MODELING, IDENTIFICATION AND CONTROL, 1993, VOL. 14, NO. 3, 133–143
doi:10.4173/mic.1993.3.2

Navigation by images

ESPEN HAGEN† and EILERT HEYERDAHL†

Keywords: *Autonomous navigation, computer vision, optical flow, Kalman filters.*

A new navigation method based on measurements of *image tokens* and Kalman filtering is presented. An image token is the central projection of a *landmark*, a point on the terrain surface. This surface being described by an elevation map, a Kalman filter processes the measurements to update estimates of camera position and orientation, and landmarks. The method has been implemented for off-line simulations of aeroplane navigation. Preliminary tests indicate a performance at least comparable to that of satellite navigation systems. The implemented algorithm also seems to have high tolerance against noise and modeling errors.

1. Introduction

Navigation, absolute positioning, is an important task in many applications. Traditional, inertial navigation systems (INS) measure *accelerations* and thus have an increasing positional uncertainty. This problem can be reduced by increasing the accuracy of the accelerometers, thus making today's high-quality INS systems very expensive.

The only possible way to *eliminate* the problem of increasing uncertainty is to employ *absolute position dependent measurements*. This is done in radio and satellite navigation systems such as Omega and Navstar GPS, but can also be done *autonomously* in image-based navigation. Research on image-based navigation has primarily been focused on *motion* estimation (Zinner *et al.* 1989), two- or three-degree-of-freedom applications (Aguirre *et al.* 1990), use of stereo imaging (Blackman 1991), and/or recognition of specific landmarks with a priori known position (Dickmanns 1988). The navigation method described in this paper is designed for use with a *single camera*, the only reference data required is an *elevation map*, and it can be used for navigation with *six degrees of freedom* (camera position and orientation unknown).

2. Basic concepts

Let each element in a set of static terrain points be called a *landmark*, and its central projection an *image token*. As the landmarks are points on the terrain surface, one component of a landmark is assumed to be a function of the other two. This function is given by an elevation map $m(\cdot)$. A token is therefore a function of camera position and orientation (called *camera state*), and two components of the landmark (called *reduced landmark*).

We define the *token flow* as a function $\mathbf{f}$: $\mathbb{R}^2 \times \mathbb{R} \rightarrow \mathbb{R}^2$ such that $\mathbf{f}(x, y; t)$ is the central projection of the landmark $(x, y, m(x, y))$, at time t. The token flow is related to the

Received 10 October 1992.
† Norwegian Defence Research Establishment (NDRE), Division for Weapons and Materiel, N-2007 Kjeller, Norway.

optical flow, often defined as the apparent motion of brightness patterns in an image (Horn and Schunk 1981).

3. Direct solution

If the imaged terrain is flat, the entire motion field (the token velocity field) can be represented by 8 so-called *essential parameters* (Tsai and Huang 1984). These parameters also relate the motion field to the camera motion (velocity by a scale factor, and angular velocity) (Zinner *et al.* 1989). As any solution for camera position can be arbitrarily translated and/or scaled, this method can not yield bounded-error position estimates. Moreover, the results in a nonlinear terrain are unpredictable.

A generalized version of the above approach, here called the *direct solution*, can be described as follows. A set of T tokens in each of F frames can be used to generate a set of $2FT$ equations describing the token flow as a function of camera states ($6F$ unknowns) and reduced landmarks ($2T$ unknowns). These equations are always nonlinear, since camera orientations are unknown. With nonlinear terrain, the map represents additional nonlinearity in the system of equations.

A closed-form solution to the equations cannot easily be obtained. However, it can be shown that under certain conditions, one being a nonlinear terrain, *isolated* solutions do exist (Hagen and Heyerdahl 1992). An isolated solution is such that a small perturbation of it does not produce a solution. If F equals two, six landmarks can be sufficient to yield isolated solutions. With more frames (and thus more measurements of the same tokens), even fewer landmarks can suffice.

Note that this approach does not involve the computation of the essential parameters, and that the equations are valid also in a nonlinear terrain. In a linear terrain, the direct solution will be equivalent to the traditional optical flow/essential parameters approach.

4. Proposed solution

Another approach towards obtaining camera position is using a Kalman filter—with a state vector including camera state and reduced landmarks, and with tokens as measurements. The Kalman filter estimates are solutions to a different estimation problem than that of the direction solution, as an estimate of the initial state is required (and not solved for). If tokens can be measured exactly, a zero error estimate requires the direct solution. The Kalman filter is advantageous when measurements are noisy or when there exists more than one solution. This is very often the case, and the Kalman filter then processes additional information in an optimal way.

