



(43) International Publication Date
19 June 2014 (19.06.2014)

(10) International Publication Number
WO 2014/092552 A2

- (51) **International Patent Classification:** Not classified
- (21) **International Application Number:**
PCT/MY2013/000266
- (22) **International Filing Date:**
13 December 2013 (13.12.2013)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
PI 2012005406 13 December 2012 (13.12.2012) MY
- (71) **Applicant:** MIMOS BERHAD [MY/MY]; Technology Park Malaysia, 57000 Kuala Lumpur (MY).
- (72) **Inventors:** ZULAIKHA, Binti Kadim; Mimos Berhad, Technology Park Malaysia, 57000 Kuala Lumpur (MY). HON, Hock Woon; Mimos Berhad, Technology Park Malaysia, 57000 Kuala Lumpur (MY). NORSHUHADA, Binti Samudin; Mimos Berhad, Technology Park Malaysia, 57000 Kuala Lumpur (MY).
- (74) **Agent:** YAP, Kah, Hong; Pyprus Sdn Bhd, Suite 8-02, 8th Floor, Plaza First Nationwide 161, Jalan Tun H.S. Lee, 50000 Kuala Lumpur (MY).
- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) **Title:** METHOD FOR NON-STATIC FOREGROUND FEATURE EXTRACTION AND CLASSIFICATION

(57) **Abstract:** The present invention provides a method for processing videos captured through a stereo camera, the videos having a plurality of frames, each frames includes at least a left and a right images. The method includes acquiring videos captured by the pair of stereo cameras (101) and acceleration information of the stereo camera (101) detected through the accelerometer (102); extracting feature points in the right and left images of a current frame; matching the feature points between the right and left images, wherein disparity value of each feature point is computed; clustering the feature points on a current frame into one or more clouds based on their appearance and distance; determining a correspondence between the clouds of the current frame and previous frame; classifying the clouds through estimating statistical disparity distribution of each cloud to identify if the respective clouds are new clouds, existing clouds or unknown clouds; and merging the clouds that are determined to be a same object based on the tracking information and changes of disparity of the clouds over time. A system carrying out the above method is also provided herein.



WO 2014/092552 A2

Method For Non-Static Foreground Feature Extraction and Classification

Field of the Invention

[0001] The present invention relates to image processing. More particularly, the
5 present invention relates to a method for extracting and classifying non-static
foreground features.

Background

[0002] The feature point extraction and classification for non-static cameras is
difficult as the computer vision is unable to differentiate the movement on the feature
10 point either it is caused by the actual moving object or by the motion of the cameras.

[0003] Many studies and research had been carried out extensively. US patent
no. 7,729,512 provides a stereo image processing system and method for detecting
moving objects. In this patent, disparity threshold is used for classifying the feature
points. This system require extensive searching and computing of all the individual the
15 feature points and their disparity information, which may lead to high processing power
needed for real-time tracking.

[0004] US Patent publication no. 2011/0134221 too suggests a system and
method of processing stereo images for extracting moving object. Similarly, feature
vectors are being extracted and matched to detect moving object.

20 **[0005]** It is recognized that feature points/vectors are widely used for detecting
moving object from a video stream. However, it is a challenge to provide a reliable

system and method to process these feature points in an effective and efficient way to extract the moving object. It is also desired that the system and method are adapted with adaptive capability to refine the detection results in real time.

Summary

5 [0006] In one aspect of the present invention, there is provided a surveillance system having a pair of stereo camera capturing videos having a plurality of frames, each frames includes at least a left and a right images, wherein the stereo camera having an accelerometer attached thereto. The surveillance system comprises a video acquisition module for acquiring the videos captured by the pair of stereo cameras and
10 acceleration information of the stereo camera; a feature point disparity computational unit adapted for extracting feature points in right and left images of a current frame and matching the feature points between the right and left images, wherein disparity value of each feature points is computed; a feature point tracking unit adapted for clustering the feature points on a current frame into one or more cloud based on their appearance
15 and distance, and determining a correspondence between the clouds of the current frame and a previous frame, a feature point classification unit for classifying the clouds through estimating the statistical disparity distribution of each cloud to identify if the respective clouds are new clouds, existing clouds or unknown clouds; and an object extraction unit for merging the clouds that are determined to be a same object based on
20 the tracking information and changes of disparity of the clouds over time.

[0007] In one embodiment, the surveillance system may further comprise an event rules unit; a display module for displaying overplayed information on live video; and a post detection module to trigger post event actions defined by the event rules unit.

