(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 July 2017 (27.07.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/127220 A1

(51) International Patent Classification:
G06T 7/246 (2017.01) *G06T 17/00* (2006.01)
G01C 11/04 (2006.01)

(21) International Application Number:
PCT/US2016/069142

(22) International Filing Date:
29 December 2016 (29.12.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
7029/CHE/2015 29 December 2015 (29.12.2015) IN
15/197,749 29 June 2016 (29.06.2016) US

(71) Applicant: **TEXAS INSTRUMENTS INCORPORATED** [US/US]; P.O. Box 655474, Mail Station 3999, Dallas, TX 75265-5474 (US).

(71) Applicant *(for JP only)*: **TEXAS INSTRUMENTS JAPAN LIMITED** [JP/JP]; 24-1, Nishi-shinjuku 6-chome, Shinjuku-ku Tokyo, 160-8366 (JP).

(72) Inventors: **NAGORI, Soyeb**; No-45, 1st Floor, 5th Main Road, Kousar Nagar, R T Nagar Post, Bangalore 560032 (IN). **MATHEW, Manu**; L-804, Brigade Metropolis, Whitefield Main Road, Bangalore 560048 (IN). **VISWANATH, Prashanth, Ramanathpur**; No. 106, 11th Cross, 7th Main, Malleshwaram, Bangalore 560003 (IN). **PODDAR, Deepak, Kumar**; C-403, 1a Cross, Sai Mitra Meadows, Kagaspura, C B Raman Nagar, Bangalore 560093 (IN).

(74) Agents: **DAVID, Michael, A., Jr.** et al.; Texas Instruments Incorporated, P.O. Box 655474, Mail Station 3999, Dallas, TX 75265-5474 (US).

(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, DJ, 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, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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.

[Continued on next page]

(54) Title: METHOD FOR STRUCTURE FROM MOTION PROCESSING IN A COMPUTER VISION SYSTEM

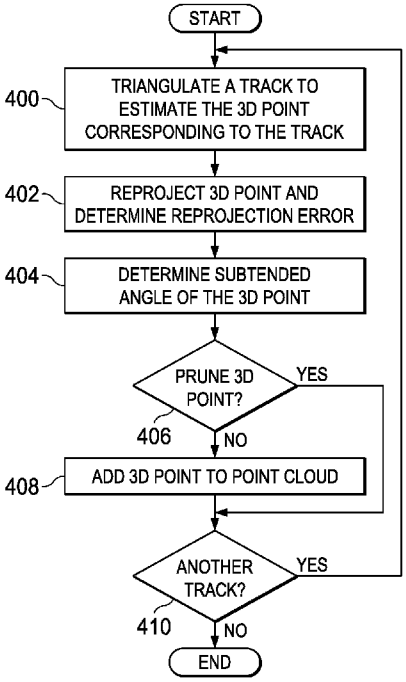

FIG. 4

(57) Abstract: In described examples, a method for structure from motion processing in a computer vision system includes receiving a plurality of tracks of estimated two dimensional (2D) locations of points in a plurality of sequential frames captured by a monocular camera, and generating (408) a cloud of three dimensional (3D) points by, for each track of the plurality tracks (410), triangulating (400) an estimated location of a 3D point based on the 2D locations in the track, and including the 3D point in the cloud when (406) a size of a subtended angle (404) of the 3D point and a projection error (402) of the 3D point indicate that the 3D point is sufficiently robust for inclusion.

WO 2017/127220 A1

(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, ST, 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:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

METHOD FOR STRUCTURE FROM MOTION PROCESSING IN A COMPUTER VISION SYSTEM

[0001] This relates generally to computer vision systems, and more particularly to structure from motion in computer vision systems.

BACKGROUND

[0002] A new class of embedded safety systems, referred to as advanced driver assistance systems (ADAS), has been introduced into automobiles to reduce human operation error. Such systems may provide functionality such as rear-view facing cameras, electronic stability control, and vision-based pedestrian detection systems. Many of these systems use monocular cameras and rely on computer vision processing to detect objects in the field of view of one or more cameras. Structure from motion (SfM) processing is a critical operation in such systems to understand the three-dimensional environment from two-dimensional images captured by the monocular camera.

SUMMARY

[0003] Described examples include methods and systems for real time structure from motion (SfM) processing in a computer vision system. In one aspect, a method for structure from motion processing in a computer vision system includes receiving tracks of estimated two dimensional (2D) locations of points in sequential frames captured by a monocular camera, and generating a cloud of three dimensional (3D) points by, for each of the tracks, triangulating an estimated location of a 3D point based on the 2D locations in the track, and including the 3D point in the cloud when a size of a subtended angle of the 3D point and a projection error of the 3D point indicate that the 3D point is sufficiently robust for inclusion. The size of the subtended angle and the projection error are tested against one or more pairs of progressively smaller respective threshold values to determine whether the 3D point is sufficiently robust. A pair of threshold values consists of an angle size threshold and projection error threshold.

[0004] In one aspect, a method for structure from motion processing in a computer vision system includes generating a cloud of three dimensional (3D) points based on sequential frames captured by a monocular camera, and merging the cloud with a previous cloud of 3D points to

generate a final cloud of 3D points. For each 3D point in the previous cloud, determining whether to include the 3D point in the final cloud includes performing point correspondence between a most recent frame in the sequential frames and a next most recent frame in the frames to attempt to locate a first two dimensional (2D) point in the most recent frame corresponding to a second 2D point in the next most recent frame corresponding to the 3D point, computing an epipolar line for the first 2D point and an epipolar line for the second 2D point if the first 2D point is located, and including or not including the 3D point in the final cloud based Euclidean distances of the first 2D point and the second 2D point to respective epipolar lines.

[0005] In one aspect, a method for structure from motion processing in a computer vision system includes receiving tracks of estimated two dimensional (2D) locations of points in N sequential frames captured by a monocular camera, and computing a three dimensional (3D) point cloud by estimating a 3D point corresponding to each of the tracks. Two most recent frames in sliding windows of N sequential frames are used for point correspondence to estimate the 2D locations in the tracks, and each track includes up to N 2D locations, where N>2.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] FIG. 1 is a high level block diagram of an example multiprocessor system-on-a-chip (SOC) configured for use in a monocular camera-based automotive safety application.

[0007] FIG. 2 is diagram illustrating data flow through an example real time structure from motion (SfM) reconstruction pipeline implemented on the example SOC of FIG. 1.

[0008] FIG. 3 is an example illustrating sliding window processing in the SfM reconstruction pipeline of FIG. 2.

[0009] FIG. 4 is a flow diagram of a method for generating a 3D point cloud in the SfM reconstruction pipeline of FIG. 2.