In a nonlinear terrain, translation and/or scaling of the reduced-landmark–camera system might give rise to different measurements. Bounded-error position estimation is therefore (at least in theory) possible.

In a linear terrain (where the direct solution approach will fail), the Kalman filter can estimate velocities with bounded uncertainties and thus position and orientation (with linearly increasing uncertainties). Furthermore, if the system does not lose track of at least three initial landmarks, and if the camera-landmarks distance is bounded, camera position can be estimated with bounded uncertainty *even in a linear terrain*: Estimates of the initial landmarks are calculated from the tokens in the first frame, the initial camera state, and the elevation map. The measured tokens in subsequent frames are thus measurements of the central projection of 'known' landmarks. (As the

landmarks are assumed to be static, their estimate error covariances are non-increasing.

An important part of the algorithm is the selection of tokens. Given a dynamic model, the selected tokens must provide satisfactory or, failing that, (sub-)optimal camera position estimate error.

When considered individually, the suitability of a potential token is determined by three criteria:

1. The *error when the token is measured (by image processing)*. The measurement error is dependent on the image function around the potential new token.
2. *The landmark estimate error*. The landmark estimate error when selected depends on the information in any extra reference data. There is, however, always an upper bound to this error, given by the camera state and its estimate error, the error in the elevation map, and the correspondence between token and landmark.
3. *The linearity in the terrain around the landmark*. As the Kalman filter ideally requires a linear measurement function, the terrain should be reasonably linear within the landmark's region of uncertainty.

Considered as a whole, the tokens should be distributed—ideally throughout the image. The terrain function normals at the landmarks should also be distributed.

By using the predicted state estimate and its error covariance, search areas can be computed to facilitate the token measuring. A landmark is typically deleted from the state vector if it is outside the field of view, or after the corresponding token has not been accepted as measured for a number of consecutive frames.

The general algorithm is outlined in Figure 1.

5. Implementation

The following sections describe an implementation of the algorithm in Fig. 1, for aeroplane navigation. The algorithm is implemented on a Teragon image processing computer using a microVax 3600 as host. The system consists of a Twin Otter aircraft and a TICM II imaging IR (8–12 μm) sensor, or a CCD TV camera, fixed to the

```
Get initial state estimate x̂ and error covariance matrix P
FOR each frame DO
  FOR each token DO
    Measure token, calculate confidence
  Calculate confidence threshold
  Measurement vector dimension:=0
  FOR each token DO
    IF measurement confidence>threshold THEN
      Include measured token in measurement vector
    ELSE
      Determine deletion of the landmark states
      Reduce x̂ and P as determined
  IF measurement vector dimension>0 THEN
    Update x̂ and P
  IF too few landmarks THEN
    Select new tokens
    Augment x̂ and P
  Predict x̂ and P for next frame time
```

Figure 1. The general algorithm.

fuselage. Camera orientation relative to the aeroplane is fixed; the camera is tilted approximately 4·5° down.

Apart from an elevation map (DTED–Digital Terrain Elevation Data), no reference data is used. When an image token is selected, an estimate of the corresponding landmark is found from an *inverse central projection function* $\mathbf{g}(\cdot)$. With camera state, the token and the elevation map as input, this function returns the nearest intersection between the projection ray and the terrain surface (Hagen 1992).

5.1. *Kalman filter*

The filter is an Iterated Extended Kalman filter with continuous-time dynamics and discrete-time measurements (Gelb 1974). A factorized mechanization is used to ensure long-term numeric stability. The system model is

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)) + \mathbf{w}(t). \tag{1}$$

The state vector $\mathbf{x}$ and the dynamics function $\mathbf{f}(\cdot)$ are shown in Table 1, $\mathbf{u} = [u_1\ u_2\ u_3]^T$ is the control vector. u_1 and u_2 are rudder and elevator deflections, u_3 is thrust. The process noise $\mathbf{w}$ is assumed to be zero mean, white, and Gaussian, with a diagonal spectral density matrix Q. Only components 7–13 in $\mathbf{w}$ are nonzero.

The aeroplane is modeled to move along its principal axis, in an atmosphere moving with the wind velocity. The dynamic model was roughly designed and not systematically tuned.

