



US 20120320162A1

(19) **United States**

(12) **Patent Application Publication**
LO et al.

(10) **Pub. No.: US 2012/0320162 A1**

(43) **Pub. Date: Dec. 20, 2012**

(54) **VIDEO OBJECT LOCALIZATION METHOD
USING MULTIPLE CAMERAS**

Publication Classification

(75) Inventors: **Kuo-Hua LO**, Kaohsiung City
(TW); **Jen-Hui CHUANG**, Hsinchu
City (TW); **Horng-Horng LIN**,
Hsinchu County (TW)

(51) **Int. Cl.**
H04N 13/02 (2006.01)

(52) **U.S. Cl.** **348/47; 348/E13.074**

(73) Assignee: **NATIONAL CHIAO TUNG
UNIVERSITY**, Hsinchu City (TW)

(57) **ABSTRACT**

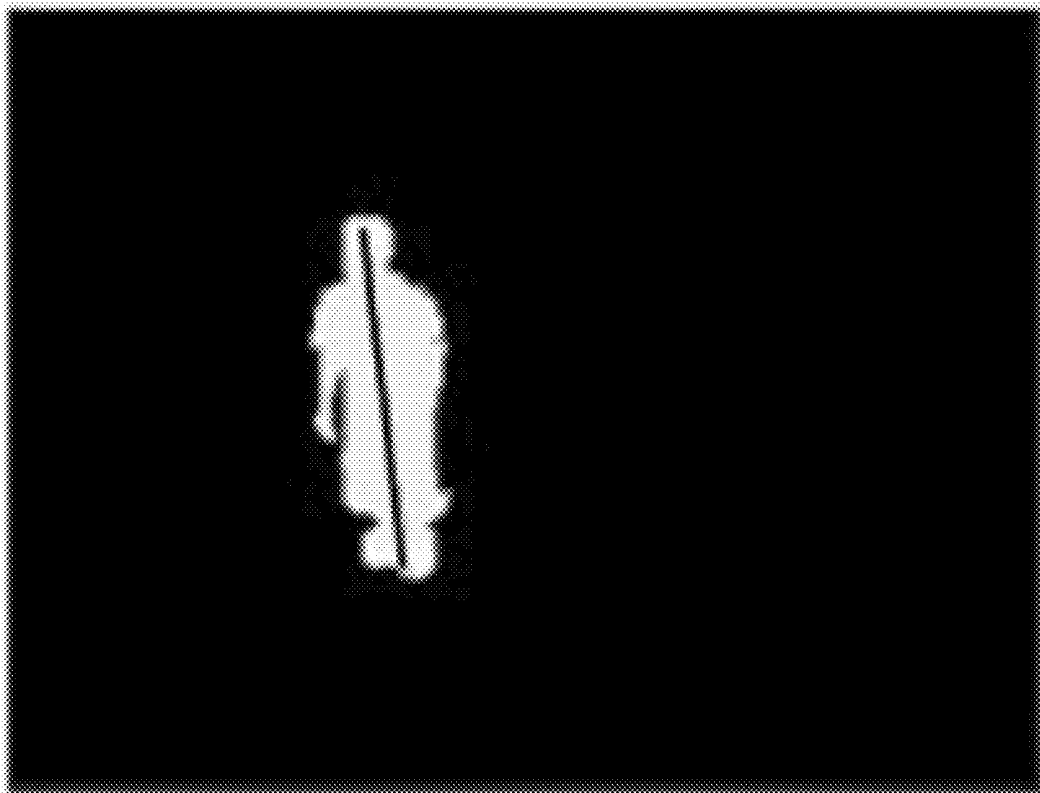
(21) Appl. No.: **13/487,264**

(22) Filed: **Jun. 4, 2012**

An efficient 3D object localization method using multiple cameras is provided. The proposed method comprises a three-dimensional object localization process that firstly generates a plurality of two-dimensional line samples originated from a pre-calibrated vanishing point in each camera view for representing foreground video objects, secondly constructs a plurality of three-dimensional line samples from the two-dimensional line samples in all the multiple camera views, and thirdly determines three-dimensional object to locations by clustering the three-dimensional line samples into object groups.

Related U.S. Application Data

(60) Provisional application No. 61/498,619, filed on Jun. 20, 2011.