[0010] FIG. 5 is an example illustrating the subtended angle for an estimated 3D point.

[0011] FIG. 6 is a flow diagram of method for determining whether 3D points in a point cloud are to be carried forward.

DETAILED DESCRIPTION OF EXAMPLE EMBODIMENTS

[0012] Like elements in the various figures are denoted by like reference numerals for consistency.

[0013] As mentioned hereinabove, structure from motion (SfM) processing is a critical operation in the computer vision processing performed in many camera-based embedded safety

systems. Many conventional SfM approaches are directed to three dimensional (3D) reconstruction from image collections in which the images are taken from various view points and possibly by different cameras. Microsoft Photo Tourism and Bundler developed at Cornell University are two examples of this SfM approach. Such SfM approaches are directed to 360 degree three-dimensional (3D) reconstruction and are typically not suitable for use in real time reconstruction. In automotive safety systems, images (frames) are generated in real time and any SfM approach used needs to process the images in real time. Example embodiments provide for real time SfM processing in a computer vision system receiving images from a monocular camera. The real time SfM processing described herein exploits constraints of the 3D environment that can be assumed for automotive applications.

[0014] FIG. 1 is a high level block diagram of an example multiprocessor system-on-a-chip (SOC) 100 configured for use in a monocular camera-based ADAS. In at least one example, the SOC 100 is an embodiment of the TDA3X SOC available from Texas Instruments, Inc. This description includes a high level description of the components of the SOC 100. More detailed descriptions of example components may be found in M. Mody, et al., "High Performance Front Camera ADAS Applications on TI's TDA3X Platform," Proceedings of 2015 IEEE 22nd International Conference on High Performance Computing, December 16-19, 2015, Bangalore, India, pp. 456-463, and "TDA3x SOC Processors for Advanced Driver Assist Systems (ADAS) Technical Brief," Texas Instruments, SPRT704A, October, 2014, pp. 1-6, which are incorporated by reference herein.

[0015] The SOC 100 includes dual general purpose processors (GPP) 102, dual digital signal processors (DSP) 104, and a vision processor 106 coupled via a high speed interconnect 122. The SOC 100 further includes a direct memory access (DMA) component 108, a camera capture component 110 coupled to a display management component 114, on-chip random access (RAM) memory 116, and various input/output (I/O) peripherals 120 all coupled to the processors via the interconnect 122. Also, the SOC 100 includes a safety component 118 that includes safety related functionality to enable compliance with automotive safety requirements. Such functionality may include support for CRC (cyclic redundancy check) of data, clock comparator for drift detection, error signaling, windowed watch-dog timer, and self testing of the SOC for damage and failures. Software implementing real-time SfM as described herein may be stored in the memory 116 and may execute on one or more programmable processors of the SOC 100.

[0016] FIG. 2 is a diagram illustrating data flow through an embodiment of a real time SfM reconstruction pipeline as implemented as part of a computer vision system implemented on the example SOC 100 of FIG. 1. The SfM reconstruction pipeline generates a 3D point cloud after each input frame is received. At a high level, the SfM reconstruction pipeline detects and tracks two dimensional (2D) locations of interest points across sequential frames, storing the 2D locations in tracks corresponding to the interest points. Thus, a track is a sequence of 2D frame coordinates of an interest point over time. Triangulation is performed on each of the tracks to generate a 3D point cloud that is then used to generate a final 3D point cloud.

[0017] Further, the pipeline operates in a sliding window fashion as illustrated in FIG. 3. A window includes N frames, the N frames being the previously processed N-1 frames and the current frame. As will be explained in more detail herein, each output point cloud is based on the frames in the current window. Further, the triangulation performed to estimate 3D points uses up to N 2D locations in N frames in the current sliding window for each interest point. Any suitable value of N may be used, and the value of N may be chosen as a tradeoff between point cloud quality (quality increases as the value of N increases) and considerations, such as computational complexity, performance and memory size. For ease of description, N = 6 is assumed in the description below.

[0018] Referring to FIG. 2, each time a frame is received from the monocular camera 124 via the capture component 110 interest point detection 200 is performed on the frame, i.e., frame (t), to detect interest points. Frame (t) and the detected interest points are stored in memory 201. For example, the memory 201 may be the on-chip RAM 116 of FIG. 1. The detection 200 of interest points, also referred to as key points or features, may be performed using any suitable technique for interest point detection. Also, for example, the interest point detection may be based on Harris corner detection or the features from accelerated segment test (FAST) detection. Harris corner detection is described in C. Harris and M. Stephens, "A Combined Corner and Edge Detector," Proceedings of Fourth Alvey Vision Conference, Manchester, UK, pp. 147-151, 1988. FAST is described in E. Rosten and T. Drummond, "Machine Learning for High Speed Corner Detection," Proceedings of 9th European Conference on Computer Vision, Vol. 1, Graz, Austria, May 7-13, 2006, pp. 430-443.

[0019] After the interest points are detected in frame (t), point correspondence 202 between the current frame (t) and the previous frame (t-1) is performed. Generally, the point correspondence

matches interest points found in the previous frame (t-1) and each of the current tracks to the current frame (t). In the latter case, the newest 2D point in a current track is used for the point correspondence. The newest 2D point in a track is a 2D point found in frame (t-1) when point correspondence was performed between frame (t-1) and frame (t-2). Also, current tracks are stored in memory 201. If a correspondence for an interest point detected in the previous frame (t-1) is not found in the current frame (t), the interest point is deleted. Also, if a correspondence for an existing track is not found in the current frame (t), the track is deleted. Any suitable technique for point correspondence may be used, such as an optical flow technique. For example, some suitable optical flow techniques are described in J. Barron, et al., "Performance of Optical Flow Techniques," International Journal of Computer Vision, Vol. 12, No. 1, pp. 43-77, 1994.

[0020] Correspondence validation 204 is then performed to validate the point correspondences between the previous frame (t-1) and the current frame (t). To perform the validation, the camera fundamental matrix is computed based on the corresponding interest points between the current frame (t) and the previous frame (t-1). Any suitable technique for computing the fundamental matrix may be used. One example technique is described in "Eight-point Algorithm", Wikipedia, available at https://en.wikipedia.org/wiki/Eight-point_algorithm on May 14, 2016. The fundamental matrix is then used to evaluate the quality of the point correspondences, i.e., to validate the point correspondences. The use of the fundamental matrix to validate point correspondences is known, so it is not further explained herein.

[0021] If a point correspondence for an interest point is not valid, the interest point is deleted. Otherwise, a new track is formed that includes the 2D location of the interest point and the 2D location of the corresponding point in frame (t). Also, if a point correspondence for an existing track is valid, the 2D location of the corresponding point is added to the track. Otherwise, the track is deleted. Also, the interest points detected in frame (t) are pruned. Accordingly, an interest point detected in frame (t) has coordinates very close to or exactly the same as any 2D location in a current track, the detected interest point is deleted.

