



- (51) **International Patent Classification:**
G06T 7/20 (2006.01)
- (21) **International Application Number:**
PCT/MY2013/000256
- (22) **International Filing Date:**
12 December 2013 (12.12.2013)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
PI 2012005403 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; 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). NORSHUHADA, Binti Samudin; 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).

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:** SYSTEM AND METHOD FOR MOVING OBJECT EXTRACTION USING STEREO VISION

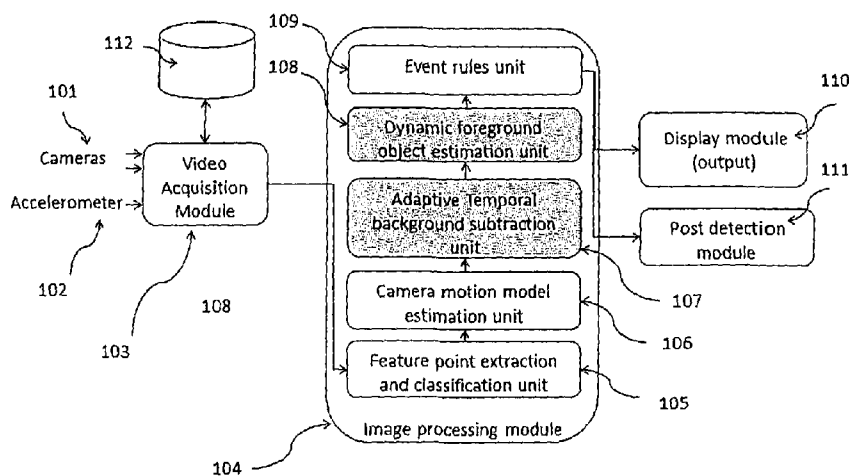


FIG. 1

(57) **Abstract:** The present invention provides a system and method for processing sequence of image frames captured through an imaging device, the imaging device includes an accelerometer. The system and method comprise an adaptive temporal background subtraction unit and a dynamic foreground object estimation unit for evaluating whether each pixel is either a moving object pixel or a non-moving object pixel based on the classification of the pixels between the first motion map and the second motion map generated thereon.

System and Method for Moving Object Extraction Using Stereo Vision

Field of the Invention

[0001] The present invention relates to image processing. More particularly, the present invention relates to a system and method for extracting and detecting moving
5 object through stereovision.

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 point either it is caused by the actual moving object or by the motion of the cameras.

10 [0003] 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

15 [0004] It is an objective of the present invention to provide a system and method for extracting moving objects from sequences of images (i.e. videos) captured on moving stereo cameras in real-time. The extracted and identified moving objects can then be further analysed through any surveillance applications such as tracking,
20 intrusion detection etc. Through the system and method of the present invention, each of the captured images are segmented into moving objects and background.

[0005] In one aspect, the present invention integrates an area-based detection and a feature-based detection extract moving objects from the images. In particular, the sequence of images, i.e. the videos, is captured through a stereo camera that comprises two imaging devices. The stereo camera may also equip with accelerometer for
5 detection movement. Such system and method can also be implemented on any mobile devices, such as smart phones, having a camera and accelerometer integrated thereon. Accordingly, the motion of the stereo camera or the mobile device can be adaptively used to determine the background from the images, which in turn provides a reference for extracting moving object(s) through area-based method.

10 **[0006]** In another aspect of the present invention, there is provided a method for processing sequence of image frames captured through an imaging device, the imaging device includes an accelerometer. The method comprises extracting feature points from the image frames; classifying each feature point as belongs to either a moving object or a non-moving object; receiving input from the accelerometer to estimate the
15 transformations between a current and a previous frames; determine relative position of current frame with respect to previous frames; extracting previous frames that relate to a relative position of the current frame; selecting previous frames that overlapped with the current frame; performing background subtraction between the current frame and the related previous frames; forming a first motion map by fusing results of the background
20 subtractions between the current frame and the respective extracted previous frames; updating precious frames list and linkage; forming a second motion map that maps out pixels as moving object pixel and non-moving object pixel based on the feature points' classification; examining each pixel of each image frame. When the examined pixel is located in an uncovered area of the first motion map, classifying the pixel as the same

classification as defined on the second motion map; otherwise, classifying the pixel as belonging to a moving object when the pixel is classified as a moving object in both the first motion map and the second motion map; or otherwise, classifying the pixel as belonging to a non-moving object.

