Vision-based Terrain Referenced Navigation for Unmanned Aerial Vehicles using Homography Relationship

Dongjin Lee · Yeongju Kim · Hyochoong Bang

Received: 11 July 2012 / Accepted: 24 July 2012 / Published online: 18 August 2012

Abstract A vision-based terrain referenced navigation (TRN) system is addressed for autonomous navigation of unmanned aerial vehicles (UAVs). A typical TRN algorithm blends inertial navigation data with measured terrain information to estimate vehicle's position. In this paper, however, we replace the low-cost inertial navigation system (INS) with a monocular vision system. The homography decomposition algorithm is utilized to estimates the relative translational motion using features on the ground with simple assumptions. A numerical integration point-mass filter based on Bayesian estimation is employed to combine the translation information obtained from the vision system with the measured terrain height. Numerical simulations are constructed to evaluate the performance of the proposed method. The results show that the precise autonomous navigation of unmanned aircrafts is achieved by the vision-based TRN algorithm.

Keywords Terrain referenced navigation · Unmanned aerial vehicle · Vision based navigation · Homography decomposition

1 Introduction

In practice, navigation systems such as an inertial navigation system (INS) and a global positioning system (GPS) are used to estimate the position of a vehicle. However, due to initial errors and measurement errors, the position estimate of the INS will drift away from the actual position of UAVs [1]. Also, the position data of the GPS can be unavailable by the blocking or the jamming of the external satellite signals. In order to overcome the drawback of the INS and the GPS, in this paper, we propose a vision-based terrain referenced navigation (TRN) system. TRN is the concept of using terrain information in order to estimate the position of vehicles such as UAVs. A typical TRN algorithm blends INS data with measured terrain information and matches with the stored digital terrain elevation database (DTED). In this framework, the accuracy and the quality of the TRN system relies on the performance of the INS. That is, low-cost UAVs that are not equipped with the high quality INS cannot obtain accurate data. For this reason, we replace the low-cost INS with

D. Lee (✉) · Y. Kim · H. Bang
Division of Aerospace Engineering, School of Mechanical, Aerospace & Systems Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Korea
e-mail: djlee@ascl.kaist.ac.kr

Y. Kim
e-mail: yjkim@ascl.kaist.ac.kr

H. Bang
e-mail: hcbang@ascl.kaist.ac.kr

a vision system using a monocular camera. Thus, the low-cost UAVs can gather the position information using a camera and a radar altimeter with the DTED map. There are several approaches to construct the TRN algorithm. For instance, an extended Kalman filter (EKF) [2], a bank of Kalman filter [3], an unscented Kalman filter (UKF) [4], a particle filter (PF) [5] and a point-mass filter (PMF) [1] have been applied to the TRN problems. In spite of computational burden, we employ the point-mass filter. The point-mass filter is a quantized version of Bayesian estimation and is robust to various terrain profiles and large initial errors. The input data of the point-mass filter is the translational motion of the UAVs and the vision system is utilized to estimate the motion.

The information obtained from camera images suffices for precise motion estimation based on visual information [6], called visual odometry. Current approaches to estimating the vehicle state through a camera system utilize the motion of feature points in an image. A geometric approach is proposed in [7, 8] that uses a series of homography relationships to estimate position and orientation with respect to an inertial pose. In this paper, we present visual odometry of monocular camera system using homography relationship as a reliable substitute for the INS on the propagation stage of the TRN filter.

The point-mass filter based TRN algorithm is stated and derived in Section 2 and the vision system using homography relationships is described in Section 3. The integration of the TRN with the vision system is implemented in Section 4 and the performance of the proposed approach is evaluated in Section 5. Finally, conclusions are drawn in the last section.

2 Terrain Referenced Navigation

2.1 System Model

The translational motion of the UAV obeys a simple linear equation

$$x_{t+1} = x_t + u_t + w_t \tag{1}$$

where x_t, u_t, w_t denote the current UAV position, the translational movement, a white Gaussian process noise respectively. And the terrain height y_t relates to the current UAV position x_t with a white Gaussian measurement noise v_t.

$$y_t = z\left(x_t\right) + v_t \tag{2}$$

where w_t and v_t are mutually independent white processes and are distributed according to the normal distribution p_{w_t} and p_{v_t}. These two Eqs. 1 and 2 yield the nonlinear estimation model.

