features extracted from the multiple target objects before the multiple target objects conflict with each other.

[0055] The third determining portion is configured to determine identification information corresponding to the reference feature matching the current feature of the target object as identification information of the target object.

[0056] According to an aspect of the present disclosure, an electronic device is provided, which includes: one or more processors; and a memory configured to store an executable instruction. The one or more processors are configured to call the executable instruction stored in the memory to execute the above method.

[0057] According to an aspect of the present disclosure, a computer-readable storage medium is provided, in which a computer program instruction is stored. When executed by a processor, the computer program instruction implements the above method.

[0058] According to an aspect of the present disclosure, a computer program is provided, which includes a computer-readable code. When the computer-readable code runs in the electronic device, a processor in the electronic device implements the above method.

[0059] In the embodiments of the present disclosure, multiple target objects conflicting with each other are determined according to the distances between the target objects in the tracking area; in response to the situation that any target object in the multiple target objects no longer conflicts with other target objects in the multiple target objects, the current feature of the target object is extracted; the reference feature matching the current feature of the target object is determined based on multiple reference features corresponding to the multiple target objects; and the identification information corresponding to the reference feature matching the current feature of the target object is determined as the identification information of the target object. In this way, the identification information of the target object can be kept consistent, that is, the identification information of the same target object can be kept consistent when the target object gets in and out of the conflict area, and thus the accuracy of multi-target tracking can be improved.

[0060] It is to be understood that the above general descriptions and detailed descriptions below are only exemplary and explanatory and not intended to limit the present disclosure.

[0061] Other features and aspects of the present disclosure will become clear as exemplary embodiments are detailed below with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0062] The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate embodiments consistent with the present disclosure and, together with the specification, serve to illustrate the technical solutions of the present disclosure.

[0063] FIG. 1 illustrates a flowchart of a target tracking method provided by some embodiments of the present disclosure.

[0064] FIG. 2 illustrates a schematic diagram of multiple target objects conflicting

with each other at the t-th moment.

[0065] FIG. 3 illustrates a schematic diagram of multiple target objects conflicting with each other at the (t+1)-th moment.

[0066] FIG. 4 illustrates a schematic diagram of multiple target objects and multiple reference features corresponding to the multiple target objects in some embodiments of the present disclosure.

[0067] FIG. 5 illustrates a schematic diagram of placing markers in a tracking area in some embodiments of the present disclosure.

[0068] FIG. 6 illustrates a block diagram of a target tracking apparatus provided by some embodiments of the present disclosure.

[0069] FIG. 7 shows a block diagram of an electronic device 800 provided by some embodiments of the present disclosure.

[0070] FIG. 8 shows a block diagram of an electronic device 1900 provided by some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0071] Each exemplary embodiment, feature and aspect of the present disclosure will be described below with reference to the drawings in detail. The same reference signs in the drawings represent components with the same or similar functions. Although each aspect of the embodiments is shown in the drawings, the drawings are not required to be drawn to scale, unless otherwise specified.

[0072] The special word "exemplary" here means "as an example, embodiment or illustration". Here, it is unnecessary to interpret any embodiment described as "exemplary" as being better than other embodiments.

[0073] In the specification, term "and/or" is only an association relationship describing associated objects and represents that three relationships may exist. For

example, A and/or B may represent three conditions: i.e., independent existence of A, existence of both A and B and independent existence of B. In addition, term "at least one" in the application represents any one of multiple or any combination of at least two of multiple. For example, including at least one of A, B and C may represent including any one or more elements selected from a set formed by A, B and C.

[0074] In addition, for describing the present disclosure better, many specific details are presented in the following specific implementation modes. It is to be understood by those skilled in the art that the present disclosure may still be implemented even without some specific details. In some examples, methods, means, components and circuits known very well to those skilled in the art are not described in detail, to highlight the subject of the present disclosure.

[0075] FIG. 1 illustrates a flowchart of a target tracking method provided by some embodiments of the present disclosure. The execution entity of the target tracking method may be a target tracking apparatus. For example, the target tracking method may be performed by a terminal device or a server or other processing devices. The terminal device may be User Equipment (UE), a mobile device, a user terminal, a terminal, a cell phone, a cordless phone, a Personal Digital Assistant (PDA), a handheld device, a computing device, a vehicle device, a wearable device and the like. In some possible implementation modes, the target tracking method may be implemented by a processor calling a computer-readable instruction stored in the memory. As shown in FIG. 1, the target tracking method may include operations S11 to S14.

[0076] At S11, multiple target objects conflicting with each other are determined according to distances between target objects in a tracking area.

[0077] The embodiments of the present disclosure may be applied to intelligent monitoring, intelligent scenario analysis, security, target detection, target tracking and other application scenarios. In some embodiments of the present disclosure, the tracking area represents an area where target tracking is required. The tracking area in the embodiments of the present disclosure may be a relatively closed area or a relatively open area. For example, the tracking area may be a field, a shopping mall, a classroom, etc. The target object represents an object needing to be tracked. For example, the target

object may be a pedestrian, a vehicle, an athlete on a sports field (such as a player on a football field), and anything else needing to be tracked. In some embodiments of the present disclosure, the target objects conflicting with each other may represent target objects that are close to each other.

[0078] In some embodiments of the present disclosure, the distances between the target objects in the tracking area may be determined respectively for each of video frames acquired by a camera, so as to determine the conflict between the target objects at the corresponding moment of the video frames. Alternatively, the video frames acquired by a camera may also not be analyzed frame by frame. For example, the distances between the target objects in the tracking area may be determined every several video frames, so as to determine the conflict between the target objects at the corresponding moment of the video frames.

[0079] In some embodiments of the present disclosure, at any moment, there may be one or more groups of target objects that conflict with each other. Alternatively, there may be no target objects that conflict with each other. FIG. 2 illustrates a schematic diagram of multiple target objects that conflict with each other at the t-th moment. As shown in FIG. 2, at the t-th moment, target objects O₁ and O₄ conflict with target object O₃, target object O₂ conflicts with target object O₅, and target object O₆ conflicts with target object O₇.

[0080] In a possible implementation mode, multiple target objects conflicting with each other may be added to the same conflict area, that is, the multiple target objects conflicting with each other may be considered to be in the same conflict area. The conflict area may be a virtual area, and the target objects in any conflict area conflict with each other. For example, in FIG. 2, at the t-th moment, conflict area 1 includes the target objects O₁, O₄ and O₃, conflict area 2 includes the target objects O₂ and O₅, and conflict area 3 includes the target objects O₆ and O₇.

[0081] At S12, in response to a situation that any target object in the multiple target objects does not conflict with other target objects in the multiple target objects, the current feature of the target object is extracted.

[0082] FIG. 3 illustrates a schematic diagram of multiple target objects conflicting with each other at the $(t+1)$ -th moment. As shown in FIG. 3, at the $(t+1)$ -th moment, the conflict area 1 includes the target objects O_1 and O_4 , the conflict area 2 includes the target objects O_2 and O_5 , and the conflict area 3 includes the target objects O_6 and O_7 . That is, at the $(t+1)$ -th moment, the target object O_1 conflicts with the target object O_4 , the target object O_2 conflicts with the target object O_5 , and the target object O_6 conflicts with the target object O_7 . It can be known by comparison based on the $(t+1)$ -th moment and the t -th moment that at the $(t+1)$ -th moment, the target object O_3 no longer conflicts with the target objects O_1 and O_4 . That is, all the target objects in the scene are dynamically clustered into multiple conflict areas at any moment.

[0083] In a possible implementation mode, the time interval between two adjacent moments may be equal to a reciprocal of a frame rate for acquiring video frames by a camera. In another possible implementation mode, the time interval between two adjacent moments may be greater than the reciprocal of the frame rate for acquiring the video frames by the camera, for example, the time interval may be equal to H times the reciprocal of the frame rate for acquiring the video frames by the camera, where H is an integer greater than 1.

[0084] In a possible implementation mode, the distance between any target object in multiple target objects conflicting with each other and at least one other target object in the multiple target objects is less than or equal to a distance threshold. In response to any target object in the multiple target objects does not conflict with the other target objects in the multiple target objects, the current feature of the target object is extracted may include that: for any target object in the multiple target objects, in response to a situation that the distances between the target object and each of the other target objects in the multiple target objects are greater than the distance threshold, the current feature of the target object is extracted.

[0085] In some embodiments, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) may be used to cluster the target objects, the distances between which and other target objects are less than the distance threshold, to form conflict areas.

[0086] In the implementation mode, a distance between any two target objects may be

calculated in any frame. If the distance between the two target objects is less than or equal to the distance threshold, it may be determined that the two target objects conflict with each other. If a distance between another target object and at least one of the two target objects is less than or equal to the distance threshold, it may be determined that the three target objects conflict with each other. For example, at the t -th moment, if the distance between the target object O_1 and the target object O_4 is less than or equal to the distance threshold, it may be determined that the target object O_1 conflicts with the target object O_4 . If the distance between the target object O_3 and the target object O_4 is less than or equal to the distance threshold, it may be determined that the target objects O_1 and O_4 conflict with the target object O_3 .