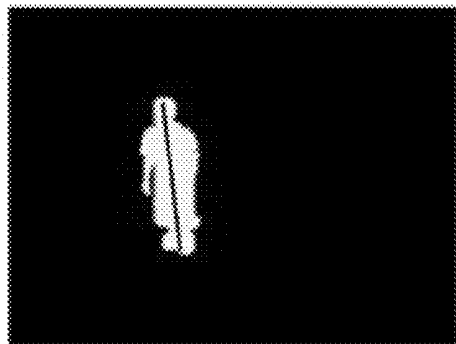


FIG. 1

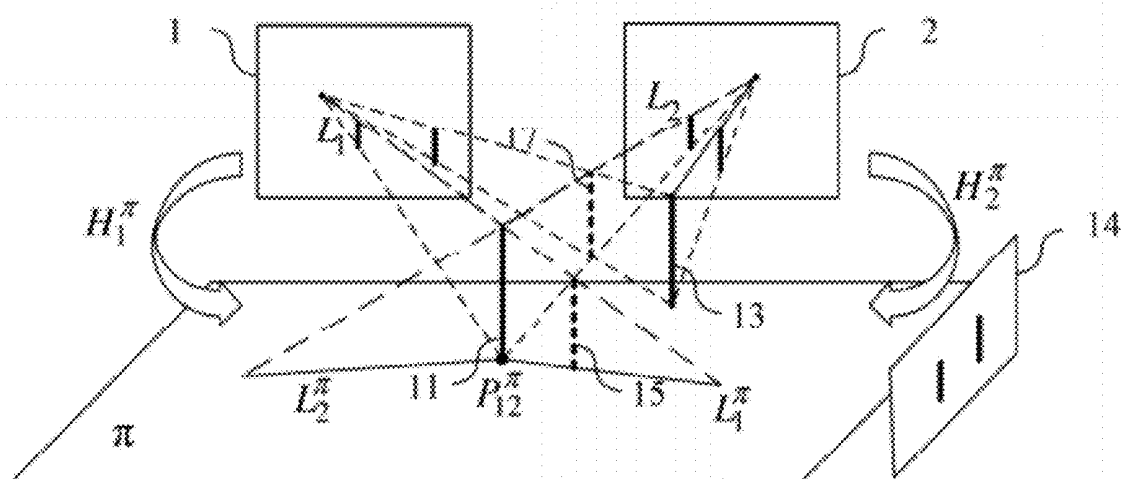


FIG. 2

3



30

FIG. 3A

3'



30'