[0022] The number of 2D locations in a track is bounded by the number of frames in the sliding window, i.e., any given track may have at most six point coordinates, five coordinates from the previous five frames and the coordinates from the current frame. If six coordinates are

already in a track when frame (t) is processed, the oldest coordinates are dropped from the track and coordinates from frame (t) are added.

[0023] Camera pose computation 206 is then performed. Camera pose consists of six degrees of freedom made up of the rotation, i.e., roll, pitch, and yaw, and 3D translation of the camera with respect the world. The output of the camera pose computation 206 is a three degree of rotation vector and a three degree of translation vector. The problem of estimating the pose of a calibrated camera given a set of n 3D points in the world and their corresponding 2D projections in a frame is referred to as Perspective-n-Point (PnP). The pose of a camera in the form of rotation and translation can be found given a set of n 3D points and their corresponding 2D image projections. Any suitable PnP solution technique may be used, such as the technique described in S. Li, et al., "A Robust O(n) Solution to the Perspective-n-Point Problem," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, Issue 7, July, 2012, pp. 1444-1450.

[0024] nP techniques require the point cloud generated in the previous iteration through the SfM reconstruction pipeline. For the first pair of frames processed in the pipeline either at startup or after failure to generate a point cloud in a previous iteration, no point cloud is available to be used. In some embodiments, a singular value decomposition (SVD) technique is used for pose estimation in such situations. The inputs to the SVD technique are the two frames, the interest point correspondences between the two frames, and the fundamental matrix. In subsequent iterations, a point cloud is available and pose estimation is performed using a PnP technique. Any suitable SVD based technique may be used. One example technique is described in R. Haralick, et al., "Pose Estimation from Corresponding Point Data," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 19, Issue 6, pp. 1426-1446, Nov/Dec 1989.

[0025] Triangulation computation 208 is then performed to generate a point cloud corresponding to the frames in the current sliding window. FIG.. 4 is a flow diagram of a method performed by the triangulation computation 208 to generate the 3D point cloud. For each track 410, triangulation 400 is performed to find the 3D position corresponding to the 2D positions in the track. As mentioned hereinabove, a track records the position of an interest point in two or more consecutive frames in the current sliding window. The triangulation 400 uses the 2D coordinates in the track find the corresponding 3D position based on the current estimated

camera pose and the current fundamental matrix. Because the sliding window size is six frames, the track for an interest point can have at most six 2D coordinates.

[0026] Any suitable triangulation technique may be used. For example, some suitable techniques are described in R. Hartley and P. Sturm, "Triangulation," Computer Vision and Image Understanding, Vol. 68, No. 2, pp. 146-157, November, 1997 ("Hartley" herein), which is incorporated by reference herein. In some embodiments, an iterative least squares technique as described in Hartley is used. Hartley described triangulation using two frames, but the technique in Hartley can be extended to more than two frames.

[0027] After the 3D point corresponding to the track is estimated, the 3D point is reprojected 402 into the frames of the two 2D points used to estimate the 3D position and the reprojection error for each 2D point is determined 402. Generally, reprojection error is the geometric error corresponding to the Euclidean distance between a projected 2D point and the corresponding measured 2D point. The reprojection error quantifies how closely an estimate of a 3D point recreates the true projection of a measured 2D point. Ideally, the reprojection error is zero, and a non-zero value indicates error in the triangulation. Thus, the reprojection error may be an indication of the robustness of the 3D point.

[0028] Also, the subtended angle of the 3D point is determined 404. FIG. 5 is an example illustrating the subtended angle for an estimated 3D point. In this example, the track for the interest point includes three 2D points corresponding to locations in each of three consecutive frames. The subtended angle is formed by the intersection of the left ray and the right ray at the 3D point. The left ray passes through the optical center and the measured 2D point location in Frame N and the right ray passes through the optical center and measured 2D point location in Frame N+2. The size of the subtended angle may be an indication of the robustness of the 3D point as smaller angles can adversely affect the quality of the triangulation.

[0029] Referring again to FIG. 4, a determination 406 is made about whether to prune the 3D point, i.e., about whether to include the 3D point in the point cloud. This decision is based on the reprojection errors of the 2D points triangulated to estimate the 3D point, and on the size of the subtended angle of the estimated 3D point. Generally, the angle size and the sum of the reprojection errors are compared to respective empirically determined thresholds and the results of the comparisons are used to decide whether the 3D point is sufficiently robust to include in the point cloud.

[0030] In many vision applications, a single subtended angle threshold and a single projection error threshold are sufficient for pruning. However, using single threshold values is problematic in automotive applications, especially in applications using front mounted cameras. When the vehicle is moving forward in a straight line, the subtended angles of points corresponding to objects in front of the car can become very small. This effect is more pronounced when the points corresponding to an object of interest fall near the optical center. Robustly detecting objects in the center is critical for taking action to avoid hitting such objects. Tests have shown that such automotive applications, setting the angle threshold to 0.25 degrees prunes too many 3D points near the optical center which lowers the density of the 3D points whereas setting the angle threshold to 0.125 degrees does not provide sufficient pruning, thus retaining too many poorly reconstructed 3D points. To yield both a good density of 3D points around the optical center while also pruning poorly reconstructed 3D points, a single angle threshold value does not exist.

[0031] In some embodiments, rather than using a single value for the subtended angle threshold and a single value for the projection error threshold, the projection error and subtended angle values are tested against multiple pairs of progressively smaller respective threshold values chosen such that when the size of the subtended angle is larger, more projection error is tolerated in the pruning decision and when the size of the subtended angle is smaller, less projection error is tolerated in the pruning decision. The use of progressively smaller threshold values ensures that otherwise robust 3D points with smaller subtended angles and acceptable projection error are not pruned.

[0032] The example pseudo code of Table 1 illustrates this approach to pruning. In this example pseudo code, subtendedAngle is the size of the subtended angle, projectionError is the sum of the reprojection errors of the two 2D points, angleTh is the base angle size threshold, projectionErrTh is the base projection error threshold, and the kx, x = 1, 2, 3, 4, are scale factors used to reduce angleTH and projectionErrTh. If either the subtended angle size or the projection error fail the respective threshold comparison test, the base thresholds are reduced by the next larger scale factor, and the threshold comparison tests are tried again. A 3D point is pruned if all of the scale factors are tried unsuccessfully. Any suitable values for the base angle size threshold, the base projection error threshold and the scale factors may be used. In some

embodiments, these values are determined empirically. Further, embodiments are possible in which more or fewer scaling factors are used.

