Integrated Position Estimation Using Aerial Image Sequences

Dong-Gyu Sim, Rae-Hong Park, Senior Member, IEEE, Rin-Chul Kim, Member, IEEE, Sang Uk Lee, Senior Member, IEEE, and Ihn-Cheol Kim

Abstract—This paper presents an integrated system for navigation parameter estimation using sequential aerial images, where navigation parameters represent the position and velocity information of an aircraft for autonomous navigation. The proposed integrated system is composed of two parts: relative position estimation and absolute position estimation. Relative position estimation recursively computes the current position of an aircraft by accumulating relative displacement estimates extracted from two successive aerial images. Simple accumulation of parameter values decreases the reliability of the extracted parameter estimates as an aircraft goes on navigating, resulting in a large position error. Therefore, absolute position estimation is required to compensate for the position error generated in relative position estimation. Absolute position estimation algorithms by image matching and digital elevation model (DEM) matching are presented. In image matching, a robust-oriented Hausdorff measure (ROHM) is employed, whereas in DEM matching the algorithm using multiple image pairs is used. Experiments with four real aerial image sequences show the effectiveness of the proposed integrated position estimation algorithm.

Index Terms—Navigation, aerial image, image matching, digital elevation model (DEM), recovered elevation map (REM), relative position estimation, absolute position estimation, robust-oriented Hausdorff measure.

1 INTRODUCTION

Navigation guides a vehicle to a spot where one wants to go. For autonomous navigation, it is important to extract the accurate navigation parameters such as the absolute position and the velocity of an aircraft, where the absolute position, represented with respect to the global coordinates of the earth, corresponds to the position component of the exterior orientation used in the photogrammetry field [1]. Using these parameters, we can adjust a velocity and a direction to the destination.

Many approaches to estimation of accurate navigation parameters have been presented [2], [3], [4], [5], [6]. Various systems such as terrain contour matching (TERCOM), inertial navigation system (INS), and global positioning system (GPS) are well-known. TERCOM navigates by matching and measuring the altitude of terrain with special radar. It cannot accurately estimate its own 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 a gyroscope has been widely used for navigation. It does not require any constraint on the flying trajectory and it cannot be affected by any external signal. But, the estimated error tends to increase as an aircraft goes on flying. Also, the GPS has been known as an effective system that can easily estimate a position with a small error, however, it can also be disturbed by an external signal.

This paper investigates the practical position estimation system of an aircraft using sequential aerial images. Because aerial image sequence is used as an input in our video-based navigation system, the navigation system has the advantage that it is not detected by enemies nor is guided by external signals, compared with other active approaches. Also, it can be easily attached to an aircraft, without any special apparatus to compensate for an attitude change between successive frames. Four test aerial image sequences used in this paper were acquired from a camera fixed on a helicopter or a light airplane, in which the optical axis of the camera varies according to the aircraft attitude.

Many computer vision and image processing techniques have been proposed for a video-based navigation system. However, they are not practical and are not focused on the construction of the overall system. An online vehicle motion estimation method using visual information was proposed, based on Kalman filter estimation with an expensive down-looking camera [4]. Another algorithm with a down-looking camera was presented [5]. A passive navigation algorithm based on optical flow estimation and optimization [7] was also proposed to estimate the instantaneous velocity of camera motion (rotation and translation) parameters. It could be unstable with a high computational load because it estimates not only translation but also rotation parameters. On the other hand, 3D reconstruction approaches have been proposed for matching between the digital elevation model (DEM) and the recovered elevation map (REM) [8], [9], [10].
However, they required a high-computational load and were not applied to various video sequences. Many image matching algorithms [3], [4], [5], [11] have been proposed; however, it is difficult to directly apply them to practical navigation systems because of their high-computational complexity and usage of a down-looking camera. The gray-level-based image matching algorithm employs geodetical calibration that requires an additional computational load [11], which can be applied to absolute position estimation for autonomous navigation. The hybrid estimation of navigation parameters has been proposed to cope with these problems [12]. In order to implement a cheap and accurate visual navigation system that can be easily attached to any aircraft, we propose an integrated navigation parameter estimation system.

This paper presents an integrated system for position estimation using sequential aerial images. The main objective is to develop an effective vision-based position estimation algorithm for real-time implementation. The proposed system is composed of two parts: relative position estimation and absolute position estimation. The former is based on the stereo modeling of two successive frames, whereas the latter is accomplished by DEM matching or by image matching with Indian remote sensing (IRS) or high-resolution images. IRS images are acquired from the satellite launched by the Department of Space, India, in 1995.

The rest of the paper is structured as follows: Section 2 presents the proposed integrated system for position estimation. Sections 3 and 4 describe the relative position estimation and absolute position estimation algorithms, respectively. Section 5 gives some experimental results and discussions. Finally, Section 6 summarizes conclusions.

2 INTEGRATED POSITION ESTIMATION

Fig. 1 shows the overall block diagram of the proposed position estimation system that consists of relative and absolute position estimation parts, where \( I_n \) is the current input aerial image, with \( \phi_n, \omega_n, \theta_n, \) and \( h_n \): roll, pitch, yaw, and altitude parameters of an aircraft at the \( n \)th frame, respectively; \( P'_n \): \( n \)th estimated current position.
error generated and accumulated in relative position estimation.

The integrated system is controlled by the switching scheme, in which absolute position estimation is incorporated with a planned trajectory of autonomous navigation. Absolute position estimation is performed using images or DEM information. Absolute position estimation by image matching uses IRS or high-resolution images as reference images. The term “high-resolution” is used to indicate that the resolution of aerial images is higher than that of IRS images. The ground resolution of IRS images is 5 m/pixel while that of aerial images is about 1-3 m/pixel. The absolute positioning system with IRS images is effective for estimation of the current position in regions containing artificial structures such as roads and buildings. Note that because US Geological Survey (USGS) digital orthophoto quadrangles (DOQs) are about 1 m/pixel resolution, they could be used as high-resolution images. If these images are available, the image matching algorithm can obtain the current position accurately, based on the robust-oriented Hausdorff measure (ROHM) [13].