5 **[0007]** In one embodiment, the imaging device is a stereo imaging device capturing the sequence of image frames, each image frame forms by a stereo image of a left image and a right image, the feature points are respectively extracted from the left image and the right image. The classification of each feature point may be derived based on matching corresponding feature points between the left and the right image.

10 **[0008]** In another embodiment, the method further comprises identifying an uncovered area on the first motion map that is not covered by any of the related previous frames. Yet, the method may further segment the image frame based on the classified feature points to generate the second motion map.

15 **[0009]** In a further aspect of the present invention, there is also provide a surveillance system having an imaging device for capturing sequence of image frames, wherein the imaging device having an accelerometer. The surveillance system is adapted to carry out the aforesaid method. the surveillance system comprises an adaptive temporal background subtraction unit for carrying out the background subtraction to generate the first motion map, wherein the first motion maps out whether
20 each pixel belongs to a moving object or a non-moving object; and a dynamic foreground object estimation unit for generating a second motion map that maps out whether each pixel belongs to a moving object or a non-moving object based on feature points classification. Each pixel is being identified as either a moving object pixel or a

non-moving object pixel based on the evaluation of the classification of the pixel between the first motion map and the second motion map.

Brief Description of the Drawings

[0010] 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;

[0011] **FIG. 1** illustrates a schematic block diagram a surveillance system in accordance with one embodiment of the present invention;

[0012] **FIG. 2A** illustrates the process adaptive temporal background subtraction in accordance with one embodiment of the present invention

[0013] **FIG. 2B** exemplifies schematically the processing on sequence of frames through the adaptive temporal background subtraction

[0014] **FIG. 3** illustrates a process of extracting foreground object in accordance with one embodiment of the present invention; and

[0015] **FIG. 4** illustrates a process for segmenting image in accordance with one embodiment of the present invention.

Detailed Description

[0016] 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 normally occur to one skilled in the art to which the invention relates.

[0017] The present invention provides a surveillance system and method to differentiate whether a pixel in an image captured on a non-static stereo camera is a moving object pixel or a static/background pixel. Accordingly, image segmentation can be realized through a plurality of cameras or a stereo camera that has accelerometer attached thereto. The method is based on an integration of area-based and feature-based analysis. The method may further integrate with feature point detection and classification is performed to refine a final output of area-based detection, and to remove errors due to misalignment of current and background image.

[0018] **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 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. Preferably, the surveillance cameras **101** equipped or attached with accelerometer **102** for detecting the movements of the respective surveillance cameras **101**. 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 database module 112 also stores presets and user configurations information thereon. 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.

[0019] In the present embodiment, a feature-based detection method is adapted to sparse feature-point classification to pixel points as moving object or background/static object. Further, an area-based detection method would give dense pixel classification into either moving pixels or not. Under these methods, current image of the video sequence will be compared with a previous one or sets of reference images to find the moving points or pixels within the image. In this case, the decision for the classification can be done on the pixels within the overlapping regions between current and the reference images. To achieve the above, as shown in **FIG. 1**, the image-processing module 104 further comprises a feature point extraction and classification unit 105, camera motion model estimation unit 106, an adaptive temporal background subtraction unit 107, a dynamic foreground object estimation unit 108, and an event rule unit 109.