[0008] In another aspect of the present invention, there is further provided a method for processing videos captured through a stereo camera, the videos having a plurality of frames, each frames includes at least a left and a right images, wherein the stereo camera having an accelerometer attached thereto. The method comprises

5 acquiring videos captured by the pair of stereo cameras and acceleration information of the stereo camera detected through the accelerometer; extracting feature points in the right and left images of a current frame; matching the feature points between the right and left images, wherein disparity value of each feature point is computed; clustering the feature points on a current frame into one or more clouds based on their appearance

10 and distance; determining a correspondence between the clouds of the current frame and previous frame; classifying the clouds through estimating statistical disparity distribution of each cloud to identify if the respective clouds are new clouds, existing clouds or unknown clouds; and merging the clouds that are determined to be a same object based on the tracking information and changes of disparity of the clouds over

15 time.

[0009] In one embodiment, the method further comprises filtering noises while matching the feature points between the left and the right images; computing distances between the matched features points; and quantizing the distance based on a current disparity value range. It is also possible that matching the feature points between the

20 left and right image comprising searching area for fining a best matched feature point based on matching feature points determined in the previous frame.

[0010] In yet another embodiment, searching area for finding the best matched feature point may comprise matching feature points within an entire image of a first

frame; and matching feature points for all subsequent frames based on areas corresponding to feature points matched in a previous frame.

[0011] In yet a further embodiment, clustering the feature points into the one or more clouds follows by the steps of tracking each cloud by determining the correspondences between clouds detected in current frame and in previous frames; 5 determining a tracking status of each cloud detected in current frame, wherein the tracking status include new object, existing moving object, existing background and existing undetermined object; performing feature points matching between one of the left and right images of the current frame and previous frame; and redistribute the 10 feature point clouds into different or same cloud based on the matching information.

[0012] In a further embodiment, classifying feature point may further comprise estimating a statistical disparity distribution of each feature point cloud through mean, mpd or standard deviation; Storing the information when the cloud is determined as a new cloud; computing the change of disparity of each clouds that are tracked between a 15 current frame and a previous frame; determining the acceleration information of the stereo camera for the instance frame. When the acceleration is substantially zero, the rate of disparity changes of each tracked cloud in the current frame over a subsequent frame is computed, the cloud is determined to be a moving object when the rate of change is not substantially zero, and the cloud is determined to be a static object or 20 background when the rate of change is substantially zero; and when the acceleration is not substantially zero, the rate of disparity change between a current frame and a previous frame is computed, and the cloud is determined to be a static object or background if the rate of change is increasing/decreasing, and the cloud is determined to be a moving object when the rate of disparity change is substantially zero.

Brief Description of the Drawings

[0013] Preferred embodiments according to the present invention will now be described with reference to the figures accompanied herein, in which like reference numerals denote like elements;

5 [0014] FIG. 1 illustrates a schematic diagram a surveillance system in accordance with one embodiment of the present invention;

[0015] FIG. 2 illustrates the process flow operationally carries out by a feature point disparity computational unit in accordance with one embodiment of the present invention;

10 [0016] FIG. 3 illustrates the process flow operationally carries out by a feature point tracking unit in accordance with one embodiment of the present invention;

[0017] FIG. 4 illustrates the process flow of a feature point classification unit in accordance with one embodiment of the present invention; and

[0018] FIG. 5 illustrates the process flow carries out by the moving object
15 extraction unit in accordance with one embodiment of the present invention.

Detailed Description

[0019] Embodiments of the present invention shall now be described in detail, with reference to the attached drawings. It is to be understood that no limitation of the scope of the invention is thereby intended, such alterations and further modifications in
20 the illustrated device, and such further applications of the principles of the invention as

illustrated therein being contemplated as would normally occur to one skilled in the art to which the invention relates.

[0020] The objective of this invention is to extract foreground feature points captured by a pair of non-static cameras and classify them into a moving pixel or static /
5 background pixel. The resultant output (sparse information) can be used to delineate moving objects from background for further processing or event analysis. The feature point extraction and classification for non-static cameras is difficult as the computer vision is unable to differentiate the movement on the feature point either it is caused by the actual moving object or by the motion of the cameras. Thus, proposed invention is
10 to solve the problem on how to differentiate between moving object and background when everything in the scene is non-static between the frames.