The measurement model is given by

$$\begin{bmatrix} z_{2i-1} \\ z_{2i} \end{bmatrix} = \frac{F}{F+Y_i}\begin{bmatrix} X_i \\ Z_i \end{bmatrix} + \begin{bmatrix} v_{2i-1} \\ v_{2i} \end{bmatrix}, \quad i = 1, 2, .., m, \tag{2}$$

Comp. no	State vector: Symbol	Description	Group	Dynamics function
1	α	Azimuth angle		$\dot{\alpha}$
2	β	Pitch angle	Platform orientation [rad]	$\dot{\beta}$
3	γ	Roll angle		$(c_1\dot{\alpha} - c_2 u_1)v/g + \dot{\gamma}_q$
4	l	Longitude		$-v\sin\alpha\cos\beta + \omega_l$
5	λ	Latitude	Platform position [m]	$v\cos\alpha\cos\beta + \omega_\lambda$
6	h	Altitude		$v\sin\beta + \omega_h$
7	$\dot{\alpha}$	Azimuth rate		$-c_1\dot{\alpha} + c_2 u_1$
8	$\dot{\beta}$	Pitch rate	Angular velocity [rad·s^{-1}]	$-c_3\dot{\beta} + c_4 u_2$
9	$\dot{\gamma}_q$	Quasi roll rate		$-c_5\gamma - c_6\dot{\gamma}_q$
10	v	Speed	[ms^{-1}]	$-c_7 v - g\sin\beta + c_8 u_3$
11	ω_l	Wind East		$-c_9\omega_l$
12	ω_λ	Wind North	Wind velocity [ms^{-1}]	$-c_9\omega_\lambda$
13	ω_h	Wind Up		$-c_9\omega_h$
14	l_1	Longitude 1	Landmark 1 [m]	0
15	λ_1	Latitude 1		0
⋮	⋮			⋮
$12+2m$	l_m	Longitude m	Landmark m [m]	0
$13+2m$	λ_m	Latitude m		0

Table 1. Description of state vector and dynamics function. $g = 9{\cdot}81\ \mathrm{ms}^{-2}$ is gravitational acceleration.

where F is the focal length of the camera, and

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \mathcal{R}(\alpha_c, \beta_c, \gamma_c)\mathcal{R}(\alpha, \beta, \gamma) \begin{bmatrix} l_i - l \\ \lambda_i - \lambda \\ m(l_i, \lambda_i) - h \end{bmatrix}. \tag{3}$$

$\mathcal{R}(\cdot)$ is a 3×3 rotation matrix, and $(\alpha_c, \beta_c, \gamma_c)$ defines the camera orientation relative to the aeroplane. $m(\cdot)$ is the elevation map. Therefore,

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{v}. \tag{4}$$

The measurement noise $\mathbf{v} \sim N(\mathbf{0}, R)$ is assumed to be white and uncorrelated with the process noise $\mathbf{w}$. The covariance matrix R is diagonal.

5.2. *Token description*

Pixel describing features are used in the measuring and selection processes. The features are *circle integrals* (Heyerdahl 1991), defined by

$$e_i(\mathbf{p}) = \int_0^{2\pi} f(p_1 + r_i \cos\theta, p_2 + r_i \sin\theta)\, d\theta \tag{5}$$

where $\mathbf{p} = (p_1, p_2)$ is a point in the image, f is the image function and r_i is a radius. If f is a step function, (5) reduces to a 2D FIR filter where each pixel weight equals the sector angle of the part of the circumference enclosed in the pixel. This is illustrated in Fig. 2.

These features are invariant with respect to translation and rotation, but not to scaling. Also, the summation of pixel values suppresses image noise effects (signal-to-noise ratio is increased).

Several circle integrals are compiled in a feature vector, denoted $\mathbf{e}(\mathbf{p})$. In the implementation, 5 circle integrals, with radii of 3, 6, 9, 12 and 15 pixels, are used.

5.3. *Token selection*

The selection of new tokens is performed in a two-stage process. Firstly, the entire image is (sub-optimally) searched for points that maximize a *uniqueness measure* (Heyerdahl 1991) under the side condition of a minimum distance to other tokens. The uniqueness measure of a pixel $\mathbf{p}$ is designed to be small if there exists a pixel on a circular disc with a given radius, centred at $\mathbf{p}$, with large spatial distance and small feature distance to $\mathbf{p}$. This first stage thus returns a list of potential new tokens.