Table 1

```
k1= 2; k2= 4; k3 = 8; k4 =16
prune = true;
if((subtendedAngle>=angleTh)&&(projectionError<=projectionErrTh)) {
    prune = false;
} else if ((subtendedAngle >= angleTh/k1 && (projectionError <= projectionErrTh/k1)) {
            prune = false;
} else if ((subtendedAngle >= angleTh/k2) && (projectionError <= projectionErrTh/k2)) {
    prune = false;
} else if ((subtendedAngle >= angleTh/k3) && (projectionError <= projectionErrTh/k3)) {
    prune = false;
} else if ((subtendedAngle >= angleTh/k4) && (projectionError <= projectionErrTh/k4)) {
     prune = false;
}
```

[0033] Referring again to FIG. 2, after triangulation computation 208, final 3D point cloud generation 210 is performed. The final 3D point cloud is generated by merging the 3D point cloud output by the triangulation computation 208 with the 3D point cloud generated for the previous sliding window. A simple merger of the two point clouds is problematic, because new points in the current point cloud may correspond to a new object that occludes an object corresponding to older points in the previous point cloud. Such older points should not be carried forward to the final point cloud as these points are not representative of the new object.

[0034] In some embodiments, before the two point clouds are merged, 3D points in the previous 3D point cloud are tested using the fundamental matrix relationship test to decide whether the 3D point should be carried forward into the final 3D point cloud. FIG. 6 is a flow diagram of an example method to decide whether a 3D point is to be carried forward. Initially, the 2D point in the previous frame (t-1) corresponding to a 3D point in the previous 3D point cloud is determined 600 from the track corresponding to the 3D point. A 2D point corresponding to the 2D point in the previous frame (t-1) is then found 602 in the current frame (t) using point correspondence. Any suitable technique for point correspondence may be used,

such as the techniques mentioned hereinabove. If the point correspondence does not find 604 a valid corresponding point in the current frame (t), the 3D point is removed 612 from the previous 3D point cloud and processing continues 614 with the next 3D point, if any.

[0035] If a valid corresponding point was found 604, the fundamental matrix relationship test is applied 606-610 to determine if the 3D point is to be carried forward. More specifically, the epipolar lines corresponding to the 2D points are computed 606 using the fundamental matrix. Accordingly, if x is the 2D point in the previous frame (t-1), x' is the corresponding 2D point in the current frame (t), and F is the fundamental matrix, l' = Fx is the epipolar line corresponding to x and l = FTx' is the epipolar line corresponding to x'. The Euclidean distance of each 2D point to the corresponding epipolar line is then computed 608. The sum of the Euclidean distances is then compared 604 to a threshold to determine whether the 3D point is to be carried forward. If the sum is sufficiently small, i.e., is less than or equal to the threshold, the 3D point is carried forward. Otherwise, the 3D point is removed 612 from the previous point cloud. In either instance, processing continues 614 with the next 3D point, if any. Any suitable value may be used for the threshold. The value may be determined empirically.

Other Embodiments

[0036] Example embodiments have been described here in which the camera pose estimate is computed using a PnP solution technique. In at least one example embodiment, an inertial measurement unit (IMU) is used to enhance the pose estimate found using a PnP solution technique, or an IMU replaces the camera pose estimation.

[0037] In another example, embodiments have been described herein in which interest point detection is performed. In at least one example embodiment, the interest point detection is not performed, and point correspondence is performed for all pixels (points) in a frame.

[0038] In at least one example embodiment, a suitable technique for bundle adjustment is added to the SfM reconstruction pipeline for joint optimization of the camera pose estimation and the triangulation.

[0039] In another example, embodiments have been described herein in which the processing performed in the SfM reconstruction pipeline is implemented as software instructions executed on processors in a multiprocessor SOC. The functionality of the SfM reconstruction pipeline may be implemented as any suitable combination of software, firmware and hardware. For

example, some of the functionality may be implemented in one or more hardware accelerators, application specific integrated circuits (ASICs) and/or field-programmable gate arrays (FPGAs).

[0040] In another example, embodiments have been described herein in which carry forward processing is performed on all the 3D points in the previous point cloud to remove any 3D points that should not be carried forward before merging the previous point cloud and the current point cloud. In yet another example, the carry forward decision criteria described herein is applied to points in the previous point cloud as the two point clouds are merged.

[0041] In another example, embodiments have been described herein in reference to automotive safety systems. Example embodiments are possible for other computer vision applications having similar environmental constraints, such as industrial applications, robotics and consumer applications (e.g., vacuum cleaners and drones).

[0042] Although method steps may be presented and described herein in a sequential fashion, one or more of the steps shown in the figures and described herein may be performed concurrently, may be combined, and/or may be performed in a different order than the order shown in the figures and/or described herein. Accordingly, embodiments are not limited to the specific ordering of steps shown in the figures and/or described herein.

[0043] In this description, the term “couple” and derivatives thereof mean an indirect, direct, optical and/or wireless electrical connection. For example, if a first device couples to a second device, that connection may be through a direct electrical connection, through an indirect electrical connection via other devices and connections, through an optical electrical connection, and/or through a wireless electrical connection.

[0044] Modifications are possible in the described embodiments, and other embodiments are possible, within the scope of the claims.

CLAIMS

What is claimed is:

1. A method for structure from motion processing in a computer vision system, the method comprising:

receiving a plurality of tracks of estimated two dimensional (2D) locations of points in a plurality of sequential frames captured by a monocular camera;

generating a cloud of three dimensional (3D) points by, for each track of the plurality tracks;

triangulating an estimated location of a 3D point based on the 2D locations in the track; and

including the 3D point in the cloud when a size of a subtended angle of the 3D point and a projection error of the 3D point indicate that the 3D point is sufficiently robust for inclusion;

wherein the size of the subtended angle and the projection error are tested against one or more of a plurality of pairs of progressively smaller respective threshold values to determine whether the 3D point is sufficiently robust, wherein a pair of threshold values consists of an angle size threshold and projection error threshold.

2. The method of claim 1, wherein an initial threshold pair of the plurality of pairs of threshold values includes a base angle size threshold value and a base projection error threshold value and threshold values in other pairs of the progressively smaller respective threshold values are determined by applying corresponding predetermined progressively larger scale factors to the base angle size threshold and the base projection error threshold.

3. The method of claim 2, wherein the pairs of progressively smaller respective threshold values are five pairs of progressively smaller respective threshold values, and the progressively larger scale factors are 2, 4, 8 and 16.