FIG. 3B

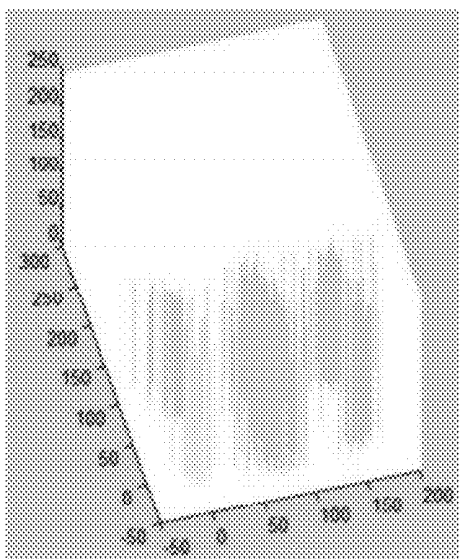


FIG. 3C

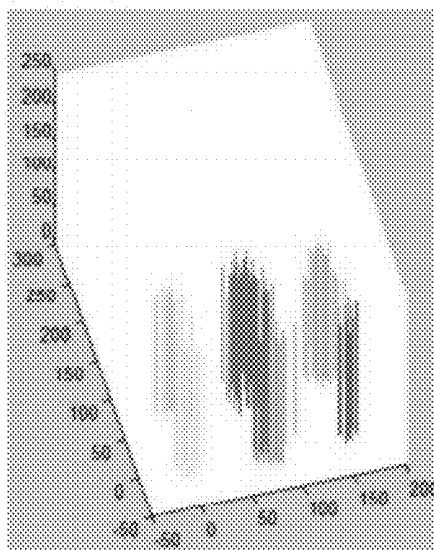


FIG. 3D

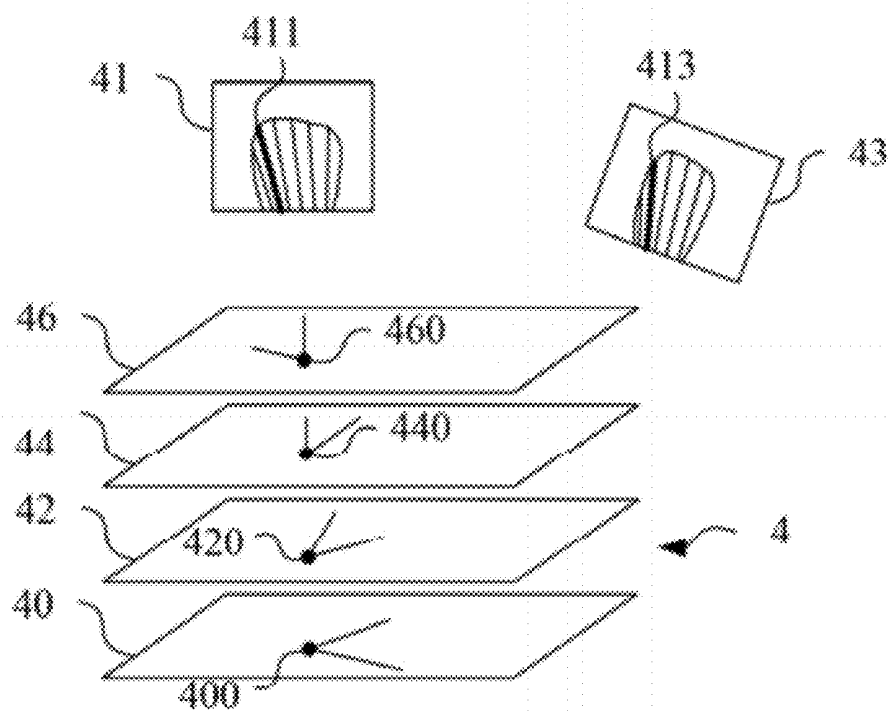


FIG. 4

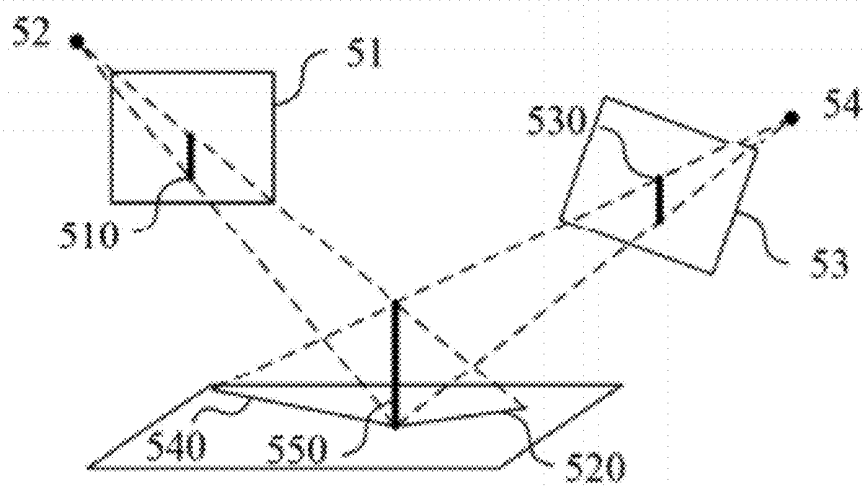


FIG. 5



FIG. 6A

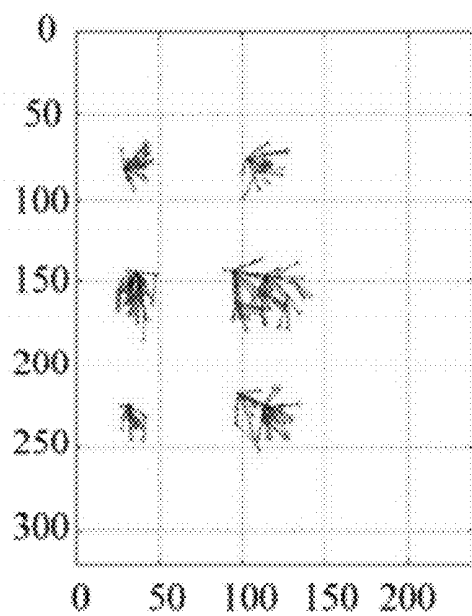


FIG. 6B

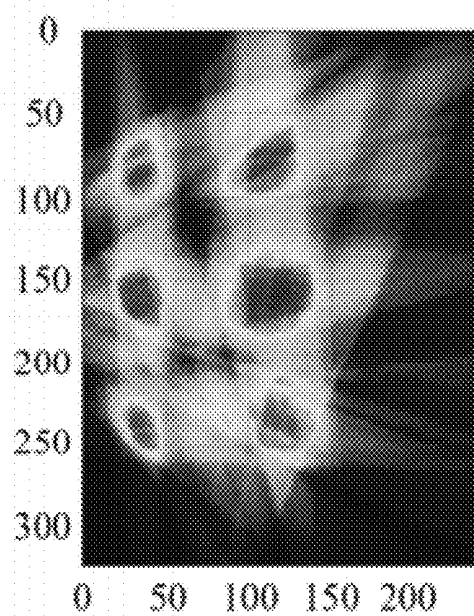


