



- (51) **International Patent Classification:**
G06T 7/00 (2006.01)
- (21) **International Application Number:**
PCT/MY2013/000264
- (22) **International Filing Date:**
10 December 2013 (10.12.2013)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
PI 2012005333 10 December 2012 (10.12.2012) MY
- (71) **Applicant:** MIMOS BERHAD [MY/MY]; Technology Park Malaysia, 57000 Kuala Lumpur (MY).
- (72) **Inventors:** BINTI KADIM, Zulaikha; c/o Mimos Berhad, Technology Park Malaysia, 57000 Kuala Lumpur (MY). HON, Hock Woon; c/o Mimos Berhad, Technology Park Malaysia, 57000 Kuala Lumpur (MY). BINTI SAMUDIN, Norshuhada; c/o 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).

[Continued on next page]

- (54) **Title:** METHOD FOR CAMERA MOTION ESTIMATION WITH PRESENCE OF MOVING OBJECT

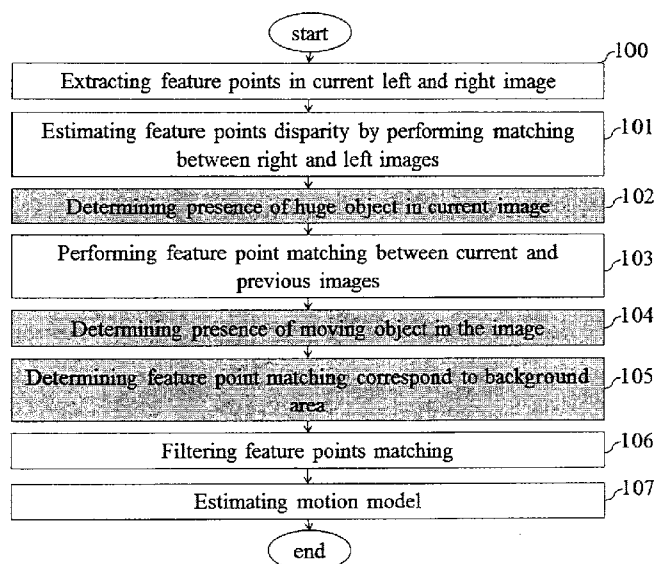


FIG. 1A

(57) **Abstract:** The present invention provides a method of processing sequence of image frames captured through an imaging device. The method comprises extracting feature points from a current frame of the sequence of the image frames; estimating disparity of the feature points; identifying an object from the current frame that occupies a large area of the current frame through the disparity of the feature points; matching feature points between the current image frame and a previous image frame to determine presence of a moving object; filtering out the matched feature points that correspond to the moving object to obtain matched feature points that correspond to background; estimating motion model based on the feature points that correspond to the background.



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))*

Method For Camera Motion Estimation With Presence of Moving Object

Field of the Invention

[0001] The present invention relates to image processing. More particularly,
5 the present invention relates to a method for motion estimation on a camera with presence of moving object.

Background

[0002] In digital image processing, it is well known in the art that camera motion model estimation can be achieved through computing geometric/affine
10 transformation between frames to define how each frames are related to each other in the sequence. Accordingly, it is possible to extract feature points from sequence of frames taken, then match the extracted feature points, then filter the matched feature points that do not correspond the actual motion (outliers) through Random Sample Consensus (RANSAC) technique, and finally use filtered matched feature
15 points to compute transformation or homography matrix that defines the model of the motion of one frame with respect to the other.

[0003] In the aforesaid mentioned method, RANSAC technique can be used to remove matched feature points that do not correspond to the actual motion. A basic assumption used in the technique to remove matched feature points is that the
20 data includes "inliers" whereby the data whose distribution can be explained by sets of model parameters, and "outliers" whereby the data that do not fit the model.

[0004] In addition to this, the data can be subjected to noise. Some of the outliers may come from extreme values of the noise or from erroneous measurements or incorrect hypotheses about the interpretation of data. Further, RANSAC technique also assumes that outlier are usually found in a small amount, i.e. most feature points shall belongs to background, which can be used to estimate the parameters of a model that optimally explains or fits this data.

[0005] In the case when majority of the matched feature points are determined belongs to moving objects, the camera motion may be incorrect. It follows that any image warping will fail, which will in turn fail subsequent process of extracting moving object from the image.

