Localization Based on DEM Matching Using Multiple Aerial Image Pairs
Dong-Gyu Sim and Rae-Hong Park

Abstract—This paper proposes the localization algorithm that estimates translation parameters of an aircraft by comparing the sampled elevation map recovered from aerial sequence images, and the digital elevation model (DEM) with the given orientation and altitude parameters obtained from a gyroscope. It consists of two stages: recovering the sampled elevation map from multiple aerial image pairs and matching of the relative recovered elevation map (REM) with the relative DEM. Computer simulations with three real aerial image sequences show the effectiveness of the proposed algorithm.

Index Terms—Aerial sequence image, attitude parameters, digital elevation model, localization, M-estimator, recovered elevation map, stereo matching, terrain.

I. INTRODUCTION

NAVIGATION guides a vehicle to a spot where one wants to go [1]. For autonomous navigation, it is important to extract accurate navigation parameters of an aircraft such as position and velocity. Using them, one can adjust a velocity or direction to the desired destination. Many approaches to navigation parameter extraction have been presented.

Navigation systems can be realized by a number of different approaches such as terrain contour matching (TERCOM) [2], inertial navigation system (INS) [1], global positioning system (GPS), and so on. The TERCOM navigates by measuring altitude of terrain with a special radar and by matching terrain contours. This system has difficulty in estimating the position in plain regions where change in elevation is small, and it can be out of control by external signals. The INS with an accelerometer and gyroscope has been widely used for navigation. It does not require any constraint on flying trajectory and is not affected by any external signal, however it has a drawback that the estimation error increases as navigation proceeds. The GPS estimates an absolute position with more than three satellites, however the signals from satellites can be disturbed and modified intentionally. Navigation parameters can also be estimated by using computer vision and image processing techniques [3]–[10], however, they are not practical for implementing real-time navigation systems with real aerial image sequences, because they require a high computational load or an expensive down-looking camera. On the other hand, three-dimensional (3-D) reconstruction algorithms have been proposed by matching the digital elevation model (DEM) and the relative recovered elevation map (REM) [5], [6].

The proposed algorithm reconstructs the wide sampled REM using multiple image pairs. It estimates the aircraft position using the relative geometrical data (REM) obtained from multiple aerial image pairs and the relative DEM, where the robust M-estimator is used. It is important to analyze the characteristics of the REM. The rest of the paper is structured as follows. The proposed algorithm is presented in Section II. Experimental results are shown in Section III to illustrate the effectiveness of the proposed algorithm. Finally, conclusions are given in Section IV.

II. PROPOSED IMAGE-BASED LOCALIZATION ALGORITHM USING THE DEM

The proposed algorithm consists of two stages. At the first stage, a sampled elevation profile is constructed by computing elevations, at sample feature points along an arbitrary horizontal line, based on stereo matching, with point correspondences established between two consecutive aerial images. In the second stage, a matching point in the REM is found by searching the location whose local terrain is the most similar to the REM, where relative REMs and DEMs are employed.

A. First Stage (Construction of a Sampled Elevation Profile)

First, the global feature point FP_{k-1} denotes the feature point having the maximum variance at the (k - 1)th (previous) frame I_{k-1} in the aerial image sequence. MP_k denotes the matching point in the kth (current) frame I_k, corresponding to FP_{k-1}. P_{k-1} and P_k denote 3-D positions of the optical centers of a camera in the (k - 1)th and kth frames, respectively. The difference between these two positions is denoted by the base line B_k

\[ B_k = r_k R (P_{k-1}, P_{k-1}) - r'_k R (MP_k, P_k) \]  

Fig. 1. Sample feature points (FP_{k-1}) and their corresponding matching points (MP_k) for construction of a sampled elevation profile, based on the relative elevation at each point with respect to that reconstructed from the global feature point (FP_{k-1}) and its corresponding matching point (MP_k), where sample feature points are selected along a line, 1 \leq j \leq M.
where $R(FP_{k-1}, P_{k-1})$ and $R(MP_k, P_k)$ represent vectors in the 3-D space, from $P_{k-1}$ to $FP_{k-1}$ and from $P_k$ to $MP_k$, respectively, which are computed with the attitude and altimeter parameters obtained from a gyroscope and an altimeter. Assuming that the elevation of the ground point corresponding to the feature point is zero, the equalities

$$r_k R(FP_{k-1}, P_{k-1}) = P_{k-1}$$

(2)

$$r_k R(MP_k, P_k) = P_{k}$$

(3)

are satisfied, where $r_k(r'_k)$ represents the distance from the feature point $FP_{k-1}(MP_k)$ to the zero-elevation ground point, and the subscript $z$ denotes the $z$ coordinate of the 3-D position. The longitude and latitude of the approximate base line can be estimated using three equations by a simple regression method with the altitude of an aircraft obtained from a gyroscope. Even though the estimated base line is not accurate, global features and shapes of the REM reconstructed with the approximate base line yields proper matching of the REM with the DEM.