[0020] The feature point extraction and classification unit 105 is adapted to process the images to classify feature points extracted from the images as either a moving object or background/static object. The camera motion model estimation unit 106 is adapted to estimate transformations between current and previous frames by

aligning the images with respect to a common coordinate system, which will be required for later processing. The common coordinate system is referring to a reference coordinate system that common to both images that are being compared. The current image frame, for example, may be suitable to be served as the common coordinate system. The camera motion model estimation unit **106** operationally performs image matching between a current and previous frames to determine the movement of the imaging device. The information regarding the movement estimation is supplied to the adaptive temporal background subtraction unit **107** accordingly. The adaptive temporal background subtraction unit **107** uses a reference image from the camera motion model estimation unit **106** to align with the current frame, a background subtraction is then carried out on these two frames. The reference image will be aligned to the current frame's coordinate system, i.e. common coordinate system. Through the background subtraction, the pixels can be identified as either belongs to a moving object or a static/background object. Each pixel is being tagged with the results on the background subtraction. Subsequently, the dynamic foreground extraction unit **108** combines the outputs from feature point extraction and classification unit **105** and adaptive temporal background subtraction unit **107** to produce a final image segmentation result. The detected moving object through the segmentation will be passed to the event rule unit **109** to analyze and determine if the moving object fits any specific event that requires attention. The event analysis will be carried out based on sets of preset rules. The rules may further define the necessary event to trigger as required. The display module **110** is provided to output sequence of the image frames captured by the cameras or the processed images or both in a superimposed manner for viewing visually. The post detection module **111** triggers alarm for post event action as required.

[0021] **FIG. 2A** illustrates the process adaptive temporal background subtraction in accordance with one embodiment of the present invention. The process is implementable on the adaptive temporal background subtraction unit **107** of **FIG. 1**. Briefly, the process provides an area-based detection method to classify the image
5 pixels into moving pixels and non-moving pixels. In one example, the area-based method includes a background subtraction, whereby a current image of the sequence is compared with a reference image to find the difference between the images. The results of the background subtraction, i.e. the differences, represents pixels of the moving object. The process starts by receiving the input from accelerometer in step **201**. The
10 acceleration information from the accelerometer is used to determine relative positions of current frame with respect to a previous or reference frames in step **202**.

[0022] **FIG. 2B** exemplifies schematically the processing on sequence of frames through the adaptive temporal background subtraction. The current frame, t , can either appear on a right frame, $t-3$, or left frame, $t-2$, of previous frames. Noted that the right
15 frame and the left frame of previous frames are not to be confused with the left image and right image of a stereo image. The left frame and the right frame are taken by the imaging device at a physical location of the current camera position with respect to the previous camera position. After a set of previous frames in relation to current frame is determined and extracted in step **203**, the camera motion can be estimated in step **204**
20 by aligning the current frame with the extracted previous frames to determine overlapping regions between the frames. Then only the previous frames that are overlapped with current image are selected in step **205** for background subtraction. In this step, all other non-overlapped image frames will be eliminated. These overlapped frames are then be warped with respect to current frame in step **206**. Background

subtraction is then carried out to each of these warped images and current image in step 207. Say there are three previous frames that are overlapped with current frame, then there will be three background subtraction with each previous frames respectively. The resultant motion map from each background subtraction steps is then fused in step 208 to derive a final motion map. To fuse different maps, one may use rules such as if the majority of motion map indicate current pixel is a moving pixel, then it will denoted as moving pixel in final motion map. After final motion map has been determined, then the area within motion map, which are not covered by any previous frames will be determined 209. This area will be denoted as an uncovered area. Finally, the current image content is used to update the previous images in the step 210. The overlapping regions found in 204 are also used to determine which pixels in previous frames are to be updated. Only the current pixels that not identified as moving object are updated in the previous images. Output from this processing unit is the motion map; a binary image which value '1' indicates moving pixel and '0' indicates background or static object. For easy reference, the output motion map is denoted herewith as motion map A.

