Ground Plane Feature Detection in Mobile Vision-Aided Inertial Navigation

Ghazaleh Panahandeh, Nasser Mohammadiha, and Magnus Jansson

Abstract—In this paper, a method for determining ground plane features in a sequence of images captured by a mobile camera is presented. The hardware of the mobile system consists of a monocular camera that is mounted on an inertial measurement unit (IMU). An image processing procedure is proposed, first to extract image features and match them across consecutive image frames, and second to detect the ground plane features using a two-step algorithm. In the first step, the planar homography of the ground plane is constructed using an IMU-camera motion estimation approach. The obtained homography constraints are used to detect the most likely ground features in the sequence of images. To reject the remaining outliers, as the second step, a new plane normal vector computation approach is proposed. To obtain the normal vector of the ground plane, only three pairs of corresponding features are used for a general camera transformation. The normal-based computation approach generalizes the existing methods that are developed for specific camera transformations. Experimental results on real data validate the reliability of the proposed method.

I. INTRODUCTION

Ground plane detection and obstacle removal are the main keys of detecting the moveable paths in, for example, vision-aided navigation systems, mobile robotics, traffic applications, and mosaicing. Vision-based ground plane detection frameworks are generally based on a single camera that can be used together with complementary sensors, such as a secondary camera (stereo-vision) [1]–[3], laser scanners [3], or inertial sensors [1], to increase the flexibility or performance of the system.

Depending on the hardware structure, vision-based ground plane detection approaches benefit from a priori knowledge of the scene or system structure; for instance, the geometric assumptions on the nature of the traffic environments [3]–[5], planar properties of indoor environments [6], color cue segmentations [7]–[9], and the assumption of the ground plane projected region on the image plane [4]. For instance, in [4] a method for road detection and obstacle detection is presented that is entirely based on stereovision. A least square fitting approach is used in disparity data for ground plane detection. However, they use assumptions of a planar road in the close proximity of the stereo couple and a known and fixed height from the ground. In addition, features are sampled from a trapezoidal-shaped region of interest in the lower part of the image. A similar traffic scenario ground plane detection approach is presented in [5] in which lane

markings painted on the road are used for computing the homography.

The plane geometrical constraints in terms of the homography or the plane normal, widely used in the literature, can be interpreted as the essential part of the vision-based plane detection approaches. However, in the existing methods either the ground plane features are manually determined to construct the initial planar homography or the camera is precisely calibrated relative to the ground plane; in addition, the camera movement is usually restricted to be planar [2], [8], [10], [11]. In the mobile robotics applications, the work of Lobo and Dias [1] is among the first researches in ground plane feature detection. In contrast to our method, their mechanical mounting system consists of an IMU and a dual-axis inclinometer coupled to a stereo camera. Furthermore among the most recent approaches, we can refer to [2] in which a modified expectation maximization approach for estimating homography of ground plane is introduced. However, their method needs an accurate initialization step from images of the ground floor to estimate the planar homography of the ground plane. Among the few single camera-used ground plane feature detection approaches, we can mention [10] in which the single camera provides a top view map of the robot's field of view in real-time. However, the robot motion is limited to be planar and the homography estimation is based on the ground plane features that are selected manually by choosing a rectangle in the lower area of the camera image.

Moreover, homography-based approaches are always suffering from “virtual plane false detection” [11]. Virtual plane false detection happens when a set of feature points that share the same homography are not lying on the same physical plane. Only few studies have addressed the effects of virtual plane false detection or how to overcome this issue [9], [11]. For instance, Zhou and Li [11] have introduced an ad-hoc method based on the color classification to avoid the virtual plane feature detection.

In this paper, we address the problem of robustly determining the ground plane features among the extracted and matched features between query images for a mobile vision system. The framework of the system consists of only one mobile monocular camera rigidly mounted on an IMU. To determine the ground plane features, first, the extracted image features are tested through the estimated planar homography. Additionally, a second-step outlier rejection method is proposed based on a new ground-plane normal computation approach using only three pairs of corresponding feature points. Furthermore, similar to [9], a

The authors are with the ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. {ghpa, nmoh, janssonm}@kth.se

simple stationarity test is performed to avoid virtual plane false detection. This additional step significantly improves the performance of the detection algorithm when the camera is stationary. The proposed algorithm is not dependent on the geometrical features of the scene nor on the geometrical position of the features in the navigation frame.

The main contributions of this work include the new plane-normal computation approach that generalizes the previously proposed method in [12]. In contrast to [12], our approach is not limited to the special case where the two cameras are aligned to the same orientation, and rather, a general transformation between cameras is considered. Moreover, the ground plane homography is not constructed manually, as it is usually done in the available systems [4], [10], and instead, an automatic approach is used from [13] that is based on a non-linear Kalman filter.