[0021] FIG. 1 illustrates a block diagram of a surveillance system 100 in accordance with one embodiment of the present invention. The surveillance system 100 is adapted for processing surveillance videos captured through a pair of surveillance
15 cameras 101. The pair of cameras is adapted for capturing stereo images/videos operationally. For the purpose of description below, the pair of cameras shall herewith below refer to as stereo cameras. Such stereo camera can be a commercially available stereo camera or a camera assembly that comprise two suitable videos that are installed on a platform, or frames. Preferably, the surveillance cameras 101 equipped or attached
20 with an accelerometer 102 for detecting the movements of the respective surveillance cameras 101b. The surveillance system 100 comprises a video acquisition module 103, database 112, an image-processing module 104, a display module 110, and a post detection module 111. The video acquisition module 103 is connected to the pair of cameras 101 to acquire images or videos captured through the cameras. The

acceleration information concerning the plurality of cameras **101** recorded through the accelerometers **102** is also transmitted to the video acquisition module **103**. The captured videos/images and the acceleration information are stored on the database module **112**, which are retrievable by the imaging-processing module **104** for processing anytime. The image-processing module **104** responsible for processing the input videos and acceleration information transmitted from the video acquisition module **103**, and analyzing the same to extract and detects moving feature points from the images/videos. Once the feature points are being detected and extracted, they are output to the display module **110** as visual outputs, and to the post detection module **111** for triggering further processing or control as necessary.

[0022] Still referring to **FIG. 1**, the image-processing module **104** further comprises a feature point disparity computational unit **105**, a feature point tracking unit **106**, a feature point classification unit **107**, a moving object extraction unit **108**, and an event rule unit **109**. The feature point disparity computational unit **105** performs feature point matching between right and left images of the stereo videos at any instance of time, and subsequently estimating the disparity values for each of the matched points. The disparity values are used to track and classify the matched points through the feature point-tracking unit **106**. The feature point tracking unit **106** tracks the matched feature points by processing subsequent stereo image frames to extract the movement information of each feature point. In accordance with the present embodiment, the feature points are tracked in a group of similar properties of the features. It is recognized that some feature points may also be represented by the same or substantially the same feature vector. The feature vector may be appearance such as e.g. color, edges etc. or the like. It can be challenging to find the correct

correspondences between the multiple feature points having same or substantially the same feature vector, as opposed to finding the correspondence between the groups of feature points between two video frames. For simplicity, the group of feature points that shares similar properties shall herein referred to as "the cloud". As the cloud of feature points are being tracked, instead of the individual feature point, the computational processing needed by the image processing module 104 and the information stored over time can be reduced.

[0023] When a stereo camera or a camera assembly that comprising two cameras disposed on a frame with a fix distance apart, the internal intrinsic parameters of the stereo camera or the camera assembly are known. These internal intrinsic parameters provide distance between the two cameras (or imaging sensors) and the rotation of one camera with respect to the other. Typically, the two cameras are also horizontally aligned with a 3D point projected in a relatively same horizontal line on both images. Thus, for any instant of two images taken by these two cameras, the matching can be done by assuming that any point in one line in one image is correspond to one point in same corresponding line in the other image. Then with these point corresponding in hand, together with the intrinsic parameter known earlier, disparity value of this 3D point from the camera can be computed.

[0024] Through the present invention however, it is cater for stereo camera or the camera assembly that is may not be properly calibrated or misaligned in the cause of operating the same. Accordingly, one line in an image may not always correspond to the same line in the other image. The proposed invention has therefore proposed the optimized matching process in light of the above.

[0025] Each feature point is being classified as either moving points or background points through the feature point classification unit 107. The classification is made up based on the information from the feature point tracking unit 106. The information includes tracking and disparity information. It is desired that the classification be done on each cloud. The moving object extraction unit 108 further group one or more clouds into a moving entity based on the feature point's classification result, tracking information and the disparity information. Accordingly, the event rule unit 109 applies an appropriate rule based on the extraction results. For example, the event rule unit 109 applies an intrusion event rule when the moving object is detected to be an intruder. These rules are predefined and they may be stored in the database 112 for retrieval. The image-processing module 104 then output the video processing results to the display module 110 for visual output and when appropriate, the post detection module 111 is triggered to activate a corresponding alarm based on the processed results.

15 [0026] Feature points that have almost the same or similar appearance to each other will be group as one cloud. For example, all the feature points that are extracted from the subject's shirt are presumably having the same or similar appearances and will therefore be group as one cloud. The appearance that includes color would have certain level of similarity. In the case where the same subject containing more than one colors, for example, wearing a pant with a color different from the shirt, then all feature points extracted from the pant part shall be group as the same cloud, which is a separate cloud from the cloud for the shirt. After monitoring this subject, i.e. formed by the two detected cloud, over subsequent frames, the motion of both clouds may form specific

corresponding patterns and therefore can further be grouped into one cloud defining the subject.