2.2 Bayesian Estimation

The estimation problem associated with the TRN is to find the best matched position in the DTED using the set of measurements (Fig. 1). Using a statistical view of the problem, the probability density function (PDF) for the states gives all information one can ask for regarding the characteristics of the states [1]. Thus, the a posteriori density function $p_{x_t|Y_t}(x)$ contains all information about the states x_t given the collected set of measurements. We can define the observation set $\mathbf{Y}_t$ collected until present time t

$$\mathbf{Y}_t = \{y_i\}_{i=0}^{t} \tag{3}$$

The distribution of a generic random variable z conditioned on another related random variable w [1] is

$$p\left(z|w\right) = \frac{p\left(z,w\right)}{p\left(w\right)} = \frac{p\left(w|z\right)p\left(z\right)}{p\left(w\right)} \tag{4}$$

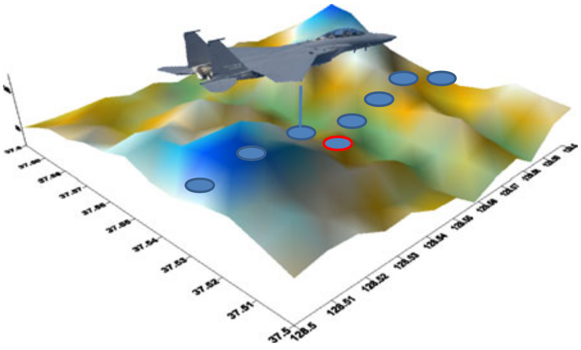

Fig. 1 Terrain referenced navigation

If we know the a priori $p(x_t|\mathbf{Y}_{t-1})$, then the posterior can be derived using Eq. 4

$$p\left(x_t|\mathbf{Y}_t\right)=\frac{p\left(y_t|x_t,\mathbf{Y}_{t-1}\right)p\left(x_t|\mathbf{Y}_{t-1}\right)}{p\left(y_t|\mathbf{Y}_{t-1}\right)} \tag{5}$$

where $p(y_t|\mathbf{Y}_{t-1})$ is a normalizing constant and the joint PDF for the state at time x_t and x_{t+1} can be calculated using Eq. 4

$$p\left(x_{t+1},x_t\right)=p\left(x_{t+1}|x_t\right)p\left(x_t\right) \tag{6}$$

By marginalizing on the state x_t, the finalized Bayesian formulas for the time update and measurement update are represented as follows

$$\begin{aligned}p\left(x_t|\mathbf{Y}_t\right)&=\alpha_t^{-1}p_{v_t}\left(y_t-z(x_t)\right)p\left(x_t|\mathbf{Y}_{t-1}\right)\\ p\left(x_{t+1}|\mathbf{Y}_t\right)&=\int p_{w_t}\left(x_{t+1}-x_t-u_t\right)p\left(x_t|\mathbf{Y}_t\right)dx_t\end{aligned} \tag{7}$$

where

$$\alpha_t=\int p_{v_t}\left(y_t-z\left(x_t\right)\right)p\left(x_t|\mathbf{Y}_{t-1}\right)dx_t \tag{8}$$

Given the posterior in Eq. 7, the estimate can be found by the minimum mean square error (MMSE) estimation

$$\hat{x}_t=\int x_t p\left(x_t|\mathbf{Y}_t\right)dx_t \tag{9}$$

2.3 Point Mass Filter

In order to implement the finalized Bayesian estimation formulas, an approximation technique such as a grid based point-mass filter is employed. We introduce N grid points over the position domain $\mathbf{R}^2$ and all integral operations in Eq. 7 to Eq. 9 are approximated by a finite sum over the gird points

$$\int f\left(x_t\right)dx_t\approx\sum_{k=1}^{N}f\left(x_t\left(k\right)\right)\delta^2 \tag{10}$$

where δ is the resolution of the grid points. As a result, we can rewrite the Bayesian formulas into approximated form

$$\begin{aligned}p\left(x_t(k)|\mathbf{Y}_t\right)&=\alpha_t^{-1}p_{v_t}\left(y_t-z\left(x_t\left(k\right)\right)\right)p\\ &\quad\times\left(x_t\left(k\right)|\mathbf{Y}_{t-1}\right)\\ p\left(x_{t+1}\left(k\right)|\mathbf{Y}_t\right)&=\sum_{n=1}^{N}p_{w_t}\left(x_{t+1}\left(k\right)-x_t\left(n\right)-u_t\right)p\\ &\quad\times\left(x_t\left(k\right)|\mathbf{Y}_t\right)\delta^2\end{aligned} \tag{11}$$

where

$$\alpha_t=\sum_{k=1}^{N}p_{v_t}\left(y_t-z\left(x_t\left(k\right)\right)\right)p\left(x_t\left(k\right)|\mathbf{Y}_{t-1}\right)\delta^2 \tag{12}$$

And the MMSE yields

$$\hat{x}_t=\sum_{k=1}^{N}x_t\left(k\right)p\left(x_t\left(k\right)|\mathbf{Y}_t\right)\delta^2 \tag{13}$$

The grid points have the form of a uniform mesh of size $(m\times n)$ over $\mathbf{R}^2$.