The paper is organized as follows. An overview of the system architecture is given in Section II. In Section III, theoretical geometric relations are derived. The proposed algorithm is described in Section IV and results are discussed in Section V. Finally, the conclusion of the study is given in Section VI.

II. SYSTEM OVERVIEW

A. Notation

In the following sections, scalars are denoted by lowercase letters (s), vectors by bold letters ($\mathbf{s}$), and matrices by bold capitals ($\mathbf{S}$).

The goal of our proposed method is to determine the ground plane features among all the extracted features in consecutive image frames received from a mobile IMU-camera sensor fusion system. Since the IMU provides the acceleration and the angular velocity of the rigid body with respect to the navigation frame $\{n\}$, this coordinate frame is considered as the reference frame. In order to simplify the treatment with different coordinate frames, we assume that the IMU coordinate frame $\{i\}$ is located in the center of the mobile body frame (moving object) $\{b\}$ (i.e., the $\{b\}$ coordinate frame is representative for both $\{i\}$ and the moving object); additionally, the transformation between the camera coordinate frame $\{c\}$ and the IMU coordinate frame is assumed to be known [14], [15].

Throughout this paper, $\mathbf{x}^y$ denotes a vector $\mathbf{x}$ expressed with respect to frame $\{y\}$, $\mathbf{R}_z^y$ is the rotation matrix rotating vectors from frame $\{z\}$ to frame $\{y\}$, and $\mathbf{p}_z^y$ is the origin of frame $\{z\}$, expressed with respect to frame $\{y\}$. The subscript i that is used with the frame indicator $\{y\}$ ($\mathbf{x}^{y_i}$) denotes the parameter $\mathbf{x}$ in the corresponding frame at time instant t_i .

B. System architecture

In recent years, IMU-camera sensor fusion systems have been widely used for motion estimation in several areas such as robot navigation and traffic applications [16], [17]. Under the inertial navigation system (INS) equations, the linear accelerations and rotational velocities provided by the IMU are integrated to generate the position, velocity, and attitude of a moving system. However, due to the integration drift,

Fig. 1: Architecture of the system. The camera ego-motion is estimated based on using the IMU data and the extracted image features. Then, the estimated motion of the system is used to reject the non-ground plane features.

the position and the attitude must be periodically corrected; otherwise, the system cannot be used over an extended period of time. The correction in the pose that is provided by the camera in the IMU-camera sensor fusion system can reduce the drift for long term operations. That is, the measurements from the camera provide salient features as keys of motion estimation on the image planes. Due to the dynamic nature of the system, it is natural to use a Kalman filter for fusion of the IMU output signals and camera image features over time for the motion estimation. Depending on the image processing tools and computer vision methods used to extract the pose information from the camera images, the state space model of the system is different in various Kalman filter framework-based methods; hence, the main differences in the current IMU-camera ego-motion methods are how to mimic the state of the system in the image plane to construct the measurement equations [13], [18].

Using an IMU-camera ego-motion estimation approach developed in [13], here we propose an automatic approach for determining image features on the ground plane. Images are captured by a mobile monocular camera mounted on an IMU. The block-diagram of the system is shown in Fig. 1. Since the details of the applied nonlinear ego-motion estimation is beyond the scope of this study, in the following only a short overview of the ego-motion estimation algorithm [13] is given.

Whenever new measurements from the IMU, linear acceleration $\mathbf{f}$ and rotational velocity $\mathbf{w}$, are received, the position and attitude of the body frame relative to the navigation frame— $\mathbf{p}_n^b$ and $\mathbf{R}_n^b$, see Fig. 1—are updated using the INS equations [19]. From each new received image, feature points are extracted and their corresponding points in the previous image are found. Then, epipolar geometry constraints between the matched features are used to correct the state of the system (function $f(\cdot)$ relates the state of system to the camera measurement) using a non-linear Kalman filter; the correction in the state is depicted by $\delta\mathbf{x}$.

The connection from the motion estimation to our feature detection approach is through the function $h(\cdot)$ in Fig. 1, which uses the corrected states of the system to construct the geometrical information required for ground plane feature detection that is explained in Section III.

Fig. 2: The IMU-camera sensor fusion system and the corresponding coordinate frames. The camera's relative rotation and translation between two frames are denoted by $\mathbf{R}$ and $\mathbf{t}$, respectively. The corresponding features located on ground plane, which are detected using the SURF method, are connected by colored lines between two images from left to right.