[0087] As an example of the implementation mode, when it is determined whether the target objects conflict with each other, whether a distance between locations (for example, fusion coordinates) of the target objects is less than or equal to the distance threshold may be only considered. For example, if the distance between the locations of any two target objects is less than or equal to the distance threshold, it may be determined that the two target objects conflict with each other. As another example of the implementation mode, not only the distance between the locations of the target objects, but also a distance between visual features (for example, reference features and/or current features) of the target objects may be considered. For example, if the distance between the locations of any two object objects is less than or equal to a first distance threshold, and the distance between the visual features of the two object objects is less than or equal to a second distance threshold, it may be determined that the two object objects conflict with each other. The distance between the visual features of the target objects may represent the similarity between the visual features of the target objects.

[0088] In the implementation mode, if the distance between one target object and another target object is less than or equal to the distance threshold, it may be determined that the two target objects conflict with each other; if the distances between a target object and each of other target objects are greater than the distance threshold, it may be determined that the target object does not conflict with any of the other target object. According to the implementation mode, a conflict relationship between the target objects is determined and target tracking is performed based on the conflict relationship, which

helps to improve the accuracy and efficiency of target tracking.

[0089] In another possible implementation mode, the distances between any target object in multiple target objects conflicting with each other and all of the other target objects in the multiple target objects are less than or equal to a distance threshold. The operation that in response to the situation that any target object in the multiple target objects does not conflict with other target objects in the multiple target objects, the current feature of the target object is extracted may include that: for any target object in the multiple target objects, when the distance between the target object and any target object in the multiple target objects is greater than the distance threshold, the current feature of the target object is extracted.

[0090] In a possible implementation mode, the operation that the current feature of the target object is extracted may include that: a sub-area where the target object is located in the tracking area is determined; a camera for extracting the current feature of the target object is determined according to confidence of multiple cameras for the sub-area where the target object is located; and the current feature of the target object is extracted according to video frames acquired by the camera for extracting the current feature of the target object.

[0091] In the implementation mode, the tracking area may include at least one sub-area, for example, the tracking area may be divided into multiple sub-areas. In the implementation mode, visual information at multiple viewing angles corresponding to multiple cameras can be obtained, and the visual information obtained from the multiple viewing angles can complement each other, which helps to improve the accuracy of target detection and tracking.

[0092] In the implementation mode, the camera for extracting the current feature of the target object is determined according to the confidence of multiple cameras for the sub-area where the target object is located, and the current feature of the target object is extracted according to the video frames acquired by the camera for extracting the current feature of the target object; in this way, the extracted current feature of the target object can have richer visual information. Therefore, by performing target matching based on the current feature of the target object extracted in the implementation mode, the

accuracy of target matching can be improved.

[0093] As an example of the implementation mode, the operation that the camera for extracting the current feature of the target object is determined according to the confidence of multiple cameras for the sub-area where the target object is located may include that: the camera for extracting the current feature of the target object is determined according to the confidence of the multiple cameras for the sub-area where the target object is located and overlapping information between a detection box of the target object and detection boxes of other target objects in the video frames acquired by the multiple cameras.

[0094] In the example, the camera for extracting the current feature of the target object is determined according to the confidence of multiple cameras for the sub-area where the target object is located and the overlapping information between the detection box of the target object and the detection boxes of other target objects in the video frames acquired by the multiple cameras, and the current feature of the target object is extracted according to the video frames acquired by the camera for extracting the current feature of the target object; in this way, the extracted current feature of the target object can have richer visual information. Therefore, by performing target matching based on the current feature of the target object extracted in the example, the accuracy of target matching can be further improved.

[0095] In an example, the overlapping information includes Intersection-over-Union (IoU) between the detection box of the target object and the detection boxes of other target objects. The operation that the camera for extracting the current feature of the target object is determined according to the confidence of multiple cameras for the sub-area where the target object is located and the overlapping information between the detection box of the target object and the detection boxes of other target objects in the video frames acquired by the multiple cameras may include that: a camera with the highest confidence for the sub-area where the target object is located is determined from the cameras that meet an overlapping condition, as the camera for extracting the current feature of the target object. The overlapping condition indicates that in the video frames acquired currently, the IoU between the detection box of the target object and the

detection boxes of other target objects is less than a predetermined threshold.

[0096] For example, the sub-area where the target object is located is sub-area A_1 , the multiple cameras include camera C_1 , camera C_2 , camera C_3 and camera C_4 , the confidence of the camera C_1 for the sub-area A_1 is Z_{11} , the confidence of the camera C_2 for the sub-area A_1 is Z_{21} , the confidence of the camera C_3 for the sub-area A_1 is Z_{31} , and the confidence of the camera C_4 for the sub-area A_1 is Z_{41} , where $Z_{11} > Z_{21} > Z_{31} > Z_{41}$. If there is no other detection box, the IoU between which and the detection box of the target object is greater than or equal to the predetermined threshold, in the video frames currently acquired by the camera C_1 , that is, in the video frames currently acquired by the camera C_1 , the IoU between the detection box of the target object and the detection boxes of the other target objects is less than the predetermined threshold, the current feature of the target object is extracted according to the video frames acquired by the camera C_1 . If there are other detection boxes, the IoU between which and the detection box of the target object is greater than or equal to the predetermined threshold, in the video frames currently acquired by the camera C_1 , it is determined whether there are other detection boxes, the IoU between which and the detection box of the target object is greater than or equal to the predetermined threshold, in the video frames currently acquired by the camera C_2 , that is, it is determined whether the IoU between the detection box of the target object and the detection boxes of other target objects is less than the predetermined threshold in the video frames currently acquired by the camera C_2 . If there is no other detection box, the IoU between which and the detection box of the target object is greater than or equal to the predetermined threshold, in the video frames currently acquired by the camera C_2 , that is, in the video frames currently acquired by the camera C_2 , the IoU between the detection box of the target object and each of the detection boxes of other target objects is less than the predetermined threshold, the current feature of the target object is extracted according to the video frames acquired by the camera C_2 , and so on. That is, according to the order of the confidence of the cameras for the sub-area, if the target object is blocked in the video frames currently acquired by the camera with the highest confidence, it is possible to switch to the camera with the second highest confidence, and so on, so that the camera for extracting the current feature of the target object is a camera with as high confidence as possible in the case of not being blocked,

thereby reducing the rate of missed detection and improving the accuracy of target matching.

[0097] In the example, for any camera in the multiple cameras, the IoU between the detection box of the target object and the detection boxes of other target objects in the video frames currently acquired by the camera may represent a proportion of the target object being blocked by other target objects. The larger the IoU between the detection box of the target object and the detection boxes of the other target objects in the video frames currently acquired by the camera is, the larger the proportion of the target object being blocked by the other target objects is. The smaller the IoU between the detection box of the target object and the detection boxes of the other target objects in the video frames currently acquired by the camera is, the smaller the proportion of the target object being blocked by the other target objects is. In the video frames currently acquired by the camera, if the proportion of the target object being blocked by other target objects is smaller, then the visual information of the target object possessed by the current feature of the target object extracted based on the video frames currently acquired by the camera is richer, for example, the visual information can have more information about feature points of the target object. In the example, a camera with the highest confidence for the sub-area where the target object is located is determined from the cameras, for which the video frames currently acquired meet the overlapping condition, in the multiple cameras as the camera for extracting the current feature of the target object, and the current feature of the target object is extracted according to the video frames acquired by the camera for extracting the current feature of the target object. In this way, the extracted current feature of the target object can have richer visual information. Therefore, by performing target matching based on the current feature of the target object extracted in the example, the accuracy of target matching can be further improved.

[0098] At S13, a reference feature matching the current feature of the target object is determined based on multiple reference features corresponding to the multiple target objects. The multiple reference features are features that are extracted from the multiple target objects before the multiple target objects conflict with each other.

[0099] FIG. 4 illustrates a schematic diagram of multiple target objects and multiple

reference features corresponding to the multiple target objects in the embodiments of the present disclosure. As shown in FIG. 4, the multiple target objects include target object O₁, target object O₂, target object O₃ and target object O₄. The reference feature of the target object O₁ is reference feature F₀₁, the reference feature of the target object O₂ is reference feature F₀₂, the reference feature of the target object O₃ is reference feature F₀₃, and the reference feature of the target object O₄ is reference feature F₀₄.

[00100] In a possible implementation mode, the reference feature of the target object may be extracted in the first frame and stored. For example, the reference features of all target objects may be extracted in the first frame and stored. In another implementation mode, if a part of the target objects are seriously blocked in the first frame, the reference features of these target objects may be extracted in the subsequent video frames, so as to obtain the reference features with higher quality and richer visual information of the target objects. After the reference feature of the target object is extracted, a corresponding relationship between the identification information of the target object and the reference feature of the target object may be stored. In some embodiments of the present disclosure, the manner of determining the camera for extracting the reference feature of the target object is similar to the above manner of determining the camera for extracting the current feature of the target object, and thus is not elaborated herein.

[00101] As an example of the implementation mode, after the reference feature of a target object is extracted, if the current feature of the target object is extracted in the subsequent target tracking process, the reference feature of the target object may be updated according to the extracted current feature of the target object, so as to improve the accuracy of target matching based on the stored reference feature. For example, a new reference feature may be obtained by weighting the stored reference feature of the target object with the current feature of the target object. Alternatively, the reference feature of the target object may also not be updated to reduce the computation amount.

[00102] In a possible implementation mode, when the target object is a person (for example, a pedestrian, a player on the field, etc.), features of the target object may be extracted through a person Re-IDentification (ReID) module. For example, the reference feature and/or the current feature of the target object may be extracted through the ReID

module. In an example, the ReID module may be realized by a convolutional neural network. Alternatively, the features of the target object may also be extracted by other feature extraction methods, as long as the extracted features of the target object can reflect the visual information of the target object.