The current position can also be estimated by matching the stored DEM with the REM that is reconstructed from multiple successive image pairs [8], [9], [10]. The proposed absolute position estimation algorithm using the DEM information consists of two stages: recovering the sampled elevations from multiple aerial image pairs and matching the relative REM with the relative DEM. The relative REM and DEM are defined with respect to the elevations at global feature points of the REM and DEM, respectively. Thus, the proposed algorithm can accurately estimate the absolute position using a wide recovered area and is much faster than conventional algorithms [8], [9]. Each part of the overall system will be presented in the following sections.

3 RELATIVE POSITION ESTIMATION

Relative position estimation sequentially computes the current position by accumulating the displacement of a camera, estimated with respect to the previous position. Fig. 2a shows the block diagram of the proposed relative position estimation algorithm, in which Z represents one-frame time delay and $MP_n'$ denotes a feature point. The relative position estimation method recursively computes the relative displacement $B_n$ between two positions ($P_{n-1}$ and $P_n$) and updates the position of an aircraft by accumulating the relative displacement $B_n$.

Since a camera is tightly attached to an aircraft in the proposed algorithm, it is necessary to compensate for the attitude change of a camera between two successive frames. Thus, the previous input image $I_{n-1}$ is transformed to $I_{n-1}$ for matching with the current input image $I_n$ according to the attitude of the aircraft, acquired from a gyroscope. $I_n$ and $I_{n-1}$ have different attitudes, thus $I_{n-1}$ is required to be mapped. $I_{n-1}$ is obtained by ray tracing from every point in $I_{n-1}$ to its corresponding point in $I_n$. Matching pairs are detected by the blockmatching algorithm (BMA) employing the normalized correlation coefficient (NCC) measure. Two-level hierarchical matching is used to reduce the computational load. The hierarchical matching is done using the Gaussian pyramid. The higher level is reduced by a factor of two from the original image. The correspondence is detected with NCC at the higher level. Then, the correspondence is found by searching with a small search range (-1 ~ +1) on the original image plane. The final matching point $MP_n'$ with subpixel accuracy is obtained by refining the coarse displacement with the fine displacement estimated by the optical flow method [14], [15], [16]. Subpixel update displacement is obtained by regression of the optical flow constraint equations at every pixel in the matching window [14], [16].

Fig. 2b shows the displacement estimation based on stereo matching, in which the proposed relative position estimation method extracts the displacement $B_n$ of an aircraft using the feature point $MP_n'$ of the current input image $I_n$ and the feature point $MP'_{n-1}$ of the previous image $I_{n-1}$. $B_n$ can be expressed as $TV_n$, where $T$ denotes the sampling interval and $V_n$ signifies the velocity of an aircraft. $(X_n, Y_n, Z_n)$ signifies the navigation coordinates, in which $X_n$ and $Y_n$ are expressed in terms of the longitude and latitude, respectively, and $Z_n$ denotes the altitude with respect to the sea level. $f$ is the focal length of a camera, and $M_n = (M_{n,x}, M_{n,y}, M_{n,z})$ is a 3D DEM position, corresponding to the feature point. $P'_n(R_n)$ denotes the vector from $P_{n-1}(P_n)$ to the 3D DEM position $M_n$, and $r_n'(r_n)$ signifies its normalized vector obtained by converting the feature (matching) point in the image space into the real navigation coordinates, with the attitude parameters acquired by the gyroscope. Note that the vector $R'_n(R_n)$ passes through the feature point $MP'_n$ in the previous frame (matching point $MP_n$ in the current frame $I_n$).

First, $R'_n$ and $R_n$ must be calculated to obtain $B_n$. $R'_n$ is given by $R'_n = (M_{n,x} - P_{n-1,x}, M_{n,y} - P_{n-1,y}, M_{n,z} - P_{n-1,z})$, where $P_{n-1}$ is given and $M_n$ is obtained by ray tracing with $P_{n-1}$ and $r_n'$. Then, $R_n = (R_{n,x}, R_{n,y}, R_{n,z})$ is computed using

$$
\frac{R_{n,x}}{r_{n,x}} = \frac{R_{n,y}}{r_{n,y}} = \frac{R_{n,z}}{r_{n,z}} = \frac{M_{n,z} - P_{n,z}}{r_{n,z}},
$$

where $r_n = (r_{n,x}, r_{n,y}, r_{n,z})$ and the altitude $P_{n,z}$ with respect to the sea level are given by a gyroscope and an altimeter, respectively. The displacement $B_n$ is defined by

$$
B_n(MP'_n, MP_n, X_{n-1}, Y_{n-1}, S_{n-1:n}) = R'_n - R_n,
$$

where $S_{n-1:n}$ represents the parameter vector consisting of roll, pitch, yaw, and altimeter parameters of two consecutive images $I_{n-1}$ and $I_n$. Then, we can estimate the current position $P_n$ of the aircraft by accumulating the estimated displacement $B_n$ to the previous position $P_{n-1}$. Also, the velocity $V_n$ of an aircraft is calculated by dividing the displacement $B_n$ by the sampling interval $T$ of the image sequence, i.e., $V_n = B_n/T$. In our system, the sampling interval $T$ is set to 1 second to guarantee that the overlapping area between two successive frames is about half the whole image size. The proposed relative positioning algorithm was implemented in real time on TMS320C80 [17].