3 Homography Decomposition

Translation of the UAV between two time instances can be obtained by constructing homography relationship from two successive images of monocular camera planted on the UAV. Consider a body-fixed coordinate frame F_c that defines the position and attitude of a camera, with respect to a navigation frame F_n. The rotation and translation of F_c, with respect to F_n, is defined as $R\in\mathbf{R}^{3\times3}$ and $T\in\mathbf{R}^3$, respectively. The camera rotation and translation of F_c between two time instances, t_0 and t_1, is denoted by R_{01}, T_{01}. During the camera motion, a collection of n (where $n\geq4$) coplanar and noncolinear static feature points are assumed to be visible in a plane. The assumption of four coplanar and noncolinear feature points is only required to simplify the subsequent analysis and is made without loss of generality. If four coplanar target points are not available, then the subsequent development can also exploit a variety of solutions for more noncoplanar points [9–12].

Consider a feature point $\mathbf{X}_i(t) = [X_i, Y_i, Z_i]^T \in \mathbf{R}^3 \forall i \in \{1, ..., n\}$ defined in F_c. Standard geometric relationships can be applied to the coordinate systems depicted in Fig. 2 to develop the following relationship

$$\mathbf{X}_i(t_1) = R_{01}\mathbf{X}_i(t_0) + T_{01} \tag{14}$$

The relationship Eq. 14 can be rewritten as

$$\begin{aligned}\mathbf{X}(t_1) &= R_{01}\mathbf{X}(t_0) + T_{01} \\ &= R_{01}\mathbf{X}(t_0) + T_{01}\frac{1}{d(t_0)}\mathbf{N}^T(t_0)\mathbf{X}(t_0) \\ &= \left(R_{01} + \frac{1}{d(t_0)}T_{01}\mathbf{N}^T(t_0)\right)\mathbf{X}(t_0) \\ &= H\mathbf{X}(t_0)\end{aligned} \tag{15}$$

where $H \in \mathbf{R}^{3\times3}$ is the Euclidean homography matrix, $\mathbf{N}^T(t_0)$ is the constant unit vector normal to the plane P from F_c at time t_0, and $d(t_0)$ is the constant distance between the plane P and F_c along $\mathbf{N}^T(t_0)$. Further details on the Euclidean homography can be found in [13].

Using standard projective geometry the Euclidean coordinate $\mathbf{X}_i(t)$ can be expressed in image-space pixel coordinates as $\mathbf{p}_i(t) = [u_i, v_i, 1]^T \in \mathbf{R}^3$. After normalizing the Euclidean coordinates as

$$\mathbf{x}_i(t) = \left[\frac{X_i}{Z_i}, \frac{Y_i}{Z_i}, 1\right]^T \tag{16}$$

the relationship Eq. 14 becomes

$$\mathbf{x}_i(t_1) = \frac{Z_i(t_0)}{Z_i(t_1)}H\mathbf{x}_i(t_0) = \alpha_i H\mathbf{x}_i(t_0) \tag{17}$$

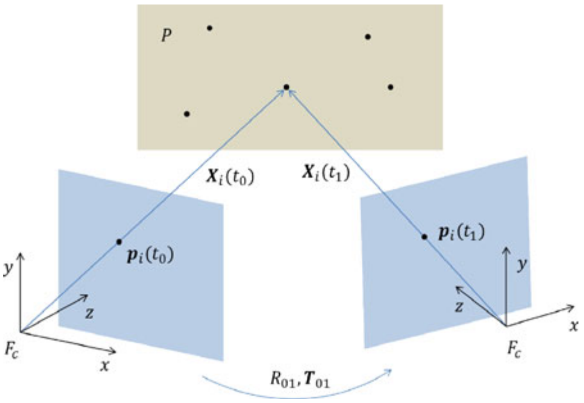

Fig. 2 Euclidean relationships between two camera poses

The projected pixel coordinates are related to the normalized Euclidean coordinates $\mathbf{x}_i(t)$ by the pin-hole camera model [13] as

$$\mathbf{p}_i(t) = A\mathbf{x}_i(t) \tag{18}$$

where A is an invertible, upper triangular camera calibration matrix defined as

$$A = \begin{bmatrix} a & -a\cos\varphi & u_0 \\ 0 & \dfrac{b}{\sin\varphi} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{19}$$

In Eq. 19, u_0 and $v_0 \in \mathbf{R}$ denote the pixel coordinates of the principal point (the image center as defined by the intersection of the optical axis with the image plane), a and $b \in \mathbf{R}$ represent scaling factors of the pixel dimensions, and $\phi \in \mathbf{R}$ is the skew angle between camera axes.