FIG. 6C

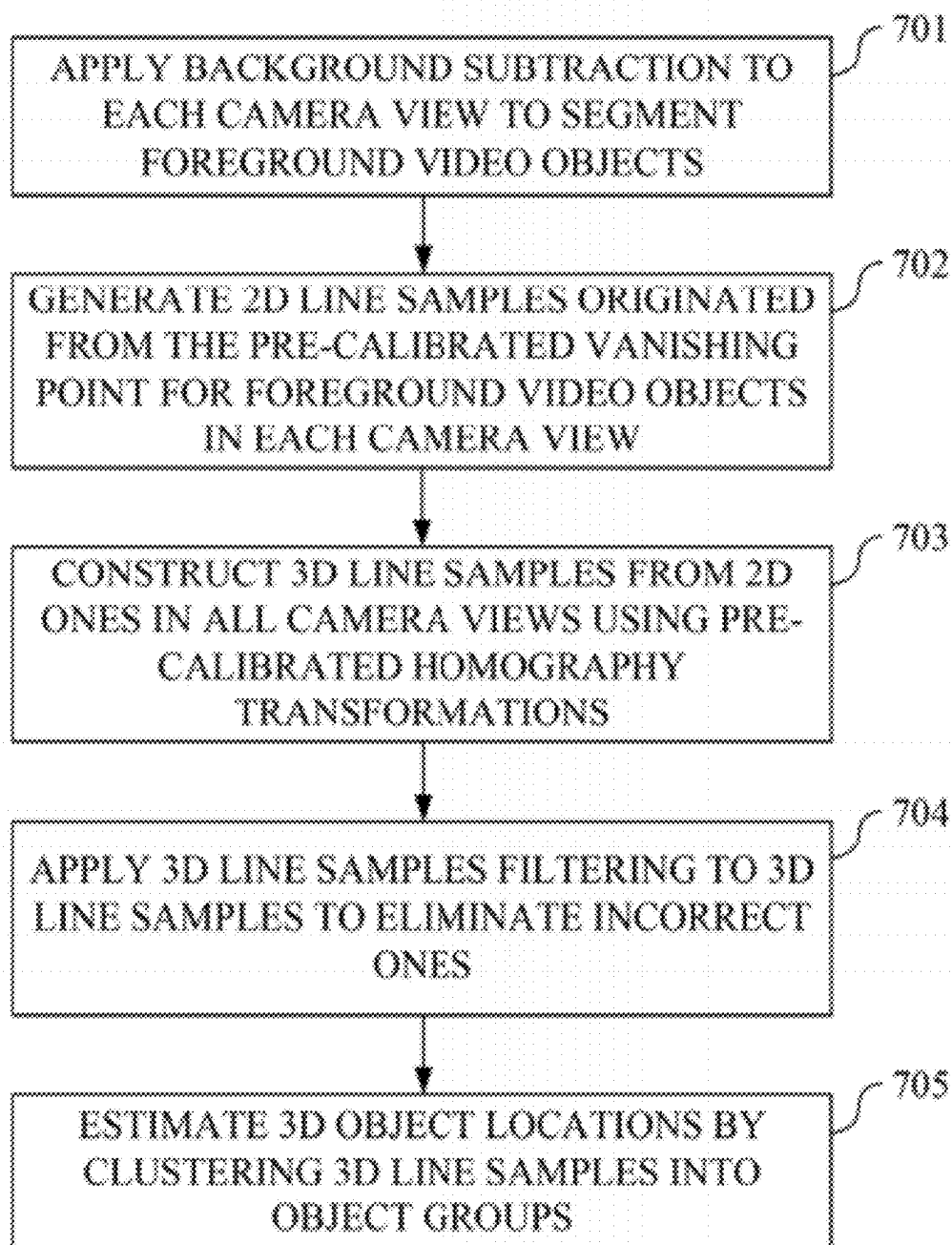
700

FIG. 7

VIDEO OBJECT LOCALIZATION METHOD USING MULTIPLE CAMERAS

RELATED APPLICATIONS

[0001] This application claims priority to U.S. provisional Application Ser. No. 61/498,619, filed Jun. 20, 2011, which is herein incorporated by reference.