[00103] In a possible implementation mode, the operation that the reference feature matching the current feature of the target object is determined based on multiple reference features corresponding to the multiple target objects may include that: the reference feature with the maximum similarity to the current feature of the target object is determined from the multiple reference features corresponding to the multiple target objects; and in response to the similarity between the reference feature with the maximum similarity and the current feature of the target object being greater than or equal to a similarity threshold, the reference feature with the maximum similarity is determined as the reference feature matching the current feature of the target object. The similarity may be a cosine similarity and so on.

[00104] For example, at the t -th moment, conflict area 1 includes target objects O_1 , O_4 , and O_3 , the reference feature corresponding to the conflict area 1 includes the reference feature F_{01} of the target object O_1 , the reference feature F_{04} of the target object O_4 , and the reference feature F_{03} of the target object O_3 . At the $(t+1)$ -th moment, when it is detected that the distance between one target object and the remaining target objects is greater than a distance threshold, it is indicated that the target object is out of the conflict area 1, that is, one target object in the conflict area 1 no longer conflicts with the other two target objects, then the current feature F_{1n} of the target object is extracted, and the similarities between the three reference features F_{01} , F_{04} and F_{03} corresponding to the conflict area 1 and the current feature F_{1n} of the target object are calculated respectively. If the reference feature that has the maximum similarity to the current feature F_{1n} of the target object is the reference feature F_{03} among the three reference features, and the similarity between the reference feature F_{03} and the current feature F_{1n} of the target object is greater than or equal to a similarity threshold, the reference feature F_{03} may be determined as the reference feature matching the current feature of the target object.

[00105] In the implementation mode, the reference feature with the maximum

similarity to the current feature of the target object is determined from the multiple reference features corresponding to the multiple target objects, and in response to the similarity between the reference feature with the maximum similarity and the current feature of the target object being greater than or equal to the similarity threshold, the reference feature with the maximum similarity is determined as the reference feature matching the current feature of the target object. In this way, when the target object no longer conflicts with other target objects in multiple target objects, the reference feature matching the current feature of the target object is first searched based on the multiple reference features corresponding to the multiple target objects, which helps to improve the speed and accuracy of searching the reference feature matching the current feature of the target object.

[00106] As an example of the implementation mode, after the reference feature with the maximum similarity to the current feature of the target object is determined, the method may also include that: in response to the similarity between the reference feature with the maximum similarity and the current feature of the target object being less than the similarity threshold, the reference feature matching the current feature of the target object is determined based on the reference feature corresponding to a target object that does not conflict with the target object and conflicts with any of the other target objects.

[00107] In the example, if the reference feature matching the current feature of the target object cannot be found among the multiple reference features corresponding to the multiple target objects, the reference feature matching the current feature of the target object may be searched in the reference features (i.e., reference features corresponding to a target object that does not conflict with the target object but conflicts with other target objects) corresponding to other conflict areas. For example, if the similarities between the multiple reference features corresponding to the multiple target objects and the current feature of the target object are less than the similarity threshold, it may be determined that the reference feature matching the current feature of the target object cannot be found among the multiple reference features corresponding to the multiple target objects. In the example, in response to the situation that the similarity between the reference feature with the maximum similarity and the current feature of the target object is less than the similarity threshold, the reference feature matching the current feature of

the target object is determined based on the reference feature corresponding to a target object that does not conflict with the target object and conflicts with any of the other target objects. In this way, the accuracy of the determined reference feature matching the current feature of the target object can be further improved on the premise of improving the speed of searching the reference feature matching the current feature of the target object.

[00108] In an example, if the reference feature matching the current feature of the target object cannot be found in all conflict areas, the reference feature matching the current feature of the target object may be found in a non-conflict area.

[00109] In another possible implementation mode, the operation that the reference feature matching the current feature of the target object is determined based on multiple reference features corresponding to the multiple target objects may include that: the reference feature with the maximum similarity to the current feature of the target object is determined from the multiple reference features corresponding to the multiple target objects; and the reference feature with the maximum similarity is determined as the reference feature matching the current feature of the target object.

[00110] At S14, identification information corresponding to the reference feature matching the current feature of the target object is determined as the identification information of the target object.

[00111] In some embodiments of the present disclosure, the identification information of the target object may be information that can be used for uniquely identifying the target object, such as ID, numbers, and name. For example, if the tracking area is a football field and the target objects include a player on the football field, then the identification information of the target object may be the team and uniform number of the player.

[00112] For example, if the reference feature matching the current feature F_{1n} of the target object is F_{03} , and the identification information corresponding to the reference feature F_{03} is O_3 , then O_3 may be determined as the identification information of the target object.

[00113] In a possible implementation mode, a target tracking result corresponding to any target object may also include location information of the target object. The behavior of the target object in the tracking area may be analyzed by using the target tracking result provided by the target tracking method provided in the embodiments of the present disclosure. For example, if the tracking area is a football field and the target objects include a player on the football field, then the behavior of the player, for example, whether the player is offside, may be analyzed according to the identification information of the player (for example, the team and uniform number of the player) and the location information of the player in at least one video frame.

[00114] In some embodiments of the present disclosure, multiple target objects conflicting with each other are determined according to the distances between the target objects in the tracking area; in response to the situation that any target object in the multiple target objects no longer conflicts with other target objects in the multiple target objects, the current feature of the target object is extracted; the reference feature matching the current feature of the target object is determined based on multiple reference features corresponding to the multiple target objects; and the identification information corresponding to the reference feature matching the current feature of the target object is determined as the identification information of the target object. In this way, the identification information of the target object can be kept consistent, that is, the identification information of the same target object can be kept consistent when the target object gets in and out of the conflict area, and thus the accuracy of multi-target tracking can be improved. The embodiments of the present disclosure may be applied to complex target tracking scenarios, for example, the sizes of the target objects in the tracking area are different, there are target objects that are small and difficult to be detected in the tracking area, the target objects in the tracking area are dense, the target objects have similar appearance, the movement complexity of the target object is high, and the target object is blocked seriously.

[00115] In a possible implementation mode, before the operation S11, the method may further include that: first coordinates of a marker are obtained; for any camera in the multiple cameras, a transformation matrix corresponding to the camera is determined according to second coordinates of the marker obtained by the camera and the first

coordinates of the marker.

[00116] In the implementation mode, when the multiple cameras are calibrated, markers may be placed at a certain density in the tracking area first. FIG. 5 illustrates a schematic diagram of placing markers in a tracking area in some embodiments of the present disclosure. In the example shown in FIG. 5, the tracking area is a football field, and the markers are white metal sheets. Alternatively, those skilled in the art may flexibly choose the markers according to requirements of an actual application scenario and/or personal preferences. Here, the markers placed at a certain density may be of different heights, sizes, colors, etc, which is not limited herein. By using the markers to calibrate, the accuracy of calibration in the tracking area without abundant corner points can be improved, thereby helping to improve the accuracy of target tracking.

[00117] As an example of the implementation mode, the first coordinates may be the coordinates in a world coordinate system, the second coordinates may be the coordinates in a pixel coordinate system, and the transformation matrix may be a homography matrix. Since the target object moves in the same plane, the transformation matrix $H_{3 \times 3}$ may be determined by formula (1):

$$[00118] \quad \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = H_{3 \times 3} \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \quad (1)$$

[00119] where (X_w, Y_w) represents the first coordinates of the marker, which may be, for example, the coordinates of the marker in the world coordinate system, (u, v) represents the second coordinates of the marker obtained by the camera, which may be, for example, the coordinates of the marker in the pixel coordinate system.

[00120] In the example, the degree of freedom of the homography matrix $H_{3 \times 3}$ is 8. Theoretically, the homography matrix can be solved by using the feature points corresponding to four markers. In order to improve the accuracy of internal and external parameters, more markers may be set as shown in FIG. 5, the least square method is adopted, isolated points are filtered, and a checkerboard calibration method is used to obtain the homography matrix. The homography matrix obtained in this way can remove the distortion in an image acquired by the camera and improve the accuracy of the

determined first coordinates.

[00121] Here, the external parameter of the camera may be obtained by means of Direct Linear Transformation (DLT) and Singular Value Decomposition (SVD), and the internal parameter of the camera may be obtained by means of checkerboard calibration, so as to remove the distortion in an image.

[00122] In the implementation mode, by determining the transformation matrix corresponding to each camera in the multiple cameras, the second coordinates of the target object obtained by the multiple cameras can be transformed to a unified coordinate system, for example, to the world coordinate system, and thus the first coordinates of the target object corresponding to each camera in the multiple cameras are obtained, thereby facilitating subsequent tracking of the target object.

[00123] In a possible implementation mode, the method may also include that: for any target object, the first coordinates of the target object corresponding to multiple cameras are determined; a sub-area where the target object is located in the tracking area is determined according to the first coordinates of the target object corresponding to the multiple cameras, the tracking area including multiple sub-areas; and according to the confidence of the multiple cameras for the sub-area where the target object is located, the first coordinates of the target object corresponding to the multiple cameras are fused to obtain fusion coordinates of the target object. Before the multiple target objects conflicting with each other are determined according to the distances between the target objects in the tracking area, the method may further include that: the distances between the target objects in the tracking area is determined according to fusion coordinates of the target objects in the tracking area.