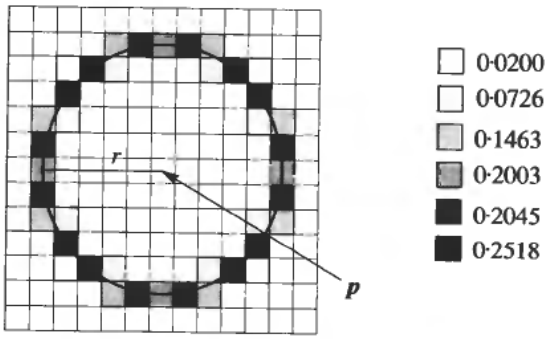

Figure 2. Illustration of the circle integral with radius $r = 5$ for pixel $\mathbf{p}$. Summation is performed over the hatched pixels, with weights as shown.

In the second part of the selection process, the landmark corresponding to each potential token is calculated, using the inverse central projection function $\mathbf{g}(\cdot)$. A potential token is rejected if the landmark is too far away from the camera, or if it is found to be near a strong non-linearity in the imaged terrain (e.g. near what is imaged as the edge of a hill). This is determined by applying $\mathbf{g}(\cdot)$ to a number of pixels near the token. Enough tokens are selected to get a required total number.

Reduced landmarks are included in the state vector. Error covariances are found by linearization of $\mathbf{g}(\cdot)$, and included in an augmented estimate error covariance matrix.

5.4. *Token measuring*

For each token $\mathbf{t}$, a search area $S(\mathbf{t})$ which contains the token with a given probability is calculated. Using the assumption that the state estimate error is Gaussian, this area is found from the predicted state estimate and error covariance by linearization of the measurement (central projection) function. $S(\mathbf{t})$ is an elliptical region in the image, centred at the predicted position of $\mathbf{t}$.

A pixel $\mathbf{p} \in S(\mathbf{t})$ which minimizes the feature distance to the predicted features $\hat{\mathbf{e}}(\mathbf{t})$, is the potential measured position of $\mathbf{t}$. The confidence $c(\mathbf{p})$ is then calculated (Heyerdahl 1991). The token $\mathbf{t}$ will be accepted as measured if $c(\mathbf{p})$ is greater than an adaptive threshold, designed to ensure that a minimum number of tokens are accepted.

$\hat{\mathbf{e}}(\mathbf{t})$ equals $\mathbf{e}(\mathbf{t})$ in the preceding frame if $\mathbf{t}$ is one frame old. $\hat{\mathbf{e}}(\mathbf{t})$ equals updated features at $\mathbf{t}$ in the preceding frame otherwise. Updated token features are calculated as a weighted sum of $\mathbf{e}(\mathbf{p})$ and $\hat{\mathbf{e}}(\mathbf{t})$.

If a token is not accepted as measured, the corresponding landmark states are deleted from the state vector. The estimate error covariance matrix is reduced accordingly.

6. Test results

A data set (image sequences) was acquired with the aforementioned aeroplane/sensor combination. The flights took place in southern Norway, mainly over rural areas with hilly or undulating terrain. The aeroplane-to-ground altitude was typically 400–700 m. For system initialization and result evaluation, the trajectory of the aircraft (position and orientation) was recorded from a combined INS/GPS navigation system. Controls were not measured.

The implemented navigation system has been tested by various off-line simulations. Due to an unexpected inconsistency between the images on one hand, and the recorded trajectories, the imaging function and the elevation map on the other, a two-step test has been devised. Firstly, the measuring process has been simulated (to determine the magnitude of the measurement errors). Secondly, the navigation system is run with simulated measurements (to determine the navigation errors).

The inconsistency is probably caused by an incorrect measurement model. More specifically, we have strong indications of (1) significant discrepancies from the central projection function, and (2) incorrect calibration of the camera as regards orientation. Correction of these errors is straightforward, but time-consuming.