[0023] FIG. 3 illustrates a process of extracting foreground object in accordance with one embodiment of the present invention. The process is implementable on the dynamic foreground object extraction unit 108 of FIG. 1. In this process, another motion map, motion map B, will be generated, and is integrated with the motion map A generated by the adaptive temporal background subtraction unit 107 to produce a final motion map. The motion map B is generated based on the classified feature points from feature point extraction and classification unit 105. The process starts with receiving feature points and their corresponding classification information from the feature point extraction and classification unit 105. These classification results are used to segment

the image into moving pixels and non-moving pixel. The segmentation outputs the motion map **B**. In step **303**, the dynamic foreground object extraction unit **108** receives the motion map **A** from the adaptive temporal background subtraction unit. Following that, each pixel in the image is examined to determine whether the pixel belongs to moving object or static/background object in step **304**. In step **305**, it is check if the pixel is within an uncovered area in motion map **A**. If the pixel is within the uncovered area in motion map **A**, the pixel is classified as the same classification (i.e. moving object or non-moving object) as in the motion map **B** in step **308**. The classification result is then output for the pixel. If the pixel is not within the uncovered area in the motion map **A**, i.e. within covered area, in step **305**, the process further determined if that pixel is classified as moving pixel in both motion map **A** and motion map **B** in step **309**. If the pixel is classified as moving pixel in the motion map **A** and the motion map **B** in the step **309**, the pixel is classified as belong to a moving object at step **311**, otherwise, the pixel is identified as static/background pixel at step **310**.

[0024] **FIG. 4** illustrates a process for segmenting image in accordance with one embodiment of the present invention. The segmentation is adaptable in the step **302** of **FIG. 3**. The segmentation is carried out based on classified feature points. The process requires to input the feature points and their label to output the motion map **B**, whereby each of the pixels in motion map indicates whether the pixel belongs to moving pixel or not. The process starts with receiving feature point or feature point clouds and their respective classification information in step **401**. In step **402**, all classified feature points are being mapped on current frame. All pixels correspond to classified feature point are labelled in step **403**. Background or static feature points are labelled with zero (0). Each pixel will be examined for similarities with respect to its k-nearest feature point

neighbors classified as moving points and background points in step **404** and labelled the same accordingly. The step **404** is provided for labeling the remaining pixels that have not yet been labeled (i.e. as moving objects or background). To label these pixels, their respective k-nearest labelled feature points are examined. It computes the
5 similarity between the examined pixel with all of its labelled feature point (within k-nearest pixels) and the examined pixel is labelled the same as the most similar labelled feature point. Thus, if its most similar feature point is the background, the examined pixel is labelled as background too, and vice versa.

[0025] The computations of the similarity is carried out recursively until all pixels are
10 labelled. The labelling process starts with pixels surrounding feature point. For example, k is equal to 5. The k-value can be preset on the system by the operator as appropriately. Then for a pixel under consideration, 5 nearest classified feature points will be examined and the similarity between pixels under consideration and these 5 neighbours pixel will be computed. Accordingly, the respective pixels under
15 consideration are labelled as the one most similar with it corresponding feature point in step **406**. Then a filling hole algorithm be applied to the resultant map, to fill up the pixels with missing label in step **406** and output a motion map **B**.

[0026] While specific embodiments have been described and illustrated, it is understood that many changes, modifications, variations, and combinations thereof
20 could be made to the present invention without departing from the scope of the invention.

Claims

1. A method for processing sequence of image frames captured through an imaging device, the imaging device includes an accelerometer, the method comprising:
- extracting feature points from the image frames;
 - 5 classifying each feature point as belongs to either a moving object or a non-moving object;
 - receiving input from the accelerometer to estimate the transformations between a current and a previous frames;
 - determine relative position of current frame with respect to previous frames;
 - 10 extracting previous frames that relate to a relative position of the current frame;
 - selecting previous frames that overlapped with the current frame;
 - performing background subtraction between the current frame and the related previous frames;
 - forming a first motion map by fusing results of the background subtractions
 - 15 between the current frame and the respective extracted previous frames;
 - updating previous frames list and linkage;
 - forming a second motion map that maps out pixels as moving object pixel and non-moving object pixel based on the feature points' classification;
 - examining each pixel of each image frame,
 - 20 wherein when the examined pixel is located in an uncovered area of the first motion map, classifying the pixel as the same classification as defined on the second motion map;
 - otherwise, classifying the pixel as belonging to a moving object when the pixel is classified as a moving object in both the first motion map and the second motion map;
 - 25 or otherwise, classifying the pixel as belonging to a non-moving object.
2. The method according to claim 1, wherein the imaging device is a stereo imaging device capturing the sequence of image frames, each image frame forms by a stereo image of a left image and a right image, the feature points are respectively
- 30 extracted from the left image and the right image, and wherein the classification of each feature point is derived based on matching corresponding feature points between the left and the right image.