[00124] In the implementation mode, the first coordinates of any target object corresponding to any camera may represent the first coordinates of the target object obtained according to the image acquired by the camera. The first coordinates may be the coordinates in a first coordinate system. For example, the first coordinate system may be the world coordinate system or other virtual coordinate systems. The fusion coordinates and the first coordinates may be coordinates in the same coordinate system. For example, the fusion coordinates and the first coordinates may both be the coordinates in the world

coordinate system.

[00125] As an example of the implementation mode, a coordinate range of each sub-area in the tracking area in the first coordinate system is predetermined. For example, a sub-area is rectangular, and the coordinate range of any sub-area in the first coordinate system may be represented by the coordinates of four vertexes of the sub-area in the first coordinate system, or the coordinate range of the sub-area in the first coordinate system may be represented by the coordinates of the upper left vertex of the sub-area in the first coordinate system and the width and height of the sub-area. Alternatively, the coordinate range of any sub-area in the first coordinate system may also be represented in other manners, which is not limited herein. In addition, the shape of any sub-area may not be rectangular, for example, it may be triangular, etc. Different sub-areas in the tracking area may have the same or different sizes.

[00126] In the implementation mode, the sub-area where the target object is located in the tracking area may be determined according to the first coordinates of the target object corresponding to the multiple cameras and the coordinate range of the sub-area in the first coordinate system. In the implementation mode, the sub-area where the target object is located in the tracking area may be determined according to the first coordinates of the target object corresponding to all or part of the multiple cameras and the coordinate range of the sub-area in the first coordinate system.

[00127] As an example of the implementation mode, the operation that the sub-area where the target object is located in the tracking area is determined according to the first coordinates of the target object corresponding to the multiple cameras may include that: a candidate sub-area where the target object is located in the tracking area is determined according to the first coordinates of the target object corresponding to any camera in the multiple cameras and the coordinate range of the sub-area in the first coordinate system; and the candidate sub-area with the highest number of votes is determined as the sub-area where the target object is located. For example, the multiple cameras include camera C_1 , camera C_2 , camera C_3 and camera C_4 . according to the first coordinates of the target object corresponding to the camera C_1 , the candidate sub-area where the target object is located in the tracking area is determined to be sub-area A_1 ; according to the first

coordinates of the target object corresponding to the camera C_2 , the candidate sub-area where the target object is located in the tracking area is determined to be sub-area A_1 ; according to the first coordinates of the target object corresponding to the camera C_3 , the candidate sub-area where the target object is located in the tracking area is determined to be sub-area A_2 ; according to the first coordinates of the target object corresponding to the camera C_4 , the candidate sub-area where the target object is located in the tracking area is determined to be sub-area A_1 . Then, the number of votes of the candidate sub-area A_1 is 3, and the number of votes of the candidate sub-area A_2 is 1. Therefore, the candidate sub-area A_1 can be determined as the sub-area where the target object is located.

[00128] As another example of the implementation mode, the operation that the sub-area where the target object is located in the tracking area is determined according to the first coordinates of the target object corresponding to the multiple cameras may include that: an average value of the first coordinates of the target object corresponding to the multiple cameras is determined according to the first coordinates of the target object corresponding to the multiple cameras; the sub-area where the average value of the first coordinates is located is determined according to the coordinate range of the sub-area in the first coordinate system, and the sub-area where the average value of the first coordinates is located is determined as the sub-area where the target object is located in the tracking area.

[00129] In the implementation mode, according to the confidence of the multiple cameras for the sub-area where the target object is located, the first coordinates of the target object corresponding to all or part of the multiple cameras may be fused to obtain the fusion coordinates of the target object. For example, if the multiple cameras include camera C_1 , camera C_2 , camera C_3 and camera C_4 , the target object is located in sub-area A_1 , the confidence of the camera C_1 for the sub-area A_1 is Z_{11} , the confidence of the camera C_2 for the sub-area A_1 is Z_{21} , the confidence of the camera C_3 for the sub-area A_1 is Z_{31} , the confidence of the camera C_4 for the sub-area A_1 is Z_{41} , the first coordinates of the target object corresponding to the camera C_1 are P_{w1} , the first coordinates of the target object corresponding to the camera C_2 are P_{w2} , the first coordinates of the target object corresponding to the camera C_3 are P_{w3} , and the first coordinates of the target object corresponding to the camera C_4 are P_{w4} , then the fusion coordinates of the target object

may

be

$$P = \frac{Z_{11}}{Z_{11}+Z_{21}+Z_{31}+Z_{41}} P_{w1} + \frac{Z_{21}}{Z_{11}+Z_{21}+Z_{31}+Z_{41}} P_{w2} + \frac{Z_{31}}{Z_{11}+Z_{21}+Z_{31}+Z_{41}} P_{w3} + \frac{Z_{41}}{Z_{11}+Z_{21}+Z_{31}+Z_{41}} P_{w4} .$$

For another example, in the multiple cameras, part of the cameras with high confidence for the sub-area where the target object is located may be selected to determine the fusion coordinates of the target object. For example, $Z_{11} > Z_{21} > Z_{31} > Z_{41}$,

$$P = \frac{Z_{11}}{Z_{11}+Z_{21}} P_{w1} + \frac{Z_{21}}{Z_{11}+Z_{21}} P_{w2} .$$

[00130] In the implementation mode, the fusion coordinates of the target object are obtained in combination with the first coordinates of the target object corresponding to the multiple cameras, and target tracking is performed based on the fusion coordinates of the target object. In this way, target tracking can be performed based on more accurate coordinates, which helps to improve the accuracy of target tracking.

[00131] In a possible implementation mode, for any target object and any camera in the multiple cameras, the first coordinates of the target object corresponding to the camera are determined according to the second coordinates of the target object obtained by the camera and the transformation matrix corresponding to the camera.

[00132] As an example of the implementation mode, the second coordinates of the target object obtained by the camera may be the coordinates of the target object in the pixel coordinate system corresponding to the camera. That is, the second coordinates may represent the coordinates in the pixel coordinate system. In the example, the first coordinates of the target object corresponding to the camera represent the coordinates of the target object in the world coordinate system, which are determined according to the second coordinates of the target object obtained by the camera. In the example, the transformation matrix corresponding to the camera may be the homography matrix.

[00133] In other examples, the second coordinates of the target object obtained by the camera may also be the coordinates of the target object in an image coordinate system and other coordinate systems corresponding to the camera. That is, the second coordinates may also be the coordinates in the image coordinate system and other

coordinate systems.

[00134] As an example of the implementation mode, target detection may be performed on an image acquired by the camera to determine a detection box of the target object in the image, and the second coordinates of the target object obtained by the camera may be determined according to the location of the detection box. In the example, any point on the detection box or inside the detection box may be taken as the second coordinates of the target object. For example, the midpoint of the bottom edge of the detection box may be taken as the second coordinates of the target object.

[00135] For example, if the transformation matrix corresponding to any camera in the multiple cameras is H , and the second coordinates of any target object obtained by the camera are P_u , then the first coordinates of the target object corresponding to the camera may be determined by formula (2):

[00136]
$$P_w = (H^T H)^{-1} H^T P_u \quad (2)$$

[00137] In a possible implementation mode, the method may further include that: for any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas is determined according to the distances between the camera and the multiple sub-areas of the tracking area.

[00138] In the implementation mode, the confidence of the multiple cameras for the multiple sub-areas may be determined according to the distances between the multiple cameras and the multiple sub-areas of the tracking area. In the implementation mode, the confidence of any camera for any sub-area is negatively correlated with the distance between the camera and the sub-area. That is, the larger the distance between the camera and the sub-area is, the lower the confidence of the camera for the sub-area is; and the smaller the distance between the camera and the sub-area is, the higher the confidence of the camera for the sub-area is. For example, if the distance between camera C_1 and sub-area A_1 is less than the distance between camera C_2 and sub-area A_1 , the confidence Z_{11} of the camera C_1 for the sub-area A_1 may be greater than the confidence Z_{21} of the camera C_2 for the sub-area A_1 . For another example, if the distance between camera C_1 and sub-area A_1 is less than the distance between camera C_1 and sub-area A_2 , the

confidence Z_{11} of the camera C_1 for the sub-area A_1 may be greater than the confidence Z_{12} of the camera C_1 for the sub-area A_2 .

[00139] In the implementation mode, for any camera in the multiple cameras, if the distance between the camera and any sub-area is smaller, the video frame of the sub-area acquired by the camera is clearer, which is advantageous for accurately identifying the information in the video frame. On the contrary, if the distance between the camera and any sub-area is larger, the video frame of the sub-area acquired by the camera is more blurred, which is not conducive to accurately identifying the information in the video frame. Therefore, in the implementation mode, for any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas are determined according to the distances between the camera and the multiple sub-areas of the tracking area, and thus the distance between a camera and a sub-area is taken as a factor to consider the confidence of the camera for the sub-area. Coordination fusion is performed and/or the camera for exacting the current feature of the target object is determined based on the confidence determined, which helps to improve the accuracy of target tracking.

[00140] In a possible implementation mode, after the confidence of the multiple cameras for the multiple sub-areas are determined, the method may further include that: for any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas are adjusted according to the average distance between the target objects in the video frames acquired by the camera.

[00141] As an example of the implementation mode, the confidence of the camera for the multiple sub-areas may be continuously adjusted in the target tracking process according to the average distance between the target objects in the video frames acquired by the camera. For example, the confidence of the camera for the multiple sub-areas may be adjusted at a preset frequency. For example, the preset frequency may be 1 second.