[0006] The problem arises when the majority of matched points are contributed by correct matched points yet incorrect to be used to define the scene motion because they belong to moving objects in the scene, instead of background. This scenario may happen when less texture in background area is lacking of texture and/or a huge object occupies the frames.

[0007] Further, there is also a question on how to check whether there is possible moving object in the image so that the elimination process has to be considered.

Summary

[0008] In one aspect of the present invention, there is provided method of processing sequence of image frames captured through an imaging device. The method comprises extracting feature points from a current frame of the sequence of

the image frames; estimating disparity of the feature points; identifying an object from the current frame that occupies a large area of the current frame through the disparity of the feature points; matching feature points between the current image frame and a previous image frame to determine presence of a moving object;
5 filtering out the matched feature points that correspond to the moving object to obtain matched feature points that correspond to background; and estimating motion model based on the feature points that correspond to the background.

[0009] In one embodiment, the imaging device is a stereo camera adapted for capturing the sequence of image frames, each image frame has a left image and a
10 right image that form a stereo image. In this case, extracting feature points and estimating disparity of the feature points from the current frame comprises extracting feature points of the left image and the right image; and matching feature points between the left image and the right image.

[0010] In another embodiment, identifying the object that occupies a large
15 area of the current frame comprises computing a percentage of feature points with disparity value less than a predefined threshold; estimating a two-dimensional (2D) Gaussian parameters to represent the spatial distribution of matched feature points when the percentage is more than the predefined threshold; concluding that a huge object is present when size of the 2D Gaussian parameters fitted the distribution is
20 more than a size threshold of the image frame size. Otherwise, it can be concluded that there is no huge object present in the current frame.

[0011] In yet another embodiment, determining presence of moving object comprises splitting the matched feature points into groups of conform motion model

vectors by recursively filter the matched feature points using RANSAC technique; and concluding that potential moving object is present in the current image, when at least two major groups of motion vectors exist, otherwise, concluding that no potential moving object present. Data associated to the matched feature points are
5 divided into inliers and outliers through a RANSAC technique, the RANSAC technique is carried out recursively to further divide outliers until no valid outliers are to be divided.

[0012] In yet a further embodiment, determining matched feature points correspond to background area comprises clustering pixels of the current image
10 frame into their corresponding group of conformed matched feature point; computing the percentage of image area correspond to each group cluster; and group of conformed matched feature point with the largest area in the image is deemed as the background points.

Brief Description of the Drawings

15 [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;

[0014] FIG. 1A illustrates a process for camera motion estimation in accordance with one embodiment of the present invention;

20 [0015] FIGs. 1B and 1C exemplify the image processing through the method provided the present invention;

[0016] FIG. 2 illustrates the process for determining presence of huge object in accordance with one embodiment of the present invention;

[0017] FIG. 3 illustrates a process of determining a presence of moving object in frame sequence in accordance with one embodiment of the present invention; and

[0018] FIG. 4 illustrates the process for determining matched feature points correspond to background area in accordance with one embodiment of the present invention.

Detailed Description

10 [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 the illustrated device, and such further applications of the principles of the invention as illustrated therein being contemplated as would

15 normally occur to one skilled in the art to which the invention relates.

[0020] The present invention provides a method for eliminating unwanted matched feature points (i.e. noise) before geometric or affine transformation or homography matrix is computed in camera motion model estimation. Such unwanted matched feature points are typically caused by the presence of moving

20 object in the image. The method includes checking for present of moving/huge object(s) in image frames through distribution of disparity information and recursive RANSAC or Maximum Likelihood Estimation Sample Consensus (MLESC);

selecting sets of matching points that correspond to background with high confidence instead of moving object by integrating area-based and feature-based in finding correspondences between subsequent image frames to define potential feature points on moving objects or background; and eliminating huge object in the
5 image using disparity information.

[0021] In another aspect, the present invention provides a method for obtaining correct camera motion model estimation by filtering unwanted matched feature points used in estimation due to huge objects and moving objects in the image frames. The motion estimation includes computing geometric or affine
10 transformation or homography matrix between subsequent frames to define how each frame is related to another. **FIG. 1B** exemplifies the step of computing affine transformation between two frames to identify the relationship between the two consecutive frames. The camera motion model estimation includes extracting feature points, matching feature points, filtering matched feature points through
15 RANSAC/MLESAC to remove the matched feature points that do not correspond to the actual motion object, i.e. outliers, and using matched feature points to compute transformation or homography matrix that defines the model of the motion of one frame with respect to the other.