By using Eq. 18, the Euclidean relationship in Eq. 17 can be expressed as

$$\begin{aligned}\mathbf{p}_i(t_1) &= \alpha_i AHA^{-1}\mathbf{p}_i(t_0) \\ &= \alpha_i G\mathbf{p}_i(t_0) = G_L\mathbf{p}_i(t_0)\end{aligned} \tag{20}$$

Sets of linear equations can be developed from Eq. 20 to determine the projective and Euclidean homography matrix G_L and H up to a scalar multiple. We need to construct the homography matrix H from the pixel measurements $\mathbf{p}_i(t_0)$ and $\mathbf{p}_i(t_1)$.

In order to eliminate the unknown scale in Eq. 20, multiplying both sides by the skew-symmetric matrix $\hat{\mathbf{p}}(t_0) \in \mathbf{R}^{3\times3}$, we obtain the equation

$$\hat{\mathbf{p}}(t_1)G_L\mathbf{p}(t_1) = 0 \tag{21}$$

This equation is called planar epipolar constraint, or also the (planar) homography constraint. Since Eq. 21 is linear in G_L, by stacking the entries of G_L as a vector,

$$\begin{aligned}\mathbf{G}_L^s \equiv &\big[G_{L11}\ G_{L21}\ G_{L31}\ G_{L12}\ G_{L22}\ G_{L32} \\ &\times G_{L13}\ G_{L23}\ G_{L33}\big]^T\end{aligned} \tag{22}$$

we may rewrite Eq. 21 as

$$a^T\mathbf{G}_L^s = 0 \tag{23}$$

where the matrix $a \in \mathbf{R}^{9\times3}$ is the Kronecker product of $\hat{\mathbf{p}}(t_1)$ and $\mathbf{p}(t_0)$,

$$a \equiv \mathbf{p}(t_0) \otimes \hat{\mathbf{p}}(t_1) \tag{24}$$

Since the matrix $\hat{\mathbf{p}}(t_1)$ is only of rank 2, so is the matrix a. Thus, even though Eq. 21 has three rows, it only imposes two independent constraints on G_L. With this notation, given n pairs of images $\{(\mathbf{p}_i(t_0), \mathbf{p}_i(t_1))\}$ from points on the same plane P by defining

$$\chi \equiv [a_1\, a_2 \ldots a_n]^T \in \mathbf{R}^{3n\times 9} \tag{25}$$

we may combine all the equations in Eq. 21 for all the image pairs and rewrite them as

$$\chi \mathbf{G}_L^s = 0 \tag{26}$$

In order to solve uniquely (up to a scalar factor) for G_L, we must have $rank(\chi) = 8$. Since each pair of image pair of image points gives two constrains, we expect that at least four point correspondences would be necessary for a unique estimate of G_L.

Thus, if there are more than for image correspondences of which no three in each image are collinear, we may apply standard linear least-squares estimation to find

$$\min \left\| \chi \mathbf{G}_L^s \right\|^2 \tag{27}$$

to recover G_L up to a scalar factor. When we compute singular value decomposition (SVD) of χ,

$$\chi = U_\chi \Sigma_\chi V_\chi \tag{28}$$

$\mathbf{G}_L^s$ is defined to be the ninth column of V_χ. G_L is obtained by unstacking the nine elements of $\mathbf{G}_L^s$ into a square 3 × 3 matrix. G_L can be recovered up to a form as

$$G = \pm A \left(R + \frac{1}{d} T \mathbf{N}^T \right) A^{-1} \tag{29}$$

by normalizing G_L by its second-largest singular value σ_2 as

$$G = \frac{G_L}{\sigma_2} \tag{30}$$

To get the correct sign, we may impose the positive depth constraint as follows

$$\mathbf{p}_i(t_1)^T G\, \mathbf{p}_i(t_0) > 0,\ \forall i = 1, \ldots, n \tag{31}$$

Lastly, the homography matrix is obtained by following relationship with the camera calibration matrix as

$$H = A^{-1} G A \tag{32}$$

Various techniques [13, 14] can be used to decompose the homography matrix to obtain α_i, $\mathbf{N}^T(t_0)$, $T_{01}/d(t_0)$ and R_{01}. The decomposition methods generally return two physically valid solutions. The correct solution can be chosen by knowledge of the correct value for the normal vector or by using an image taken at a third pose. The distance $d(t_0)$ can be measured separately by the radar altimeter.