[00142] In some embodiments of the present disclosure, the confidence of multiple cameras for the sub-area where the target object is located may be weighted, and then the first coordinates of the target object corresponding to the multiple cameras are fused according to the weighted confidence of the multiple cameras for the sub-area where the target object is located.

[00143] For any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas are adjusted according to the average distance between each two of all target objects in the video frames acquired by the camera. For example, if the multiple cameras include camera C_1 , camera C_2 , camera C_3 and camera C_4 , the average distance between the target objects in the video frames acquired by the camera C_1 is D_1 , the average distance between the target objects in the video frames acquired by the camera C_2 is D_2 , the average distance between the target objects in the video frames acquired by the camera C_3 is D_3 , and the average distance between the target objects in the video frames acquired by the camera C_4 is D_4 , where $D_1 > D_2 > D_3 > D_4$, then the confidence of the camera C_1 for each of the multiple sub-areas may be increased by B_1 , the confidence of the camera C_2 for each of the multiple sub-areas may be increased by B_2 , the confidence of the camera C_3 for each of the multiple sub-areas may be increased by B_3 , and the confidence of the camera C_4 for each of the multiple sub-areas is kept constant, where $B_1 > B_2 > B_3 > 0$, for example, $B_1=1.5$, $B_2=1$, $B_3=0.5$.

[00144] In the implementation mode, for any camera in the multiple cameras, if the average distance between the target objects in the video frames acquired by the camera is smaller, the target objects in the video frames acquired by the camera are more likely blocked by each other, which is not conducive to performing feature extraction on the target object and thus not conducive to performing target tracking. On the contrary, if the average distance between the target objects in the video frames acquired by the camera is larger, the target objects in the video frames acquired by the camera are less likely blocked by each other, which is advantageous for performing feature extraction on the target object and thus is advantageous for performing target tracking. In the implementation mode, for any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas is dynamically adjusted according to the average distance between the target objects in the video frames acquired by the camera, and coordination fusion is performed and/or the camera for exacting the current feature of the target object is determined based on the confidence determined, which helps to improve the accuracy of target tracking.

[00145] In a possible implementation mode, the method may further include that: for

any target object, in response to a situation that the target object does not conflict with other target objects at either the current moment or the previous moment of the current moment, target objects that do not conflict with other target objects at the previous moment and do not match other target objects at the current moment are determined as the remaining target objects at the previous moment; or, in response to a situation that the target object conflicts with the same other target object at the current moment and the previous moment, target objects that do not match other target objects of the current moment in a conflict area to which the target object belongs at the previous moment is determined as the remaining target objects at the previous moment; and a target object closest to the target object at the current moment is determined from the remaining target objects at the previous moment, as the target object matching the target object at the current moment, and the identification information of the matched target object is taken as the identification information of the target object.

[00146] For example, a certain target object does not conflict with other target objects at both the current moment and the previous moment of the current moment, and there are 15 target objects that do not conflict with other target objects at the previous moment, which are respectively target objects O_1 to O_{15} . If target objects O_1 and O_2 at the previous moment have matched other target objects at the current moment, that is, identification information O_1 and O_2 has been taken as the identification information of the target objects at the current moment, then target objects O_3 to O_{15} may be determined as the remaining target objects at the previous moment. After the remaining target objects O_3 to O_{15} at the previous moment are determined, a target object closest to the target object at the current moment is determined from the remaining target objects O_3 to O_{15} at the previous moment, as the target object matching the target object at the current moment. For example, if among the remaining target objects O_3 to O_{15} at the previous moment, O_6 is closest to the target object at the current moment, then the target object O_6 at the previous moment is determined as the target object matching the target object at the current moment, and the identification information of the target object at the current moment is determined as O_6 .

[00147] For another example, a certain target object conflicts with the same other target objects at the current moment and at the previous moment, for example, the conflict area

to which the target object belongs at the previous moment includes target object O_1 , target object O_2 , target object O_3 , and target object O_4 . If in the conflict area to which the target object belongs at the previous moment, the target objects O_1 and O_2 have matched other target objects at the current moment, that is, identification information O_1 and O_2 has been taken as the identification information of the target objects at the current moment, then target objects O_3 to O_4 may be determined as the remaining target objects at the previous moment. After the remaining target objects O_3 and O_4 at the previous moment are determined, a target object closest to the target object at the current moment is determined from the remaining target objects O_3 and O_4 at the previous moment, as the target object matching the target object at the current moment. For example, if among the remaining target objects O_3 and O_4 at the previous moment, O_4 is closest to the target object at the current moment, then the target object O_4 at the previous moment is determined as the target object matching the target object at the current moment, and the identification information of the target object at the current moment is determined as O_4 .

[00148] In the implementation mode, for any target object, when the target object does not conflict with other target objects at either the current moment or the previous moment of the current moment or the target object conflicts with the same other target objects at the current moment and the previous moment, a target object closest to the target object at the current moment is determined from the remaining target objects of the previous moment, as the target object matching the target object at the current moment, and the identification information of the matched target object is taken as the identification information of the target object. In this way, a distance-based greedy algorithm is used to determine the identification information of the target object without considering the visual feature of the target object, and thus the computation amount can be reduced, time cost can be reduced to meet a real-time requirement of target tracking.

[00149] In a possible implementation mode, the target tracking method provided by the embodiments of the present disclosure may be implemented by using a target tracking model. For example, the target tracking model may adopt a Faster Recurrent Convolutional Neural Network (Faster-RCNN) or a Fast Recurrent Convolutional Neural Network (Fast-RCNN). A backbone network of the target tracking model may adopt ResNet-18 and other structures. In order to improve the processing speed of the target

tracking model, a channel pruning method may be used to compress the target tracking model, and at the same time, a multi-layer feature detection method may be used to improve the accuracy of the model, so as to improve the speed of model without reducing the accuracy of model.

[00150] In a possible implementation mode, the method may further include that: a corresponding relationship between the identification information and the coordinates of the target objects is output. In an example, the corresponding relationship between the identification information and the coordinates may be output according to the order of the identification information. For example, if the identification information includes uniform numbers 1 to 11, the coordinates of the corresponding target objects may be output in the ascending order of the uniform numbers.

[00151] In a possible implementation mode, a motion trajectory of a target object may be obtained according to the coordinates of the target object at multiple moments (for example, multiple continuous moments). In an example, Kalman filtering algorithm may also be used to process the motion trajectory of the target object to make the motion trajectory of the target object smoother.

[00152] It can be understood that each method embodiment mentioned in the present disclosure may be combined to form combined embodiments without departing from principles and logics. For saving the space, elaborations are omitted in the present disclosure.

[00153] It can be understood by those skilled in the art that, in the method of the specific implementation modes, the writing sequence of each step does not mean a strict execution sequence and is not intended to form any limit to the implementation process and a specific execution sequence of each step should be determined by functions and probable internal logic thereof.

[00154] In addition, the present disclosure also provides a target tracking apparatus, an electronic device, a computer-readable storage medium, and a program, which may all be used to implement any target tracking method provided in the present disclosure. For the corresponding technical solution and description, reference may be made to the

corresponding record in the part of the method and will not be repeated.

[00155] FIG. 6 illustrates a block diagram of a target tracking apparatus provided by some embodiments of the present disclosure. As shown in FIG. 6, the target tracking apparatus includes: a first determining portion 61, an extracting portion 62, a second determining portion 63 and a third determining portion 64.

[00156] The first determining portion 61 is configured to determine multiple target objects conflicting with each other according to distances between target objects in the tracking area.

[00157] The extracting portion 62 is configured to extract, in response to the situation that any target object in the multiple target objects does not conflict with other target objects in the multiple target objects, a current feature of the target object.

[00158] The second determining portion 63 is configured to determine a reference feature matching the current feature of the target object based on multiple reference features corresponding to the multiple target objects. The multiple reference features are features extracted from the multiple target objects before the multiple target objects conflict with each other.

[00159] The third determining portion 64 is configured to determine identification information corresponding to the reference feature matching the current feature of the target object as identification information of the target object.

[00160] In a possible implementation mode, the distance between any target object in the multiple target objects conflicting with each other and at least one of the other target objects in the multiple target objects is less than or equal to a distance threshold.

[00161] The extracting portion 62 is configured to:

[00162] for any target object in the multiple target objects, in response to the distances between the target object and other target objects in the multiple target objects being greater than the distance threshold, extract the current feature of the target object.

[00163] In a possible implementation mode, the apparatus may further include: a third

determining portion 64, a fourth determining portion, a fusing portion, and a fifth determining portion.

[00164] The third determining portion 64 is configured to determine, for any target object, first coordinates of the target object corresponding to multiple cameras. The first coordinates of the target object corresponding to any camera are used for representing first coordinates of the target object obtained according to an image acquired by the camera.

[00165] The fourth determining portion is configured to determine a sub-area where the target object is located in the tracking area according to the first coordinates of the target object corresponding to the multiple cameras. The tracking area includes multiple sub-areas.

[00166] The fusing portion is configured to fuse, according to the confidence of the multiple cameras for the sub-area where the target object is located, the first coordinates of the target object corresponding to the multiple cameras to obtain fusion coordinates of the target object.

[00167] The fifth determining portion is configured to determine the distances between the target objects in the tracking area according to fusion coordinates of the target objects in the tracking area.