[0027] FIG. 2 illustrates the process flow operationally carries out by the feature point disparity computational unit 105 of FIG. 1 in accordance with one embodiment of the present invention. The feature point disparity unit 105 derives the disparity information of feature points between the two stereo (i.e. left and right images) of the current frame. Briefly, the process flow includes extracting feature points in currently left and right images in step 201; matching feature points in current left and right images in step 202; filtering noises in step 203; computing distance between the respective matched feature points in step 204; and quantizing the distance based on current disparity value in step 205.

[0028] In the step 201, the process starts with extracting feature points right and left images. In one embodiment of the present invention, salient points are extracted as the feature points in the images. These salient points can be corner and edges points that can be detected through any known techniques. For example, Harris corner point detector, scale-invariant feature transform (SIFT), speeded up robust feature (SURF) etc. can be utilized for extracting the desired salient points.

[0029] Once the feature points (i.e. the salient points) are extracted, they are matched between right and left images of the same frame in the step 202. On a stereo camera where two imaging means are aligned to be coplanar, searching and matching can be simplified to within a single or one dimensional (1D) space because the horizontal line in one image can be presumed to be in parallel to the same horizontal line in the other image. In some camera configurations, searching for matching or

corresponding feature points may requires searching in two-dimensional (2D) space, as two imaging means may not be perfectly aligned. It is more so when the imaging means are two independent cameras. One possible approach for example is to perform image rectification between right and left cameras before matching. Such image
5 rectification for processing stereo images is well known in the art. Another desired approach that can be taken is to adaptively refine the 2D searching area for matching, of which no rectification is required. In this adaptive refinery process, previous matching information is used to determine the search area in the current frame.

[0030] In one illustrative example for matching feature points on the right image
10 that correspond to one feature point on the left image, the each point on the whole horizontal lines or rows of the right image of the first frame will be considered as the searching area. Once the matching is done for all feature points, the results will be analyzed for range of possible matching horizontal lines on the right image for searching through, given a point on certain horizontal line in left image is determined.
15 Accordingly, when the next frame is to be searched, the searching area will be constraint within these ranges. That could effectively refine the number of horizontal lines to be search over time.

[0031] Once the feature points matching between the right and left images of the current frame is done, matching noises are removed in the step **203**. Methods such as
20 RANSAC can be used in the step. In the step **204**, the distance or disparity value of these matched feature points is computed. The distance or disparity value provides a relative distance of these points in the real world with respect to focal point of the surveillance camera. Further, in the step **205**, the distance or disparity values are quantized into discrete levels. Accordingly, the feature points extracted from the

right/left image of the current frame can be outputted by the feature point disparity computational unit 105 along with the respective disparity values.

[0032] FIG. 3 illustrates the process flow operationally carries out by the feature point tracking unit 106 of FIG. 1 in accordance with one embodiment of the present invention. The feature point tracking unit 106 receives input from the feature point disparity computational unit 105. The input includes either one of the stereo image (i.e. left or right image) of the current frame and previous frames, the extracted feature points from current and previous frames, and previous tracking information of feature points in previous frames. The output from the feature point tracking unit 106 provides the correspondences between feature points in current frame and previous frames. Briefly, the process includes clustering the feature points in the current frame into clouds in step 301; tracking each cloud in step 302; determining tracking status of the clouds in step 303; matching feature point between the right/left image of the current frame and previous frame in step 304; and redistributing the feature points into the cloud in step 305.

[0033] The feature point tracking unit 106 primarily forms and group the feature points into clouds and track the same accordingly. The process starts with clustering the feature points in current frame into clouds in the step 301. The clustering is done based on the similarity in appearance and proximity within the image of the current frame. Once the feature points are clustered into clouds, the feature point tracking unit 106 tracks the clouds by determining correspondences of all the resultant clouds in current frame with respect to previous frame in the step 302. The correspondences are determined based on similarity of the clouds with reference to the appearance and shape of the feature point distribution. Various correspondences determination methods are

widely available, and they can be desired for the present invention. Therefore, no further detail is provided herewith. In the step 303, a tracking status is being identified for each cloud. The tracking status includes new cloud, existing moving cloud, existing background cloud and existing unknown cloud. A new cloud assigned to one that has
5 no correspondence found between current cloud and any of previous clouds. Current and previous clouds, i.e. from current and previous frames, are referring to clouds in current and previous frame respectively. In the step 304, the feature point distributions (i.e. into clouds) are further refined by feature point matching within certain clouds. The matching results that defines which points in the current frames is matched with the
10 points in the previous frame can be used to redistribute feature point initially assigned to one cloud to another cloud in the step 305.