BACKGROUND

[0002] 1. Technical Field

[0003] The present disclosure relates to an image processing method.

[0004] Particularly, the present disclosure relates to a computer-vision-based three-dimensional object localization method using multiple cameras.

[0005] 2. Description of Related Art

[0006] In recent years, vision-based localization and tracking have shifted from monocular approaches to multi-camera approaches since more image information provided by multiple cameras can be used to resolve object occlusions. However, most conventional multi-camera approaches apply pixel-based processing to compute three-dimensional (3D) object locations. They are thus not quite suitable for real-time applications because they often need costly coordinate transformations of all foreground pixels in multiple camera views to different 3D reference planes. Accordingly, we propose a new 3D object localization method that can largely reduce the computation of coordinate transformations in multiple camera views for the 3D localization of video objects.

SUMMARY

[0007] The present disclosure is to introduce an object localization method using vanishing point-based line sampling for video surveillance using multiple cameras. Specifically, the present disclosure utilizes a plurality of static cameras to capture video sequences of a scene of interest. An aspect of the present disclosure comprises a 3D object localization process that firstly generates a plurality of two-dimensional (2D) line samples originated from a pre-calibrated vanishing point in each camera view, secondly constructs a plurality of 3D line samples from the 2D line samples in all the camera views, and thirdly determines 3D object locations by clustering the 3D line samples.

[0008] It is to be understood that both the foregoing general description and the following detailed description are by examples, and are intended to provide further explanation of the disclosure as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] The present disclosure can be more fully understood by reading the following detailed description of the embodiment, with reference made to the accompanying drawings as follows:

[0010] FIG. 1 is a diagram depicting an example of a foreground object and its 2D major axis estimated by principal component analysis in a camera view;

[0011] FIG. 2 is a diagram illustrating the main idea of the present disclosure;

[0012] FIG. 3A and FIG. 3B are diagrams depicting two camera views of a scene of interest and the generated 2D line samples by an embodiment of the present disclosure;

[0013] FIG. 3C is a diagram depicting the 3D line samples constructed from the 2D line samples in FIG. 3A and FIG. 3B in an embodiment of the present disclosure;

[0014] FIG. 3D is a diagram depicting the result of clustering the 3D line samples in FIG. 3C into six objects;

[0015] FIG. 4 is a diagram depicting the generation of a 3D line sample by four intersection points in 3D reference planes in an embodiment of the present disclosure, wherein the intersection points are computed from two 2D line samples in different camera views;

[0016] FIG. 5 is a diagram depicting the generation of a 3D line sample by the intersection of two 3D planes in an embodiment of the present disclosure, wherein the two 3D planes are computed from two 2D line samples in different camera views;

[0017] FIG. 6A and FIG. 6B are diagrams depicting, respectively, an example of a 2D camera view and the clustering results of 3D line samples in top view in an embodiment of the present disclosure;

[0018] FIG. 6C is a diagram depicting a synergy map derived by using S. M. Khan and M. Shah, "Tracking multiple occluding people by localizing on multiple scene planes"; and

[0019] FIG. 7 is a flowchart of the 3D object localization process of an embodiment of the present disclosure.

DETAILED DESCRIPTION

[0020] Reference will now be made in detail to the present embodiments of the disclosure, examples of which are illustrated in the accompanying drawings. Wherever possible, the same reference numbers are used in the drawings and the description to refer to the same or like parts.

[0021] Vision-based object localization and tracking have shifted in recent years from monocular approaches to multi-camera approaches because multiple cameras provide more information than a single camera in resolving occlusions. In multiple camera approaches, region information is often applied to match video objects, say people, in multiple views. For instance, in S. M. Khan and M. Shah, "Tracking multiple occluding people by localizing on multiple scene planes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 505-519, 2009, the Khan and Shah propose a multi-camera approach to track occluded people by locating them on multiple 3D reference planes.

[0022] While the above-mentioned multi-camera approaches deal well with scenes of a dense crowd, they are not, in general, suitable for real-time applications because these approaches often need costly coordinate transformations of all foreground pixels in multiple camera views to different 3D reference planes. To enhance efficiency, we propose a new three-dimensional object localization method that can largely reduce the computation of coordinate transformations in multiple camera views for the 3D localization of video objects.

[0023] Two processes, (1) the 3D object localization process and (2) the camera calibration process, are included in the proposed method of 3D video object localization. In the following presentations of the proposed method, we start with the descriptions of the 3D object localization process, followed by that of the camera calibration process.