[00168] In a possible implementation mode, the extracting portion 62 is configured to:

[00169] determine the sub-area where the target object is located in the tracking area;

[00170] determine a camera extracting the current feature of the target object according to the confidence of multiple cameras for the sub-area where the target object is located;

[00171] extract the current feature of the target object according to video frames acquired by the camera extracting the current feature of the target object.

[00172] In a possible implementation mode, the extracting portion 62 is configured to:

[00173] determine the camera extracting the current feature of the target object

according to the confidence of multiple cameras for the sub-area where the target object is located and overlapping information between a detection box of the target object and detection boxes of other target objects in the video frames acquired by the multiple cameras.

[00174] In a possible implementation mode, the overlapping information includes the IoU between the detection box of the target object and the detection boxes of the other target objects.

[00175] The extracting portion 62 is configured to:

[00176] determine, from the cameras that meet an overlapping condition, a camera with the highest confidence for the sub-area where the target object is located as the camera extracting the current feature of the target object. The overlapping condition indicates that in the video frames acquired currently, the IoU between the detection box of the target object and the detection boxes of other target objects is less than the predetermined threshold.

[00177] In a possible implementation mode, the apparatus may further include: a sixth determining portion.

[00178] The sixth determining portion is configured to determine, for any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas according to the distances between the camera and the multiple sub-areas of the tracking area.

[00179] In a possible implementation mode, the apparatus may further include: an adjusting portion.

[00180] The adjusting portion is configured to adjust, for any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas according to the average distance between the target objects in the video frames acquired by the camera.

[00181] In a possible implementation mode, the second determining portion 63 is configured to:

[00182] determine the reference feature with the maximum similarity to the current

feature of the target object from multiple reference features corresponding to the multiple target objects; and

[00183] in response to the situation that the similarity between the reference feature with the maximum similarity and the current feature of the target object is greater than or equal to the similarity threshold, determine the reference feature with the maximum similarity as the reference feature matching the current feature of the target object.

[00184] In a possible implementation mode, the apparatus may further include: a seventh determining portion.

[00185] The seventh determining portion is configured to determine, in response to the similarity between the reference feature with the maximum similarity and the current feature of the target object being less than the similarity threshold, the reference feature matching the current feature of the target object based on the reference feature corresponding to a target object that does not conflict with the target object and conflicts with any of the other target objects.

[00186] In a possible implementation mode, the apparatus may further include: an eighth determining portion and a ninth determining portion.

[00187] The eighth determining portion is configured to: for any target object, in response to the situation that the target object does not conflict with other target objects at either the current moment or the previous moment of the current moment, determine target objects that do not conflict with other target objects at the previous moment and do not match other target objects at the current moment as the remaining target object at the previous moment; or, in response to the situation that the target object conflicts with the same other target objects at the current moment and the previous moment, determine target objects that do not match other target objects at the current moment in a conflict area to which the target object belongs at the previous moment as the remaining target object of the previous moment.

[00188] The ninth determining portion is configured to determine, from the remaining target objects at the previous moment, a target object closest to the target object at the current moment as the target object matching the target object at the current moment, and

take the identification information of the matched target object as the identification information of the target object.

[00189] In the embodiments of the present disclosure and other embodiments, the "portion" may be a part of circuit, a part of processor, a part of program or software, etc., and alternatively, it may be a unit, or it may also be a module or non-modular.

[00190] In the embodiments of the present disclosure, multiple target objects conflicting with each other are determined according to the distances between target objects in the tracking area; in response to any target object in the multiple target objects no longer conflicts with other target objects in the multiple target objects, the current feature of the target object is extracted; the reference feature matching the current feature of the target object is determined based on multiple reference features corresponding to the multiple target objects; and the identification information corresponding to the reference feature matching the current feature of the target object is determined as the identification information of the target object. In this way, the identification information of the target object can be kept consistent, that is, the identification information of the same target object can be kept consistent when the target object gets in and out of the conflict area, and thus the accuracy of multi-target tracking can be improved.

[00191] In some embodiments, functions of or modules contained in the apparatus provided in the embodiments of the present disclosure may be used to perform the method described in the above method embodiments, and for implementations thereof, reference may be made to the description of the above method embodiments, and the details are not repeated herein for simplicity.

[00192] The embodiments of the present disclosure also provide a computer-readable storage medium, having stored thereon a computer program instruction. The computer program instruction, when executed by the processor, implements the above method. The computer-readable storage medium may be a nonvolatile computer-readable storage medium or a volatile computer-readable storage medium.

[00193] The embodiments of the present disclosure also provide a computer program product, which includes a computer readable code. When the computer readable code

runs in a device, the processor in the device executes instructions for implementing the target tracking method provided in any of the above embodiments.

[00194] The embodiments of the present disclosure further provide another computer program product for storing a computer readable instruction that, when executed, enables a computer to perform the operations of the target tracking method provided in any of the above embodiments.

[00195] The embodiments of the present disclosure further provide an electronic device, which includes: one or more processors, and a memory configured to store an executable instruction. The one or more processors are configured to call the executable instruction stored in the memory to execute the above method.

[00196] The electronic device may be provided as a terminal, a server or other forms of devices.

[00197] FIG. 7 shows a block diagram of an electronic device 800 provided by an embodiment of the present disclosure. For example, the electronic device 800 may be a mobile phone, a computer, a digital broadcast terminal, a messaging device, a gaming console, a tablet, a medical device, exercise equipment, a personal digital assistant and other terminals.

[00198] Referring to FIG. 7, the electronic device 800 may include one or more of the following components: a processing component 802, a memory 804, a power component 806, a multimedia component 808, an audio component 810, an Input/Output (I/O) interface 812, a sensor component 814, and a communication component 816.

[00199] The processing component 802 typically controls overall operations of the electronic device 800, such as the operations associated with display, telephone calls, data communications, camera operations, and recording operations. The processing component 802 may include one or more processors 820 to execute instructions to perform all or part of the steps in the above method. Moreover, the processing component 802 may include one or more modules which facilitate interactions between the processing component 802 and the other components. For instance, the processing component 802 may include a multimedia module to facilitate interactions between the

multimedia component 808 and the processing component 802.

[00200] The memory 804 is configured to store various types of data to support the operation of the electronic device 800. Examples of such data include instructions for any application programs or methods operated on the electronic device 800, contact data, phonebook data, messages, pictures, video, etc. The memory 804 may be implemented by any type of volatile or non-volatile memory devices, or a combination thereof, such as a Static Random Access Memory (SRAM), an Electrically Erasable Programmable Read-Only Memory (EEPROM), an Erasable Programmable Read-Only Memory (EPROM), a Programmable Read-Only Memory (PROM), a Read-Only Memory (ROM), a magnetic memory, a flash memory, and a magnetic or optical disk.

[00201] The power supply component 806 provides power for various components of the electronic device 800. The power component 806 may include a power management system, one or more power supplies, and other components associated with generation, management and distribution of power for the electronic device 800.

[00202] The multimedia component 808 includes a screen providing an output interface between the electronic device 800 and a user. In some embodiments, the screen may include a Liquid Crystal Display (LCD) and a Touch Panel (TP). If the screen includes the TP, the screen may be implemented as a touch screen to receive an input signal from the user. The TP includes one or more touch sensors to sense touches, swipes and gestures on the TP. The touch sensors may not only sense a boundary of a touch or swipe action but also detect a duration and pressure associated with the touch or swipe action. In some embodiments, the multimedia component 808 includes a front camera and/or a rear camera. The front camera and/or the rear camera may receive external multimedia data when the electronic device 800 is in an operation mode, such as a photographing mode or a video mode. Each of the front camera and the rear camera may be a fixed optical lens system or have focusing and optical zooming capabilities.

[00203] The audio component 810 is configured to output and/or input an audio signal. For example, the audio component 810 includes a Microphone (MIC), and the MIC is configured to receive an external audio signal when the electronic device 800 is in the operation mode, such as a call mode, a recording mode and a voice recognition mode.

The received audio signal may further be stored in the memory 804 or sent through the communication component 816. In some embodiments, the audio component 810 further includes a speaker configured to output the audio signal.

[00204] The I/O interface 812 provides an interface between the processing component 802 and a peripheral interface module, and the peripheral interface module may be a keyboard, a click wheel, a button and the like. The button may include, but not limited to: a home button, a volume button, a starting button and a locking button.

[00205] The sensor component 814 includes one or more sensors configured to provide status assessment in various aspects for the electronic device 800. For example, the sensor component 814 may detect an on/off status of the electronic device 800 and relative positioning of components, such as a display and small keyboard of the electronic device 800, and the sensor component 814 may further detect a change in a position of the electronic device 800 or a component of the electronic device 800, presence or absence of contact between the user and the electronic device 800, orientation or acceleration/deceleration of the electronic device 800 and a change in temperature of the electronic device 800. The sensor component 814 may include a proximity sensor configured to detect presence of an object nearby without any physical contact. The sensor component 814 may also include a light sensor, such as a Complementary Metal Oxide Semiconductor (CMOS) or Charge Coupled Device (CCD) image sensor, configured for use in an imaging application. In some embodiments, the sensor component 814 may also include an acceleration sensor, a gyroscope sensor, a magnetic sensor, a pressure sensor or a temperature sensor.

[00206] The communication component 816 is configured to facilitate wired or wireless communication between the electronic device 800 and another device. The electronic device 800 may access a communication-standard-based wireless network, such as a Wireless Fidelity (WiFi) network, a 2nd-Generation (2G) network, a 3rd-Generation (3G) network, a 4th-Generation (4G)/Long Term Evolution (LTE) network, a 5th-Generation (5G) network or a combination thereof. In an exemplary embodiment, the communication component 816 receives a broadcast signal or broadcast associated information from an external broadcast management system through a broadcast channel.