Only a single correspondence is used for a frame to estimate the displacement. The displacement could be
estimated by regression with multiple correspondences, however, no improvement was empirically observed because the possibility of mismatching increases as the number of correspondences increases. Additionally, because some correspondences have large errors, linear estimation based on Gaussian noise assumption could yield worse results due to outliers. The error in navigation parameter estimation, accumulated by relative position estimation, increases with time, thus the position has to be compensated by absolute position estimation.

In relative position estimation, the position error is roughly in proportion with the error in altitude, so small altitude error is desirable for good performance. Change in roll values is the most dominant error among other factors, because the acquired image could be distorted by a small change in roll measurements. Actually, the roll parameter of an aircraft changes easily because of wind, affecting the accuracy of relative position estimation.

4 Absolute Position Estimation

We propose the absolute position estimation algorithm that employs two approaches: matching using reference images and DEM information. In the case of image matching [13], if the position obtained by relative position estimation is located within the effective range (e.g., 400 m) from the reference position prespecified, the absolute position estimation method using high-resolution aerial images or
IRS images is activated. For absolute position estimation using DEM information, the effective range is set to 200m.

Generally, navigation systems have been developed under the assumption that the aircraft flies along a predetermined trajectory. Our system employs two different reference images: high-resolution aerial images and IRS images. If high-resolution images along the trajectory are available as reference images, the absolute position can be estimated from them [18]. Otherwise, the absolute position is estimated from the IRS images.

The absolute position is estimated using the ROHM [13] with the reference images (high-resolution or IRS images) when the aerial input image contains distinct geometric (artificial) structures such as roads, buildings, stadium, and interchange. For mountain regions, it is difficult to use feature-based image matching. Thus, change of terrain elevation is used for absolute position estimation if the regions show large or abrupt changes in elevation. In the case of absolute position estimation using DEM data, we employ an estimation method, in which multiple image pairs are used to construct a wide REM and the absolute position is estimated using the REM and DEM information.

4.1 Absolute Position Estimation
Based on the ROHM

The position error by relative position estimation is compensated by absolute position estimation based on the ROHM, where an input image is matched with stored reference images. If the reference image contains distinct artificial structures, the absolute position can be accurately estimated. The ROHM is introduced by replacing the distance measure by the accumulation of matching points, in which model parameters are estimated by accumulation operations in the parameter space as used in the Hough transform (HT) [13]. The ROHM algorithm is robust to severe noise embedded in an input image. The gradient information at two corresponding points in the aerial input and reference images is used to validate the correct matching. Edge images and distance transform (DT) images are constructed, from which the ROHM is computed. Fig. 3a shows the block diagram of the ROHM computation, in which gradient images, edge images, and DT images [19], [20], [21], [22] are used. Note that the DT image is the distance map of an edge image, containing a distance value [19] of the image B at position a. The weight based on the orientation difference is used to avoid inconsistent correspondences with different orientations. The symmetric proximity function, $\rho_T(x)$, shown in Fig. 3b is employed. Because the proposed algorithm accumulates the transformed distance values of all edge pixels, with the weight specified by the gradient orientation information, it counts the weighted number of matching points as the HT does. As a result, the ROHM algorithm becomes robust to noisy data. Furthermore, because the orientation information is used, the point in the input aerial image, with the orientation different from that of its matching point in the reference image, can be effectively removed. In Fig. 3c, the robustness and usefulness of orientation information are shown, where open and closed circles represent edges in the input and reference images, respectively. The similarity of the left example is the largest. Two input points correspond to two points having the same directions, even though one point considered as an outlier finds no correspondence. In the center example, all of three points have corresponding points with different directions. The similarity of the right example is larger than that of the left one with respect to linear estimation, whereas the left example case is better than the right one in the context of robust statistics.

First, the gradient image $A_G$ ($B_G$) is obtained by applying the Sobel edge operator to the gray-level image $A$ ($B$) and the edge image $A_E$ ($B_E$) is constructed by thresholding the gradient image $A_G$ ($B_G$). Note that edge points in $A_E$ ($B_E$) yields the edge point set $A_S$ ($B_S$).

Each pixel value of the gradient image is stored in vector form. For example, the gradient value at point $a$ in image $A$ is expressed as $A_G(a) = (G_{a,x}, G_{a,y})$, where $G_{a,x}, G_{a,y}$ represents the derivative at point $a$, with respect to $x$ ($y$). Gradient magnitude is thresholded to generate the edge image, and orientation information is used to eliminate incorrect correspondences. Here, the threshold for edge images is experimentally set to 100.

The DT maps $A_D$ and $B_D$ are constructed from the edge images $A_E$ and $B_E$, respectively, in which Borgefors' two-pass algorithm [19] is used. The two-pass algorithm generates the DT maps by applying two filters along opposite directions. Note that the computational complexity is $O(N_x \times N_y)$, where $N_x$ and $N_y$ denote the numbers of column and row pixels, respectively. The processing time of the HD algorithm can be reduced using the DT maps constructed by the time-efficient two-pass algorithm with a low computational load.

The directed ROHM

$$h_{ROHM}(A_G, A_E, B_D) = \left( h_{ROHM}(B_G, B_E, A_D) \right)$$

is computed with the gradient image $A_G$ ($B_G$), edge image $A_E$ ($B_E$), and DT image $B_D$ ($A_D$). It is computed by accumulating the weighted distance value at each edge pixel, where the weight is defined by the dot product of the gradient images $A_G$ and $B_G$ computed at two matching points.