6.1. *Measuring simulation*

Simulations of the measuring process have been performed on both infrared and TV image sequences in the data set. For the token measuring, a simple prediction (constant image-plane velocity for each token individually) and circular search areas

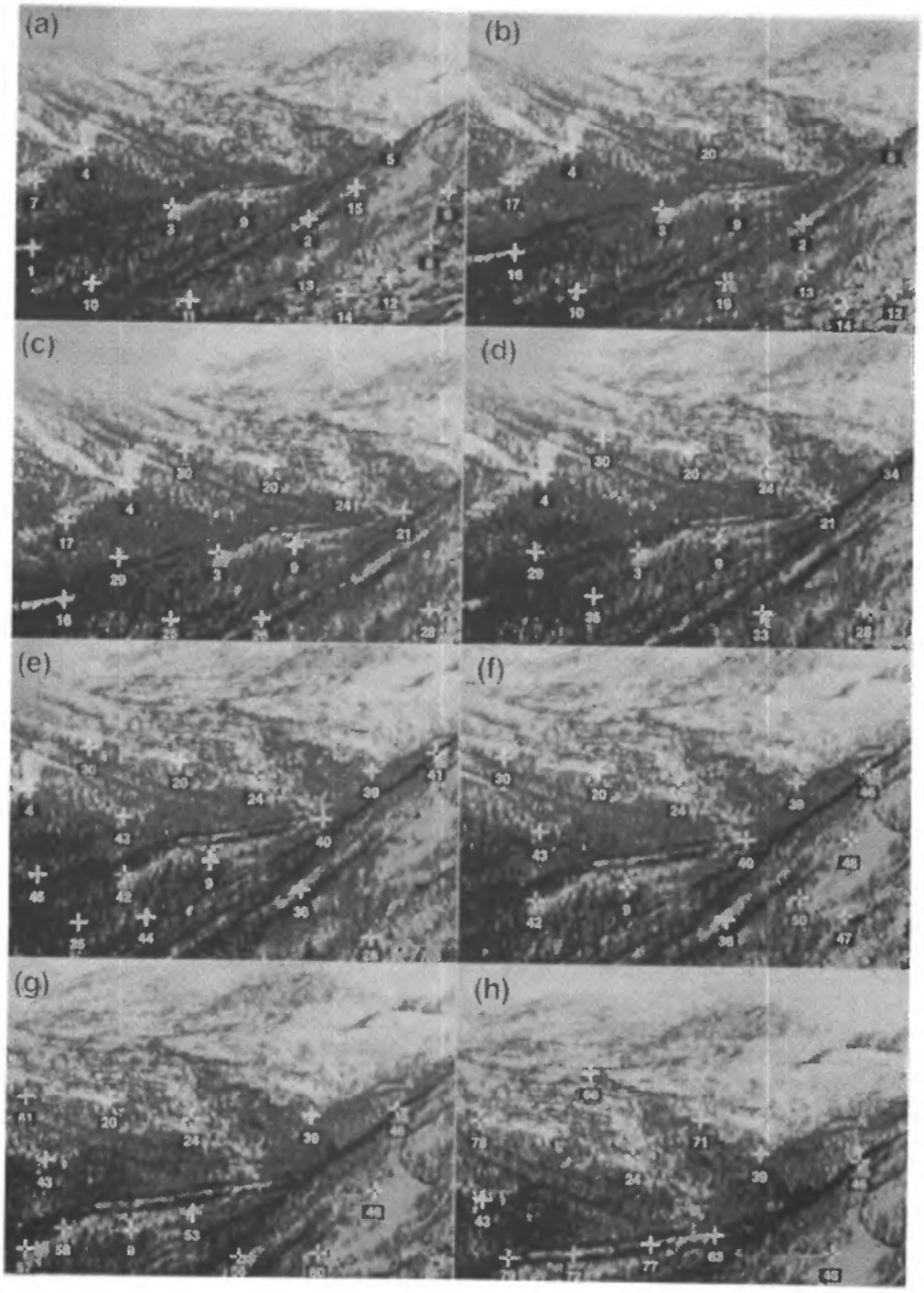

Figure 3. Measuring simulation (CCD TV images). Every 38th frame is displayed. (*a*) Start, (*b*)–(*h*) after 3, 6, 9, 12, 15, 18, 21 seconds. The images are contrast enhanced for this display.

(radius 15 pixels) were used. The resulting measurement error is approximately 1 pixel in medium-quality images. Tokens that were actually in the image were accepted as measured with more than 99·5% probability. With poorer images (e.g. in light fog), the error typically rises to 2–3 pixels, and the acceptance probability drops to 94–98%.

Figure 3 shows selected TV images from one simulation. The frame rate was 12·5 Hz, and the displayed images are 3 s apart. Image size is 500×720 pixels. By manual inspection, the mean measurement error was found to be less than 1 pixel. Similar results are achieved with IR images (Hagen 1992).