III. GEOMETRY OF THE SYSTEM

Planar homography and normal of a plane can be considered as the fundamental geometrical constraints that describe the relation between the coplanar points in different views [20]. In this section, we derive the geometrical constraints of the ground plane using the estimated camera pose from the non-linear Kalman filter (Fig. 1) and matched feature points between sequences of images. The obtained relations are then used in Section IV for determining ground plane features.

A. Planar homography of the ground plane

Given a static point $\pi^n \in \mathbb{R}^3$ in the navigation frame, its position relative to the camera frame is

$$\pi^c = \mathbf{R}_b^c (\mathbf{R}_n^b (\pi^n - \mathbf{p}_b^n) - \mathbf{p}_c^b), \quad (1)$$

where $\mathbf{R}_b^c$ is the body to camera rotation matrix and $\mathbf{p}_c^b$ is the position of the camera frame in body frame, which are assumed to be known [14], [15]. In a mobile vision system, the transformation of π^c between two time instances t_1 and t_2 is described by

$$\pi^{c2} = \mathbf{R} \pi^{c1} + \mathbf{t}, \quad (2)$$

where $\mathbf{R}$ and $\mathbf{t}$ are the relative camera rotation and translation, respectively, between the two time instances. In detail, this transformation is a function of the state variables estimated by the non-linear Kalman filter at two time instances t_1 and t_2 , Fig. 1, [13]:

$$\mathbf{R} = \mathbf{R}_b^c \mathbf{R}_n^{b2} (\mathbf{R}_b^c \mathbf{R}_n^{b1})^\top, \quad (3a)$$

$$\mathbf{t} = \mathbf{R}_b^c \mathbf{R}_n^{b2} (\mathbf{p}_{b1}^n + \mathbf{R}_n^{b1} \mathbf{p}_c^b - \mathbf{p}_{b2}^n) - \mathbf{R}_b^c \mathbf{p}_c^b. \quad (3b)$$

Without loss of generality, it is assumed that the navigation frame, n , is located in the ground plane as shown in Fig. 2. In this case, the normal of the ground plane in the camera coordinate frame is

$$\mathbf{n}^c = \mathbf{R}_b^c \mathbf{R}_n^b \mathbf{e}_z \quad (4)$$

where $\mathbf{e}_z = [0 \ 0 \ 1]^\top$ is the normal of the ground plane with respect to the navigation frame. Accordingly, it can be claimed that the scalar product of this normal with the points

lying on the ground plane is the orthogonal distance between the camera and the ground plane, d^c [20]:

$$\mathbf{n}^{c\top} \pi_j^c = d^c \quad \text{or} \quad \frac{1}{d^c} \mathbf{n}^{c\top} \pi_j^c = 1 \quad \forall \pi_j^c \in \text{ground plane}. \quad (5)$$

In principle, $d^c = \mathbf{e}_z^\top \mathbf{p}_c^n$ where $\mathbf{p}_c^n = \mathbf{p}_b^n + \mathbf{R}_n^b \mathbf{p}_c^b$, more details can be found in [18]. Finally by imposing (5) in (2), we obtain

$$\pi^{c2} = \mathbf{R} \pi^{c1} + \frac{\mathbf{t}}{d^{c1}} \mathbf{n}^{c1\top} \pi^{c1} = (\mathbf{R} + \frac{\mathbf{t} \mathbf{n}^{c1\top}}{d^{c1}}) \pi^{c1} = \mathbf{G} \pi^{c1}, \quad (6)$$

where $\mathbf{G}$ is the planar homography of the ground plane, which is the common transformation for all the points lying on the ground plane at two consecutive time instances t_1 and t_2 in which images are captured.

B. Transformations of points and lines

Using the pinhole camera model [20], the normalized projective coordinates of any point $\pi_j^c = [\pi_{x,j}^c, \pi_{y,j}^c, \pi_{z,j}^c]^\top$ will be referred to as

$$\mathbf{m}_j = \frac{1}{\pi_{z,j}^c} \pi_j^c \quad \forall j. \quad (7)$$

Hence, the homogeneous image coordinates $\mu_j = [u_j, v_j, 1]^\top$ (in pixels) for each point is $\mu_j = \mathbf{K} \mathbf{m}_j$, where $\mathbf{K}$ is the upper triangular matrix containing the camera intrinsic parameters [20]. From (7) and (2), the pixel correspondences in two images of the same plane satisfy the homography constraint as

$$\mathbf{m}_j^2 = \mathbf{H} \mathbf{m}_j^1 \quad \text{where} \quad \mathbf{H} = \gamma \mathbf{G} \in \mathbb{R}^{3 \times 3}, \quad (8)$$

and $\mathbf{m}_j^i$ represents the image of the j -th ground plane feature in the i -th image, and the equality holds up to a scale γ , based on the universal scale ambiguity [20].