In an exemplary embodiment, the communication component 816 further includes an NFC module to facilitate short-range communication. For example, the NFC module may be implemented based on a Radio Frequency Identification (RFID) technology, an Infrared Data Association (IrDA) technology, an Ultra-WideBand (UWB) technology, a Bluetooth (BT) technology and another technology.

[00207] In an exemplary embodiment, the electronic device 800 may be implemented by one or more Application Specific Integrated Circuits (ASICs), Digital Signal Processors (DSPs), Digital Signal Processing Devices (DSPDs), Programmable Logic Devices (PLDs), Field Programmable Gate Arrays (FPGAs), controllers, micro-controllers, microprocessors or other electronic components, and is configured to execute the above method.

[00208] In an exemplary embodiment, there is further provided a nonvolatile computer-readable storage medium, for example, a memory 804 including a computer program instruction. The computer program instruction may be executed by a processor 820 of the electronic device 800 to complete the above method.

[00209] FIG. 8 shows a block diagram of an electronic device 1900 provided by an embodiment of the present disclosure. For example, the electronic device 1900 may be provided as a server. Referring to FIG. 8, the electronic device 1900 includes a processing component 1922, further including one or more processors, and a memory resource represented by a memory 1932, configured to store an instruction executable by the processing component 1922, for example, an application. The application stored in the memory 1932 may include one or more than one module of which each corresponds to a set of instructions. In addition, the processing component 1922 is configured to execute the instruction to perform the above method.

[00210] The electronic device 1900 may further include a power component 1926 configured to execute power management of the electronic device 1900, a wired or wireless network interface 1950 configured to connect the electronic device 1900 to a network and an I/O interface 1958. The electronic device 1900 may be operated based on an operating system stored in the memory 1932, for example, Windows Server®, Max OS X®, Unix®, Linux®, FreeBSD® or the like.

[00211] In an exemplary embodiment, a nonvolatile computer-readable storage medium is also provided, for example, a memory 1932 including a computer program instruction. The computer program instruction may be executed by a processing component 1922 of the electronic device 1900 to complete the above method.

[00212] The present disclosure may be a system, a method and/or a computer program product. The computer program product may include a computer-readable storage medium, having stored thereon a computer-readable program instruction configured to enable a processor to implement each aspect of the present disclosure.

[00213] The computer-readable storage medium may be a tangible device capable of retaining and storing an instruction used by an instruction execution device. For example, the computer-readable storage medium may be, but not limited to, an electric storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device or any appropriate combination thereof. More specific examples (non-exhaustive list) of the computer-readable storage medium include a portable computer disk, a hard disk, a RAM, a ROM, an EPROM (or a flash memory), an SRAM, a Compact Disc Read-Only Memory (CD-ROM), a Digital Video Disk (DVD), a memory stick, a floppy disk, a mechanical coding device, a punched card or in-slot raised structure with an instruction stored therein, and any appropriate combination thereof. Herein, the computer-readable storage medium is not explained as a transient signal, for example, a radio wave or another freely propagated electromagnetic wave, an electromagnetic wave propagated through a wave guide or another transmission medium (for example, a light pulse propagated through an optical fiber cable) or an electric signal transmitted through an electric wire.

[00214] The computer-readable program instruction described here may be downloaded from the computer-readable storage medium to each computing/processing device or downloaded to an external computer or an external storage device through a network such as the Internet, a Local Area Network (LAN), a Wide Area Network (WAN) and/or a wireless network. The network may include a copper transmission cable, optical fiber transmission, wireless transmission, a router, a firewall, a switch, a gateway computer and/or an edge server. A network adapter card or network interface in each

computing/processing device receives the computer-readable program instruction from the network and forwards the computer-readable program instruction for storage in the computer-readable storage medium in each computing/processing device.

[00215] The computer program instruction configured to execute the operations of the present disclosure may be an assembly instruction, an Instruction Set Architecture (ISA) instruction, a machine instruction, a machine related instruction, a microcode, a firmware instruction, state setting data or a source code or target code edited by one or any combination of more programming languages, the programming language including an object-oriented programming language such as Smalltalk and C++ and a conventional procedural programming language such as "C" language or a similar programming language. The computer-readable program instruction may be completely executed in a computer of a user or partially executed in the computer of the user, executed as an independent software package, executed partially in the computer of the user and partially in a remote computer, or executed completely in the remote server or a server. Under the condition that the remote computer is involved, the remote computer may be connected to the computer of the user through any type of network including a Local Area Network (LAN) or a Wide Area Network (WAN), or, may be connected to an external computer (for example, connected by an Internet service provider through the Internet). In some embodiments, an electronic circuit such as a programmable logic circuit, a Field Programmable Gate Array (FPGA) or a Programmable Logic Array (PLA) may be customized by use of state information of a computer-readable program instruction, and the electronic circuit may execute the computer-readable program instruction, thereby implementing each aspect of the present disclosure.

[00216] Herein, each aspect of the present disclosure is described with reference to flowcharts and/or block diagrams of the method, device (system) and computer program product according to the embodiments of the present disclosure. It is to be understood that each block in the flowcharts and/or the block diagrams and a combination of each block in the flowcharts and/or the block diagrams may be implemented by computer-readable program instructions.

[00217] These computer-readable program instructions may be provided for a universal

computer, a dedicated computer or a processor of another programmable data processing device, thereby generating a machine to further generate a device that realizes a function/action specified in one or more blocks in the flowcharts and/or the block diagrams when the instructions are executed through the computer or the processor of the other programmable data processing device. These computer-readable program instructions may also be stored in a computer-readable storage medium, and through these instructions, the computer, the programmable data processing device and/or another device may work in a specific manner, so that the computer-readable medium including the instructions includes a product including instructions for implementing each aspect of the function/action specified in one or more blocks in the flowcharts and/or the block diagrams.

[00218] These computer-readable program instructions may further be loaded to the computer, the other programmable data processing device or the other device, so that a series of operating steps are executed in the computer, the other programmable data processing device or the other device to generate a process implemented by the computer to further realize the function/action specified in one or more blocks in the flowcharts and/or the block diagrams by the instructions executed in the computer, the other programmable data processing device or the other device.

[00219] The flowcharts and block diagrams in the drawings illustrate probably implemented system architectures, functions and operations of the system, method and computer program product according to multiple embodiments of the present disclosure. On this aspect, each block in the flowcharts or the block diagrams may represent part of a module, a program segment or an instruction, and part of the module, the program segment or the instruction includes one or more executable instructions configured to realize a specified logical function. In some alternative implementations, the functions marked in the blocks may also be realized in a sequence different from those marked in the drawings. For example, two continuous blocks may actually be executed substantially concurrently, or may also be executed in a reverse sequence sometimes, which is determined by the involved functions. It is further to be noted that each block in the block diagrams and/or the flowcharts and a combination of the blocks in the block diagrams and/or the flowcharts may be implemented by a dedicated hardware-based system

configured to execute a specified function or operation or may be implemented by a combination of a special hardware and a computer instruction.

[00220] The computer program product may be specifically realized by means of hardware, software or a combination thereof. In an optional embodiment, the computer program product is specifically embodied as a computer storage medium, and in another optional embodiment, the computer program product is specifically embodied as software products, such as a Software Development Kit (SDK).

[00221] Various embodiments of the present disclosure have been described above. The above descriptions are exemplary, non-exhaustive and also not limited to each disclosed embodiment. Many modifications and variations are apparent to those of ordinary skill in the art without departing from the scope and spirit of each described embodiment of the present disclosure. The terms used herein are selected to explain the principle and practical application of each embodiment or improvements in the technologies in the market best or enable others of ordinary skill in the art to understand each embodiment disclosed herein.

INDUSTRIAL APPLICABILITY

[00222] In the embodiments of the present disclosure, multiple target objects conflicting with each other are determined according to distances between target objects in the tracking area; in response to the situation that any target object in the multiple target objects no longer conflicts with other target objects in the multiple target objects, a current feature of the target object is extracted; a reference feature matching the current feature of the target object is determined based on multiple reference features corresponding to the multiple target objects; and identification information corresponding to the reference feature matching the current feature of the target object is determined as identification information of the target object. In this way, the identification information of the target object can be kept consistent, that is, the identification information of the same target object can be kept consistent when the target object gets in and out of the conflict area, and thus the accuracy of multi-target tracking can be improved.

CLAIMS

1. A target tracking method, comprising:

determining multiple target objects conflicting with each other according to distances between target objects in a tracking area;

in response to a situation that any target object in the multiple target objects does not conflict with other target objects in the multiple target objects, extracting the current feature of the target object;

determining a reference feature matching the current feature of the target object based on multiple reference features corresponding to the multiple target objects, wherein the multiple reference features are features extracted from the multiple target objects before the multiple target objects conflict with each other; and

determining identification information corresponding to the reference feature matching the current feature of the target object as identification information of the target object.

2. The method of claim 1, wherein a distance between any target object in the multiple target objects conflicting with each other and at least one of the other target objects in the multiple target objects is less than or equal to a distance threshold;

wherein in response to the situation that any target object in the multiple target objects does not conflict with other target objects in the multiple target objects, extracting the current feature of the target object, comprises:

for any target object in the multiple target objects, in response to a distance between the target object and each of the other target objects in the multiple target objects is greater than the distance threshold, extracting the current feature of the target object.