4. The method of claim 1, wherein two most recent frames in sliding windows of N sequential frames are used for point correspondence to estimate the 2D locations in the plurality of tracks, and each track includes up to N 2D locations, wherein a number of frames in the plurality of sequential frames is N and N>2.

5. The method of claim 4, wherein N = 6.

6. The method of claim 1, further comprising:

merging the cloud with a previous cloud of 3D points to generate a final cloud of 3D points;

wherein, for each 3D point in the previous cloud, determining whether to include the 3D point in the final cloud comprises: performing point correspondence between a most recent frame in the plurality of sequential frames and a next most recent frame in the plurality of frames to attempt to locate a first 2D point in the most recent frame corresponding to a second 2D point in the next most recent frame corresponding to the 3D point; computing an epipolar line for the first 2D point and an epipolar line for the second 2D point if the first 2D point is located; and including or not including the 3D point in the final cloud based Euclidean distances of the first 2D point and the second 2D point to respective epipolar lines.

7. The method of claim 1, wherein the points are interest points.

8. The method of claim 1, wherein the computer vision system is part of an automotive safety system.

9. A method for structure from motion processing in a computer vision system, the method comprising:

generating a cloud of three dimensional (3D) points based on a plurality of sequential frames captured by a monocular camera; and

merging the cloud with a previous cloud of 3D points to generate a final cloud of 3D points;

wherein, for each 3D point in the previous cloud, determining whether to include the 3D point in the final cloud comprises: performing point correspondence between a most recent frame in the plurality of sequential frames and a next most recent frame in the plurality of frames to attempt to locate a first two dimensional (2D) point in the most recent frame corresponding to a second 2D point in the next most recent frame corresponding to the 3D point; computing an epipolar line for the first 2D point and an epipolar line for the second 2D point if the first 2D point is located; and including or not including the 3D point in the final cloud based Euclidean distances of the first 2D point and the second 2D point to respective epipolar lines.

10. The method of claim 9, wherein generating a cloud of 3D points comprises:

receiving a plurality of tracks of estimated 2D locations of points in the plurality of sequential frames; and

computing the point cloud by estimating a 3D point corresponding to each track of the plurality of tracks;

wherein two most recent frames in sliding windows of N sequential frames are used for point correspondence to estimate the 2D locations in the tracks, and each track includes up to N 2D locations, wherein a number of frames in the plurality of frames is N and N>2.

11. The method of claim 10, wherein N = 6.

12. The method of claim 9, wherein generating a point cloud of 3D points comprises:

receiving a plurality of tracks of estimated two dimensional (2D) locations of points in the plurality of sequential frames;

generating a cloud of three dimensional (3D) points by, for each track of the plurality tracks;

triangulating an estimated location of a three dimensional (3D) point based on the 2D locations in the track; and

including the 3D point in the cloud when a size of a subtended angle of the 3D point and a projection error of the 3D point indicate that the 3D point is sufficiently robust for inclusion;

wherein the size of the subtended angle and the projection error are tested against one or more of a plurality of pairs of progressively smaller respective threshold values to determine whether the 3D point is sufficiently robust, wherein a pair of threshold values consists of an angle size threshold and projection error threshold.

13. The method of claim 12, wherein an initial threshold pair of the plurality of pairs of threshold values includes a base angle size threshold value and a base projection error threshold value and threshold values in other pairs of the progressively smaller respective threshold values are determined by applying corresponding predetermined progressively larger scale factors to the base angle size threshold and the base projection error threshold.

14. The method of claim 13, wherein the pairs of progressively smaller respective threshold values are five pairs progressively smaller respective threshold values, and the progressively larger scale factors are 2, 4, 8 and 16.

15. A method for structure from motion processing in a computer vision system, the method comprising:

receiving a plurality of tracks of estimated two dimensional (2D) locations of points in N sequential frames captured by a monocular camera; and

computing a three dimensional (3D) point cloud by estimating a 3D point corresponding to each track of the plurality of tracks;

wherein two most recent frames in sliding windows of N sequential frames are used for point correspondence to estimate the 2D locations in the tracks, and each track includes up to N 2D locations, wherein N>2.

16. The method of claim 15, wherein N = 6.

17. The method of claim 15, further comprising:

merging the 3D point cloud with a previous 3D point cloud to generate a final 3D point cloud;

wherein, for each 3D point in the previous cloud, determining whether to include the 3D point in the final cloud comprises: performing point correspondence between a most recent frame in the plurality of sequential frames and a next most recent frame in the plurality of frames to attempt to locate a first 2D point in the most recent frame corresponding to a second 2D point in the next most recent frame corresponding to the 3D point; computing an epipolar line for the first 2D point and an epipolar line for the second 2D point if the first 2D point is located; and including or not including the 3D point in the final cloud based Euclidean distances of the first 2D point and the second 2D point to respective epipolar lines.

18. The method of claim 15, wherein generating a 3D point cloud comprises:

generating the 3D point cloud by, for each track of the plurality of tracks;

triangulating an estimated location of a three dimensional (3D) point based on the 2D locations in the track; and

including the 3D point in the 3D point cloud when a size of a subtended angle of the 3D point and a projection error of the 3D point indicate that the 3D point is sufficiently robust for inclusion;

wherein the size of the subtended angle and the projection error are tested against one or more of a plurality of pairs of progressively smaller respective threshold values to determine whether the 3D point is sufficiently robust, wherein a pair of threshold values consists of an angle size threshold and projection error threshold.

19. The method of claim 18, wherein an initial threshold pair of the plurality of pairs of threshold values includes a base angle size threshold value and a base projection error threshold value and threshold values in other pairs of the progressively smaller respective threshold values

are determined by applying corresponding predetermined progressively larger scale factors to the base angle size threshold and the base projection error threshold.

20. The method of claim 19, wherein the pairs of progressively smaller respective threshold values are five pairs progressively smaller respective threshold values, and the progressively larger scale factors are 2, 4, 8 and 16.