3. The method according to claim 1, wherein examining each pixel of each image frame comprises
- 5 computing similarity of k-nearest surrounding pixels of each classified feature point;
- classifying each pixel as either moving pixel or non-moving pixel based on the computed similarity; and
- examining the all of the pixels recursively until all pixels are classified.
- 10 4. The method according to claim 1, further comprising identifying an uncovered area on the first motion map that is not covered by any of the related previous frames.
5. The method according to claim 1, further comprising segmenting the image frame based on the classified feature points to generate the second motion map.
- 15 6. A surveillance system having an imaging device for capturing sequence of image frames, wherein the imaging device having an accelerometer, the surveillance system is adapted to carry out the method of claim 1, the surveillance system comprising:
- 20 an adaptive temporal background subtraction unit for carrying out the background subtraction to generate the first motion map, wherein the first motion maps out whether each pixel belongs to a moving object or a non-moving object; and
- a dynamic foreground object estimation unit for generating a second motion map that maps out whether each pixel belongs to a moving object or a non-moving object
- 25 based on feature points classification,
- wherein each pixel is being identified as either a moving object pixel or a non-moving object pixel based on the evaluation of the classification of the pixel between the first motion map and the second motion map.
- 30 7. The surveillance system according to claim 6, wherein the imaging device is a stereo imaging device capturing the sequence of image frames, each image frame forms by a stereo image of a left image and a right image, the feature points are respectively

extracted from the left image and the right image, and wherein the classification of each feature point is derived based on matching corresponding feature points between the left and the right image.

- 5 8. The surveillance system according to claim 6, wherein each pixel of each image frame is examined recursively by the dynamic foreground object estimation unit until all pixels are being classified to compute similarity of k-nearest surrounding pixels of each classified feature point; and classifying each pixel as either moving pixel or non-moving pixel based on the computed similarity.

10

9. The surveillance according to claim 6, wherein the adaptive temporal background subtraction unit is operably to identify an uncovered area on the first motion map that is not covered by any of the related previous frames.

- 15 10. The surveillance according to claim 6, wherein the dynamic foreground object estimation unit is operably to segment the image frame based on the classified feature points to generate the second motion map.