3. The method of claim 1 or 2, further comprising:

for any target object, determining first coordinates of the target object corresponding to multiple cameras, wherein first coordinates of the target object corresponding to any of the cameras are used for representing first coordinates of the target object obtained according to an image acquired by the camera;

determining a sub-area where the target object is located in the tracking area according to the first coordinates of the target object corresponding to the multiple cameras, wherein the tracking area comprises multiple sub-areas; and

fusing, according to confidence of the multiple cameras for the sub-area where the target object is located, the first coordinates of the target object corresponding to the multiple cameras to obtain fusion coordinates of the target object;

wherein before determining the multiple target objects conflicting with each other according to the distances between the target objects in the tracking area, the method further comprises:

determining the distances between the target objects in the tracking area according to the fusion coordinates of the target objects in the tracking area.

4. The method of any one of claims 1 to 3, wherein extracting the current feature of the target object comprises:

determining a sub-area where the target object is located in the tracking area;

determining a camera extracting the current feature of the target object according to confidence of multiple cameras for the sub-area where the target object is located; and

extracting the current feature of the target object according to video frames acquired by the camera extracting the current feature of the target object.

5. The method of claim 4, wherein determining the camera extracting the current feature of the target object according to the confidence of the multiple cameras for the sub-area where the target object is located comprises:

determining the camera extracting the current feature of the target object according to the confidence of the multiple cameras for the sub-area where the target object is located and overlapping information between a detection box of the target object and detection boxes of other target objects in the video frames acquired by the multiple cameras.

6. The method of claim 5, wherein the overlapping information comprises Intersection-over-Union (IoU) between the detection box of the target object and the detection boxes of the other target objects;

wherein determining the camera extracting the current feature of the target object according to the confidence of the multiple cameras for the sub-area where the target object is located and the overlapping information between the detection box of the target object and the detection boxes of other target objects in the video frames acquired by the multiple cameras comprises:

determining, from cameras that meet an overlapping condition, a camera with a highest confidence for the sub-area where the target object is located as the camera extracting the current feature of the target object, wherein the overlapping condition indicates that in the video frames acquired, the IoU between the detection box of the target object and the detection boxes of the other target objects is less than a predetermined threshold.

7. The method of any one of claims 3 to 6, further comprising:

for any camera in the multiple cameras, determining confidence of the camera for multiple sub-areas according to distances between the camera and the multiple sub-areas of the tracking area.

8. The method of claim 7, wherein after determining the confidence of the multiple cameras for the multiple sub-areas, the method further comprises:

for any camera in the multiple cameras, adjusting the confidence of the camera for the multiple sub-areas according to an average distance between the target objects in the video frames acquired by the camera.

9. The method of any one of claims 1 to 8, wherein determining the reference feature matching the current feature of the target object based on the multiple reference features corresponding to the multiple target objects comprises:

determining a reference feature with a maximum similarity to the current feature of the target object from the multiple reference features corresponding to the multiple target objects;
and

in response to a similarity between the reference feature with the maximum similarity and the current feature of the target object being greater than or equal to a similarity threshold, determining the reference feature with the maximum similarity as the reference feature matching the current feature of the target object.

10. The method of claim 9, wherein after determining the reference feature with the maximum similarity to the current feature of the target object, the method further comprises:

in response to the similarity between the reference feature with the maximum similarity and the current feature of the target object being less than the similarity threshold, determining the reference feature matching the current feature of the target object based on a reference feature corresponding to a target object that does not conflict with the target object and conflicts with any of the other target objects.

11. The method of any one of claims 1 to 10, further comprising:

for any target object, in response to a situation that the target object does not conflict with other target objects at a current moment or at a previous moment of the current moment, determining target objects that do not conflict with other target objects at the previous moment and do not match other target objects at the current moment as remaining target objects at the previous moment; or, in response to a situation that the target object conflicts with same other target objects at the current moment and the previous moment, determining target objects that do not match other target objects at the current moment in a conflict area to which the target object belongs at the previous moment as the remaining target objects at the previous moment; and

determining, from the remaining target objects at the previous moment, a target object closest to the target object at the current moment as a target object matching the target object at the current moment, and taking identification information of the target object matched as the identification information of the target object.

12. A target tracking apparatus, comprising:

a first determining portion, configured to determine multiple target objects conflicting with each other according to distances between target objects in a tracking area;

an extracting portion, configured to extract, in response to a situation that any target object in the multiple target objects does not conflict with other target objects in the multiple target objects, a current feature of the target object;

a second determining portion, configured to determine a reference feature matching the current feature of the target object based on multiple reference features corresponding to the multiple target objects, wherein the multiple reference features are features extracted from the multiple target objects before the multiple target objects conflict with each other; and

a third determining portion, configured to determine identification information corresponding to the reference feature matching the current feature of the target object as identification information of the target object.

13. The apparatus of claim 12, wherein the extracting portion is configured to:

for any target object in the multiple target objects, in response to a distance between the target object and each of the other target objects in the multiple target objects are greater than a distance threshold, extract the current feature of the target object.

14. The apparatus of claim 12 or 13, further comprising:

a third determining portion, configured to determine, for any target object, first coordinates of the target object corresponding to multiple cameras, wherein the first coordinates of the target object corresponding to any of the cameras are used for representing first coordinates of the target object obtained according to an image acquired by the camera;

a fourth determining portion, configured to determine a sub-area where the target object is located in the tracking area according to the first coordinates of the target object corresponding to the multiple cameras, wherein the tracking area comprises multiple sub-areas;

a fusing portion, configured to fuse, according to confidence of the multiple cameras for the sub-area where the target object is located, the first coordinates of the target object corresponding to the multiple cameras to obtain fusion coordinates of the target object; and

a fifth determining portion, configured to determine the distances between the target objects in the tracking area according to the fusion coordinates of the target object in the tracking area.

15. The apparatus of any one of claims 12 to 14, wherein the extracting portion is configured to:

determine a sub-area where the target object is located in the tracking area;

determine a camera extracting the current feature of the target object according to confidence of multiple cameras for the sub-area where the target object is located; and

extract the current feature of the target object according to video frames acquired by the camera extracting the current feature of the target object.

16. The apparatus of claim 15, wherein the extracting portion is configured to:

determine the camera extracting the current feature of the target object according to the confidence of the multiple cameras for the sub-area where the target object is located and overlapping information between a detection box of the target object and detection boxes of other target objects in the video frames acquired by the multiple cameras.

17. The apparatus of claim 16, wherein the overlapping information comprises Intersection-over-Union (IoU) between the detection box of the target object and the detection boxes of other target objects;

wherein the extracting portion is configured to:

determine, from cameras that meet an overlapping condition, a camera with a highest confidence for the sub-area where the target object is located as the camera extracting the current feature of the target object, wherein the overlapping condition indicates that in the video frames acquired, the IoU between the detection box of the target object and the detection boxes of the other target objects is less than a predetermined threshold.

18. The apparatus of any one of claims 14 to 17, further comprising:

a sixth determining portion, configured to determine, for any camera in the multiple cameras, confidence of the camera for the multiple sub-areas according to distances between the camera and the multiple sub-areas of the tracking area.

19. The apparatus of claim 18, further comprising:

an adjusting portion, configured to adjust, for any camera in the multiple cameras, the confidence of the camera for the multiple sub-areas according to an average distance between the target objects in the video frames acquired by the camera.

20. The apparatus of any one of claims 12 to 18, wherein the second determining portion is configured to:

determine a reference feature with a maximum similarity to the current feature of the target object from the multiple reference features corresponding to the multiple target objects; and

in response to a similarity between the reference feature with the maximum similarity and the current feature of the target object being greater than or equal to a similarity threshold, determine the reference feature with the maximum similarity as the reference feature matching the current feature of the target object.

21. The apparatus of claim 20, further comprising:

a seventh determining portion, configured to determine, in response to the similarity between the reference feature with the maximum similarity and the current feature of the target object being less than the similarity threshold, the reference feature matching the current feature of the target object based on a reference feature corresponding to a target object that does not conflict with the target object and conflicts with any of the other target objects.

22. The apparatus of any one of claims 12 to 21, further comprising:

an eighth determining portion, configured to: for any target object, in response to a situation that the target object does not conflict with other target objects at a current moment or at a previous moment of the current moment, determine target objects that do not conflict with other target objects at the previous moment and do not match other target objects at the current moment as remaining target objects at the previous moment; or, in response to a situation that the target object conflicts with same other target objects at the current moment and the previous moment, determine target objects that do not match other target objects at the current moment in a conflict area to which the target object belongs at the previous moment as the remaining target objects at the previous moment; and

a ninth determining portion, configured to determine, from the remaining target objects at the previous moment, a target object closest to the target object at the current moment as a target object matching the target object at the current moment, and take identification information of the target object matched as the identification information of the target object.

23. An electronic device, comprising;

one or more processors;

a memory configured to store an executable instruction;

wherein the one or more processors are configured to call the executable instruction stored in the memory to implement the method of any one of claims 1 to 11.

24. A computer-readable storage medium, having stored thereon a computer program instruction, wherein the computer program instruction, when executed by a processor, causes to implement the method of any one of claims 1 to 11.

25. A computer program, comprising a computer-readable code, wherein the computer-

readable code, when running in an electronic device, causes a processor in the electronic device to implement the method of any one of claims 1 to 11.