3D Object Localization Process

[0024] The main idea of the proposed 3D object localization process is to apply efficient projections of 2D line

samples originated from pre-calculated vanishing points in camera views. Specifically, we project 2D line samples of foreground regions from all camera views to 3D reference planes at different heights via homography transformation, and generate 3D line samples. By filtering out noise of the 3D line samples and grouping the remaining ones into object clusters, stable localization results of video objects can be obtained by cluster centers. Since the proposed video object localization method is based on the processing of line samples originated from vanishing points, it is much faster than the video object localization methods based on the processing of foreground pixels, and does not sacrifice the localization accuracy.

Axis-Based Video Object Localization

[0025] For better understanding of the basic idea of the proposed vanishing point-based line sampling in the 3D object localization process, we begin by illustrating how to localize video objects, say people, given their 2D major axes (2D MAs). However, in real cases, the estimation of 2D major axes of objects in images may be unstable for crowded scenes. Therefore, we introduce vanishing point-based line samples to replace 2D MAs of video objects. The adoption of vanishing point-based line samples gives promising accuracy and efficiency in video object localization.

(i) Finding a 3D Major Axis for a Video Object

[0026] As shown in FIG. 1, we assume that the foreground region of a person derived by background subtraction is given so that we can estimate its 2D MA via principal component analysis.

[0027] By further assuming that the correspondences between the 2D MAs in two camera views are known, the 3D major axes (3D MAs) of the person can be reconstructed. For example, as shown in FIG. 2, the 2D MAs of the same person, denoted by L_1 and L_2 in camera View 1 and camera View 2, respectively, can be projected to a reference plane π by homographic matrices $H_{1\pi}$ and $H_{2\pi}$. The intersection of the two projections L_1^π and L_2^π on π can be denoted by P_{12}^π . By computing the intersection points of the projections of L_1 and L_2 on different reference planes of different heights using homography transformations, the 3D MA of the video object 11 in FIG. 2, can be formed. Specifically, a 3D MA computed from View 1 and View 2 can be denoted by an axis set A of intersection points, i.e., $A_{12}^{h_b h_t} = \{P_{12}^{h_b}, \dots, P_{12}^{h_t}\}$, where h_b and h_t are the heights of bottom and top end points of the 3D MA, respectively.

(ii) Finding 3D Major Axes for Multiple Video Objects

[0028] Now, we will extend the 3D MA estimation for one video object to that for multiple video objects. As shown in FIG. 2, a 3D MA obtained from the i th 2D MA in camera View 1 and the j th 2D MA in camera View 2 can be denoted by

$$A_{1i,2j}^{h_b h_t} = \{P_{1i,2j}^{h_b}, \dots, P_{1i,2j}^{h_t}\}.$$

[0029] Given M 2D MAs in View 1 and N 2D MAs in View 2, since we do not have correspondences of 2D MAs in the two views actually, the total number of possible 3D MAs in reconstruction will be MN . We then apply a procedure of filtering process (to be detailed later) to filter out incorrect 3D MAs. FIG. 2 gives an illustration of two incorrect 3D MAs 15

and 17 that can be ruled out by the filtering process. The remaining 3D MAs will correspond to the detected video objects in the scene.

Vanishing Point-Based Line Sampling

[0030] For real scenarios, many objects may appear in the scene at the same time so that each segmented foreground region may contain more than one video objects. Under such situations of object occlusions, the 2D MA of each video object is hard to be detected correctly. Therefore, instead of estimating the 2D MAs for video objects, we propose to use 2D line samples of foreground regions in the present disclosure to construct 3D line samples. Then, by clustering the constructed 3D line samples into object groups, the 3D locations for video objects can be derived.

(i) Generating 2D Line Samples from Vanishing Points

[0031] Owing to the observation that video objects of interest, like people, are often perpendicular to the ground plane in most scenarios of video surveillance, these objects can be represented by 2D line samples that are originated from the vanishing point corresponding to vertical 3D lines of the scene of interest. For each camera view, such a vanishing point corresponding to 3D vertical lines of the scene is determined by the camera calibration process, as will be described later.

[0032] Examples of afore-mentioned 2D line samples in two camera views 3 and 3' are given in FIG. 3A and FIG. 3B, respectively. A plurality of 2D line samples that are originated from the vanishing points 30 and 30' and pass through the foreground regions are generated.