[0022] **FIG. 1A** illustrates a process for camera motion estimation in
20 accordance with one embodiment of the present invention. The process can be implemented on a system that acquires video or image sequences from an imaging device. The imaging device may be a stereo camera adapted for capturing stereo images. In particular, the stereo is able to capture video that contains sequence of

image frames, each image frame contains a stereo image that made up of at least a left and a right image. The process acquires a sequence of frames of stereo images and processes the same in accordance with the embodiment of the present invention, and outputs a homography matrix or an affine transformation matrix that defines
5 how the imaging device is moved from previous to current state. In this document, this matrix will be referred as transformation matrix for simplicity. To obtain the transformation matrix, it is important to get matching feature points that belong to background on the frames. Unwanted matched feature points pairs (of the stereo images) that belong to objects, such as huge objects, based on disparity information
10 of current frames shall be remove.

[0023] Accordingly, the process comprises extracting feature points in stereo image of a current frame at step 100; estimating feature points' disparity through matching the same between the left and right images of the stereo image at step 101; determining presence of huge object in the current frame at step 102; matching
15 feature points between the current frame and a previous frame at step 103; determining presence of moving object in the current frame at step 104; determining matched feature points that correspond to background of the current frame at step 105; filtering matched feature points at step 106; and estimating motion model of the imaging device at step 107.

20 [0024] Returning to the step 100, the feature point extraction is carried on both left and right images of the current frame. It is understood to a skilled person that any feature point extraction techniques or methods can be adapted herein and therefore, no further information is provided herewith. Once the feature points are

extracted, the process estimate the feature point disparity on each feature point by matching between the left and right images of the current frame in the step **101**. The feature point disparity estimation also determines a distance between the matched feature points and the imaging device in 3D space. Based on the feature point disparity information, in the step **102**, the process determines the presence of huge object in current frame, if any. Following that, the process further matches feature points between the current frame and the previous frame in the step **103**. In the step, all feature points from the right or left image can be used. The matched feature points between the current frame and the previous frame is then used as the basis for determining presence of moving object in sequence of frames in the step **104**. Subsequently, the matched feature points that correspond to the background can be determined by taking the remaining matched feature points that are not regarded as the moving object as the background feature points in the step **105**. The matched feature points that belong to the moving object and huge object are further filtered out in the step **106** for estimating the imaging device motion in the step **107**.

[0025] **FIG. 2** illustrates the process for determining presence of huge object in accordance with one embodiment of the present invention. The process can be adapted in the step **102** of FIG 1. As provided in the process of **FIG. 1**, the feature points of the left and right images and their respective disparity information are already extracted for processing herein. Basically when huge object presents in the image sequence, most of image area is covered by the huge object, instead of background. Besides the object's physical size, an object located too close to the capturing area of the imaging device will also appear as huge object in the frame sequence. When the object occupies more than half of the image of the frame, the

background area is hidden behind thus very little features can be extracted from the background area. In that case, the process may choose to ignore this frame, thus no transformation matrix will be computed. The detected object(s) occupies less half of the image of the frame, a transformation matrix will be computed once the matched
5 feature points that correspond to the detected object will be removed. Based on the above, in the step **200**, the process computes the percentage of matched feature points with disparity value less than a preset threshold. As mentioned, this step is used to determine whether the object is too near to the imaging device or physically too big by itself. In **FIG. 2**, the preset threshold is set at 50%. This threshold can be
10 determined and preset before the process is carried out. Depending on a nearest acceptable distance of object from the imaging device, such threshold value can be changed accordingly when desired. When the percentage is less than 50% (or not larger than 50%) at the process concludes that there is no huge object in **205**. When the percentage is larger than 50%, the process continues to estimate two-
15 dimensional (2D) Gaussian parameters that represent spatial distributions of the matched feature points with disparity less than threshold in step **202**. The 2D Gaussian parameters include mean x and y ; (x_0, y_0) , which indicate the center of the distribution; and the σ_x and σ_y which indicate how the distribution is spreading in x and y directions respectively. These parameters shall then be used to define how
20 much has the huge object occupies the image or relative size of the object in the image. Similarly, a size threshold can be preset as 50% for example. When the size of the object is determined to be more than the size threshold, the process concludes that the object is a huge object in step **204**, which otherwise (i.e. less than 50%), it

shall conclude that the object present in the image of the frame is not a huge object in the step 205.