[0034] FIG. 4 illustrates the process flow of the feature point classification unit 107 of FIG. 1 in accordance with one embodiment of the present invention. The process is adapted for classifying each cloud detected by the feature point tracking unit
15 106. It classifies the clouds into moving objects, background/static objects or unknown objects. The feature point classification unit 107 receives inputs such as the clouds, tracking information and disparity information from the feature point tracking unit 106 for classification. The clouds are classified based on the disparity tracking information of each cloud. Assuming that the camera is moving in constant speed, then the rate of
20 change of the cloud disparity information over subsequent frames would be substantially constant as well. For the cloud that is new, it will not be classified in current frame. The global disparity information for this cloud will be stored. The statistical information such as the mean or mod or the standard deviation of the disparity

distribution with a cloud can be used to represent the global disparity information of the cloud.

[0035] With the above in mind, the process flow starts with estimating a statistical disparity distribution of each cloud (i.e. mean, mpd, standard deviation etc.)

5 in current frame in step 401. Then each of the cloud will be examined whether it is new in step 402. If the cloud is new, the disparity information will be stored in step 403 and the process ends accordingly. Returning to the step 402, if the cloud is determined not being a new cloud, the cloud is further determined whether it is an existing cloud or an unknown cloud in step 404. If it is an existing cloud at the step 404, the feature point
10 classification unit 107 computes the change in disparity of each cloud between the current and previous frame in step 405. Following that, the feature point classification unit 107 further retrieves the acceleration information of the surveillance camera recorded by the accelerometer in step 406. If the acceleration is zero or substantially zero in the step 407, it provides that the surveillance camera is moving at a constant
15 speed. The rate of change of disparity information of each cloud between the current frame and subsequent frames is computed in step 408. At least three frames would be required to compute the rate of change. This computation takes into consideration the changes of disparity information tracked over subsequent frames. If the rate of change is determined to be zero or substantially zero in the step 409, it can conclude that the
20 cloud belongs to background or static object in step 413. If the rate of change is determined to be non-zero in the step 409, it can then conclude that the cloud is a moving object in step 410. Returning to the step 407, the surveillance camera is determined to be accelerating, i.e. not zero acceleration, it can either be accelerating or decelerating. The feature point classification unit 107 further computes the rate of disparity change

over subsequent frames in the step 411. The rate of disparity change of the cloud over the subsequent frames is in conformance with the movement of the surveillance camera in step 412, the cloud shall be classified as background or static object in the step 413. Otherwise, it is classified as moving object in the step 410.

5 [0036] FIG. 5 illustrates the process flow carries out by the moving object extraction unit 108 of FIG. 1 in accordance with one embodiment of the present invention. The moving objection extraction unit 108 receives the clouds and the corresponding classifications from the feature point classification unit 107. In step 501, it merges all the clouds determined belonging to the same object based on the tracking
10 information and changes of disparity collected over certain time period. Through the series of tracking and grouping, the processing units are able to effectively determine moving objects from a stereo surveillance camera, and track the same accordingly.

[0037] In an illustrative example, one object may be defined by two or more independent clouds when the video is processed initially. For example, a human who is
15 wearing red shirt and black pant may be defined by two different clouds since the object comprises two groups of different appearances. After tracking these clouds over subsequent frame, it is expected that the two clouds that have similar tracking information, e.g. moving path, relative position between the two clouds over subsequent frames and tracking disparity information. Thus these similarities can be exploited to
20 assign all these clouds to belong to a same object entity in the step 501. As the entire process illustrated above is performed on the clouds that are classified as moving object, the output provides the group of feature point clouds, which belong to the same moving object.

[0038] While specific embodiments have been described and illustrated, it is understood that many changes, modifications, variations, and combinations thereof could be made to the present invention without departing from the scope of the invention.