(ii) Generating 3D Line Samples

[0033] The 2D line samples in each pair of camera views, as shown in FIG. 3A and FIG. 3B for example, are then used to construct 3D line samples, as shown in FIG. 3C, via an embodiment of the present disclosure. Specifically, as shown in FIG. 4, a pair of 2D line samples 411 and 413 in the camera view 41 and 43, respectively, is projected to the four reference planes 40, 42, 44 and 46 via homography transformations, and form four intersection points 400, 420, 440 and 460. By linking these intersection points on the reference planes, a 3D line sample can be constructed. Note here that only the intersection points on a vertical line will generate a 3D line sample. If the intersection points are not collinear (i.e., along a vertical line), they will not be considered as 3D line samples.

[0034] In another embodiment of the present disclosure, a 3D line sample can also be obtained by resolving the intersection of two 3D planes derived from a pair of 2D line samples in two different camera views. As shown in FIG. 5, two 2D line samples 510 and 530 in the camera view 51 and 53, respectively, are presented. By projecting the 2D line sample 510 (530) to the reference plane via homography transformation, the projection line 520 (540) can be determined. A 3D plane corresponding to the projection line 520 and the camera center 52 can then be computed. The other 3D plane corresponding to the projection line 540 and the camera center 54 can be computed as well. By resolving the intersection of the two 3D planes, a 3D line sample 50 can then be obtained.

[0035] It is clear for now that, given each pair of 2D line samples from two different camera views, a 3D line sample can be derived. By using such a technique of vanishing point-based line sampling, the required computation of coordinate

transformations of foreground pixels in all the camera views for object localization can be largely reduced.

[0036] Furthermore, since each of the 3D line samples is constructed by observations from two views only, some incorrect 3D line samples may be derived and need to be removed. A procedure of 3D line sample filtering is then applied to filter out a plurality of incorrect 3D line samples. In an embodiment, the following filtering rules are developed to rule out incorrect 3D line samples:

[0037] 1. The length $|h_t - h_b|$ of a 3D line sample is shorter than T_{len} , where h_t and h_b , are the heights of top and bottom end points of the 3D line sample, respectively.

[0038] 2. The bottom height h_b , of a 3D line sample is higher than T_{bot} .

[0039] 3. The foreground coverage of a 3D line sample by projecting the sample points of the 3D line sample back to all camera views is lower than T_{fg} .

While the first two rules are intuitive, only the third rule is discussed in further detail here. To check the foreground coverage of a 3D line sample, the sample points of the 3D line sample are back projected to all camera views. For an object do appear in the monitored scene, these back-projected points should be covered by the corresponding foreground regions in the camera views. If all the back-projected points in all camera views are of foreground, the average foreground coverage rate (AFCR) of the 3D line sample is defined as 100%. A 3D line sample with an AFCR lower than a preset threshold T_{fg} will be removed. FIG. 3D is a diagram showing the results of applying the 3D line sample filtering to FIG. 3C, wherein each individual person can now be better identified visually.

(iii) Computing 3D Object Locations

[0040] After the 3D line sample filtering, the remaining 3D line samples are clustering into 3D object groups using the breadth first search (BFS) in an embodiment of the present disclosure. Specifically, if the 2D horizontal distance between two 3D line samples on the top view of the scene is closer than a threshold T_{com} , an edge between the two 3D line samples is established in an undirected graph. FIG. 6A and FIG. 6B are diagrams showing an image frame of one camera view and the reconstructed top view of a resultant graph of BFS, respectively, in an embodiment of the present disclosure. In FIG. 6B, each graph node on the top view denotes a 3D line sample. Six clusters corresponding to six people are found by the BFS grouping. The center position and the average height of each cluster can then be estimated (via various average schemes) to represent the object location and its height.

[0041] In FIG. 6C, as a reference, we show the synergy map obtained by using the method proposed in S. M. Khan and M. Shah, "Tracking multiple occluding people by localizing on multiple scene planes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 505-519, 2009. It is clear that the distributions of the derived clusters by the proposed approach in FIG. 6B is similar to that of the compared approach in FIG. 6C, which validate the effectiveness of the proposed approach. Regarding the comparisons of computational efficiency, the proposed approach is more than twenty-five times faster than the compared approach in our experimental simulations using the video sequence of FIG. 6 as a test bench on Windows 7 with, 4 GB RAM and a 2.4G Intel Core2 Duo CPU.

Camera Calibration Process

[0042] In an embodiment, a camera calibration process can be made prior to the 3D object localization process for each

camera to estimate (1) a specific vanishing point in the camera view and (2) homography transformations from the camera view to multiple 3D reference planes of different heights of the scene by using a certain number of vertical pillars. For example, four pillars that are perpendicular to the ground plane of the scene can be utilized as markers for the calibration of a vanishing point in each camera view. By estimating the intersection point of the projection lines of the four pillars in a camera view, a specific vanishing point of the camera view can be determined.