The final ROHM $H_{ROHM}$ is determined by

$$H_{ROHM} = \min(h_{ROHM}(A_G, A_E, B_D), h_{ROHM}(B_G, B_E, A_D)),$$

where $h_{ROHM}(A_G, A_E, B_D)$ and $h_{ROHM}(B_G, B_E, A_D)$ denote the directed ROHMs. The directed ROHM $h_{ROHM}(A_G, A_E, B_D)$ is defined as

$$h_{ROHM}(A_G, A_E, B_D) = \sum_{a \in A_S} d_{A_G(a)} \cdot d_{B_D(a)} \rho_T(B_D(a))$$

$$= \sum_{a \in A_S} s(a) \rho_T(B_D(a)),$$

where an orientation vector $d_{A_G(a)}$ ($d_{B_D(a)}$) represents a unit gradient vector of a gray-level image $A$ ($B$) at position $a$, $s(a) = d_{A_G(a)} \cdot d_{B_D(a)}$ denotes the dot product of two gradient vectors obtained from two images $A$ and $B$, and $B_D(a) = \min_{b \in B_S} \| a - b \|$ is the distance map value [19] of the image $B$ at position $a$. The weight based on the orientation difference is used to avoid inconsistent correspondences with different orientations. The symmetric proximity function, $\rho_T(x)$, shown in Fig. 3b is employed. Because the proposed algorithm accumulates the transformed distance values of all edge pixels, with the weight specified by the gradient orientation information, it counts the weighted number of matching points as the HT does. As a result, the ROHM algorithm becomes robust to noisy data. Furthermore, because the orientation information is used, the point in the input aerial image, with the orientation different from that of its matching point in the reference image, can be effectively removed. In Fig. 3c, the robustness and usefulness of orientation information are shown, where open and closed circles represent edges in the input and reference images, respectively. The similarity of the left example is the largest. Two input points correspond to two points having the same directions, even though one point considered as an outlier finds no correspondence. In the center example, all of three points have corresponding points with different directions. The similarity of the right example is larger than that of the left one with respect to linear estimation, whereas the left example case is better than the right one in the context of robust statistics.
The absolute position is detected by finding the position having the maximum proximity of the ROHM value between the input frames and the reference image. By sliding the reference image with respect to the input video frame, the matching point is detected using the ROHM value and the absolute position of the input frame is calculated with the known position of the reference image. It is assumed that the attitude and altitude parameters of the reference image are given for matching of the reference image and the input video frame.
Pixel-based matching is performed with the reference image and input video images. Thus, the error is determined by the ground resolution of images. As mentioned before, the ground resolution depends on the altitude of an aircraft. Usually, the ground resolution of high-resolution images ranges from 0.5 to 3 m/pixel. For IRS satellite images, it is 5m/pixel. In pixel-based image matching, the maximum error corresponds to half-pixel. Then, the final position estimation error in ground is also determined with the attitude and altitude parameters of a camera as in relative position estimation.

### 4.2 Absolute Position Estimation Using DEM Matching

In the previous section, the image matching algorithm is employed to match artificial structures in scenes. However, it is not effective for mountain areas where there is no artificial object having distinct gray-level changes. In such a case, matching based on elevations is effective, thus the DEM based absolute position estimation is proposed.

The proposed absolute position estimation algorithm using DEM matching localizes an aircraft, based on the matching between the relative REM and the relative DEM [10]. The relative REM and DEM are obtained with respect to the REM and DEM at a global feature point, \( MP_n \), respectively. The global feature point is the point at which the variance of gray level is the maximum, as used in relative position estimation. In this algorithm, the wide REM is reconstructed by consecutively combining multiple sets of sampled REMs acquired along the 1D lines, each of which is obtained over the region overlapped by two successive images. Note that two consecutive frames are enough to obtain the wide REM by combining the 1D elevations, with the assumption that the aircraft flies with the prespecified altitude and velocity. Then, the position of an aircraft is estimated by matching the relative REM with the relative DEM. The proposed algorithm consists of two stages: recovering the elevations at sample points and matching the relative REM with the relative DEM. At the first stage, the elevations at sample feature points along the line are calculated based on the stereo matching method, with point correspondences established between two consecutive aerial images. In the second stage, a matching point is found by searching for the location in the DEM whose local terrain is the most similar to the reconstructed REM, where relative REMs and DEMs are employed for establishing correspondences.

At first, the global feature point that yields the maximum local variance is extracted in the previous frame, where a 31 × 31 local window is used for variance computation. It is used as a reference point over the whole frame in computing the relative REMs and DEMs. Feature point correspondence is established by block-based matching, using the NCC measure.

In Fig. 4, \( FP_{n-k-1}^j \) denotes the global feature point having the maximum variance at the \((n - k - 1)\)th (previous) frame \( I_{n-k-1} \) in an aerial image sequence. \( MP_{n-k} \) represents the matching point in the \((n - k)\)th (current) frame \( I_{n-k} \), corresponding to the global feature point \( FP_{n-k-1}^j \). \( P_{n-k-1} \) and \( P_{n-k} \) denote 3D positions of the optical centers of a camera attached on the aircraft in the previous and current frames, respectively. The difference between these two positions is denoted by \( B_{n-k} \). Similarly, \( FP_{n-k-j}^j \), \( 1 \leq j \leq J \), is the \( j \)th sample feature point in the previous frame, where \( J \) denotes the total number of sample feature points employed. The sample feature points \( FP_{n-k-1}^j \) 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. 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 in the upper half region of the image, where an aircraft is assumed to fly upwards in the
image plane. $MP_{n-k}^j$ of the current frame is the corresponding feature point detected by the NCC measure. Note that these matching points are detected for $(N-1)$ image pairs ($k = N - 2$ to 0).