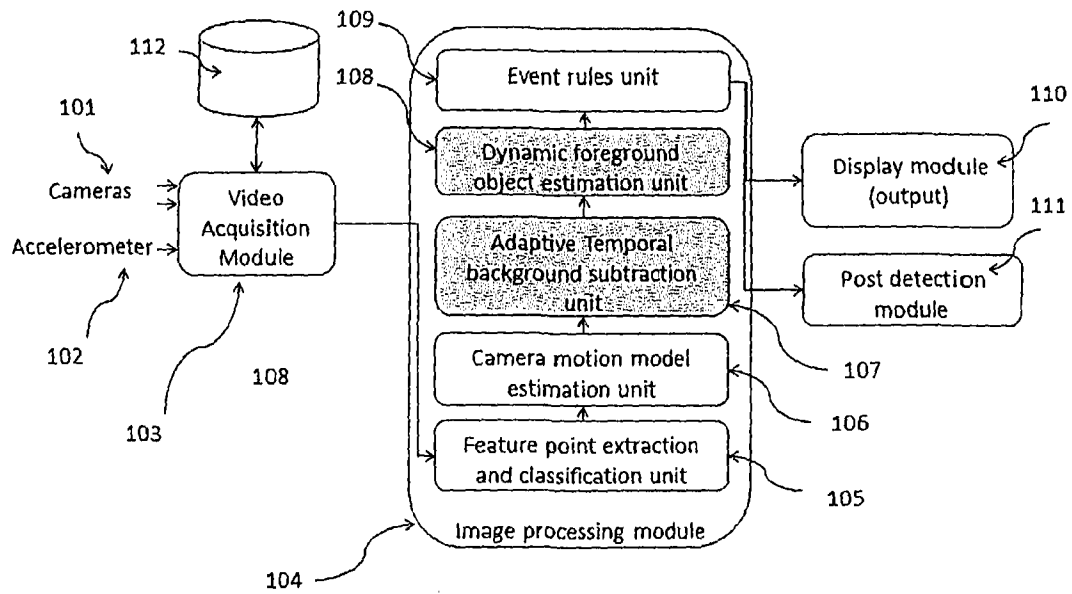


FIG. 1

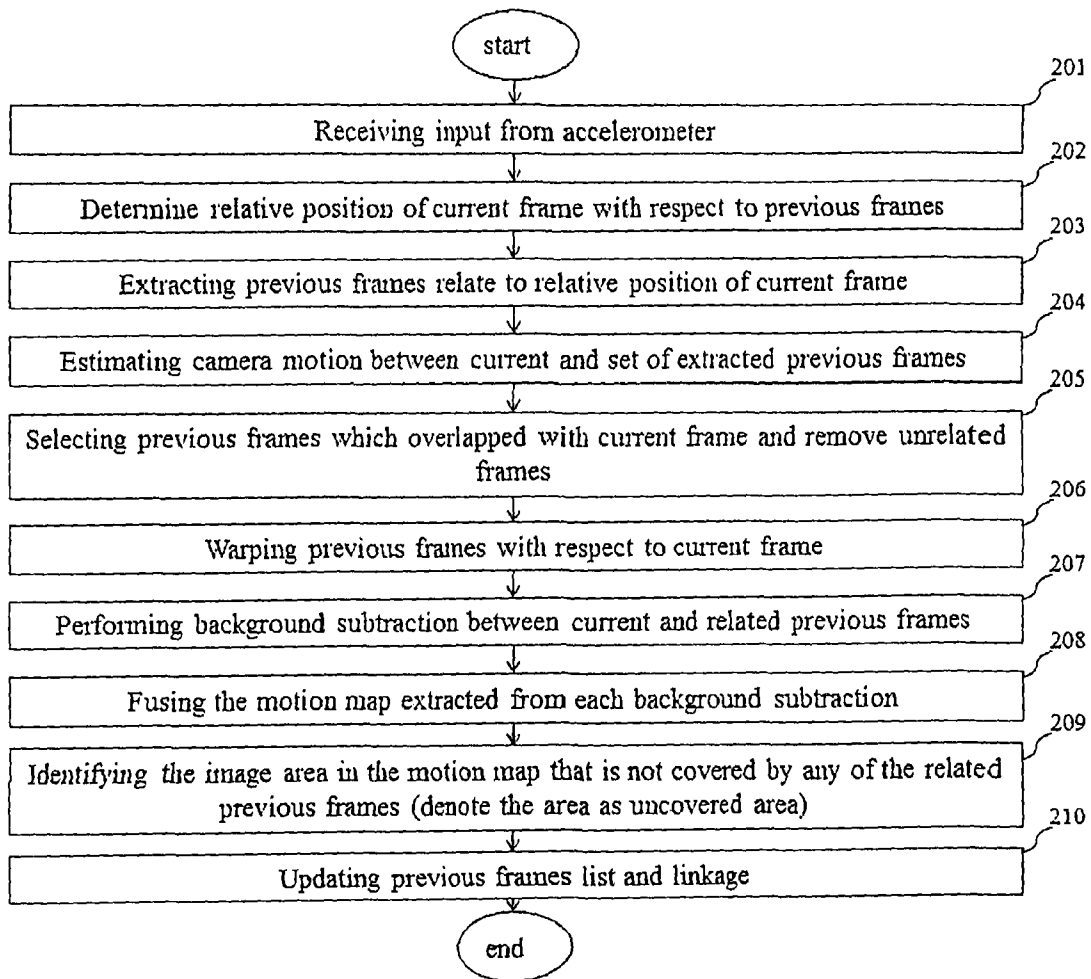