Definition 1: Point $\mathbf{m}_j$ lies on a line $\mathbf{l}_k$ iff $\mathbf{m}_j \cdot \mathbf{l}_k = 0 \quad \forall j, k$ where $\{\cdot\}$ represents the dot product operator.

If a point $\mathbf{m}_j^1$ lies on a line $\mathbf{l}_j^1$, then the transformed point $\mathbf{m}_j^2$, under a projective transformation, lies on the line $\mathbf{l}_j^2$ [20] as

$$\mathbf{l}_j^1 = \mathbf{H}^\top \mathbf{l}_j^2. \quad (9)$$

Definition 2: The bi-intersection point of two corresponding lines, not proportional to each other, is equal (up to scale) to their cross product: $\mathbf{s}_{jk} \sim \mathbf{l}_j \times \mathbf{l}_k \quad \forall j, k$.

Definition 3: The line through any two points is given by their cross product: $\mathbf{l}_{jk} = \mathbf{m}_j \times \mathbf{m}_k \quad \forall j, k$.

In [12], Piazzini and Prattichizzo derived the normal vector of a plane using static stereo camera images for a special case where the two cameras are aligned to the same orientation $\mathbf{R} = \mathbf{I}$. In this paper, we generalize this approach to obtain the normal vector of a plane given three corresponding features for any alignment between cameras.

Theorem 1: Consider a plane and its planar homography $\mathbf{H}$ between a pair of images captured by a pair of cameras with relative rotation $\mathbf{R}$ and relative translation $\mathbf{t}$. Consider a set of lines living in the same plane. Define $\mathbf{l}_j^1$ as the projections of the lines in the first image, and $\mathbf{l}_j^2$ as the

projections of the lines in the second image that are rotated by $\mathbf{R}^\top$, as $\check{\mathbf{l}}_j^2 = \mathbf{R}^\top \mathbf{l}_j^1$. The intersection of $\mathbf{l}_j^1$ and $\check{\mathbf{l}}_j^2$ lives on the normal of the plane.

Proof: Denote the normal of the ground plane with respect to the first camera as $\mathbf{n}^{c1}$. Let $[\mathbf{n}^{c1}]_\times$ be the skew-symmetric matrix² corresponding to $\mathbf{n}^{c1}$. From (8), $\mathbf{H}^\top \sim \mathbf{R}^\top + \frac{\mathbf{n}^{c1} \mathbf{t}^\top}{d^{c1}}$, then we can write

$$[\mathbf{n}^{c1}]_\times \mathbf{H}^\top \sim [\mathbf{n}^{c1}]_\times (\mathbf{R}^\top + \frac{\mathbf{n}^{c1} \mathbf{t}^\top}{d^{c1}}) = [\mathbf{n}^{c1}]_\times \mathbf{R}^\top. \quad (10)$$

Then, multiplying both sides of (10) by the rotation matrix $\mathbf{R}$ in (3a) yields

$$\mathbf{R}[\mathbf{n}^{c1}]_\times \mathbf{H}^\top \sim \mathbf{R}[\mathbf{n}^{c1}]_\times \mathbf{R}^\top = [\mathbf{R}\mathbf{n}^{c1}]_\times. \quad (11)$$

This implies that $\mathbf{R}[\mathbf{n}^{c1}]_\times \mathbf{H}^\top$ is skew-symmetric, and for any skew symmetric matrix we can write

$$\mathbf{l}_j^{2\top} \mathbf{R}[\mathbf{n}^{c1}]_\times \mathbf{H}^\top \mathbf{l}_j^2 = 0. \quad (12)$$

Using (9), this is equivalent to

$$\mathbf{l}_j^{2\top} \mathbf{R}[\mathbf{n}^{c1}]_\times \mathbf{l}_j^1 = (\mathbf{l}_j^1 \times \mathbf{R}^\top \mathbf{l}_j^2)^\top \cdot \mathbf{n}^{c1} = 0. \quad (13)$$

Based on Definition 1 and 2, it can be concluded that the intersection of $\mathbf{l}_j^1$ and $\check{\mathbf{l}}_j^2$ lives in the normal of the ground plane. ■

Theorem 2: Consider three non collinear points π_1, π_2 , and π_3 viewed by two cameras with the relative rotation and translation $\mathbf{R}$ and $\mathbf{t}$, respectively. Denote their projections in the first and the second image plane as $\mathbf{m}_j^1$ and $\mathbf{m}_j^2$ for $j = 1, 2, 3$, respectively. Then the normal $\mathbf{n}$ of the plane characterized by these three pairs of projected points (if not coincident and not collinear) is given by the right null vector of the matrix

$$\mathbf{S} = \begin{bmatrix} (\mathbf{m}_1^1 \times \mathbf{m}_2^1) \times (\mathbf{R}^\top (\mathbf{m}_1^2 \times \mathbf{m}_2^2)) \\ (\mathbf{m}_2^1 \times \mathbf{m}_3^1) \times (\mathbf{R}^\top (\mathbf{m}_2^2 \times \mathbf{m}_3^2)) \\ (\mathbf{m}_3^1 \times \mathbf{m}_1^1) \times (\mathbf{R}^\top (\mathbf{m}_3^2 \times \mathbf{m}_1^2)) \end{bmatrix}^\top. \quad (14)$$