4 Visions-based Terrain Referenced Navigation

The vison-based TRN algorithm can be yielded by replacement of the INS with the vision system. The estimation model is rewritten by

$$\begin{aligned} x_{t+1} &= x_t + T_{01_t} + w_t^{vision} \\ y_t &= z(x_t) + v_t \end{aligned} \tag{33}$$

where T_{01_t} is the translational solution of the homography decomposition. In this framework, since the feature points are taken from the different terrain height we set the nominal feature height to the terrain height below the UAV. This assumption yields errors to T_{01_t}. However, the proposed TRN algorithm bounds the errors and derives the best estimate. Figure 3 shows the diagram of the proposed vision based TRN algorithm.

5 Numerical Simulation

5.1 Simulation Condition

In order to evaluate the proposed algorithm, we construct two scenarios using different terrain

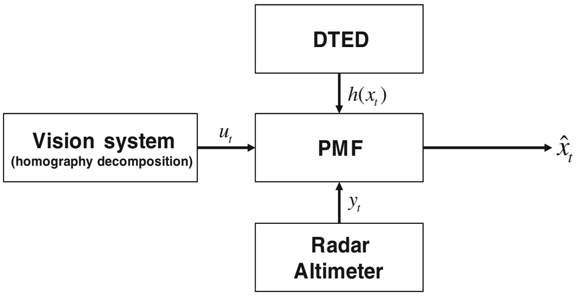

Fig. 3 Vision-based TRN architecture

Table 1 Simulation condition

Object	Parameters	Value
UAV	Crusing altitude	2 km
	Speed	30 m/s
DTED	Resolution	3 arcsec
	Size	5 km × 5 km
Camera	Field of view	45 deg
	Sampling	1 Hz
	Pixel random noise (1σ)	0.5 pixel
Radar altimeter	Random noise (1σ)	5 m
	Bias	0.7 m
INS	Accelerometer random noise (1σ)	$0.22\ m/s/\sqrt{hr}$
	Accelerometer bias	1 mg
	Gyro random noise (1σ)	$0.125\ \deg/\sqrt{hr}$
	Gyro bias	0.1 deg/hr

profiles: 1) smooth terrain and 2) rough terrain. In each scenario, we compare the error characteristic of the pure INS, vision only and vision-based TRN. The feature points are generated randomly and the average number of the points is 5 in each 100 m × 100 m area. The INS model is referred to [2] (Table 1).

5.2 Scenario I

This scenario is composed with the smooth terrain profile.

Figure 4 shows the smooth terrain map used in scenario I. The UAV flies over the map from the south to the north. Also we perform the Monte Carlo simulation in 1000 times. From the results, we can find that the proposed vision-based TRN algorithm bounds the error of the vision system (Figs. 5, 6, 7 and 8). During the Monte Carlo simulation, the RMS error is around 30 m and