FIG. 2A

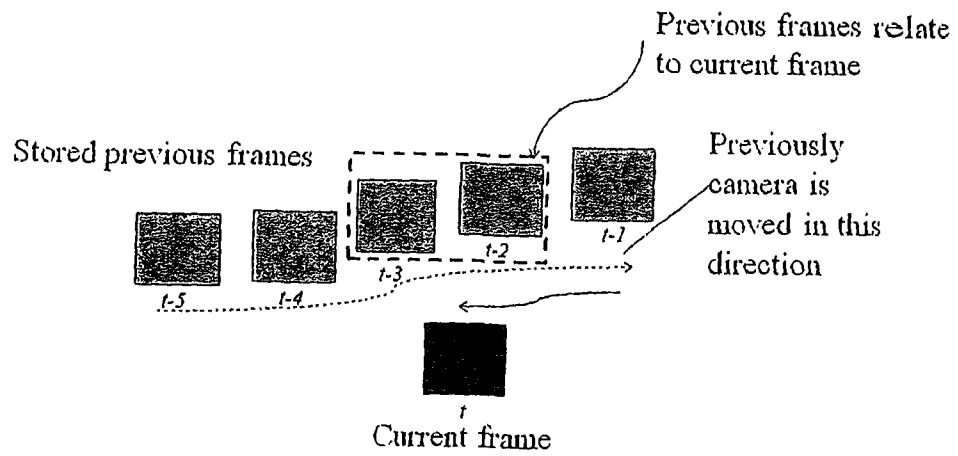


FIG. 2B