Proof: By Definition 3 the line passing through $\mathbf{m}_j^i$ and $\mathbf{m}_k^i$ in the i -th image is $\mathbf{l}_{jk}^i = \mathbf{m}_j^i \times \mathbf{m}_k^i$. Applying Theorem 1 to each corresponding pair of lines in the first and second image, and then rewriting it in matrix form we get

$$\mathbf{S}\mathbf{n} = 0 \quad \in \mathbb{R}^3 \quad \text{where} \quad \mathbf{S} = \begin{bmatrix} \mathbf{l}_{12}^1 \times \check{\mathbf{l}}_{12}^2 \\ \mathbf{l}_{23}^1 \times \check{\mathbf{l}}_{23}^2 \\ \mathbf{l}_{31}^1 \times \check{\mathbf{l}}_{31}^2 \end{bmatrix}^\top, \quad (15)$$

which implies that the right null space of the matrix $\mathbf{S}$ is the normal of the corresponding plane. A comprehensive study on the rank of matrix $\mathbf{S}$ is given in [12]. ■

²The cross product of two vectors $\mathbf{a}$ and $\mathbf{b}$ can be written as $\mathbf{a} \times \mathbf{b} = [\mathbf{b}]_\times \mathbf{a}$. In this paper, the following properties of the cross product skew-symmetric matrix are used: $[\mathbf{a}]_\times \mathbf{a} = 0$; $[\mathbf{A}\mathbf{a}]_\times = \mathbf{A}[\mathbf{a}]_\times \mathbf{A}^\top$; $\mathbf{b}^\top [\mathbf{a}]_\times \mathbf{b} = 0$; $\mathbf{b}[\mathbf{a}]_\times \mathbf{c} = (\mathbf{b} \times \mathbf{c})^\top \mathbf{a}$; $\forall \mathbf{A} \in \mathbb{R}^{3 \times 3}, \{\mathbf{a}, \mathbf{b}, \mathbf{c}\} \in \mathbb{R}^{3 \times 1}$.

IV. OUTLIER REJECTION AND GROUND PLANE FEATURE CLASSIFICATION

In this section, we present our method for classifying detected image feature points into ground or non-ground. When a new image is captured by the camera, its feature points are extracted and their corresponding matches in the previous image are obtained using the SURF feature detector and matching. A two-step algorithm is implemented to remove the non-ground feature points from the obtained set of feature points that consists of homography-based outlier rejection and normal-based outlier rejection.

A. Homography-based outlier rejection

In the homography-based outlier rejection, the estimated planar homography of the ground plane is applied to all the matched feature points between two images. The planar homography is calculated using the estimated camera transformation from the non-linear Kalman filter between any two consecutive images. The features that satisfy the homography constraint (8) are selected as the candidates of ground plane features. Since the estimated homography is erroneous, for most of the feature points the constraint (8) is satisfied up to an error. In this step, we select only a fraction of the feature points yielding the smallest errors to be processed in the next step.

B. Normal-based outlier rejection

A second step of outlier rejection is then applied to the detected features using the normal-based outlier rejection. Based on Theorem 2, any three corresponding feature points between two images can be used to obtain the normal of the plane that contains them. For this purpose, the matrix $\mathbf{S}$ in (15) is formed and the normal $\mathbf{n}$ of the plane is found by the singular value decomposition (SVD) as

$$\mathbf{n} \sim \mathbf{V}(:, 3) \quad \text{where} \quad \mathbf{S} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top. \quad (16)$$

That is, $\mathbf{n}$ is selected as the right singular vector of $\mathbf{S}$ corresponding to the smallest singular value. In general using a triangulation technique, the given set of feature points can be triangulated in one image. Then for each triangle and its corresponding triangle in the other image, the normal of the plane is computed using (16) and compared with the normal of the ground plane $\mathbf{n}^c$ (4). The vertices of the triangles whose normals are similar to the ground plane normal are finally selected as the ground plane feature points. We used cosine similarity to check the similarity between two normals; two normal vectors are similar if the calculated angle between them is less than a threshold T_{ang} . An important issue is to choose a good set of feature points for triangulation so that proper triangles can be obtained. A proper triangle should not be very small (to avoid effects of small errors in the matching), and should not be very large (so that the obtained triangles belong to physical planes in the navigation frame), see e.g. [8]. For this purpose, a method based on the Euclidean type of adjacency information is implemented, which efficiently uses all the candidate feature points. Hence, a neighboring set is obtained for each feature

point in which the mutual distances between all the members are between a minimum and a maximum threshold. The second step of our ground plane feature detection algorithm is summarized in Algorithm 1.