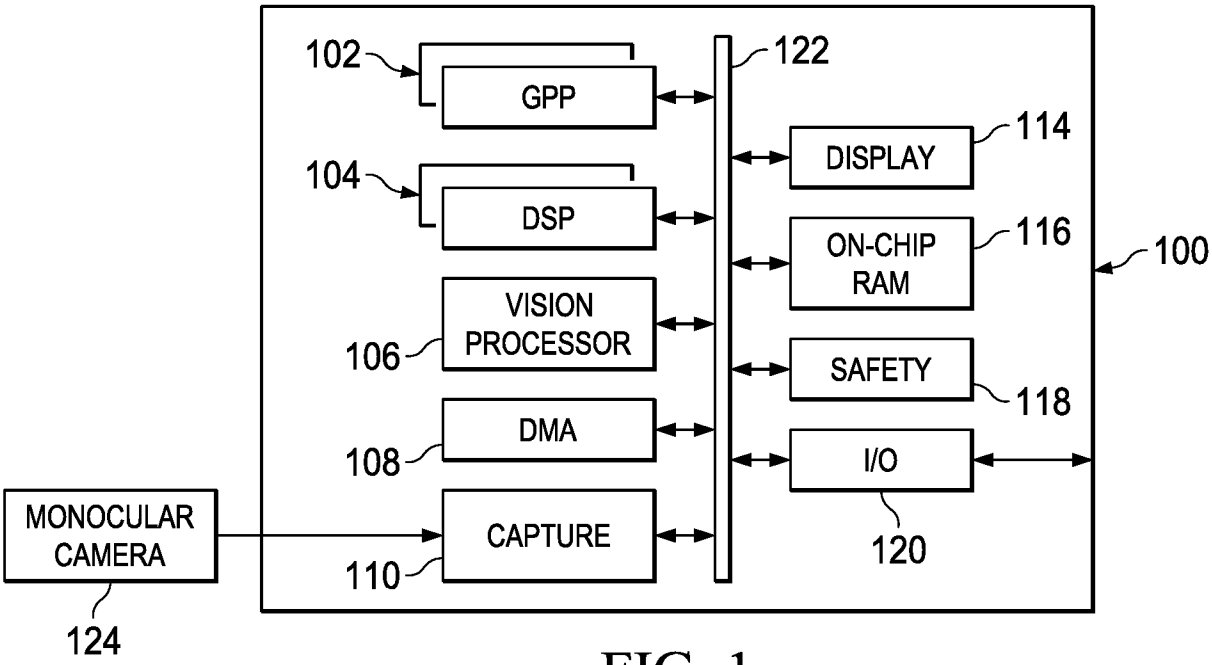
102
GPP
104
DSP
VISION PROCESSOR
106
DMA
108
CAPTURE
110
MONOCULAR CAMERA
124
122
DISPLAY
114
ON-CHIP RAM
116
SAFETY
118
I/O
120
100

FIG. 1

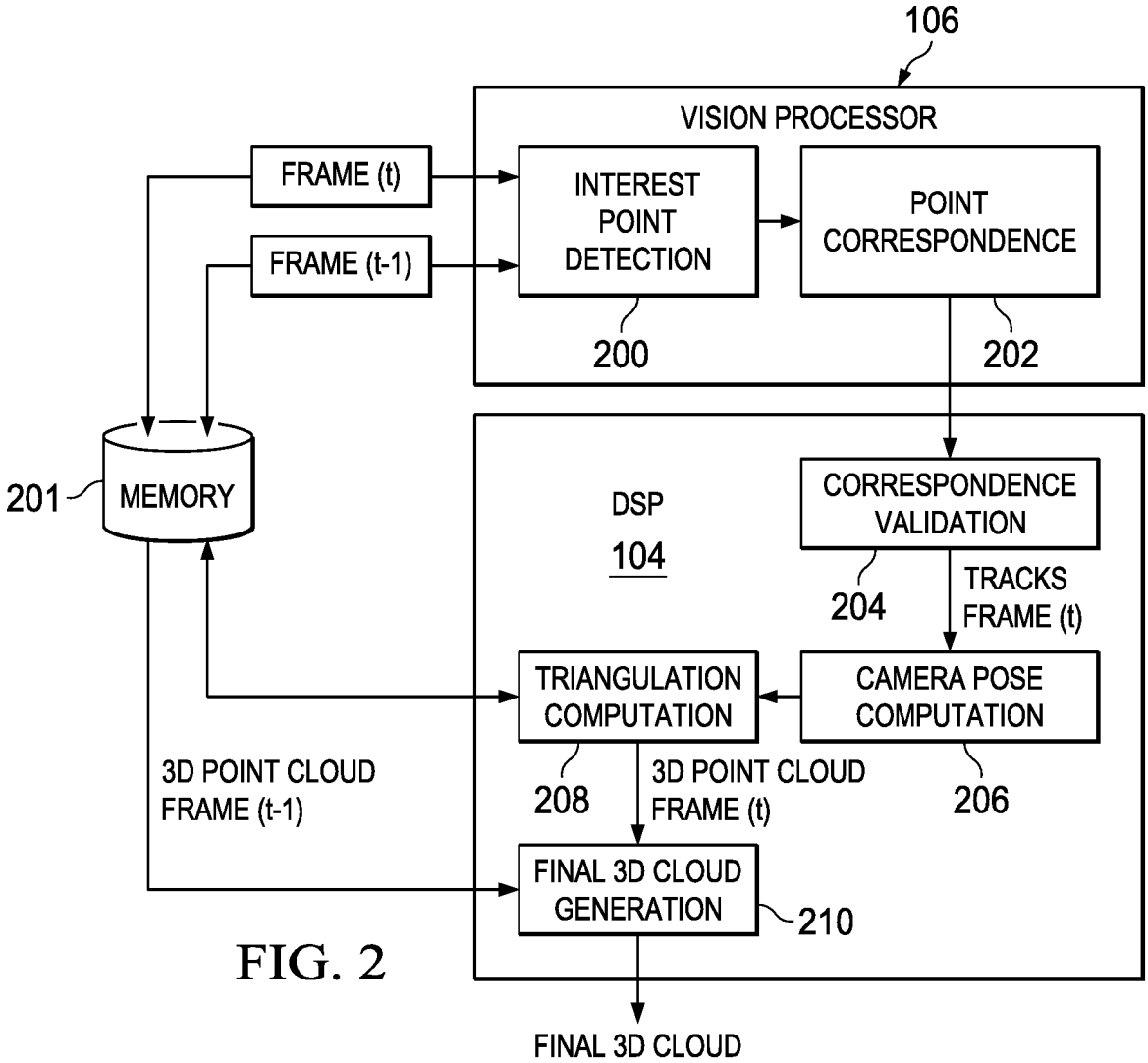
106
VISION PROCESSOR
FRAME (t)
FRAME (t-1)
INTEREST POINT DETECTION
200
POINT CORRESPONDENCE
202
201
MEMORY
DSP
104
CORRESPONDENCE VALIDATION
204
TRACKS FRAME (t)
TRIANGULATION COMPUTATION
CAMERA POSE COMPUTATION
208
206
3D POINT CLOUD FRAME (t-1)
3D POINT CLOUD FRAME (t)
FINAL 3D CLOUD GENERATION
210
FINAL 3D CLOUD