Claims

1. A surveillance system having a pair of stereo camera capturing videos having a plurality of frames, each frames includes at least a left and a right images, wherein the stereo camera (101) having an accelerometer (102) attached thereto, the surveillance
5 system comprising:
 - a video acquisition module (103) for acquiring the videos captured by the pair of stereo cameras (101) and acceleration information of the stereo camera;
 - a feature point disparity computational unit (105) adapted for extracting feature points in right and left images of a current frame and matching the feature points
10 between the right and left images, wherein disparity value of each feature points is computed;
 - a feature point tracking unit (106) adapted for clustering the feature points on a current frame into one or more cloud based on their appearance and distance, and determining a correspondence between the clouds of the current frame and a previous
15 frame,
 - a feature point classification unit (107) for classifying the clouds through estimating the statistical disparity distribution of each cloud to identify if the respective clouds are new clouds, existing clouds or unknown clouds; and
 - an object extraction unit (108) for merging the clouds that are determined to be a
20 same object based on the tracking information and changes of disparity of the clouds over time.
2. The surveillance system according to claim 1, further comprising:
 - an event rules unit (109);
 - 25 a display module for displaying overlaid information on live video; and
 - a post detection module to trigger post event actions (111) defined by the event rules unit (109).
3. A method for processing videos captured through a stereo camera, the videos
30 having a plurality of frames, each frames includes at least a left and a right images, wherein the stereo camera (101) having an accelerometer attached thereto, the method comprising:

acquiring videos captured by the pair of stereo cameras (101) and acceleration information of the stereo camera (101) detected through the accelerometer (102);

extracting feature points in the right and left images of a current frame;

5 matching the feature points between the right and left images, wherein disparity value of each feature point is computed;

clustering the feature points on a current frame into one or more clouds based on their appearance and distance;

determining a correspondence between the clouds of the current frame and previous frame;

10 classifying the clouds through estimating statistical disparity distribution of each cloud to identify if the respective clouds are new clouds, existing clouds or unknown clouds; and

merging the clouds that are determined to be a same object based on the tracking information and changes of disparity of the clouds over time.

15

4. The method according to claim 3, further comprising:

filtering noises while matching the feature points between the left and the right images;

computing distances between the matched features points; and

20 quantizing the distance based on a current disparity value range.

5. The method according to claim 4, wherein matching the feature points between the left and right image comprising searching area for finding a best matched feature point based on matching feature points determined in the previous frame.

25

6. The method according to claim 5, wherein searching area for finding the best matched feature point comprising:

matching feature points within an entire image of a first frame; and

30 matching feature points for all subsequent frames based on areas corresponding to feature points matched in a previous frame.

7. The method according to claim 3, wherein clustering the feature points into the one or more clouds follows by the steps comprising:

tracking (302) each cloud by determining the correspondences between clouds detected in current frame and in previous frames;

determining (303) a tracking status of each cloud detected in current frame, wherein the tracking status include new object, existing moving object, existing

5 background and existing undetermined object;

performing feature points matching (304) between one of the left and right images of the current frame and previous frame; and

redistribute (305) the feature point clouds into different or same cloud based on the matching information.

10

8. The method according to claim 3, wherein classifying feature point further comprising:

estimating a statistical disparity distribution of each feature point cloud through mean, mpd or standard deviation;

15 storing the information when the cloud is determined as a new cloud;

computing the change of disparity of each clouds that are tracked between a current frame and a previous frame;

determining the acceleration information of the stereo camera for the instance frame,

20 wherein when the acceleration is substantially zero, the rate of disparity changes of each tracked cloud in the current frame over a subsequent frame is computed, the cloud is determined to be a moving object when the rate of change is not substantially zero, and the cloud is determined to be a static object or background when the rate of change is substantially zero; and when the acceleration is not substantially zero, the rate
25 of disparity change between a current frame and a previous frame is computed, and the cloud is determined to be a static object or background if the rate of change is increasing/decreasing, and the cloud is determined to be a moving object when the rate of disparity change is substantially zero.