Algorithm 1 Normal-based ground plane feature detection

- 1) Calculate the adjacency matrix $\mathbf{A}$ for all the N candidate points, where A_{ij} is the Euclidean distance between points i and j .
 - 2) Label all the points as non-ground and set $k = 1$, set a minimum and a maximum threshold as $T_{\min}$ and $T_{\max}$.
 - 3) Consider k th point, if $\mathbf{m}_k$ is labeled as non-ground then:
 - a) Create a neighboring set for $\mathbf{m}_k$ as $\mathfrak{F}$; a point $\mathbf{m}_i$ is a member of $\mathfrak{F}$ if $\{i = k\}$ or $\{T_{\min} < A_{ki} < T_{\max} \text{ and there is not a point } \mathbf{m}_j \in \mathfrak{F} \text{ such that } A_{ij} < T_{\min}\}$.
 - b) Apply Delaunay triangulation to set $\mathfrak{F}$, [21].
 - c) For each obtained triangle and its corresponding match in the other image, calculate the normal using (16).
 - d) Check the cosine similarity between the obtained normal and the normal of the ground plane; if the normals are similar, label all the vertices of the triangle as ground.
 - 4) $k = k + 1$.
 - 5) If $k < N$ go to 3.
-

C. Virtual plane false detection

Until now, the necessary geometrical conditions for detecting ground plane features have been derived. Regardless of the motion estimation uncertainties, which are nonlinearly propagated to the applied geometrical constraints, the dynamics of the mobile vision system may cause conditions for which the preceding constraints are not sufficient for correct detection; that is, image features that do not lie in the same physical plane share the same planar homography which in turn will result in an incorrect detection, i.e., the virtual plane false detection [9], [11]. In short, the virtual plane false detection happens when the camera relative translation $\mathbf{t}$ is close to zero. So, Equation (6) loses its dependency on the plane descriptors $\{\mathbf{n}^{c1}, d^{c1}\}$. Moreover, the normal-based outlier rejection is not able to obtain a valid estimate of the ground plane normal vector because the matrix $\mathbf{S}$ in (15) approaches the zero matrix. Based on the estimated motion parameters, we perform a simple stationarity test by examining the camera displacement among the consecutive frames, similar to [9]. Hence whenever a stationary phase is encountered ($\mathbf{t} \approx 0$), the previous image is replaced with a preceding image for which the camera relative displacement is not close to zero.

V. EXPERIMENTAL RESULT AND DISCUSSION

The performance of the proposed algorithm has been evaluated using real data recorded by a wheeled IMU-camera sensor fusion system. A MicroStrain 3DMGX2 IMU, with sampling rate of 250 Hz, and an internally calibrated AVT

Fig. 3: (a): An AVT Guppy monochrome camera, with resolution 752×480 that is rigidly mounted on top of a MicroStrain 3DMGX2 IMU. (b): A sample estimated trajectory (provided by [13]) together with the floor map of the building in which the measurement campaign took place.

Guppy monocular camera, with resolution 752×480 sampled at 10 Hz, have been used for the experiment; see Fig. 3a.

The Matlab OpenSURF Toolbox v0.1c (Dirk-Jan Kroon) was used for feature detection and matching. Sequential data processing was done offline. In this experiment, only 10% of the feature points that result in the smallest errors in the homography constraint were selected for the normal-based outlier rejection step. This number was set empirically depending on the texture of the ground floor, i.e., if the ground floor has a rich texture, this number can be increased. It should be mentioned that the aim of the proposed algorithm is to minimize the number of incorrectly detected ground features rather than maximizing the number of the selected ground plane features. The threshold of the similarity angle, T_{ang} , was set to 10° ; while the performance of the detection algorithm was not found to be sensitive to small changes in the threshold and fairly similar performances were obtained for $5^\circ < T_{\text{ang}} < 10^\circ$. $T_{\min}$ and $T_{\max}$ were set to represent around 3cm and 30cm, which in our setup resulted in 10 and 80 pixels, respectively.