$FP_{k-1}, 1 \leq j \leq M$, is the $j$th sample feature point in the $(k-1)$th frame, where $M$ denotes the total number of sample feature points employed. The sample feature points $FP_{k-1}$ are selected along a fixed row with equal intervals. In experiments, they are selected along the 80th row, with 25 pixels apart from each other, where the image size is 320 x 240. It is assumed that the size of the area overlapped by two successive images is greater than half the image size. The fixed line is arbitrarily chosen, for simplicity of implementation, in the upper half region of the image, where an aircraft is assumed to fly upwards in the image plane. $MP_k$ of the $k$th frame are corresponding feature points detected by the NCC measure. Note that these matching points are detected for $(N-1)$ image pairs. A sampled REM is recovered with the consecutive sets of feature points and matching points, i.e., the REM at $M_1$ is calculated by the stereo matching method [6], [9], [11], with $FP_{k-1}$ and $M_0$ as well as attitude parameters $\Phi, \omega, \nu$ and $\psi$ obtained from a gyroscope and flying altitude from an altimeter. The equality

$$r'_k R(FP'_{k-1}, P_{k-1}) = r'_k R(MP'_k, P_k) = B_k$$

(4)

is satisfied from the triangle $(P_{k-1}, P_k, B_k)$ in Fig. 1, from which $r'_k$ and $r'_k$ can be calculated [6]. $M'_1$ is defined by $P_{k-1} + r'_k R(FP'_{k-1}, P_{k-1})$ or $B_k + r'_k R(MP'_k, P_k)$. The relative REM $\text{Rel}_{\text{REM}}^j$ at $MP'_k$ is defined by subtracting the REM of $M_1$ from that of $M'_1$, where the REM at $M_1$ is estimated with $FP_{k-1}$ and $MP_k$. Similarly the relative DEM $\text{Rel}_{\text{DEM}}^j$ is defined. The wide REM is generated by combining several sets of sampled elevation profiles from multiple image pairs, each of which is obtained by stereo matching of an image pair.

**B. Second Stage (Matching of the Relative REM and the Relative DEM)**

The localization algorithm searches the area, with 5 pixels x 5 pixels interval, in order to find the matching point in the DEM. Assuming that the resolution of the DEM is about 2 m/pixel, the real search interval is approximately equal to 10 m x 10 m. The DEM matching algorithm is described as follows:

$$M_i = \text{Infinite}$$

For $dx = x_0$ to $x_1$ step $\delta x$

$$dx = 0$$

For $k = N-2$ to 0 step $-1$

For $j = 1$ to $M$ step 1

$$\text{Rel}_{\text{REM}}^j = M'_1 - M_{k, h}$$

where $M$ represents the total number of feature points, along a line for each image pair, $N$ signifies the number of image frames used, and $\delta x$ and $\delta y$ represent the search intervals along the $x$ and $y$ directions, respectively. $dx$ and $dy$ represent the compensated distances with respect to the initial latitude and longitude, respectively. $M_k = (M_{h,k}, M_{v,k}, M_{l,k})$ signifies the ground point corresponding to the global feature point $FP_{k-1}$, with $M_{l,k}$ corresponding to the REM at $M_k$. Similarly, $M'_k = (M'_{h,k}, M'_{v,k}, M'_{l,k})$ is the ground point corresponding to the sample feature point $FP'_{k-1}$ in the $(k-1)$th frame and their matching point $MP'_k$ in the $k$th frame, and $(x_1 - x_0)(y_1 - y_0)$ denotes the size of the search area in the DEM. In experiments, $\text{DEM}(x, y)$ represents an elevation at position $(x, y)$, DEM$_k$ and DEM$_{k'}$ denote the elevations at the global feature point $FP_{k-1}$ and the $j$th sample feature point $FP'_{j-1}$, respectively. $\text{Rel}_{\text{DEM}}^j$ (Rel$_{\text{REM}}^j$) signifies the relative DEM (REM) at $M'_{k'}$ with respect to that at $M_k$. For computational simplicity, the error in elevation estimation is assumed to be the same for all feature points in a frame. So, relative elevations instead of absolute elevations are used in computing the distance for magnitude-invariant matching. The relative DEM matching algorithm also employs a robust cost function $\rho(x)$ based on M-estimation, by which some outliers could be rejected [12], [13]. The cost function is defined by

$$\rho(x) = \begin{cases} \tau^2, & x < \tau \\ \tau^2, & x \geq \tau \end{cases}$$

(5)

where $\tau$ is set to 20. The matching position is determined by finding the point having the smallest distance $E$, assuming that the matching point is in the search area.