A sampled REM is recovered with the consecutive sets of feature points and matching points, i.e., the REM at $M_{n-k}^j$ is calculated by the stereo matching method [8], [9] with $FP_{n-k-1}^j$ and $MP_{n-k}^j$ as well as attitude parameters $\varphi$, $\omega$, and $\psi$. From the triangle $P_{n-k-1}^jP_{n-k}M_{n-k}^j$ in Fig. 4,

$$B_{n-k} = P_{n-k} - P_{n-k-1} = (M_{n-k} - P_{n-k-1}) - (M_{n-k} - P_{n-k})$$

$$= p(FP_{n-k-1} - P_{n-k-1})q(MP_{n-k} - P_{n-k})$$

is given, where $p$ and $q$ can be computed from $P_{n-k-1}^j$, $P_{n-k}^j$, $FP_{n-k-1}^j$, and $MP_{n-k}^j$. The relative REM $RelREM_{n-k}^j$ at $M_{n-k}^j$ is defined by subtracting the REM at $M_{n-k}$ from that at $M_{n-k}^j$, where the REM at $M_{n-k}$ is estimated from the global feature point $FP_{n-k-1}$ and its matching point $MP_{n-k}$. Similarly, the relative DEM $RelDEM_{n-k}^j$ is defined.

In the second stage, the proposed localization algorithm searches the area, with 5 pixels interval, 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 absolute position is estimated by finding the position having the minimum cost in matching the relative REM and the relative DEM, where M-estimator [13] is used for computing the cost. Note that M-estimator is one of robust estimators, defined by replacing the square operation of least squares by the symmetric cost function that limits the large error. Note that the Huber metric is used for the symmetric cost function. The initial position is given by relative position estimation and the absolute position is obtained by searching around the initial position. The proposed DEM matching algorithm is described in [10].

$M_{n-k} = (M_{n-k-x}, M_{n-k-y}, M_{n-k-z})$ signifies the ground point corresponding to the global feature point $FP_{n-k-1}$, with $M_{n-k-z}$ corresponding to the REM at $M_{n-k}$. Similarly,

$$M_{n-k}^j = (M_{n-k-x}^j, M_{n-k-y}^j, M_{n-k-z}^j)$$

is the ground point corresponding to the sampled feature point $FP_{n-k-1}^j$ in the previous frame and its matching point $MP_{n-k}^j$ in the current frame, and $(x_1 - x_0) \times (y_1 - y_0)$ denotes the size of the search area in the DEM. In experiments, $DEM(x, y)$ represents an elevation at position $(x, y)$. $DEM_{n-k}(DEM_{n-k}^j)$ signifies the elevation at the global feature point $FP_{n-k-1}$ (the $j$th sampled feature point $FP_{n-k-1}^j$). $RelDEM_{n-k}(RelDEM_{n-k}^j)$ signifies the relative DEM (REM) at $M_{n-k}$ with respect to that at $M_{n-k}$. The error in elevation estimation is assumed to be the same for all feature points in a fixed frame. Thus, relative elevations instead of absolute elevations are used in computing the matching distance for magnitude-invariant matching.

For DEM matching, the error is largely dependent on the resolution of REM and DEM. As addressed before, the resolution of DEM is approximately 2m/pixel and the sampling interval is five pixels to match DEM and REM. Additionally, the resolution of REM depends on the altitude of an aircraft, with the minimum accuracy of estimation is about 10m.

### 5 EXPERIMENTAL RESULTS AND DISCUSSIONS

Experiments with four sets of real sequential aerial images were conducted to show the effectiveness of the proposed algorithm. The first aerial image sequence (test sequence I) was taken by a camera attached to a helicopter, over Daejon in Korea. The second, third, and fourth aerial image sequences (test sequences II, III, and IV) were taken over Kongju, in Korea, by a camera attached to a light airplane. The test sequences I, II, III, and IV consist of about 1,400, 1,770, 1,360, and 1,570 frames (sampled at one frame/second), with the total flight trajectories corresponding to about 54, 27, 126, and 124 km long, respectively. The trajectories are composed of a number of curved orbits and straight paths. The test sequence I was acquired by a $\beta$-cam video camera with the field of view equal to $42.4^\circ \times 54.7^\circ$, whereas the test sequences II, III, and IV were obtained by a Hi-8mm video camera with the field of view equal to $5.4^\circ \times 2.0^\circ$, $38.4^\circ \times 25.7^\circ$, and $18.0^\circ \times 22.5^\circ$, respectively. Attitude and altitude parameters are obtained from a gyroscope and an altimeter, respectively. These equipments were mounted with the camera and synchronized to correctly measure the parameters specific to an appropriate frame. Characteristics of the four test image sequences are

<table>
<thead>
<tr>
<th>Test sequence</th>
<th>Test sequence</th>
<th>Test sequence</th>
<th>Test sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aircraft</td>
<td>Helicopter</td>
<td>Light airplane</td>
<td>Light airplane</td>
</tr>
<tr>
<td>Camera</td>
<td>$\beta$-cam</td>
<td>Hi-8m video</td>
<td>Hi-8m video</td>
</tr>
<tr>
<td>Altitude (ft)</td>
<td>2,000 - 5,000</td>
<td>5,000</td>
<td>6,000</td>
</tr>
<tr>
<td>FOV (degree$^2$)</td>
<td>42.7 x 54.7</td>
<td>5.4 x 2.0</td>
<td>38.4 x 25.7</td>
</tr>
<tr>
<td>Area</td>
<td>Daejon</td>
<td>Kongju</td>
<td>Kongju</td>
</tr>
<tr>
<td>Frame size</td>
<td>320 x 240</td>
<td>240 x 320</td>
<td>240 x 320</td>
</tr>
<tr>
<td>Number of frames</td>
<td>1,400</td>
<td>1,770</td>
<td>1,360</td>
</tr>
<tr>
<td>Total trajectory (km)</td>
<td>54</td>
<td>27</td>
<td>126</td>
</tr>
</tbody>
</table>
summarized in Table 1. Note that the ground resolution (m/pixel) is given by \(2h \tan(\theta_r/2)/N_c\), where \(h\), \(\theta_r\), and \(N_c\) are the altitude of a camera, field of view along the row direction of the input frame, and the number of columns, respectively, assuming that the camera is down-looking. Usually, the ground resolution ranges approximately from 0.5 to 2 m/pixel. We use universal transverse Mercator (UTM) as the global coordinates, and Kongju and Daejon are located in 52 zone. Window size \((W_x \times W_y)\) is set to 31 \times 15 for NCC-based block matching. Search area \((U_x \times U_y)\) is set to 175 \times 113 at the first frame and reduced to 60 \times 80 for the next consecutive frames by making use of the previously estimated displacement.