Even better performance must be expected with the complete system, as the superior token prediction offered by the Kalman filter makes smaller search areas acceptable.

6.2. *Navigation simulation*

In the tests utilizing simulated measurements, the first part of the 'token' selection was governed by the side condition only. The corresponding 'landmarks' were found by the inverse central projection function. Measurements were simulated by computing the central projections of the landmarks, and adding Gaussian noise. In the projection processes the recorded aeroplane trajectories and the elevation map were used. A 100% measurement acceptance probability was assumed for tokens in the image. Because control measurements were not available, the constants c_2, c_4, and c_8 were set to zero. The other constants are listed in Table 2. The nonzero elements of the process noise spectral density matrix Q are shown in Table 3.

Results from one such simulation are shown in Figs 4 and 5. The duration is 160 seconds, i.e. 4000 frames at full video rate. An $8 \times 12°$ field of view discretized into 1000×1000 pixels was simulated. 12 tokens are 'measured', with a measurement error standard deviation (SD) of 1·2 pixels. The terrain-map deviation is modeled as Gaussian with SD = 2·5 metres.

The estimated position is seen to be within a few metres from the recorded. More important still, the errors do not increase with time. This performance is at least comparable to that of high-quality satellite navigation systems. Estimated camera orientation is correct within 0·05° for azimuth and pitch, and 0·4° for roll. Other simulations have shown satisfactory results with lower image resolution, lower frame rates (as low as 2 Hz), larger measurement noise, and larger initial errors (Hagen 1992).

7. Possible applications

The implemented algorithm was designed for aeroplane navigation, in that the dynamic model was assumed to roughly describe a trajectory of a light aircraft. For

Constant	c_1, c_3	c_5	c_6	c_7	c_9
Value	0·2 s^{-1}	0·04 s^{-2}	0·4 s^{-1}	0	0·01 s^{-1}

Table 2. Model constants.

Diagonal element no.	7	8	9	10	11, 12, 13
Value	0·0004 $rad^2 s^{-4}$	0·0002 $rad^2 s^{-4}$	0·0016 $rad^2 s^{-4}$	0·36 $m^2 s^{-4}$	0·04 $m^2 s^{-4}$

Table 3. Nonzero elements of the process noise spectral density matrix.