1/5

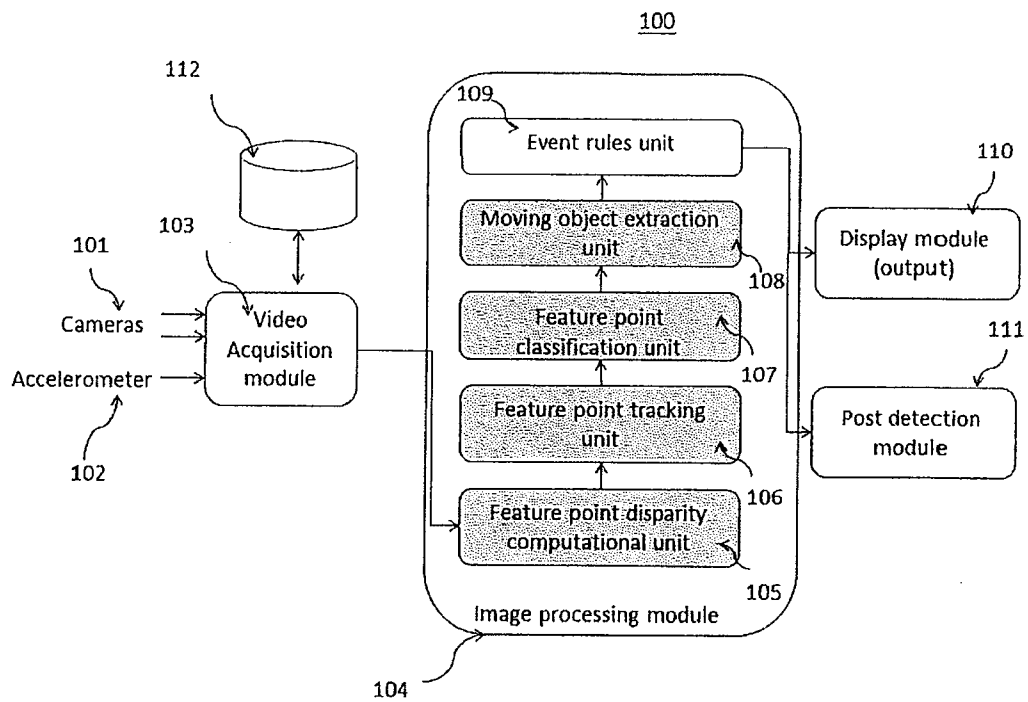


FIG. 1

2/5

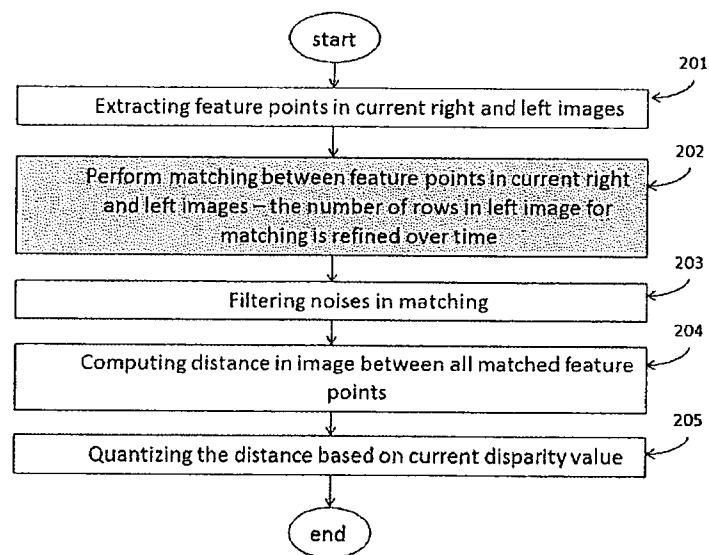


FIG. 2

3/5

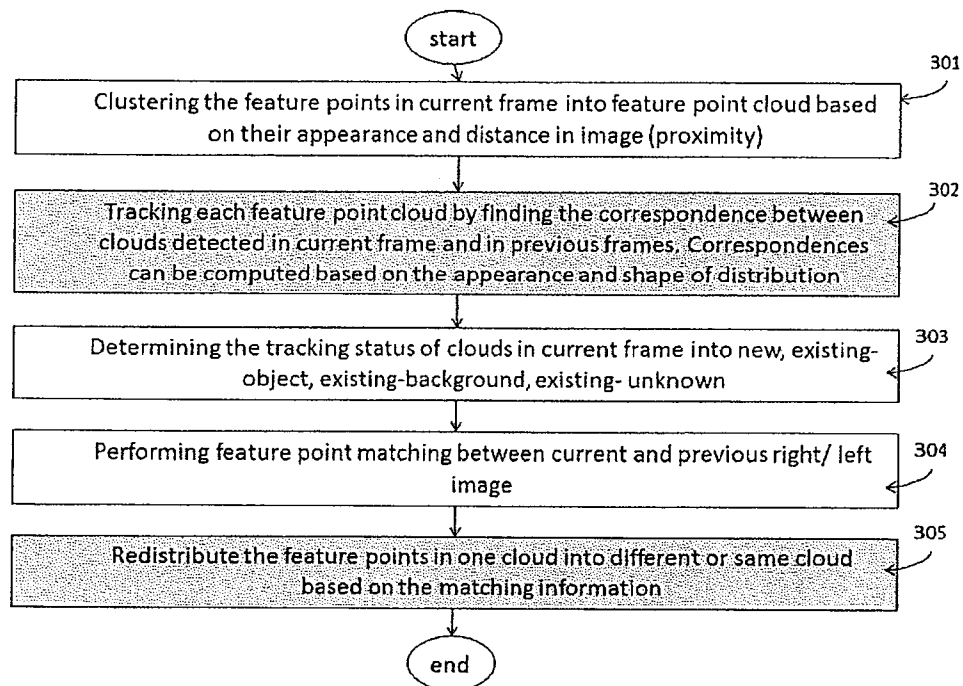