An example of the estimated trajectory, provided by [13], for the mobile system is plotted in Fig. 3b over the floor map of the building for about 110 seconds data acquisition. No calibration objects were used on the ground and the experiment was done in a usual lighting situation in corridors with poor texture of plastic parquet flooring. In the quantitative performance analysis, three different scene scenarios are selected as a representative for clutter regions and complex environments with different lighting situations. The red circle marked by P1, P2, and P3 in Fig. 3b indicate approximately the three different scene scenarios in which the results of the implemented algorithm are illustrated in Fig. 4. The detection results on the first row of Fig. 4 (a), (b), and (c) show the candidate ground plane features after applying the homography test, and images in the second row, (d), (e), and (f) show the detected ground plane features after the final outlier rejection. The detected features are labeled based on the error of the homography constraint. As the results show, the homography-based outlier rejection

is not sufficient for determining the ground plane features. In particular, the number of false detections are significantly reduced from Fig. 4 (b) to (e) in which very close obstacles limit the camera's field of view.

A careful study of Fig. 4 (a), (b), and (c) shows that most of the erroneously detected ground feature points are located around the edges where the ground plane meets a perpendicular plane. In addition to the motion estimation error, this might be due to the observation that the density of the feature points is usually high around the edges (in fact, due to the second order derivative approximations and feature point detection measures of the SURF method, many false feature points are detected along lines in the images), and also when the ground homography constraint is applied to these points, a relatively small error is obtained; hence, a large portion of these points can not be rejected in the homography-based outlier rejection. Although most of these points are rejected in the second stage (see Fig. 4 (d), (e), and (f)), still a portion of the incorrect detected feature points are located close to the edges. This is also a trade-off which occurs in the triangulation approach; since the remaining feature points in the perpendicular planes are quite close together, they are not permitted to form triangles which correspond to a real plane. Instead, these points might construct triangles with the ground feature points for which the normal vectors are similar to the normal vector of the ground plane. As a result, a potential improvement of the proposed algorithm might be addressed regarding this issue.

An example of the virtual plane false detection is shown in Fig. 5. Subfigure (a) illustrates the result of the detection when the estimated relative camera displacement between the two consecutive images was on the order of one millimeter; subfigure (b) shows the results on the same image frame after applying the stationarity test, i.e., the previous image was replaced with a preceding image for which the camera relative displacement was larger than one millimeter.

Table I provides a statistical measure of the achieved performance. Due to the unavailability of the ground truth, data collection was done manually; 10% of the 1000 image frames with the same setting were used for the evaluation. The sequence covers very different scene scenarios, e.g., narrow corridors, shadowed regions, and complex environments with a poor lighting situation. The table provides the results in two cases: the homography-based approach (in the first row), and homography-based together with the normal-based approaches (in the second row). Positive predictive value (PPV=number of the correctly detected ground features/total number of detected features) is used to evaluate the detection performance. The result shows that PPV is increased by about 10% when the homography-based detection algorithm is followed by the normal-based outlier rejection. The evaluation shows that after the normal-based outlier rejection the total number of the remaining features is decreased; this is the given penalty to achieve higher correct detection rate; this number of detected features suffices in most of the related applications (e.g. [2], [8], [18]).

It should be mentioned that if the algorithm, Algorithm

Fig. 5: Comparison of feature verification when the estimated camera relative displacement between the current and previous image frames is less than 1mm. Subfigure (a), virtual plane false detection using the two successive images. Subfigure (b), detected features after applying the stationarity test.

TABLE I: Statistics of the proposed method

	Total number of detected features	Number of incorrectly detected features	PPV
Homography-based	56,96	8,18	85%
Homography-based+normal-based	26,22	1,34	95%

1, starts by processing the feature points with the smallest homography errors, many of the ground feature points will be recognized just after a few iterations, and since step 3 of the algorithm is performed only for non-ground feature points, the computational complexity will be reduced considerably. Our Matlab implementation for the two-stage ground plane feature detection (except for the SURF feature extraction and matching and motion estimation) takes around 0.2 seconds using an Intel(R) Core(TM)2 Duo processor computer at 2.4 GHz with 4 GB of memory. We believe that a more efficient implementation of step 3 of the proposed Algorithm 1 would reduce the computational complexity enough for real-time applications. Video clips and supplemental materials of the experiment are available at IEEE Xplore and <http://www.ee.kth.se/~ghpa/gpd>.

VI. CONCLUSION

An image processing approach is proposed for extracting ground plane features among the detected features from sequential images received from a mobile vision system. The hardware of the system is composed of a monocular camera and an IMU, which are rigidly mounted together. Exploiting the complementary nature of the IMU-camera

Fig. 4: Subfigure (a), (b), and (c) represent the candidate feature points after the homography-based outlier rejection. Subfigure (d), (e), and (f) show the final detected ground feature points after the normal-based outlier rejection.

sensor fusion system for estimating the camera translation and rotation, the developed algorithm consists of two parts, namely homography-based and normal-based outlier rejection. In contrast to the existing methods, our proposed approach is not scene dependent. Additionally, no restriction on the camera movement is used in the algorithm; hence, the proposed method can be used both for mobile robot and pedestrian navigation systems. Currently, we are investigating the integration of the proposed ground plane feature detection algorithm to assist the navigation solution to improve the accuracy of the motion estimation and the results will be reported in a future publication.