Additionally, numerical error analysis of the proposed system is shown. It is difficult to analyze mathematically because the system is nonlinear. Thus, the robustness and sensitivity of the proposed system are shown through experiments.

Fig. 5. Sample test images. (a) Test sequence I (320 \times 240). (b) Test sequence II (240 \times 320). (c) Test sequence III (240 \times 320). (d) Test sequence IV (320 \times 240). (e) IRS (256 \times 256).
5.1 Estimated Trajectories for Four Aerial Image Sequences

Figs. 5a, 5b, 5c, and 5d show sample aerial images of four test image sequences I, II, III, and IV, respectively. All of them were quantized to eight bits. Fig. 5a (320 × 240) was acquired at 2,540 feet high on the helicopter. In parenthesis, the image size is represented by \( N_c \times N_r \), where \( N_c \) and \( N_r \) denote the numbers of column and row pixels, respectively. Figs. 5b and 5c (240 × 320) were acquired at 5,000 and 6,000 feet high on the light airplane, respectively. Fig. 5d

Fig. 6. Attitude parameters of the four test sequences as a function of the frame index. (a) Roll. (b) Pitch. (c) Yaw.
(320 × 240) was obtained at 6,000 feet on the light airplane. Fig. 5e shows the IRS image (256 × 256), which was quantized to six bits with 5 m/pixel ground resolution, covering the same region as the aerial image shown in Fig. 5a. Even though both images show the same region containing a number of apartments, they look different because they were acquired with different attitude parameters by different image sensors. While the resolution of an input frame is almost the same as that of high-resolution reference images such as Figs. 5a, 5b, 5c, and 5d, the satellite image looks different from the input video frame because the resolution and image characteristics are different. Thus, image compensation is necessary for effective matching of two images.

Each image in Fig. 5 shows the image at the position at which absolute position estimation was performed. Figs. 5a, 5b, 5c, and 5d correspond to the reference images at the position whose symbol indices are given by S2 (in Fig. 7), H2 (in Fig. 8), D1 (in Fig. 9), and H3 (in Fig. 10), respectively. Fig. 5e is used as a reference image at the position, whose symbol index is given by S2 (in Fig. 7).
Positions at which absolute position estimation is performed are marked with symbols consisting of two characters. The first character represents the type of absolute position estimation methods: S, H, and D for matching using the IRS satellite image, the high-resolution image, and DEM information, respectively. The second numeric character denotes the sequential index. Because it is assumed that the trajectory of the aircraft is planned before flying, the spots at which absolute position estimation is performed are determined in advance. Image matching is employed for the images that contain distinct artificial structures. Especially, the satellite image is used for the region where a high-resolution reference image is not available. DEM matching is used for mountain regions showing a large change rate of elevation.

Figs. 6a, 6b, and 6c show the roll, pitch, and yaw parameters as a function of the frame index for the four test sequences, respectively. The outline of the trajectory is determined by yaw and roll parameters, and the height of an aircraft is affected by the pitch parameter. The variance of roll values becomes large by strong wind. As shown in Fig. 6a, the roll angle of the test sequences III and IV shows a large variation because of strong wind.

Figs. 7a, 8a, 9a, and 10a show the experimental results of relative and absolute position estimations with the test sequences I, II, III, and IV, respectively. In the trajectory representation, thick and thin lines represent the real (correct) trajectory and the trajectory estimated by the proposed integrated system, respectively. The dot line denotes the trajectory obtained by manually identifying the position of an aircraft on a 5000:1 map, in which the
inevitable error is observed. It is very difficult to show the accuracy analytically; however, the error in the image space is less than ten pixels, which corresponds to approximately 30 m in the worst case. Note that only the positions having distinct features are manually estimated for accuracy. Because the relative position estimation method is based on a recursive approach, the estimated error increases with the frame number. The average position error defined as
is used as a performance measure, where \((X'_k, Y'_k)\) and \((X_k, Y_k)\) represent positions estimated by manual and relative/integrated position estimations, respectively, and \(K\) denotes the total number of frames manually estimated.

For the test image sequence I, the average position error by relative position estimation alone is 291 m. The average position error by the proposed system with integrated
position estimation is 175 m. The maximum errors by relative and integrated position estimations are 1,211 and 593 m, respectively. Experiments show that the estimated error is greatly reduced by absolute position estimation. Note that absolute position estimation pulls the trajectory, deviated by error accumulation in relative position estimation, to the real trajectory. It is confirmed that absolute position estimation is effective. Fig. 7b shows that absolute position estimation reduces the position error and that absolute position estimation using DEM information provides a good performance in mountain areas, in terms of the position error.

Fig. 8b shows the experimental results of relative and integrated position estimation algorithms with the test sequence II. Because the relative position estimation method is based on a recursive approach, the estimated error increases with time, similar to the case of the test sequence I. The average position error by relative position estimation alone is 538 m, whereas that by the proposed system with integrated position estimation is 105 m. The maximum errors by relative and integrated position estimations are 1,040 and 367 m, respectively. The estimated error is greatly reduced by absolute position estimation. Because this test sequence is acquired on the relatively straight flying trajectory, absolute position estimation is very effective. In the test sequence II, the angle of view is small (see Table 1). As a result, the ground coverage of the image is too small to match the relative DEM with the image matching algorithm is used as an absolute position estimation method. Experiments with the test sequence II also show that the proposed system yields a reasonable performance by effectively combining both relative and absolute position estimation methods.