FIG. 3

4/5

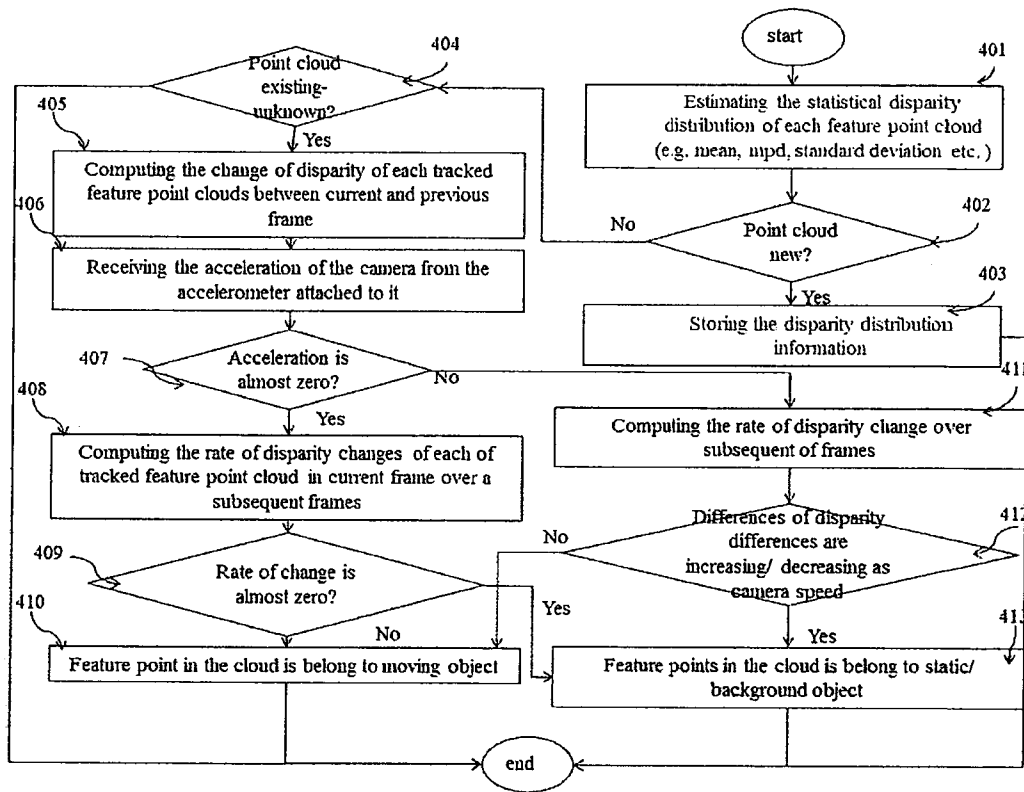


FIG. 4

5/5

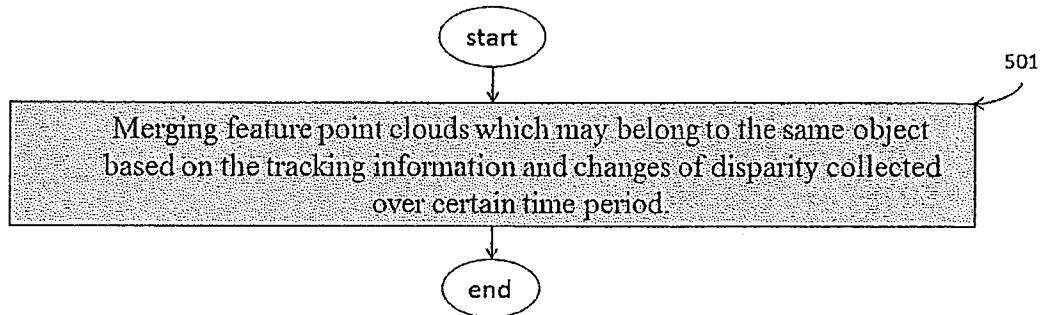


FIG. 5

Moving object extraction unit (108)