[0026] FIG. 3 illustrates a process of determining a presence of moving object in frame sequence in accordance with one embodiment of the present invention. This process can be adapted in the step 104 of FIG. 1. This process is provided to determine whether there are any moving objects present in a current frame. The process monitors the number of groups of similar motion vectors in the frame. Group of similar motion vectors may belong to a same object or a same background. Thus, when more than one group exist, it can be concluded that there is a likelihood of moving object(s) present in the frame. In the further step (i.e. the step 105 of FIG. 1), the matched feature points that correspond to these moving objects will be disregarded to obtain the features point that correspond to background.

[0027] As shown in FIG. 3, the process comprises splitting the matched feature points into groups of conformed motion vectors by recursively filtering the matched feature points in step 300. This can be achieved by using a RANSAC-based technique. It is known that RANSAC technique can be used to remove the matched feature points that do not correspond to an actual motion. These matched points are herein referred as the outlier. The actual motion contains a group of motion vectors that correspond to a majority. Accordingly, perform recursive the RANSAC-based technique to delineate matched points to their conform motion. The RANSAC-based technique includes dividing the matched feature points into two groups, i.e. inliers and outliers. The RANSAC-based technique will be carried

out to process the outliers, which are derived through the previous RANSAC process. The outliers will then be further divided into inlier and outliers through the subsequent RANSAC process. The recursive steps will be repeated until no valid outliers are to be divided. Outliers are considered valid to be divided if the numbers
5 is at least more than 20% of total number of matched points. Then, the process determines the number of major group of motion vectors at step **301**. A group is considered as a major group when the number of matched feature points in the group is more than 20% of total matched feature points or at least 8 matched points, whichever is more. If at least 2 major groups exist in step **302**, then it is concluded
10 that there is potential moving objects in the current frame. If not, it is concluded otherwise in step **304**.

[0028] **FIG. 4** illustrates the process for determining matched feature points correspond to background area in accordance with one embodiment of the present invention. This process can be adapted in the step **105** of **FIG. 1**. This process is
15 provided for delineating matched points that correspond to background from moving objects, so that only these points are considered in transformation matrix computation to determine the imaging device's movement. This process may be required only if it is concluded that there is a potential moving object in the image. In the step **400**, image pixels of the image are clustered into corresponding group of
20 conformed matched feature points. In this case, any known clustering techniques can be adapted herein. The major group's properties (e.g. appearance, centroid, average motion direction etc.) can be used for indicating whether other pixels belong to their group. Then the percentage of image area correspond to each group segment is computed in step **401**. This can be computed by dividing total number of pixels

correspond to the group divided by total number of pixels in the image. The group with largest area will be regarded as background points in step 402. Accordingly, all other matched points not correspond to background can be filtered as the step 106 of FIG. 1, leaving only matched points correspond to background to be used in
5 computing the camera motion model as the step 107.

[0029] 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 method of processing sequence of image frames captured through an imaging device, the method comprising:
 - extracting feature points from a current frame of the sequence of the image
 - 5 frames;
 - estimating disparity of the feature points;
 - identifying an object from the current frame that occupies a large area of the current frame through the disparity of the feature points;
 - matching feature points between the current image frame and a previous
 - 10 image frame to determine presence of a moving object;
 - filtering out the matched feature points that correspond to the moving object to obtain matched feature points that correspond to background;
 - estimating motion model based on the feature points that correspond to the background.
 - 15
2. The method according to claim 1, wherein the imaging device is a stereo camera adapted for capturing the sequence of image frames, each image frame has a left image and a right image that form a stereo image.
- 20 3. The method according to claim 2, wherein extracting feature points and estimating disparity of the feature points from the current frame comprising:
 - extracting feature points of the left image and the right image; and
 - matching feature points between the left image and the right image.
- 25 4. The method according to claim 1, wherein identifying the object that occupies a large area of the current comprises:
 - computing a percentage of feature points with disparity value less than a predefined threshold;
 - estimating a two-dimensional (2D) Gaussian parameters to represent the
 - 30 spatial distribution of matched feature points when the percentage is more than the predefined threshold;

concluding that a huge object is present when size of the 2D Gaussian parameters fitted the distribution is more than a size threshold of the image frame size.

5 otherwise, concluding that there is no huge object present in the current frame.