Whereas the conventional algorithm needs $K \times L$ elevation computations to reconstruct the dense REM [5], [6], the proposed algorithm requires only $M \times (N-1)$, where $K$ and $L$ are the numbers of rows and columns of the input image, respectively, $M$ is the number of sample feature points in an image pair, and $N$ is the number of image pairs employed. For example, with $K = 320$, $L = 210$, $M = 11$, and $N = 1 = 8$, the conventional and proposed algorithms require 76800 and 88 elevation computations, respectively. In other words, the conventional algorithm requires elevation calculations 873 times greater than the proposed algorithm. The number of additions for matching the sampled relative REM with the relative DEM is equal to $\sum_{i=1}^{N} (x_i - x_0)(y_i - y_0)/(|\delta x| \delta y)$. According to computer simulations on Pentium 400, the proposed algorithm takes 3 s whereas the conventional algorithm requires about 1400 s.

**III. EXPERIMENTAL RESULTS AND DISCUSSION**

The effectiveness of the proposed algorithm is shown by computer simulations with three test aerial image sequences (320 x 240) that were acquired with a CCD camera attached to an aircraft. The first image sequence was acquired by a helicopter with a $\beta$-cam video camera near Yusung in Korea. Each of them was digitized at 1 frame/s and quantized to eight bits/pixel. The three test image sequences consist of 1320, 1360, and 1570 frames. Universal transverse Mercator (UTM) as a global reference coordinate was used, and Yusung and Kongju are located in 52 zone.

Fig. 2 shows nine consecutive images sampled at 1 frame/s, where the test sequence was acquired from the helicopter at spot 1, the number indicated below denotes the frame index, the white cross (+) marks
represent the sample feature points $\mathbf{FP}_k^{i-1}$ in the previous frame for elevation estimation, and the box (□) marks denote the matching points $\mathbf{MP}_k^i$ in the current frame corresponding to $\mathbf{FP}_k^{i-1}$. Note that a helicopter flew upwards, and that in the first and last frames only the sample feature points and the matching points are shown, respectively.

In experiments, 11 sample feature points are used in stereo matching and they are well corresponded as shown in Fig. 2. In experiments, 654 × 896 DEM is used, and the matching result is shown in Fig. 3, where the darker (brighter) region represents the lower (higher) terrain and the black dot (●) marks denote the sample feature points $\mathbf{FP}_k^{i-1}$. Whereas the asterisk (+) mark represents the real position in the last frame which is obtained by manually matching the aerial image and a map, the circle (○) and diamond (◇) marks signify the positions estimated by the conventional [5] and proposed algorithms, respectively. The DEM is generated by interpolating the contours that are manually drawn with 25 m interval. In the case of the first test image sequence, the ground area overlapped by two images is about 300 m × 200 m. The REM reconstructed over such a small overlapped area may not contain enough number of reliable feature points, resulting in a large probability that the DEM matching fails. The area (300 m × 380 m) recovered by the proposed algorithm is wide enough to match correctly the DEM, in which multiple pairs of images (nine frames) are used.

Table I shows the estimation error by the conventional [5] and proposed algorithms, where the estimation error is calculated in terms of the Euclidean distance between the real and estimated positions. Three image sequences are used, and absolute positions at six spots, two spots for each sequence, are estimated. The ground area recovered with nine images is about 300 m × 380 m. The estimation error of the proposed algorithm at spot 1 is 26 m whereas that by the conventional algorithm is 351 m, as listed in Table I.

Fig. 4 shows the estimation error of the proposed algorithm at six spots, as a function of the number of sample feature points along the line in elevation computation. Simulations are performed with one test image sequence (Helicopter) acquired from a helicopter and two test image sequences (Light airplane I & II) obtained from a light plane, with the number indicating the spot index. Fig. 4 shows that at six test spots the estimation error is small enough if the number of sample feature points is greater than nine.

### IV. Conclusions

In this paper, an image-based localization algorithm is proposed by estimating the translation parameters, in which the relative elevations recovered from multiple image pairs are compared with the relative DEM. The effectiveness of the proposed system is shown via several experiments. Further research will focus on the integration of the proposed algorithm into a practical visual navigation system.
REFERENCES