FIG. 2

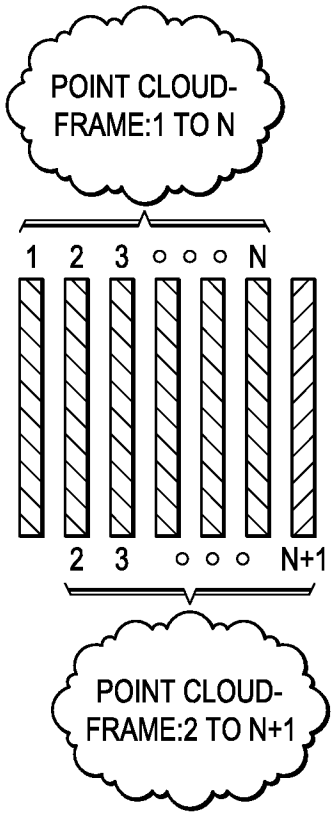
POINT CLOUD-
FRAME:1 TO N
1 2 3 ○ ○ ○ N
2 3 ○ ○ ○ N+1
POINT CLOUD-
FRAME:2 TO N+1

FIG. 3

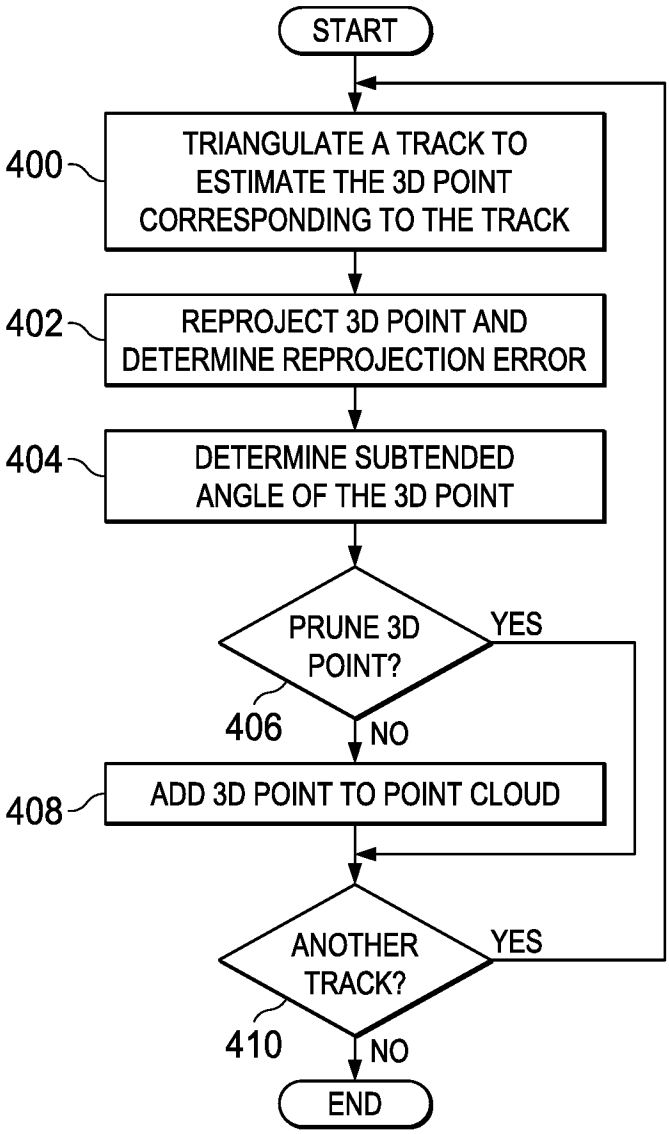
START
400
TRIANGULATE A TRACK TO ESTIMATE THE 3D POINT CORRESPONDING TO THE TRACK
402
REPROJECT 3D POINT AND DETERMINE REPROJECTION ERROR
404
DETERMINE SUBTENDED ANGLE OF THE 3D POINT
PRUNE 3D POINT?
YES
406
NO
408
ADD 3D POINT TO POINT CLOUD
ANOTHER TRACK?
YES
410
NO
END

FIG. 4

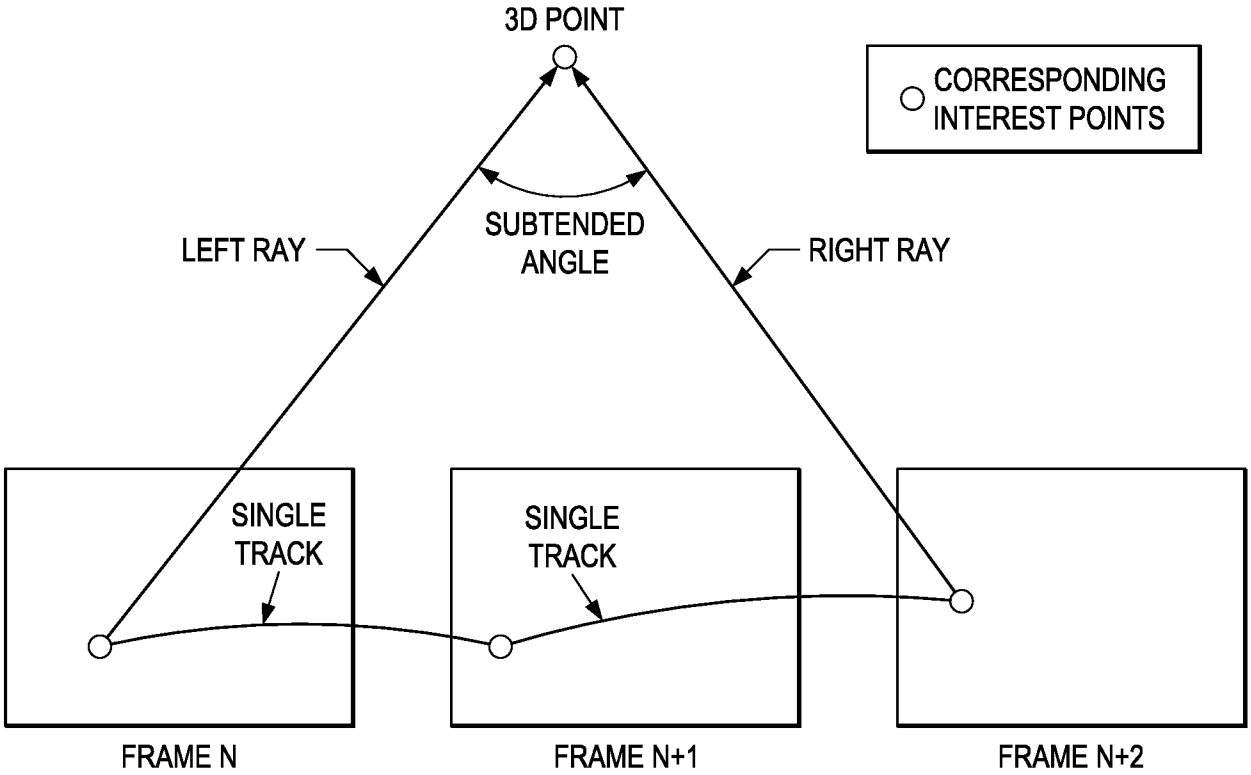
3D POINT
CORRESPONDING INTEREST POINTS
SUBTENDED ANGLE
LEFT RAY
RIGHT RAY
SINGLE TRACK
SINGLE TRACK
FRAME N
FRAME N+1
FRAME N+2