Fig. 9b shows the experimental results of relative and integrated position estimation methods with the test sequence III. The average position error by relative position estimation alone is 456 m, whereas that by the proposed system with integrated position estimation is 309 m. The maximum errors by relative and integrated position estimations are 1,136 and 977 m, respectively. The estimated error is reduced by absolute position estimation. Even though the total flight distance is long compared with that of the test sequence II, the estimation error is small because this test image sequence is acquired with relatively stable attitude parameters.

Fig. 10b shows the experimental results of relative and integrated position estimation methods with the test sequence IV. The average position error by relative position estimation alone is 757 m, whereas that by the proposed system with integrated position estimation is 209 m. The maximum errors by relative and integrated position estimations are 1,521 and 1,495 m, respectively. The estimated error is greatly reduced by absolute position estimation. Experiments with the test sequence IV also show that the proposed system greatly reduces the error accumulated in relative position estimation. When this test sequence was acquired, the attitude of the aircraft was very unstable because of strong wind, which gives large fluctuations in roll parameter values. As a result, the estimation error for this test sequence is larger than that of the test sequence III.

Experimental results with four real test sequences show that relative position estimation yields reasonable estimates. However, the estimation error tends to increase as the aircraft navigates, which can be reduced by absolute position estimation. High-resolution image matching is more effective than other algorithms because resolution of reference images is high. IRS image matching is needed for the spots where high-resolution images are not available. DEM matching is suitable for mountainous regions with significant elevation change.

### 5.2 Numerical Error Analysis

The sensitivity of the proposed system against error is shown through experiments. There are two kinds of error sources: internal and external. The internal error is due to incorrect correspondences, whereas the external error includes measurement error and inaccuracy of the reference information. The error sensitivity of the relative position estimation and measurement error is shown experimentally with the test sequence I.

The derivatives of the displacement with respect to the correspondence is given by

\[
\frac{\partial B_{n,x}}{\partial M_{P_{n,x}}} = -(M_{n,z} - P_{n,z})
\]

\[
\frac{\partial B_{n,z}}{\partial M_{P_{n,y}}} = \frac{(R_{11}R_{32} - R_{12}R_{31})M_{P_{n,y}} + (R_{11}R_{33} - R_{13}R_{31})f}{(R_{31}M_{P_{n,x}} + R_{32}M_{P_{n,y}} + R_{33}f)^2}
\]

\[
\frac{\partial B_{n,y}}{\partial M_{P_{n,z}}} = \frac{(R_{12}R_{32} - R_{11}R_{32})M_{P_{n,x}} + (R_{12}R_{33} - R_{13}R_{32})f}{(R_{31}M_{P_{n,x}} + R_{32}M_{P_{n,y}} + R_{33}f)^2}
\]

From these derivatives, the sensitivities of the displacement along the \( x \) and \( y \) directions are defined by

\[
S_x = E(S_{n,x}) = E \left( \sqrt{\frac{\partial B_{n,x}}{\partial M_{P_{n,x}}} M_{P_{n,x}} + \frac{\partial B_{n,z}}{\partial M_{P_{n,y}}} M_{P_{n,y}}} \right)
\]

\[
S_y = E(S_{n,y}) = E \left( \sqrt{\frac{\partial B_{n,y}}{\partial M_{P_{n,z}}} M_{P_{n,z}} + \frac{\partial B_{n,z}}{\partial M_{P_{n,y}}} M_{P_{n,y}}} \right)
\]

where \( E(\cdot) \) represents expectation. The sensitivities \( S_x \) and \( S_y \) along \( x \) and \( y \) directions computed over the entire frames are experimentally obtained as 3.4 and 1.6 percent, respectively, assuming that the correspondence error is a half-pixel. On the other hand, the error caused by a large mismatching is approximately given by \( 2h \tan(\theta_f/2) d/N_c \), where \( h, \theta_f, d, \) and \( N_c \) represent the altitude of a camera, field of view along the row direction of the input frame,
Fig. 11. Error sensitivity of the proposed system. (a) Average relative error of $B_x$ as a function of the standard deviation of the Gaussian noise added in attitude parameters. (b) Average relative error of $B_y$ as a function of the standard deviation of the Gaussian noise added in attitude parameters. (c) Average relative errors in the attitude parameter. (d) Estimation additive Gaussian noise (standard deviation = 1 degree). (e) Estimation error as a function of the frame index with noisy attitude parameters contaminated by the additive Gaussian noise (standard deviation = 5 degrees). (f) Estimation error as a function of the frame index with the noisy altitude parameter contaminated by the additive Gaussian noise (standard deviation = 2 and 6 percent of the altitude value).
disparity error, and the number of columns, respectively, assuming that the camera is down-looking. The error is about 3.23 m/pixels when H is 1000 m. As a result, the error caused by a large mismatching (1 to 10 pixels) ranges from 3.23 m to 32.3 m. Such a large estimation error should be compensated by absolute position estimation.

Average relative errors of the displacements $B_x$ and $B_y$ as a function of the standard deviation of the Gaussian noise added in attitude parameters are shown in Figs. 11a and 11b, respectively. Also, those as a function of the standard deviation of the Gaussian noise added in the altitude parameter are shown in Fig. 11c. Assuming that the displacements calculated with acquired attitude parameters are correct, the average relative errors are obtained with noisy measurements. The displacement error largely depends on the accuracy of roll parameter values among other parameters.