[0043] The four pillars can also be utilized to estimate the homography transformations from a camera view to multiple 3D reference planes of different heights of the scene. Specifically, as indicated in R. Eshel and Y. Moses, "Tracking in a Dense Crowd Using Multiple Cameras," *International Journal of Computer Vision*, vol. 88, no. 1, pp. 129-143, 2010, one can use four vertical pillars placed in the scene, with marker points at three known heights on each of them, to establish the homographies between a camera view and any 3D reference planes at desired heights. Since a new reference point at any height along a pillar can be identified in the images using the cross-ratio along that pillar, the above homographic relationship can actually be established for planes at arbitrary height. Note that the derived homography transformations can also be used to back-project points on a 3D reference plane to the camera views as required by other computational processes.

[0044] Flowchart of the Video Object Localization Method **[0045]** FIG. 7 is a flowchart of an object localization method 700 of an embodiment of the present disclosure. The object localization method 700 comprises the steps as outlined in the sentences that follow.

[0046] In step 701, background subtraction is applied to the captured image from each camera to segment foreground video objects.

[0047] In step 702, 2D line samples that are originated from the pre-calibrated vanishing point are generated for the foreground video objects in each camera view.

[0048] In step 703, 3D line samples are constructed from the 2D line samples in all the camera views using pre-calibrated homography transformations.

[0049] In step 704, 3D line sample filtering is applied to the 3D line samples to eliminate incorrect ones.

[0050] In step 705, 3D object locations are estimated by clustering the remaining 3D line samples into object groups.

[0051] Using 2D line sampling in camera views for object localization can avoid the complicated axis analysis of video objects and can effectively handle partial occlusion problems. Furthermore, the method of the present disclosure enhances the efficiency of object localization because only a few samples of foreground pixels are processed in the computation of coordinate transformations.

[0052] It will be apparent to those skilled in the art that various modifications and variations can be made to the structure of the present disclosure without departing from the scope or spirit of the disclosure. In view of the foregoing, it is intended that the present disclosure cover modifications and variations of this disclosure provided they fall within the scope of the following claims.

What is claimed is:

1. An object localization method using multiple cameras comprising at least one of the steps of:

obtaining a plurality of two-dimensional (2D) line samples in a camera view, wherein the 2D line samples is origi-

nated from a vanishing point in the camera view for representing foreground video objects;
 constructing a plurality of three-dimensional (3D) line samples on at least one 3D reference plane from the 2D line samples in multiple camera views; and
 to determining 3D locations of foreground objects by clustering the 3D line samples into object groups.

2. The object localization method of claim 1, wherein the vanishing point is determined according to a plurality of calibration pillars in parallel in the scene of interest.

3. The object localization method of claim 1, wherein a 3D reference plane is the plane of a preset height in the scene of interest.

4. The object localization method of claim 1, wherein the step of constructing 3D line samples according to 2D line samples comprises:

mapping each pair of the 2D line samples in different camera views to a plurality of 3D reference planes to generate a plurality of intersection points; and
 constructing a 3D line sample by the derived intersection points that are collinear or approximately collinear.

5. The object localization method of claim 1, wherein the step of constructing 3D line samples according to 2D line samples comprises:

generating two 3D planes from a pair of 2D line samples in different camera views, wherein each 3D plane is formed by the projection line of one of the paired 2D line

samples on a 3D reference plane and the 3D camera center corresponding to that 2D line sample; and
 constructing a 3D line sample by resolving the intersection of the two 3D planes.

6. The object localization method of claim 1, further comprising a filtering process for ruling out a plurality of incorrect 3D line samples in 3D line samples.

7. The object localization method of claim 6, wherein the step of the filtering process comprises discarding the 3D line samples whose lengths are shorter than a specific threshold.

8. The object localization method of claim 6, wherein the step of the filtering process comprises discarding the 3D line samples whose bottom heights are higher than a specific threshold.

9. The object localization method of claim 6, wherein the step of the filtering process comprises discarding the 3D line samples whose foreground coverage rates are lower than a specific threshold, wherein the foreground coverage rate of a 3D line sample is calculated based on the number of foreground pixels on the projections of the 3D line sample back to all the camera views.

10. The object localization method of claim 1, wherein the step of determining 3D locations of foreground objects comprises a step of determining a major axis of each foreground object from its clustered 3D line samples.

* * * * *