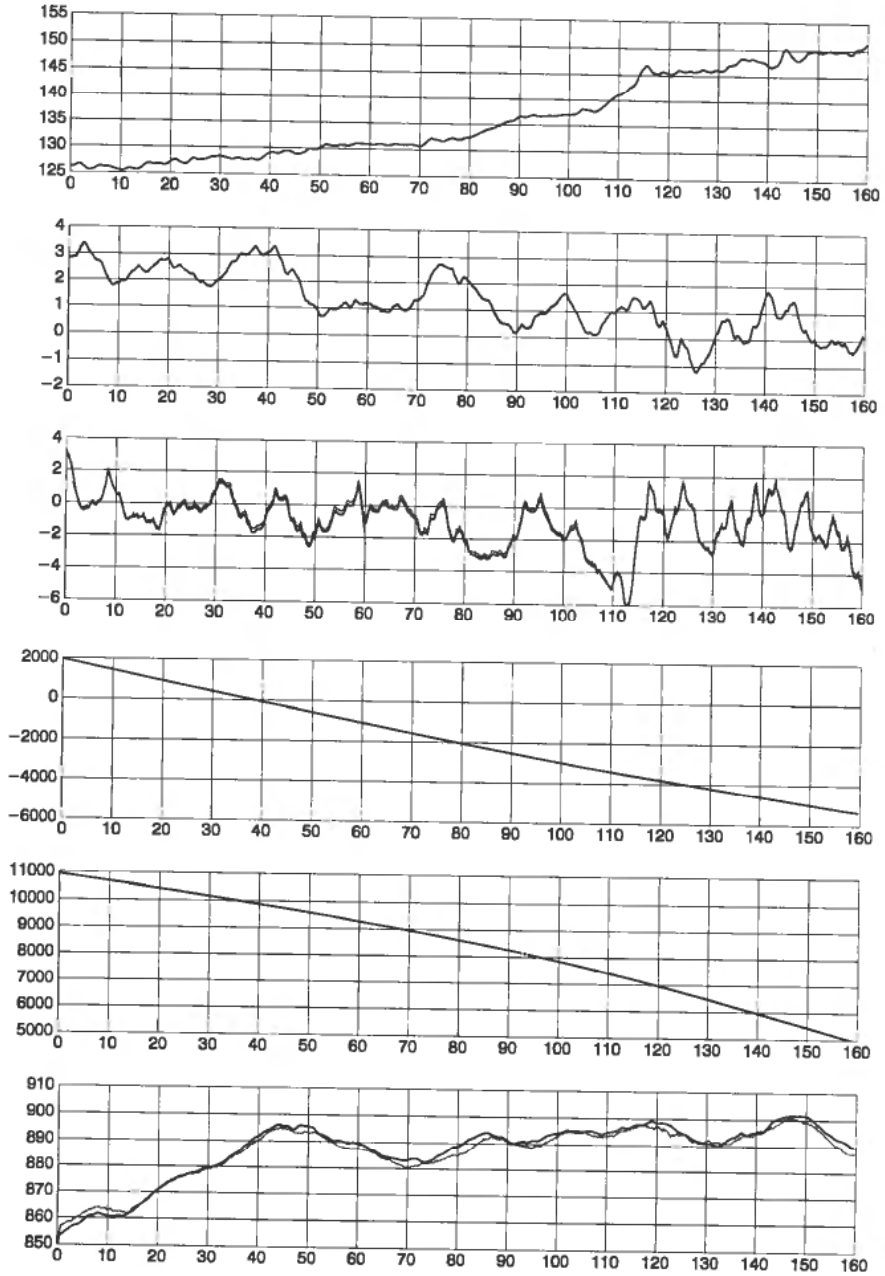

Figure 4. Results from the navigation simulation. (*a*) Azimuth angle, (*b*) pitch angle, (*c*) roll angle, (*d*) longitude relative to 11°E, (*e*) latitude relative to 60°N, and (*f*) altitude relative to mean sea level. Horizontal axis: Time in seconds. Vertical axis: Recorded (black) and estimated (gray) orientation and position. Angles in degrees, positions in metres.