Figs. 11d and 11e show the average position error as a function of the frame index for the noisy altitude parameter contaminated by the additive Gaussian noise with the standard deviation equal to 1 and 5 degrees, respectively, with the assumption that the trajectory estimated by the algorithm yields good estimates on the whole in short navigation. As an aircraft navigates, the estimation error is accumulated. To reduce the estimation error, absolute positioning algorithms based on image and DEM matching are employed. Experiments with real aerial image sequences show the effectiveness of the proposed integrated position estimation system, in terms of the average and maximum errors. Further research will focus on a real-time implementation of the proposed system on the TMS320C80 digital signal processor.

6 CONCLUSIONS

This paper proposes an integrated location estimation algorithm using aerial image sequences. The proposed integrated system consists of the relative and absolute positioning parts. The relative position estimation algorithm yields good estimates on the whole in short navigation. As an aircraft navigates, the estimation error is accumulated. To reduce the estimation error, absolute positioning algorithms based on image and DEM matching are employed. Experiments with real aerial image sequences show the effectiveness of the proposed integrated position estimation system, in terms of the average and maximum errors. Further research will focus on a real-time implementation of the proposed system on the TMS320C80 digital signal processor.

APPENDIX

LIST OF ACRONYMS

- BMA: Block matching algorithm
- DEM: Digital elevation model
- DT: Distance transform
- GPS: Global positioning system
- HT: Hough transform
- INS: Inertial navigation system
- IRS: Indian remote sensing
- NCC: Normalized correlation coefficient
- REM: Recovered elevation map
- ROHM: Robust oriented Hausdorff measure
- TERCOM: Terrain contour matching
- USGS DOQs: US Geological Survey digital ortho-photo quadrangles
- UTM: Universal transverse Mercator

ACKNOWLEDGMENTS

This work was supported in part by the Agency for Defense Development and by ACRC (Automatic Control Research Center), Seoul National University.

REFERENCES


Dong-Gyu Sim received the BS and MS degrees in electronics engineering from Sogang University, Seoul, Korea, in 1993 and 1995, respectively. He also received the PhD degree at the image processing laboratory of the same university, in 1999. He was at Hyundai Electronics Co., Ltd from 1999 to 2000, where was involved in MPEG-7 standardization. He was a senior research engineer at Vario Vision Co., Ltd., working on MPEG-4 applications. Currently, he works at Sogang University in the Department of Electronic Engineering. His current research interests are image processing, computer vision, pattern recognition, and MPEG-4/7.

Rae-Hong Park (S ’76-M ’84-SM ’99) received the BS and MS degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1976 and 1979, respectively, and the MS and PhD degrees in electrical engineering from Stanford University, Stanford, California, in 1981 and 1984, respectively. He joined the faculty of the Department of Electronic Engineering at Sogang University, Seoul, Korea, in 1984, where he is currently a professor. In 1990, he spent his sabbatical year at the Computer Vision Laboratory of the Center for Automation Research, University of Maryland, College Park, as a visiting associate professor. He received a postdoctoral fellowship from the Korea Science and Engineering Foundation (KOSEF) in 1990. He received the Academic Award in 1987 from the Korea Institute of Telematics and Electronics (KITE) and the Haedong Paper Award in 2000 from the Institute of Electronics Engineers of Korea (IEEK). Also, he received the First Sogang Academic Award in 1997 and the Professor Achievement Excellence Award in 1999, from Sogang University. He served as editor for the KITE Journal of Electronics Engineering in 1995-1996. His current research interests are computer vision, pattern recognition, and video communication. He is a senior member of the IEEE.

Rin-Chul Kim (M ’95) received the BS, MS, and PhD degrees all in control and instrumentation engineering from Seoul National University, in 1985, 1987, and 1992, respectively. From 1992 to 1994, he was with the Daewoo Electronics Co., LTD, and worked on the development of the HDTV and DBS systems. From 1994 to 1999, he was an assistant professor with the School of Information and Computer Engineering at Hankuk University, Seoul. In 1999, he joined the School of Electrical and Computer Engineering at the University of Seoul as an assistant professor. His current research interests include image processing, target tracking, image data compression, and VLSI signal processing. He is a member of the IEEE.

Sang Uk Lee (S ’75-M ’80-SM ’99) received the BS degree from Seoul National University, Seoul, Korea, in 1973, the MS degree from Iowa State University, Ames, in 1976, and PhD degree from the University of Southern California, Los Angeles, in 1980, all in electrical engineering. In 1980-1981, he was with the General Electric Company, Lynchburg, Virginia, working on the development of digital mobile radio. In 1981-1983, he was a member of technical staff, M/A-COM Research Center, Rockville, Maryland. In 1983, he joined the Department of Control and Instrumentation Engineering at Seoul National University as an assistant professor, where he is now a professor in the School of Electrical Engineering. Currently, Dr. Lee is also affiliated with the Automation and Systems Research Institute and the Institute of New Media and Communications at Seoul National University. His current research interests are in the areas of image and video signal processing, digital communication, and computer vision. He served as an editor-in-chief for the Transaction of the Korean Institute of Communication Science from 1994 to 1996. He is a member of the editorial board of the Journal of Visual Communication and Image Representation and an associate editor for IEEE Transactions on Circuits and Systems for Video Technology. He is a member of Phi Kappa Phi. He is a senior member of the IEEE.

Ihn-Cheol Kim received the BS and MS degrees in electronics engineering from Dankook University, Seoul, Korea, in 1985 and 1987, respectively. He is a senior research engineer at the Agency for Defense Development of Korea, where he is working in the areas of computer vision and image understanding.

For more information on this or any other computing topic, please visit our Digital Library at http://computer.org/publications/dlib.