FIG. 5

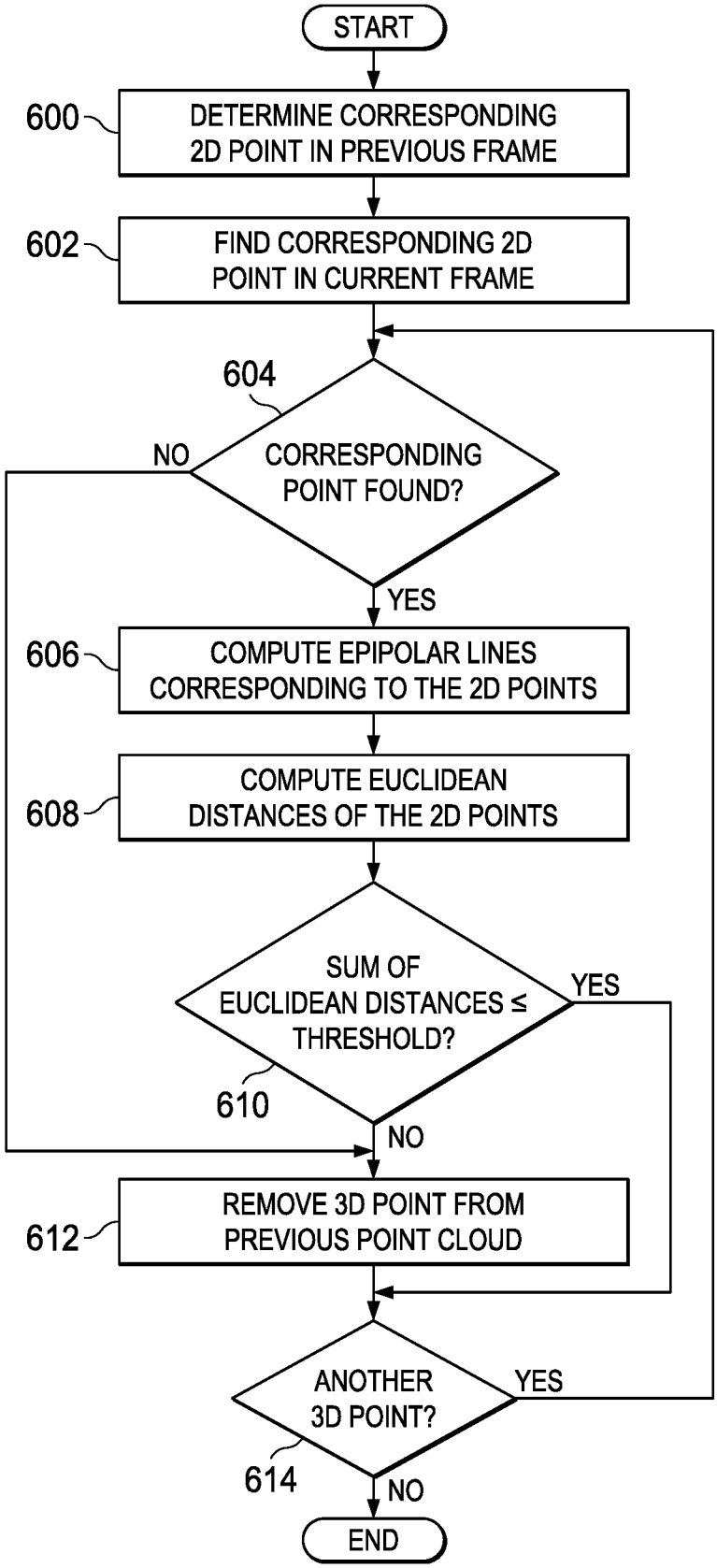
START
600
DETERMINE CORRESPONDING 2D POINT IN PREVIOUS FRAME
602
FIND CORRESPONDING 2D POINT IN CURRENT FRAME
604
CORRESPONDING POINT FOUND?
NO
YES
606
COMPUTE EPIPOLAR LINES CORRESPONDING TO THE 2D POINTS
608
COMPUTE EUCLIDEAN DISTANCES OF THE 2D POINTS
SUM OF EUCLIDEAN DISTANCES ≤ THRESHOLD?
YES
610
NO
612
REMOVE 3D POINT FROM PREVIOUS POINT CLOUD
ANOTHER 3D POINT?
YES
614
NO
END

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 2016/069142

A. CLASSIFICATION OF SUBJECT MATTER

G06T 7/246 *(2017.01)*
G01C 11/04 *(2006.01)*
G06T 17/00 *(2006.01)*

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G01C 11/00, 11/04-11/34, G06F 17/00, G06K 9/00, 9/20, 9/36, 9/46-9/50, G06T 1/00, 5/00, 7/00, 7/20-7/292, 7/70-7/80, 17/00-17/30, 19/00-19/20, H04N 13/00-13/04

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatSearch (RUPTO internal), USPTO, PAJ, K-PION, Esp@cenet, Information Retrieval System of FIPS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y A	US 2014/0010407 A1 (MICROSOFT CORPORATION) 09.01.2014, abstract, paragraphs [0002], [0005], [0008], [0023]-[0025], [0027]-[0028], [0031]-[0032], [0037]-[0039], [0042], [0046], [0048], [0053]	15 1-14, 16-20
Y A	US 2014/0139635 A1 (NEC LABORATORIES AMERICA, INC.) 22.05.2014, abstract, paragraphs [0009], [0031], [0033], [0038]-[0040], [0044], [0047], [0051]	15 1-14, 16-20
A	US 8442307 B1 (GOOGLE INC.) 14.05.2013	1-20
A	US 2014/0037189 A1 (QUALCOMM INCORPORATED) 06.02.2014	1-20

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 April 2017 (20.04.2017)	27 April 2017 (27.04.2017)

Name and mailing address of the ISA/RU:	Authorized officer
Federal Institute of Industrial Property, Berezhkovskaya nab., 30-1, Moscow, G-59, GSP-3, Russia, 125993 Facsimile No: (8-495) 531-63-18, (8-499) 243-33-37	M. Demchenko Telephone No. (499) 240-25-91