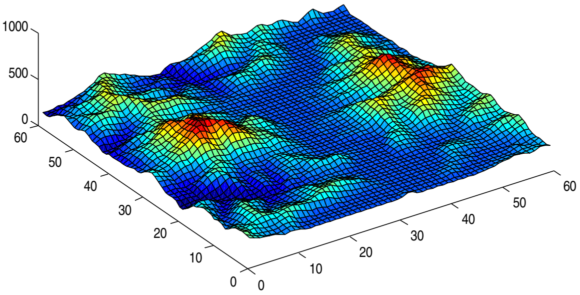

Fig. 4 Smooth DTED map

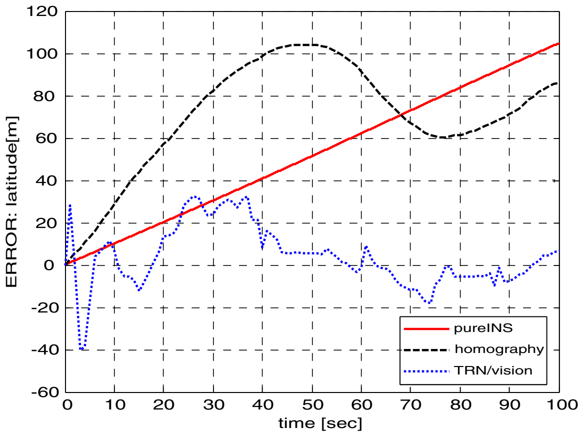

Fig. 5 Latitude error

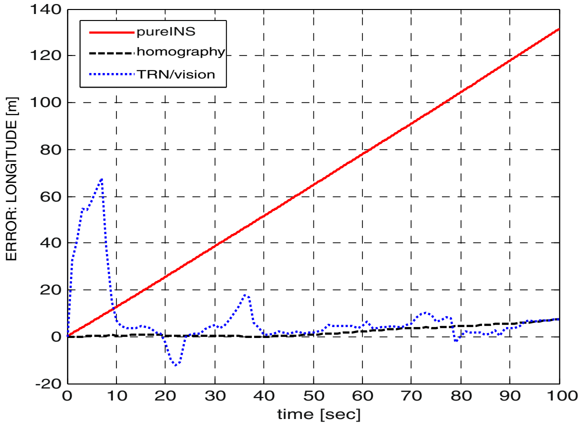

Fig. 6 Longitude error

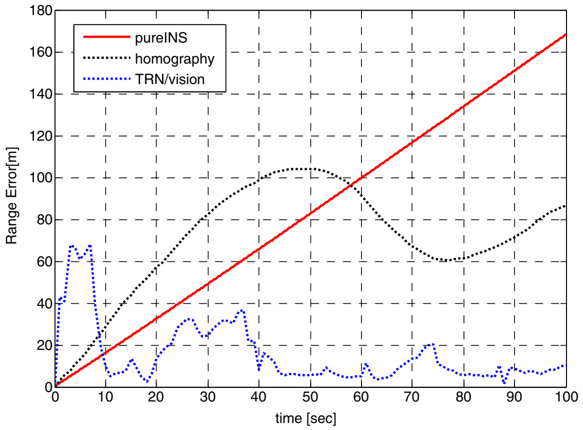

Fig. 7 Range error

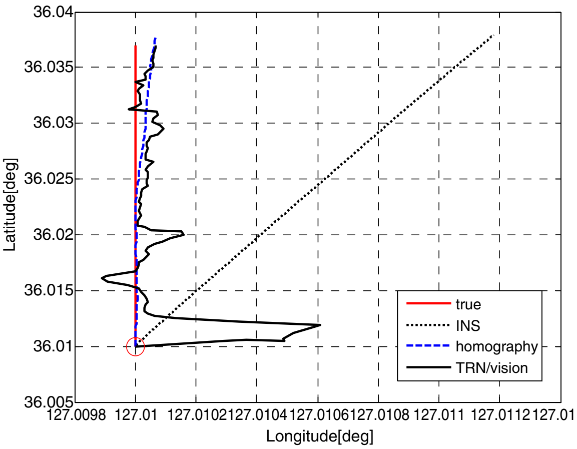

Fig. 8 Horizontal trajectory

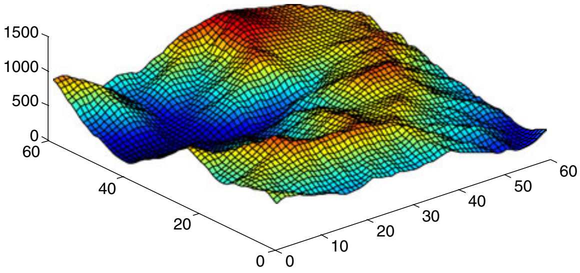

Fig. 11 Rough DTED map

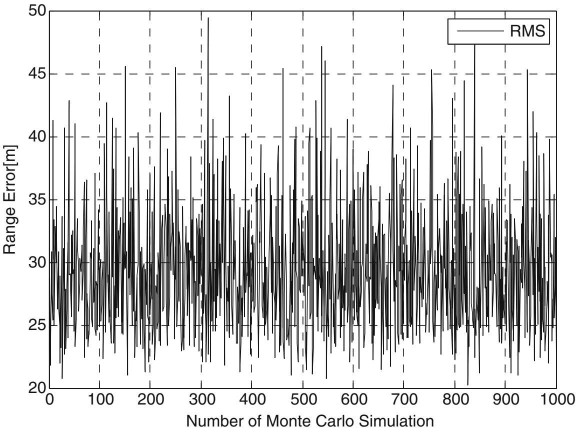

Fig. 9 RMS error

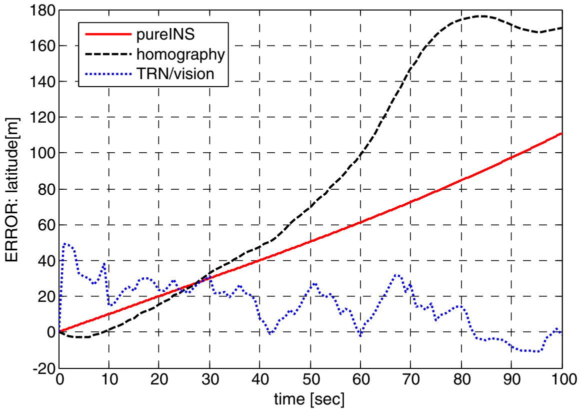

Fig. 12 Latitude error

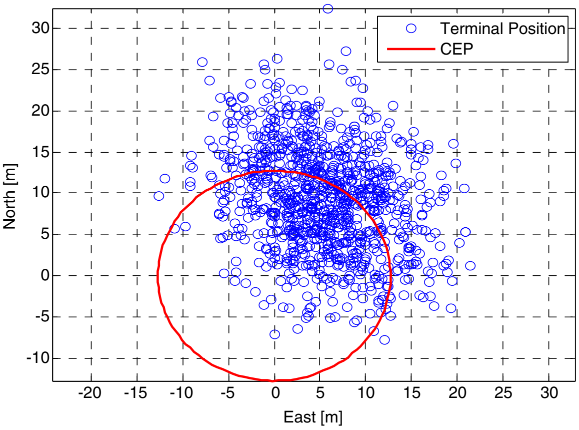

Fig. 10 Termial position error and circular error probable

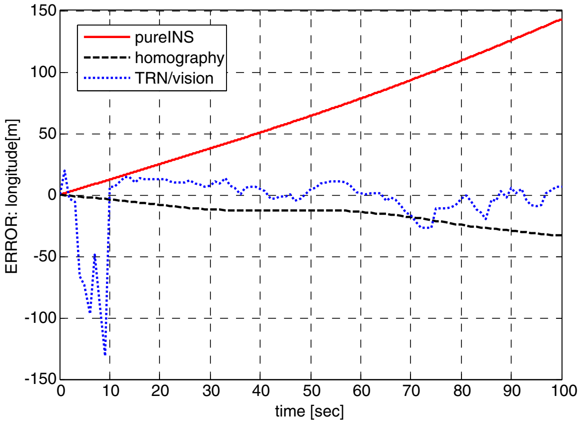

Fig. 13 Longitude error

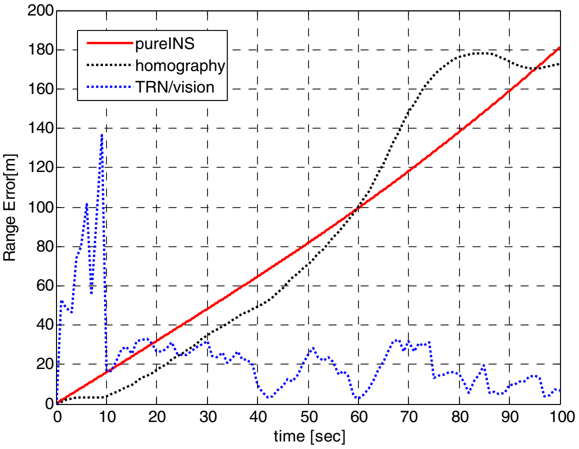

Fig. 14 Range error

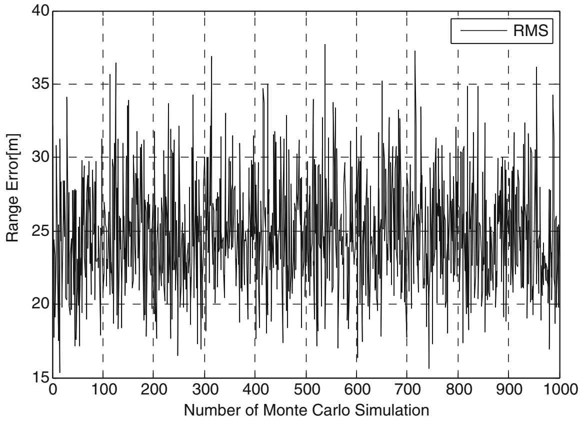

Fig. 16 RMS error

the CEP is about 13 m (Figs. 9 and 10). The mean of the terminal error is biased since the terrain profile is smooth. The error of the INS and the homography relation drifts away without the terrain aided navigation. We assumed that there are no initial errors in the vision system. After 40 s, there is a trend that filtered position follows the solution of homography decomposition. This is the limit of the TRN algorithm over the smooth area. In other words, the terrain information is not observable.

5.3 Scenario II

This scenario is constructed with the rough terrain. As mention previously, the TRN algorithm estimate more accurate position in the rough terrain than the smooth terrain. And the errors from the vision system increase in the rough terrain. Figure 11 shows the rough terrain map used in scenario II. The UAV flies over the map from the south to the north. In this scenario, also, we perform the Monte Carlo simulation in 1000 times. The results show that the proposed algorithm works properly (Figs. 12, 13, 14 and 15).

Compared to the results of scenario I, the accuracy of the estimate is improved. The RMS error is reduced to 25 m and the CEP is 7 m (Figs. 16 and 17). And the mean of the terminal position error is unbiased and there is no trend of following the trajectory of the vision system. This is affected by