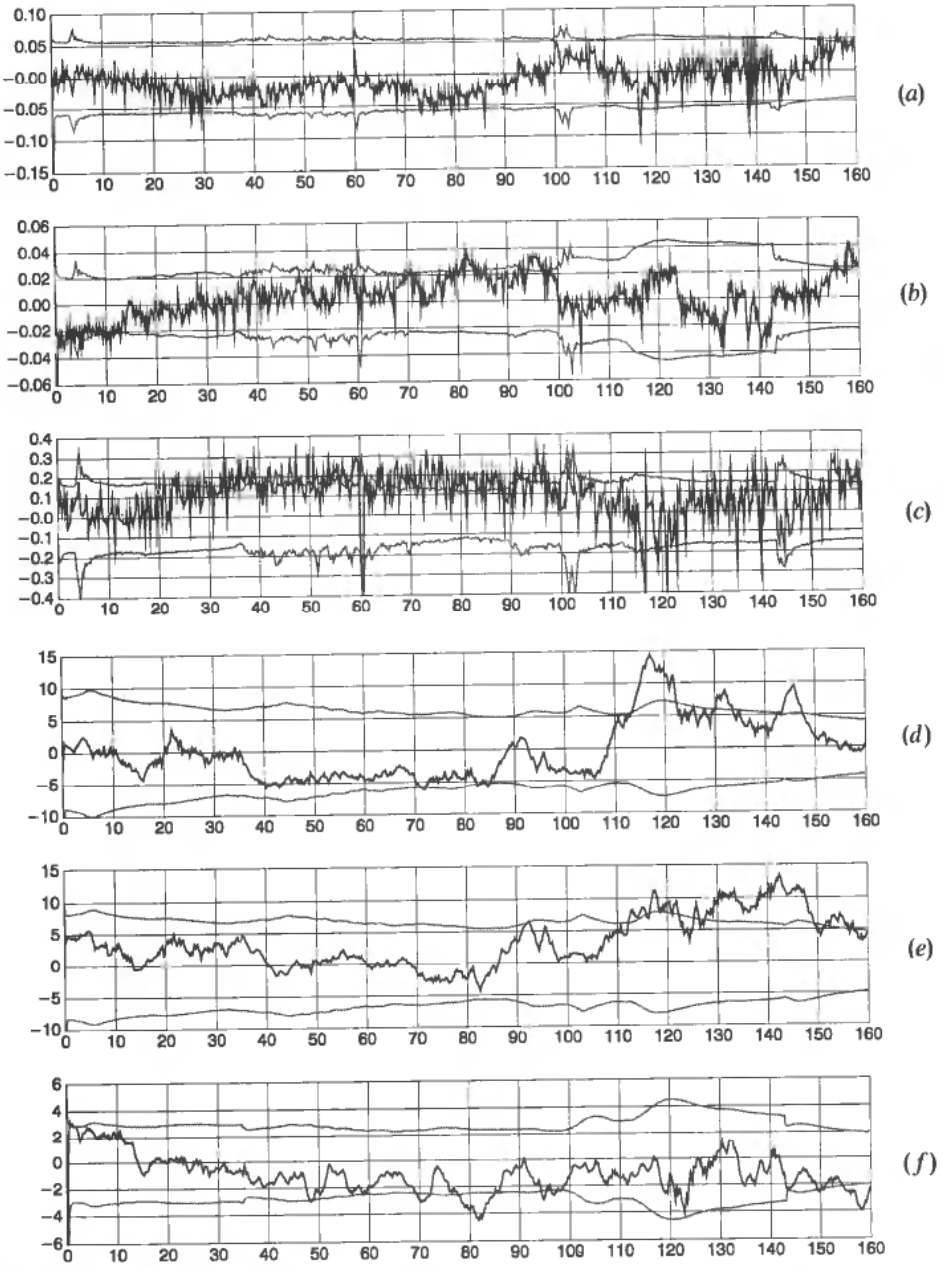

Figure 5. Results form the navigation simulation. (*a*) Azimuth angle, (*b*) pitch angle, (*c*) roll angle, (*d*) longitude, (*e*) latitude and (*f*) altitude. Horizontal axis: Time in seconds. Vertical axis: Orientation and position estimate error (black) and ± 1 filter estimate error standard deviation (gray). Angles in degrees, positions in metres.

the general algorithm, a number of other applications are conceivable—missiles, unmanned air vehicles, mobile robots, and ships in coastal waters. A navigation system employing the presented method can be run with no other measurement sources than a camera. Inclusion of other measurement sources, such as inertial platforms or altimeters, will improve system performance.

For adequate handling of abrupt manoeuvring, actuator information will normally be necessary. This information will, however, be readily available in most applications.

8. Conclusion

A new method for navigation (absolute positioning) based on image processing and Kalman filtering has been presented. The method has been implemented for aircraft navigation. Preliminary test results from this implementation indicate performance comparable to, or better than, that of high-quality satellite navigation systems, and high tolerance against noise and modeling errors. Several other applications of the method are conceivable, including navigation in missiles, mobiles robots, and ships.

References

AGUIRRE, F., BOUCHER, J. M., and JACQ, J. J. (1990), Underwater navigation by video sequence analysis. *Proceedings 10th ICPR*, Atlantic City, NJ, 537–539.

BLACKMAN, C. P. (1991), Robot vision for an autonomous land vehicle. *Proceedings NATO DRG Seminar on Battlefield Robotics*, Paris, France, 1991, 330–341.

DICKMANNS, E. D. (1988), An integrated approach to feature based dynamic vision. *Proceedings CVPR '88*, Ann Arbor, Mich., 820–825.

GELB, A. (ed.) (1974), *Applied Optimal Estimation* (M.I.T. Press, Cambridge, Mass).

HAGEN, E. (1992), BASIS image based navigation—method, implementation and results (in Norwegian). FFI/Rapport–92/4008, NDRE.

HAGEN, E. and HEYERDAHL, E. (1991–92), On algebraic solutions to the navigation problem (in Norwegian). Internal notes and discussions, NDRE.

HEYERDAHL, E. (1991), Circle integrals—simple, noise robust, translation and rotation invariant features (in Norwegian). *Proceedings NOBIM 1991*, Skedsmo, Norway, 43–50.

HORN, B. K. P. and SCHUNK, B. G. (1981), Determining optical flow. *Artificial Intelligence*, **17**, 185–203.

TSAI, R. Y. and HUANG, T. S. (1984), Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 13–27.

ZINNER, H., SCHMIDT, R. and WOLF, D. (1989), Navigation of autonomous air vehicles by passive imaging sensors. *Guidance and Control of Unmanned Air Vehicles, AGARD Conference Proceedings No. 436*, 34: 1–14.