5. The method according to claim 4, further comprising processing the next frame when a huge object is present in the current frame.

10 6. The method according to claim 1, wherein determining presence of moving object comprises

splitting the matched feature points into groups of conform motion model vectors by recursively filter the matched feature points using RANSAC-based technique;

15 concluding that potential moving object is present in the current image, when at least two major groups of motion vectors exist, otherwise, concluding that no potential moving object present.

20 7. The method according to claim 6, wherein data associated to the matched feature points are divided into inliers and outliers through a RANSAC-based technique, the RANSAC technique is carried out recursively to further divide outliers until no valid outliers are to be divided.

25 8. The method according to claim 1, wherein determining matched feature points correspond to background area comprising

clustering pixels of the current image frame into their corresponding group of conformed matched feature point;

computing the percentage of image area correspond to each group cluster;
and

group of conformed matched feature point with the largest area in the image
is deemed as the background points.

1/5

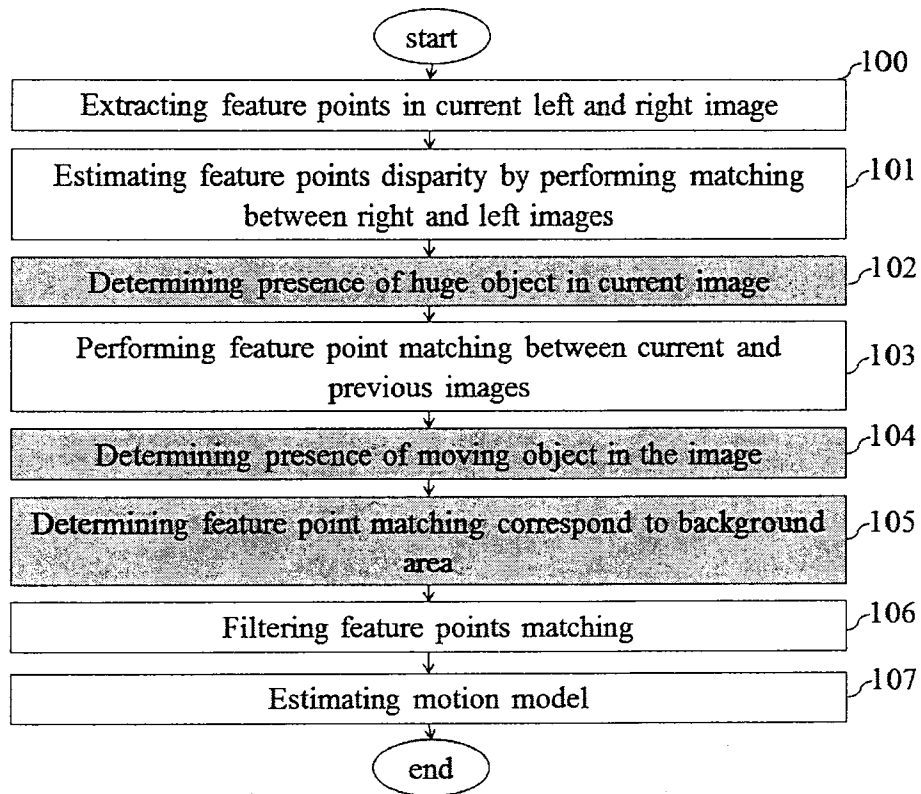


FIG. 1A

2/5

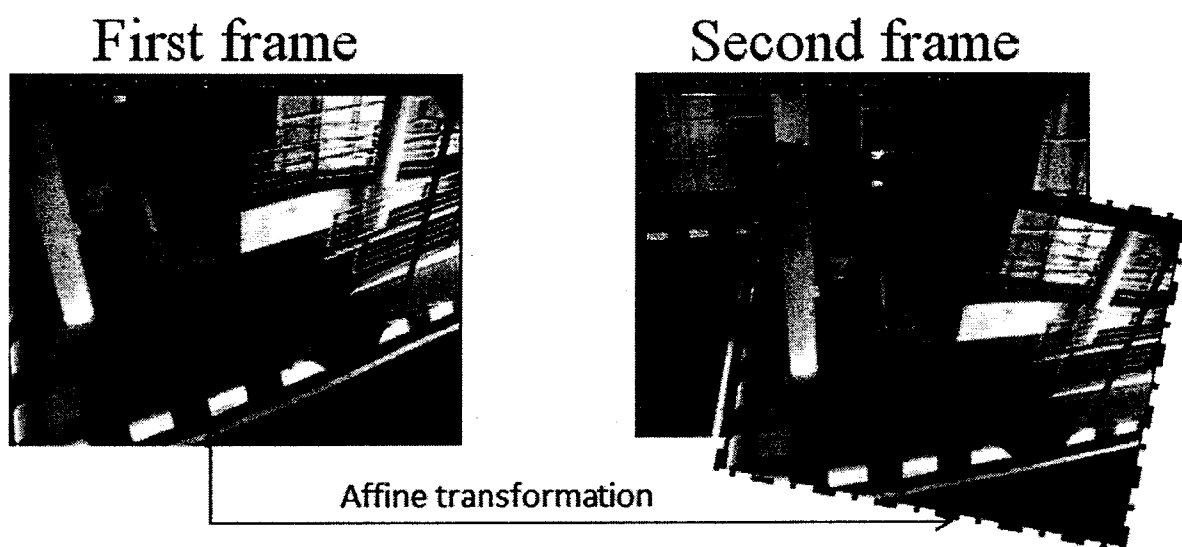


FIG. 1B

3/5

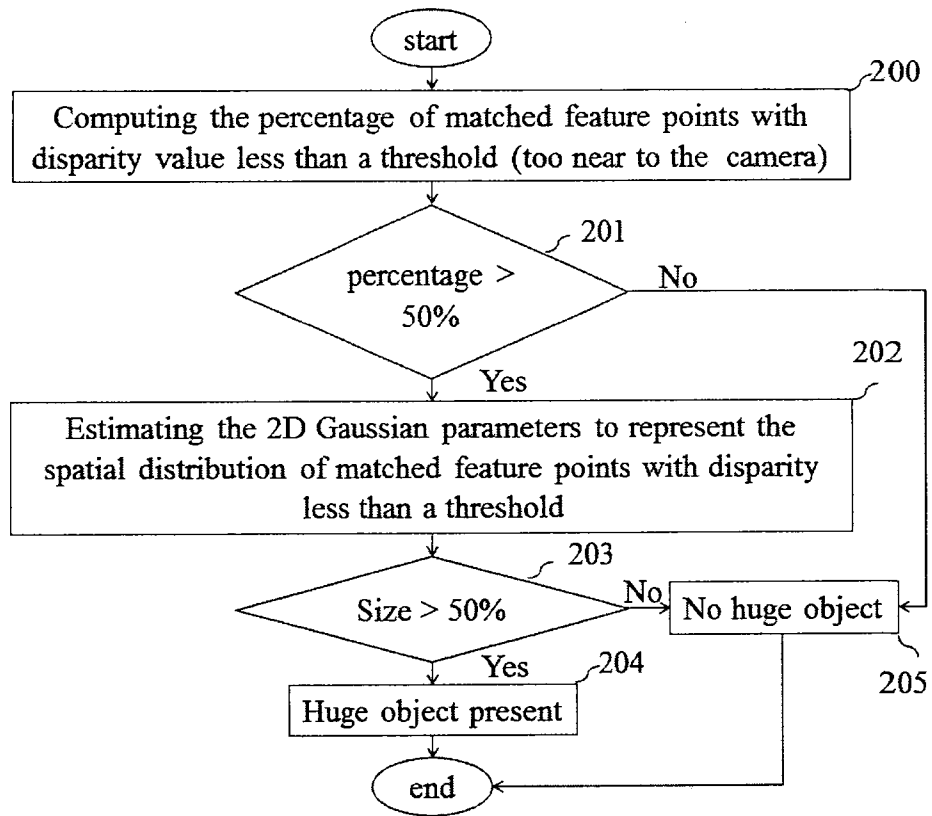


Fig. 2

4/5

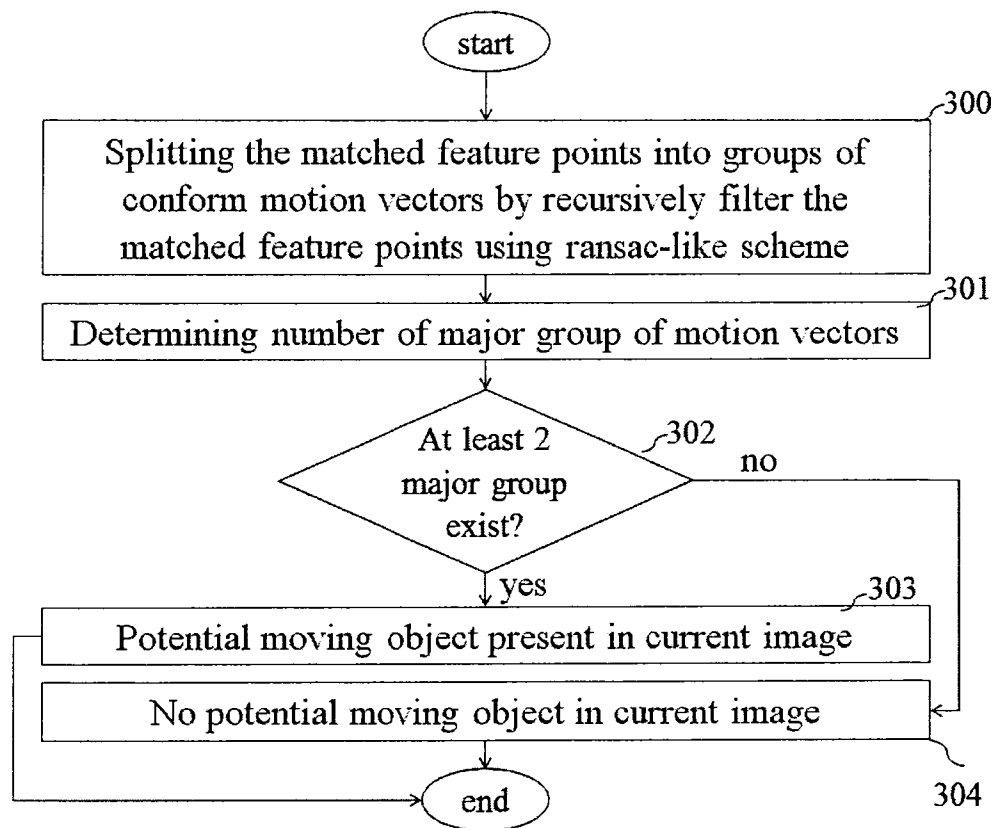


Fig. 3

5/5

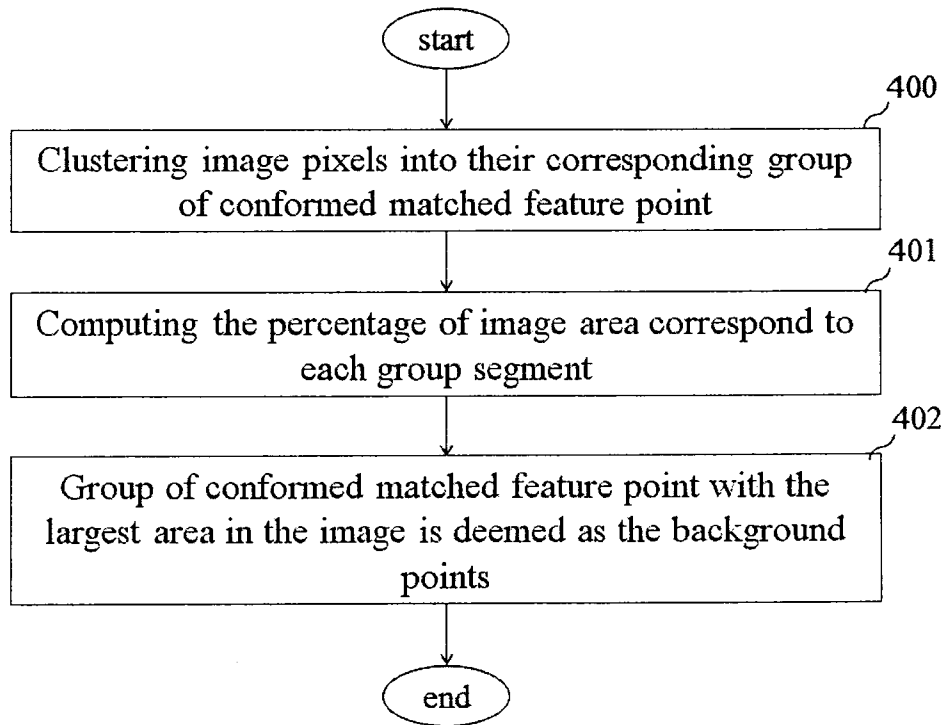


FIG. 4