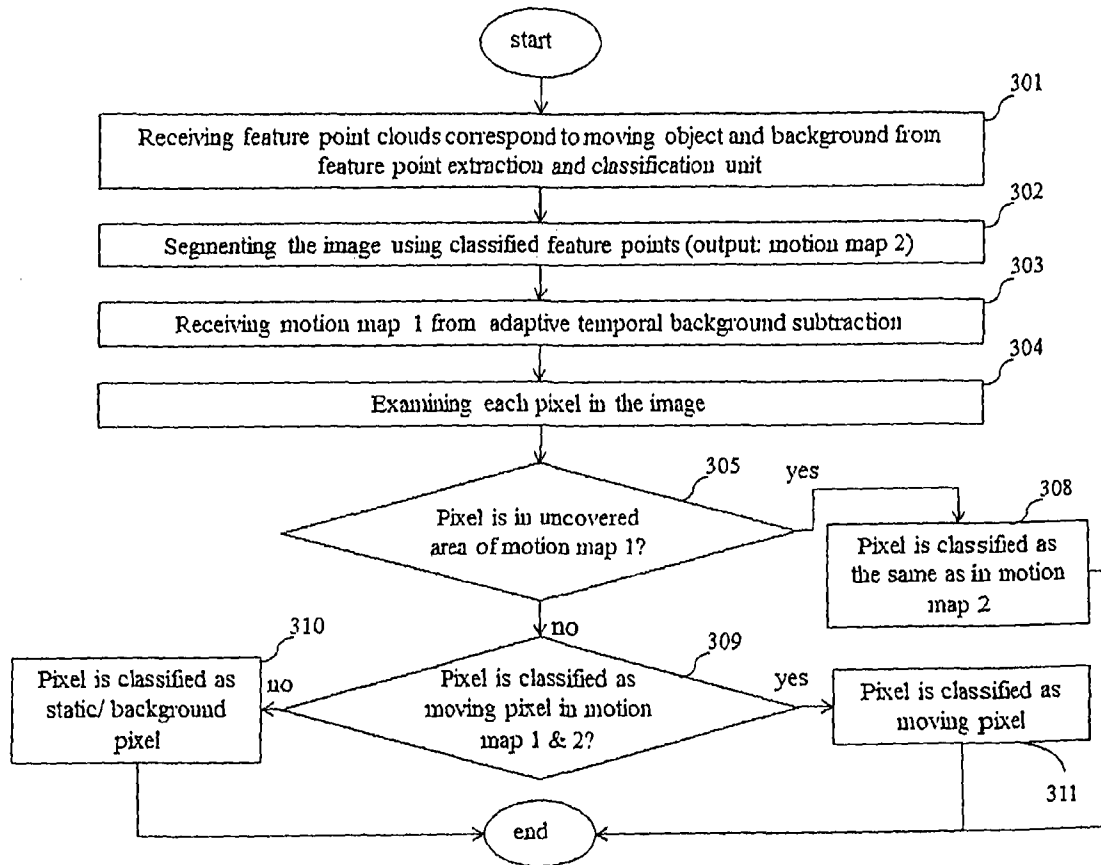


FIG. 3

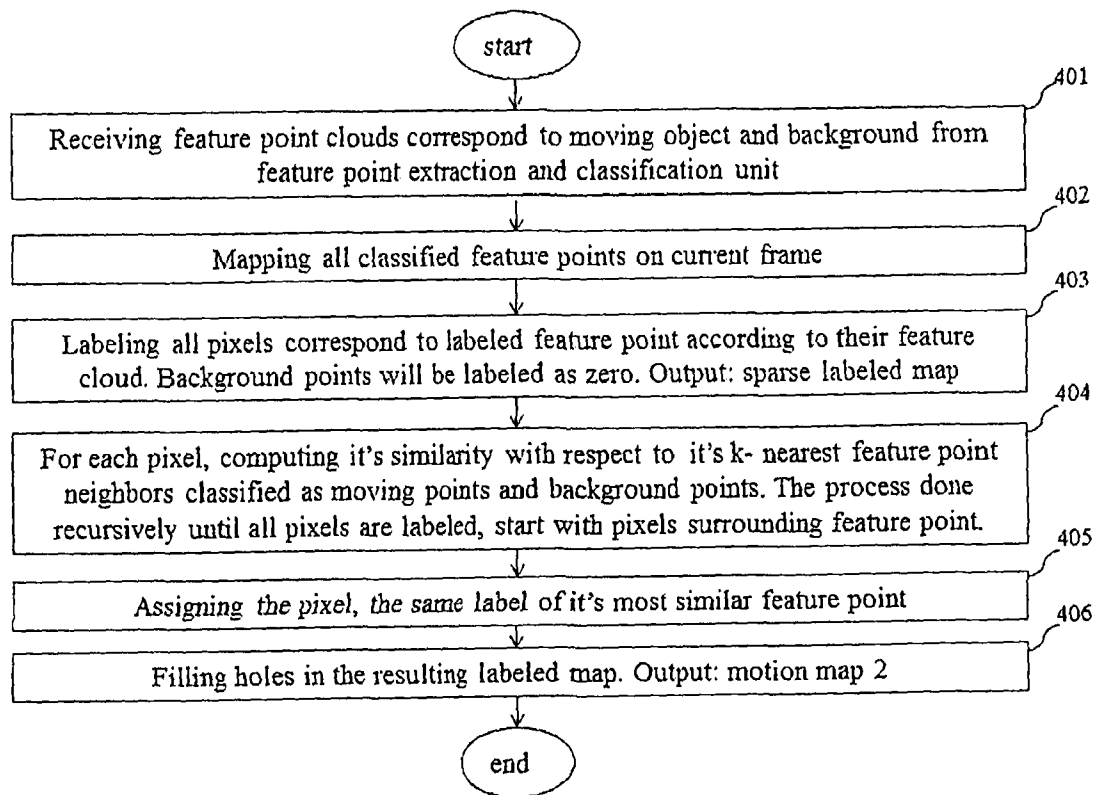


FIG. 4