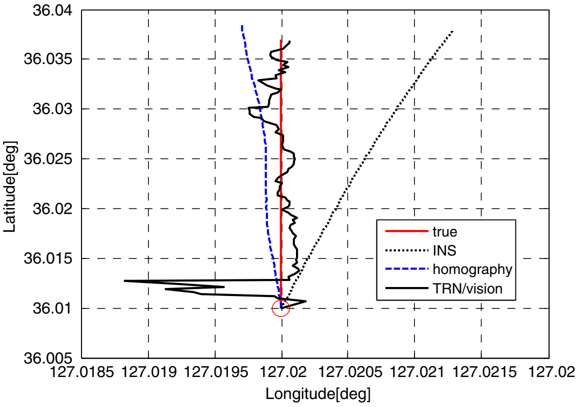

Fig. 15 Horizontal trajectory

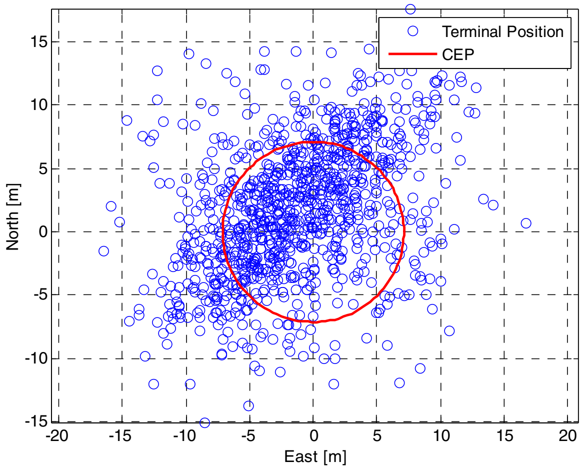

Fig. 17 Termial position error and circular error probable

the rough terrain profile. If the terrain the UAV flies over is smooth, there are small differences in heights between different locations. It indicates that the terrain profile has low observability for estimating precise position.

In this study, we have not considered the error of a digital terrain map explicitly. We assume that it is included in the error of the radar altimeter.

6 Conclusion

We proposed a vision-based TRN system to replace the low-cost INS and to overcome the drawback of the vision system using homography relationship. The optimal Bayesian estimation formulas were derived and the point-mass filter was applied to approximate the Bayesian algorithm. The vision system was considered, which uses the monocular camera and employs the homography decomposition in order to calculate the translational motion. The numerical simulation was performed to evaluate the proposed algorithm and the results showed that we can acquire the accurate navigation information with a camera, a radar altimeter and a DETD map. As a future research, we can consider about the effect of the DTED map error and the radar altimeter accuracy over water or wet area.

References

1. Bergman, N.: Recursive Bayesian estimation: navigation and tracking applications. Ph.D. Dissertation, Linköping University, Linköping, Sweden (1999)
2. Hostetler, L., Andreas, R.: Nonlinear Kalman filtering techniques for terrain-aided navigation. IEEE Trans. Automat. Contr. **28**(3), 315–323 (1983)
3. Oliveira, P.: MMAE terrain reference navigation for underwater vehicles using eigen analysis. In: Proc. of the 44th IEEE CDC/ECC 2005, Seville, Spain (2005)
4. Nonsen, K.B., Hallingstad, O.: Sigma point Kalman filter for underwater terrain-based navigation. In: Proc. Of IFAC Conference of Control Appli. Marine Syst. (2007)
5. Nonsen, K.B., Hallingstad, O.: Terrain aided underwater navigation using point mass and particle filters. In: Proc. of IEEE/ION Position Location Navigat. Symp., pp. 1027–1035 (2006)
6. Agrawal, M., Konolige, K.: Real-time localization in outdoor environments using stereo vision and inexpensive gps. In: Proc. of the 18th International Conference on Pattern Recognition, pp. 1063–1068 (2006)
7. Kaiser, K., Gans, N., Dixon, W.: Localization and control an aerial vehicle through chained, vision-based pose reconstruction. In: Proc. of the American Control Conference, pp. 5934–5939 (2007)
8. Kaiser, K., Gans, N., Dixon, W.: Vision-based estimation for guidance, navigation, and control of an aerial vehicle. IEEE Trans. Aerosp. Electron. Syst. **46**(3) 1064–1077 (2010)
9. Trucco, E., Verri A.: Introductory Techniques for 3-D Computer Vision. Prentice-Hall, Englewood Cliffs, NJ (1998)
10. Longguet-Higgins, H.: A computer algorithm for reconstructing a scene from two projections. Nature **293**, 133–135 (1981)
11. Boufama, B., Mohr, R.: Epipole and fundamental matrix estimation using virtual parallax. In: Proc. of the IEEE International Conference on Computer Vision, pp. 1030–1036 (1995)
12. Nister, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. Pattern Anal. Mach. Intell. **26**, 79–97 (2004)
13. Ma, Y., Soatto, S., Koseck, J., Sastry, S.: An Invitation to 3-D Vision. Springer, New York (2004)
14. Zhang, Z., Hanson, A.: 3D Reconstruction Based on Homography Mapping. ARPA Image Understanding Workship, Palm Springs, CA (1996)