VII. ACKNOWLEDGMENTS

The authors would like to thank Dave Zachariah for kindly providing us with the ego-motions estimated signals from their algorithm [13].

REFERENCES

- [1] Jorge Lobo and Jorge Dias, "Ground plane detection by fusing visual and inertial information," in *Int. Workshop on Advanced Motion Control*, 1998.
- [2] D. Conrad and G.N. DeSouza, "Homography-based ground plane detection for mobile robot navigation using a modified EM algorithm," in *Proc. of ICAR*, May. 2010, pp. 910–915.
- [3] Christoph Stiller, J. Hipp, C. Rssig, and A. Ewald, "Multisensor obstacle detection and tracking," *Image and Vision Computing*, vol. 18, pp. 389–396, 2000.
- [4] Paolo Lombardi, Michele Zanin, and Stefano Messelodi, "Unified stereovision for ground, road, and obstacle detection," in *IEEE Conf. on Intelligent Vehicles Symposium*, Jun. 2005, pp. 783–788.
- [5] J. Arrospe, L. Salgado, M. Nieto, and R. Mohedano, "Homography-based ground plane detection using a single on-board camera," *Intelligent Transport Systems, IET*, vol. 4, no. 2, pp. 149–160, Jun. 2010.
- [6] Young geun Kim and Hakil Kim, "Layered ground floor detection for vision-based mobile robot navigation," in *Proc. of ICAR*, 2004, vol. 1, pp. 13–18.
- [7] Luis J. Manso, Pablo Bustos, Pilar Bachiller, and Jose Moreno, "Multi-cue visual obstacle detection for mobile robots," *Journal of physical agents*, vol. 4, Jan. 2010.
- [8] Nick Pears and Bojian Liang, "Ground plane segmentation for mobile robot visual navigation," in *Proc. of ICAR*, 2001.
- [9] Chia How Lin, Sin Yi Jiang, Yueh Ju Pu, and Kai-Tai Song, "Robust ground plane detection for obstacle avoidance of mobile robots using a monocular camera," in *Proc. of IROS*, Oct. 2010, pp. 3706–3711.
- [10] Tilman Wekel, Olaf Kroll-Peters, and Sahin Albayrak, "Vision based obstacle detection for wheeled robots," in *Int. Conf. on Control, Automation and Systems*, Oct. 2008, pp. 1587–1592.
- [11] Jin Zhou and Baoxin Li, "Robust ground plane detection with normalized homography in monocular sequences from a robot platform," in *IEEE Int. Conf. on Image Processing*, Oct. 2006.
- [12] Jacopo Piazzi and Domenico Prattichizzo, "Plane detection with stereo images," in *Proc. of ICAR*, 2006.
- [13] Dave Zachariah and Magnus Jansson, "Camera-aided inertial navigation using epipolar points," in *Proc. of PLANS*, 2010.
- [14] F.M. Mirzaei and S.I. Roumeliotis, "A kalman filter-based algorithm for imu-camera calibration," in *Proc. of IROS*, 2007.
- [15] Ghazaleh Panahandeh and Magnus Jansson, "IMU-camera self-calibration using planar mirror reflection," in *Proc. of IPIN*, 2011.
- [16] Peter Corke, Jorge Lobo, and Jorge Dias, "An introduction to inertial and visual sensing," *Int. Journal of Robotics Research*, vol. 26, no. 6, pp. 519–535, 2007.
- [17] Francisco Bonin-Font, Alberto Ortiz, and Gabriel Oliver, "Visual navigation for mobile robots: A survey," *Journal of Intelligent and Robotic Systems*, 2008.
- [18] Ghazaleh Panahandeh and Magnus Jansson, "Vision-aided inertial navigation using planar terrain features," in *IEEE Int. Conf. on Robot, Vision and Signal Processing*, Sep. 2011.
- [19] Jay A. Farrell and Matthew Barth, *Global Positioning System, Inertial Navigation and Integration*, McGraw-Hill Companies, 1999.
- [20] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521623049, 2000.
- [21] D. T. Lee and B. J. Schachter, "Two algorithms for constructing a delaunay triangulation," *Int. Journal of Computer and Information Sciences*, vol. 9, no. 3, pp. 219–242, 1980.